Transportation Research Part C 167 (2024) 104802

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Lateral conflict resolution data derived from Argoverse-2: Analysing safety and efficiency impacts of autonomous vehicles at intersections

Guopeng Li *, Yiru Jiao, Simeon C. Calvert, J.W.C. (Hans) van Lint

Transport and Planning, Civil Engineering and Geosciences, Delft University of Technology, Mekelweg 5, 2628 CD Delft, Netherlands

ARTICLE INFO

Keywords:
Autonomous vehicles
Trajectory dataset
Conflict resolution
Safety assessment
Efficiency assessment

ABSTRACT

With the increased deployment of autonomous vehicles (AVs) in mixed traffic flow, ensuring safe and efficient interactions between AVs and human road users is important. In urban environments, intersections have various conflicts that can greatly affect driving safety and traffic efficiency. This study uses road test data to examine the possible safety and efficiency impacts of intersection conflict resolution involving AVs. The contribution comprises two main aspects. Firstly. we prepare and open a high-quality lateral conflict resolution dataset derived from the Argoverse-2 data, specifically targeting urban intersections. A rigorous data processing pipeline is applied to extract pertinent scenarios, rectify anomalies, enhance data quality, and annotate conflict regimes. This effort yields 5000+ AV-involved and 16000 AV-free cases, covering rich conflict regimes and balanced traffic states. Secondly, we employ surrogate safety measures to assess the safety impact of AVs on human-driven vehicles (HVs) and pedestrians. In addition, a novel concept of Minimum Recurrent Clearance Time (MRCT) is proposed to quantify the traffic efficiency impacts of AVs during conflict resolution. The results show that, for AV–HV and HV–HV conflict resolution processes, the differences in selected safety and efficiency measures for human drivers are statistically insignificant. In contrast, pedestrians demonstrate diverse behaviour adjustments. Some pedestrians behave more conservatively when interacting with AVs than with HVs. Notably, the efficiency of AV-involved conflict resolution is significantly lower than in AV-free instances due to the conservative driving style of AVs. This efficiency gap is particularly large when AVs pass through the conflict point after human drivers in unprotected left turns. These observations offer a perspective on how AVs potentially affect the safety and efficiency of mixed traffic. The processed dataset is openly available via https://github.com/RomainLITUD/conflict_resolution_dataset.

1. Introduction

High-level Autonomous Vehicles (AVs) are expected to revolutionize urban traffic systems, promising safer roads, enhanced mobility, and increased traffic efficiency (Duarte and Ratti, 2018). Before reaching fully automated traffic, AVs will co-exist with human road users in the foreseeable future (Calvert et al., 2017). In such a mixed-traffic environment, AVs introduce novel interactions between participants in traffic and may further change the collective characteristics of traffic flow (Yu et al., 2021). Therefore, evaluating the safety and efficiency impacts of AVs on urban traffic is essential.

* Corresponding author.
E-mail address: g.li-5@tudelft.nl (G. Li).

https://doi.org/10.1016/j.trc.2024.104802
Received 3 November 2023; Received in revised form 25 June 2024; Accepted 2 August 2024
Available online 18 August 2024

For urban traffic, intersections are safety-critical scenarios (Shirazi and Morris, 2016) and are often traffic efficiency bottlenecks (Briz-Redón et al., 2019). Traffic flows intersect and re-distribute at intersections, thus bringing some of the most diverse and complex conflicts. How these conflicts are resolved has significant impacts on both driving safety and traffic efficiency. Conflict resolution refers to a process by which two or more road users cooperate to prevent accidents or disruptions. Unsuccessful conflict resolution can lead to near-collisions or even accidents. Meanwhile, conflict resolution efficiency influences intersection capacity (Brilon and Wu, 2001; Li et al., 2024). In mixed traffic, conflict resolution between AVs and human road users encapsulates both the control of AVs and the behaviour adjustments of humans. These two intertwined sides together determine the safety and efficiency outcomes of a conflict resolution process. Therefore, investigating behaviour changes of human road users and further quantifying the safety and efficiency measures of AV-involved conflicts at intersections is critical for comprehensively understanding the implications of automated driving on mixed traffic flow.

In this paper, we aim to study two research questions using real-world AV road test data:

- *What behaviour differences do human drivers and pedestrians demonstrate when resolving conflicts with AVs at intersections, compared to human traffic?*
- *What are the AV impacts on the safety measures of conflicting human road users and the efficiency measure of traffic flow at intersections?*

Answering these two questions requires a reliable conflict resolution dataset and appropriate measures of safety and efficiency performances.

1.1. AV-involved conflict resolution data

Reliable motion data of AV-involved and AV-free interactions is a prerequisite for assessing the impacts of AVs. In principle, such data can be collected or generated in several ways, such as experiments in driving simulators (like CARLA Dosovitskiy et al., 2017), controlled field tests (e.g., Zhao et al., 2020), and AV road tests. Recently, with the fast development of sensing technology, employing AVs to gather trajectory data in naturalistic driving environments has shown its great advantages in realisticness, diversity, and scalability. Several companies have made their road test datasets public, such as Waymo (Sun et al., 2020), Argoverse (Chang et al., 2019; Wilson et al., 2023), Lyft (Houston et al., 2021), Shifts (Malinin et al., 2021), and nuScenes (Caesar et al., 2020). These open datasets were initially released for motion prediction, which greatly expedites studies on interactions between AVs and human road users. However, the open data are insufficient for more specific assessments. For instance, driving scenarios are unlabelled, not all involve conflicts, and trajectories have anomalies and noise. Thus, categorizing datasets into well-annotated subsets and improving data quality is crucial for evaluating AV impacts in specific scenarios.

Currently, most scenario-specific open AV datasets focus on car following. The OpenACC dataset (Makridis et al., 2021) stands out as one of the earliest real-world field test collections, which encompasses different vehicle models and controlled leading speed profiles for comparison studies. Hu et al. (2022) processed a car-following dataset from the Waymo v1.0 motion dataset, providing 196 AV-following-HV pairs, 274 HV-following-AV pairs, and 1032 HV-following-HV pairs. Li et al. (2023) extracted a large dataset from Lyft level-5. The resulting data contains 29k+ HV–AV pairs and 42k+ HV–HV pairs. These datasets expedite comprehensive evaluations of the impact of AVs on longitudinal car-following.

Compared to car-following, lateral interactions are more complex, as they are two-dimensional and may involve different road user types, such as human-driven vehicles (HV), cyclists, and pedestrians. However, to the best of our knowledge, there is no open dataset focusing on AV-involved conflict resolution at intersections yet. Some drone-based datasets, such as pNEUMA (Barmpounakis and Geroliminis, 2020), INTERACTION (Zhan et al., 2019), and InD (Bock et al., 2020), provide trajectories of human road users at intersections only. The lack of well-annotated, high-quality, and diverse data hinders researchers from evaluating the impacts of AVs in lateral conflict resolutions at intersections (Curtis et al., 2021).

1.2. Safety and efficiency assessment of AV-involved conflict resolution

Besides appropriate motion data, how to measure the safety and efficiency performances of conflict resolution (or in general, interactions) is also important. This subsection briefly overviews pertinent studies and then identifies the research gap.

The safety of conflict resolution is usually assessed by Surrogate Safety Measures (SSMs), such as Time-to-Collision (TTC), Post-Encroachment-Time (PET), Deceleration Rate to Avoid Crash (DRAC), etc. The review by Wang et al. (2021) categorizes SSMs into 3 groups, namely time-based, deceleration-based, and energy-based measures. The review also summarizes the applicable scenarios for each SSM. Specifically for lateral conflicts at intersections, only PET, Gap Time (GT), and Proportion of Stopping Distance (PSD) are suitable. We refer the reader to Wang et al. (2021) for more information.

In terms of the safety impact of AVs, car-following is the most widely studied scenario due to the availability of data. For example, Hu et al. (2023) analyse the Waymo car-following dataset and report that no significant differences are observed for HVs when following an AV versus an HV, except for smaller jam spacing. Wen et al. (2022), using the same dataset, find that human drivers exhibit reduced driving volatility, shorter time headway, and longer time-to-collision (TTC) when following AVs. The authors thus argue that the rear-end collision risk is lower for HVs following AVs. Additionally, Wang et al. (2023) analyse the car-following behaviours when vehicles are approaching stopping lines or crosswalks. The results show that HVs significantly adapt their decelerating and accelerating behaviours to leading AVs.

In contrast to longitudinal interactions, lateral conflict safety of AV-involved cases is barely directly assessed through measured SSMs due to insufficient open data. Instead, behaviour modelling, simulation, and simulator experiments are used. For instance, Reddy et al. (2022) employ driving simulators to investigate AV–HV interactions at T-intersections and find that the recognizability and behaviours of AVs do influence the SSMs of humans. Virdi et al. (2019) conduct simulations and use the SSM to evaluate safety in virtual environments. Their results indicate that a low penetration rate combined with low headway can potentially create more risky conflicts at intersections. However, these approaches cannot replace real-world road tests because the diversity and complexity of human behaviours are often over-simplified. The results may be biased to the used assumptions in the simulation. Simulator experiments are costly to generate many safety-critical scenarios as well as diverse human reactions (requiring many participants). Therefore, directly analysing SSMs from real-world AV-involved conflict datasets at intersections is an important but unaddressed issue.

Compared to driving safety, traffic efficiency is a macroscopic concept, usually measured by the maximum traffic flow (capacity) on the fundamental diagram (Loder et al., 2019), intersection throughput, total travel time, or average delay (Cheng et al., 2016). As AVs have not been widely deployed, the efficiency impact of AVs is usually evaluated by simulation. Again, car-following is the focus. Lu et al. (2020) is a representative example using a pre-defined Intelligent Driver Model (IDM) to describe AV behaviours. Under this assumption, results show traffic efficiency improvement with the growth of AV penetration. Huang et al. (2023) use a similar methodology but the IDMs are calibrated from real-world AV data. They report an increase in maximum production. Calvert et al. (2017) also use a simulation-based approach to explore the influence of vehicle automation level and the penetration rate on traffic capacity. The results suggest that low-level automated vehicles in mixed traffic will initially have a small negative effect on road capacities. As for intersections, Şentürk Berktaş and Tanyel (2020) calibrate the driving model of AVs from real-world data and observe significantly lower intersection capacities in simulation. In the robotics domain, carrying out Monte-Carlo simulation against heterogeneous HVs is a usual method to assess the average passing time (Althoff and Mergel, 2011; Wang et al., 2019), and further assess the efficiency impact. For an in-depth exploration of this domain, we refer readers to Yu et al. (2021), Matin and Dia (2022) for reviews.

In comparison to simulation-based approaches, directly measuring conflict resolution efficiency at intersections can better utilize real-world data. Nevertheless, besides data shortage, there are two methodological challenges. First, microscopic efficiency measures rely on intersection layouts because determining where interactions start is indispensable. This is a controversial subject that triggers extensive debates. In the robotics domain, most studies focus on one intersection and fix the route to calculate e.g. passing time. But AV datasets are ego-vehicle-focused. Conflicts are observed at arbitrary intersections. How to find a uniform measure regardless of layouts has not been discussed. Second, most microscopic efficiency measures focus on one individual conflict but ignore continuous incoming vehicles. Thus, they cannot be related to macroscopic measures. Therefore, we need a novel conflict resolution efficiency measure that can be applied to any intersection and can further facilitate macroscopic traffic efficiency measures estimation.

In summary, the discussion above identifies two research gaps. First, there is a pressing need for an AV-involved conflict resolution dataset focusing on intersections that enables direct analysis of AV impacts and human road users' behaviour adjustments. Second, a microscopic conflict resolution efficiency measure that can be applied to arbitrary road configurations and give traffic flow estimation is missing. This paper aims to fill these two gaps and provide possible answers to the main research questions.

1.3. Contributions and outline

This paper presents a high-quality and well-annotated lateral conflict resolution dataset at urban intersections based on Argoverse-2 (Wilson et al., 2023). The safety impact of AVs is assessed through surrogate safety measures and a new measure is proposed to assess the traffic efficiency impact. The major contributions are listed below:

- Provide a guideline for AV impact assessment using motion data collected by AVs as probe vehicles.
- Propose a pipeline to extract, process, enhance and annotate an AV-involved lateral conflict resolution trajectory dataset at intersections from open Argoverse-2.
- Assess the safety and efficiency impacts of autonomous vehicles in different types of lateral conflicts at intersections through selected measures.
- Investigate if human road users behave differently when resolving lateral conflicts with an AV and an HV at intersections.

The rest of this paper is organized as follows. Firstly, Section 2 describes the used data source and Section 3 gives a guideline for impact assessment using AV motion data. Next, Section 4 describes how to prepare the conflict resolution dataset. Using this derived dataset, Section 5 analyses and compares the safety and efficiency performances of AV-involved and AV-free cases. Then, Section 6 summarizes the main findings and discusses the limitations. Finally, Section 7 concludes and proposes several further research directions.

2. Dataset description

Argoverse-2 motion data. To prepare a lateral conflict resolution dataset that comprises both AV-involved and AV-free cases at intersections, choosing an appropriate data source that meets the following requirements is necessary:

- AVs must be labelled in the dataset.
- AVs have a recognizable appearance that distinguishes them from HVs.

Table 1
Comparison of open AV motion datasets.

Properties	Waymo (Sun et al., 2020)	Argoverse (Chang et al., 2019)	Lyft (Houston et al., 2021)	Shifts (Malinin et al., 2021)	nuScenes (Caesar et al., 2020)	Argoverse-2 (Wilson et al., 2023)
AV labelled	✗	✓	✓	✗	✓	✓
Automated mode	✓	?	✓	?	✗	✓
Conflict diversity	✓	✓	✗	✗	✗	✓
Track duration (s)	9	5	25+	10	8	11
Frequency (Hz)	10	10	10	5	2	10

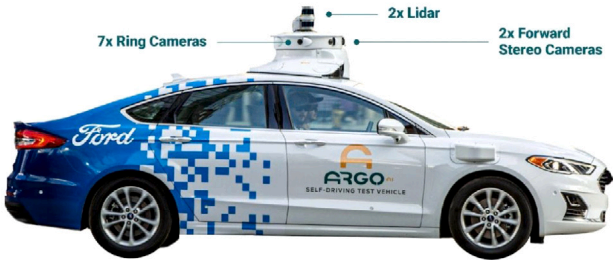

Fig. 1. The appearance of the automated driving test vehicles in Argoverse 2.

- The automated driving mode of AVs must be on.
- Diverse lateral conflicts at intersections must be included.
- The tracking duration should be long enough to cover the entire conflict resolution process.
- The sampling frequency should be high enough to provide behavioural details.

Table 1 compares 6 widely-used open AV motion datasets. The recently-released Argoverse-2 (Wilson et al., 2023) motion dataset fulfils all these requirements. An automated fleet collects this dataset in 6 cities. The appearance of automated driving test vehicles is shown in Fig. 1.[1] The logo and the sensors mounted on the top of vehicles make them recognizable. Argoverse-2 contains 250k non-overlapping scenarios specifically focusing on safety-critical long-tailed situations. The duration of each scenario is 11 s with a uniform 10 Hz sampling rate (0.1 s time interval). High-definite maps, including vectorized maps and drivable areas, are also provided. Considering all these properties, this paper chooses Argoverse-2 as the data source.

3. Guideline for AV impact assessment using road test data

Compared to vehicle trajectories collected by road-side cameras or drones (e.g., NGSIM US Department of Transportation – FHWA (2008) and pNEUMA Barmpounakis and Geroliminis (2020)), motion data collected by instrumented probe vehicles is distinct in two aspects. First, errors and noise of the ego vehicle and surrounding road users are heterogeneous. For the ego vehicle, its motion state can be precisely recovered from CAN Bus data. However, the motion of surrounding agents must be reconstructed from sensors (like Radar in SHRP 2 Hankey et al. (2016)). Specifically for using AVs as probe vehicles, data-driven machine learning methods are widely used to reconstruct trajectories from sensors in the industry. The derived trajectories can be internally inconsistent (for example, between positions and velocities) and result in different data quality for the AVs and their surrounding HVs. This inconsistency was reported for Waymo (Hu et al., 2022) and Lyft level-5 (Li et al., 2023). Second, as the AV follows a specific driving model which might be different from average human drivers, behaviour biases might be induced. In addition, because the number of surrounding vehicles is correlated with traffic states, the distribution of traffic states for the AV and other agents can be significantly different. Therefore, directly comparing the distribution of some measures without balancing confounders cannot give reliable findings. For example, Jiao et al. (2024) systematically investigate possible alternative explanations for observed shorter headways when HVs follow AVs. The authors argue that biases such as traffic states in the collected data, as well as driving variability and driving characteristics of the leading vehicles, can lead to the observation of shorter average headway — not necessarily human behaviour changes. In summary, using AV road test data requires particular attention to heterogeneous data quality, behaviour biases, and potential confounders in data.

To address the concerns raised above, we propose a generic guideline for impact assessment using motion data collected by AVs. This guideline shown in Fig. 2 is also the research framework of this study. It has two related parts: data preparation and impact assessment. For data preparation, pertinent scenarios are first selected from the initial dataset based on the interested

[1] This picture is from https://www.argoverse.org/av2.html

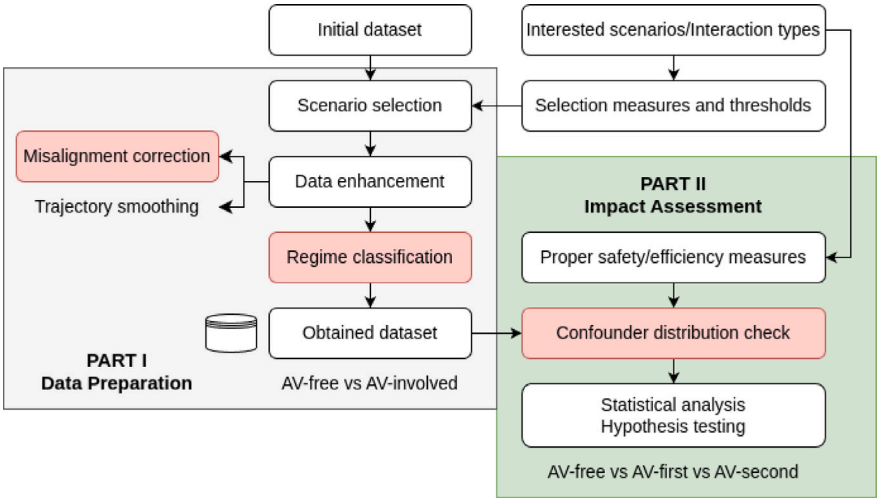

Fig. 2. Research guideline/framework. The steps in red are common pitfalls.

interaction types, selected measures, and the associated thresholds. In general, we choose relatively looser thresholds to preserve enough samples for further refinement. Notice that this selection process can combine several conditions by logic. Next, raw data are enhanced through two steps: misalignment correction and trajectory smoothing. We emphasize that ego vehicles and surrounding agents must be processed separately and different types of motion data must be aligned consistently. Then, conflict regimes are classified and annotated for in-depth analysis. This process yields a high-quality conflict resolution dataset. For impact assessment, the first step is checking if confounders are balanced between AV-involved and AV-free (human agents only) cases in each conflict regime, such as the distribution of traffic states and passing directions (from the left or right side). If yes, then statistical analysis and hypothesis testing methods can be used to assess the impacts of AVs.

The three steps marked in red are commonly ignored pitfalls that may affect the reliability of impact assessment. Following this guideline can help avoid them. Specific to this study on lateral conflict resolution at urban intersections, for safety assessment, surrogate safety measures and deceleration profiles of AV-involved cases are compared against AV-free cases. For efficiency assessment, we introduce the novel Minimum Recurrent Clearance Time (MRCT) concept and then use MRCT-based traffic flow to quantify the efficiency impact of AV on traffic. In the following sections, all steps will be explained in detail.

4. Data preparation

This section describes the data preparation process. We will exemplify how to address the aforementioned issues when using open AV motion data.

4.1. Scenario selection and raw data assessment

First of all, we define the term "conflict". In general, a conflict refers to a situation in which two or more road users create a risk of collision, disruption, or interference with each other's intended paths or movements. In this paper, we focus on two specific types of lateral conflicts at intersections:

- **Crossing:** One vehicle and another human road user with different coming lanes and different intended lanes pass the same location at a short time interval or a close distance.
- **Merging:** Two vehicles from non-adjacent lanes enter the same lane at a short time interval or a close distance.

To select scenarios that involve conflict *resolution* behaviours, the following rules are further considered:

- **Behaviour constraints:** At least one of the two conflicting road users accelerates or decelerates (changes their speeds) before passing the conflict point.
- **Isolated conflicts:** No other road users pass the conflict area earlier than any of the two conflicting agents in the recorded duration.

In the definitions above, 3 key hyper-parameters must be determined to select scenarios of interest:

- **Critical distance = 8 m:** The critical distance between the centroids of two road users is set as 8 m. This value is approximately the acceptable approaching distance between laterally interacting vehicles in urban traffic (Jiao et al., 2023).

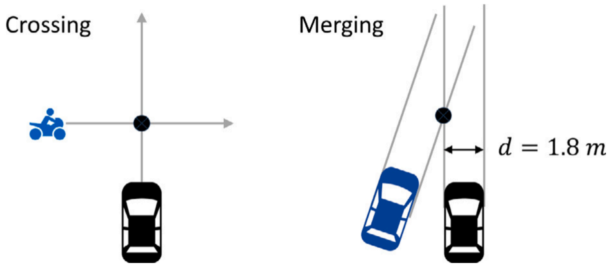

Fig. 3. Conflict points in crossing and merging. PET is defined as the time difference between two road users' centroids passing (or being closest to) the conflict point.

Table 2
Selected scenarios from Argoverse-2.

Category	Crossing (AV)	Crossing (AV-free)	Merging (AV)	Merging (AV-free)
Nb. of scenarios	5337	16 094	514	283

- **Critical time interval = 5 s:** The time interval between two road users passing the same area resembles the definition of Post-Encroachment-Time (PET). PET is originally defined as the time difference between an encroaching road user leaving the conflict area and another road user entering the same area (Allen et al., 1978). For merging, we use a variant of PET, the so-called Gap Time (GT), which is the time difference between the entries into the conflict area of two road users. However, computing PET and GT requires vehicle lengths, which are not provided in Argoverse-2. In this study, we calculate the time difference between *two road users' centroids* passing the conflict point. For crossing, the conflict point is the intersection between two trajectories. For merging, we assume vehicle width is 1.8 m and the conflict point is where the trajectories of the broadsides facing each other intersect, as illustrated in Fig. 3. For convenience, we still call it PET. Notice that this PET in merges is larger than the original PET in crossing conflicts. In the literature, the selection of critical PET for safety analysis ranges from 1.5 s (Peesapati et al., 2013) to 5 s (Shekhar Babu and Vedagiri, 2018), depending on scenarios. To preserve more samples for further refinement, we choose the largest value 5 s.
- **Minimum (de-)acceleration** = $\pm$ 1 m/s^2: At least one of the road users have a maximum absolute value of acceleration larger than 1 m s^{-2} before reaching the conflict point. This assumption is rooted in the belief that, if neither of the two users changes speed, there are no conflict resolution behaviours at all.

Consequently, we select scenarios with [PET < 5 s **or** a minimum distance < 8 m] **and** [with significant accelerations]. Table 2 lists the number of selected scenarios for AV-involved vs. AV-free crossing and merging cases.

The quality of raw data must be carefully examined before use. As discussed in Section 3, AV and other agents' trajectory qualities should be assessed separately. Argoverse-2 provides time t (0.1 s interval), location (x, y), velocity (v_x, v_y), and heading ϕ for each non-static agent. Next, we examine the internal consistency between positions and speeds. The so-called *position-based speed*, for example at moment t, is given by:

$$
\begin{aligned}
v_p &= \frac{s_{t-1,t} + s_{t,t+1}}{2\Delta t} \\
s_{t-1,t} &= \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}
\end{aligned}
\tag{1}
$$

It is compared with the provided speeds in Argoverse-2 for different road user types. Fig. 4 gives examples of the internal consistency for 4 different road users in one scenario. The comparison shows that the raw speed and position data in Argoverse-2 are **not** internally consistent. The first and the last 1.5 s of positions are not correctly processed. The position-based speed has sharp drops at the beginning and the end of each 11 s scenario. The drop magnitude is always around half of the nearest peak value. This error can be produced by using improper processing methods, such as fitting a polynomial with boundary constraints, using over-fitted wavelet denoising smoothers, or just using a pre-trained neural network. The same inconsistency is observed for **every** non-static agent in Argoverse-2, which is problematic for further analysis. Meanwhile, we also observe that, in the middle of each trajectory, the internal consistency is high for AVs, but relatively low for other agents.

The observed internal inconsistency brings a critical question: *Speed or position, which one should we trust?* We notice that given speeds are noisier than position-based speeds, especially for AVs. For surrounding agents, some missing speed data was filled by 0 (this is also reported in Lyft level-5 raw data (Li et al., 2023)). Given speeds do not have strange drops but only isolated outlier segments induced by 0-value padding and filled missing values. Therefore, we infer that *provided speeds are more reliable than positions.* This is the foundational assumption for the following data enhancement procedure. Next, we use the provided speed data as references to reconstruct every agent's motion data in selected scenarios and thus eliminate the internal inconsistency.

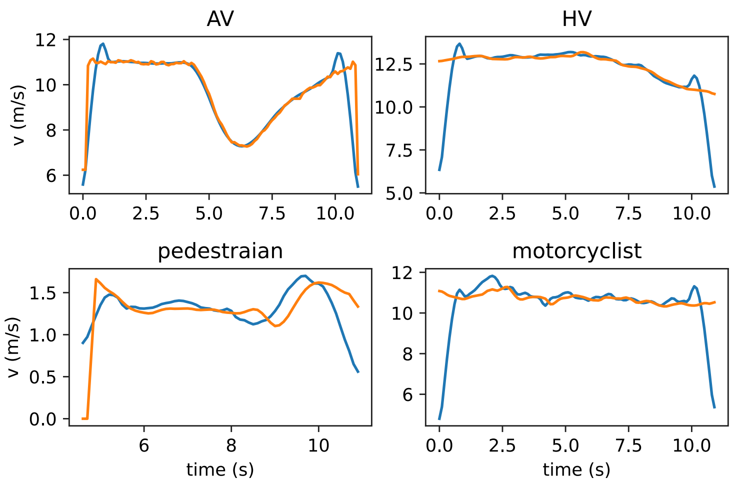

Fig. 4. An example: comparison of given speeds and position-based speeds for 4 different types of road users in one scenario.

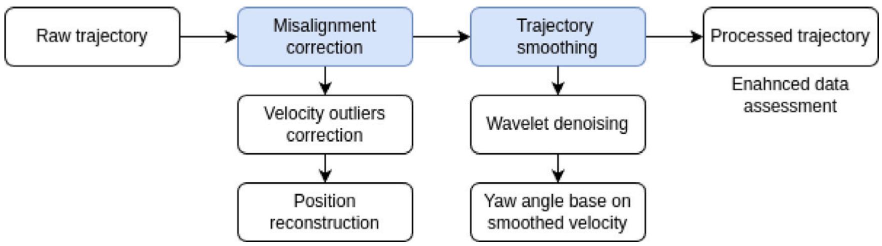

Fig. 5. Trajectory processing and enhancement.

4.2. Data enhancement

The data enhancement follows the two steps presented in Fig. 2. Detailed steps are shown in Fig. 5

Misalignment correction. The first step is eliminating the inconsistency between position and speed data. We first apply velocity outliers correction. At moment t, the speed $v(t)$ is an outlier if the absolute value of acceleration is abnormally high, or any speed value in the neighbouring time interval is filled with 0 (which means missing values):

$$\begin{aligned}&|\frac{dv(t)}{dt}| > 10\,\mathrm{m\,s^{-2}} \quad \text{or}\\&\exists\ t^* \in [t-0.3\,\mathrm{s}, t+0.3\,\mathrm{s}],\ \text{s.t.}\quad v(t^*)=0\end{aligned} \tag{2}$$

An empirical threshold of $0.3\,\mathrm{s}$ is chosen because almost all outliers are isolated points or short segments lasting less than $0.3\,\mathrm{s}$ in Argoverse 2. These outliers are thus fitted using observations within the nearest $1\,\mathrm{s}$ window by a constant-jerk polynomial model. The corrected speed is denoted as $\hat{v}(t)$. The second step is position reconstruction. For AVs, because internal consistency is high, only the first and the last $1.5\,\mathrm{s}$ segments of position data are reconstructed from the corrected velocity $\hat{v}$ by integration using Simpson's rule (Filon, 1930). For other road users, because their positions and speeds are inconsistent throughout the entire trajectory (see Fig. 4), positions are completely reconstructed. We keep the main trend of the given trajectory geometry but optimize the segmentation of the spline to minimize the total internal inconsistency. However, not all lost information can be corrected in this way. In the following situations, we do not enhance trajectories but preserve the raw data, even if the quality is poor. Note that these rules are only applied to non-conflicting background road users, which do not affect impact assessment.

- The duration is shorter than $5\,\mathrm{s}$. The unreliable percentage is too high so we cannot correct the data.
- The length of the trajectory is shorter than $8\,\mathrm{m}$. The perception fluctuation is too large.
- The trajectory length inconsistency is larger than $2\,\mathrm{m}$, which means the perception error is too large.

Trajectory smoothing. Next, the aligned consistent velocity data is smoothed, from which we derive accelerations and heading directions. For AVs, considering that speeds are noisy, the wavelet denoise method (Pan et al., 1999) is used to smooth the speed and derive acceleration by computing gradient (Hu et al., 2022). The noise level is set to $\sigma = 0.5\,\mathrm{m\,s^{-1}}$ by trial-and-error. The wavelet mode is the Daubechies family with 6 vanishing moments ('db6'). The soft threshold method is used and the maximum wavelet decomposition level is 3. For non-AV road users, corrected speeds are smooth enough already, so we directly compute acceleration.

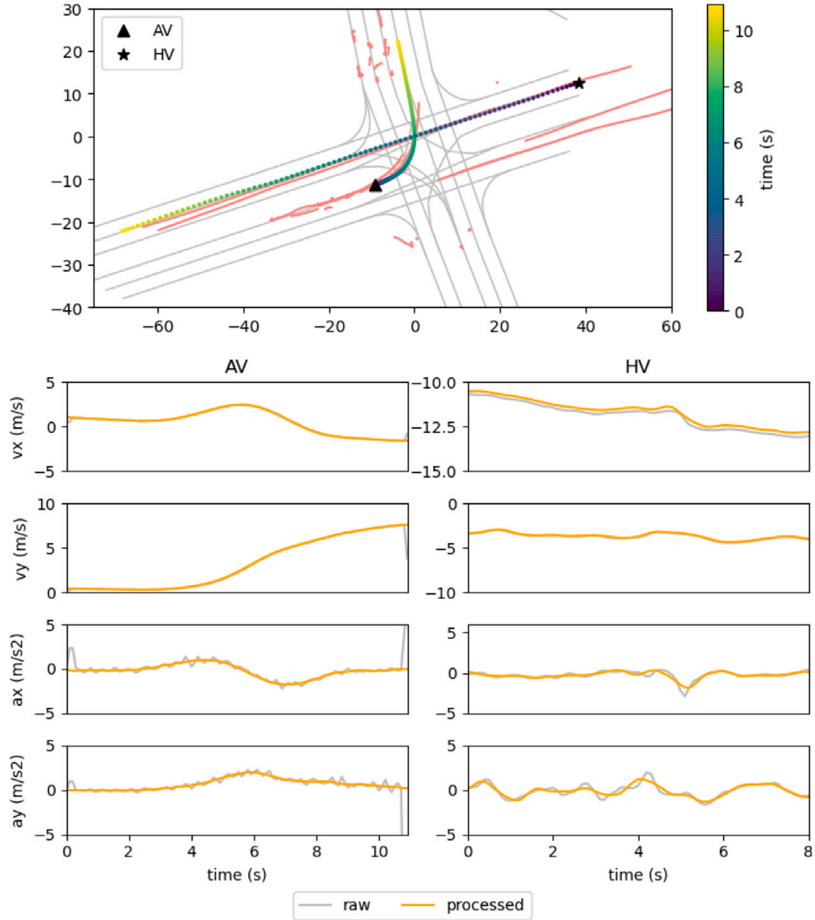

Fig. 6. An example that compares raw and processed motion data.

The heading is defined as the tangent direction along the trajectory. As an example, Fig. 6 compares raw and processed motion data for an AV and an HV in an unprotected left turn. We see that speed outliers are removed. Speed and acceleration data are smoother after processing.

Many other trajectory smoothing methods can also be used for this step. For example, Zhao et al. (2024b) propose an optimization-based 2D trajectory reconstruction and soothing method. This method is applied to the pNEUMA dataset (Barmpounakis and Geroliminis, 2020) and gives dynamics-compatible results. This method works well for high-frequency cases (25 Hz in pNEUMA) but suffers from instability in low-frequency cases (10 Hz) due to its original constant-speed dynamics assumption (between adjacent time intervals). If the dynamics model is extended to consider constant acceleration and wheelbase, it can give almost the same speed, acceleration, and yaw angle as our approach. Zhao et al. (2024b) also summarize other trajectory smoothing methods in the literature. We refer the readers to that paper for a comprehensive review and our code reproduction[2] for comparisons.

Enhanced data assessment. Now we assess the quality of smoothed trajectories. In addition to the internal inconsistency measured by the average mean-absolute-error (MAE) between speeds and position-based speeds, three constraints on acceleration a and jerk j proposed by Punzo et al. (2011) are also considered:

- $a \in [-8, 5]\ \mathrm{m/s^2}$
- $j \in [-15, 15]\ \mathrm{m/s^3}$
- The jerk's sign cannot reverse more than once in 1 s (*Jerk Sign Inversion*, JSI).

Violating these constraints is considered an "anomaly". Table 3 presents the speed inconsistency (Δv) and the duration percentages of the 3 types of anomalies for vehicles. Raw data is compared with enhanced data. Pedestrians' and cyclists' motion is more flexible so we do not assess their data quality here. The result clearly shows that the quality of the enhanced data is

[2] https://github.com/RomainLITUD/conflict_resolution_dataset

Table 3
Anomaly assessment for vehicle trajectories.

	Δv ($\mathrm{m\,s^{-1}}$)	Acc %	Jerk %	JSI %
AV-enhanced	0.01	0.01	0.01	0.3
HV-enhanced	0.30	0.01	0.03	0.9
AV-raw	0.32	0.21	0.23	83.6
HV-raw	3.30	0.14	0.19	13.9

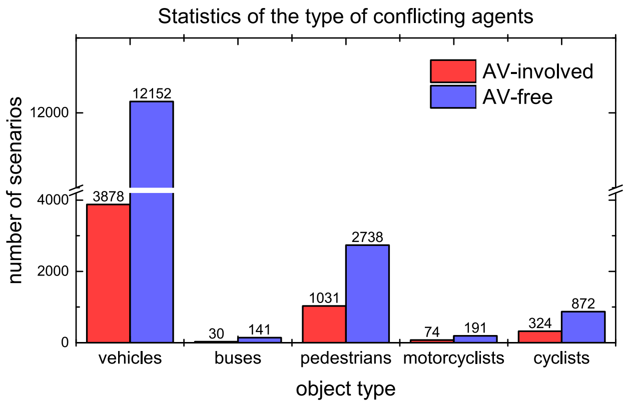

Fig. 7. Distributions of different types of conflicting agents (with an AV or an HV) in crossing conflicts.

significantly improved. After enhancement, motion data are internally consistent and better conform to vehicle dynamics constraints. The enhanced data is ready for behavioural analysis.

4.3. Conflict regime classification

In this subsection, conflict regimes are further classified and annotated to support deeper studies on each type of conflict. We first show the diversity of conflicting road users in the obtained dataset. Fig. 7 shows the distribution of different types of road users in the case of crossing conflicts with an AV or with an HV. We can see that diverse road users are included. Particularly, vehicle–vehicle and vehicle–pedestrian conflicts compose the majority of the dataset. Merging conflicts are not shown here because both road users are vehicles (see the definition in Section 4.1).

Considering the functionality of redistributing traffic flow for intersections, vehicle–vehicle crossing conflict regimes are classified based on the relative movement of the two agents, i.e., whether they run on parallel lanes (**P**), cross (**C**), or on opposite lanes (**O**) before and after reaching the conflict point. Consequently, 9 combinations of crossing conflicts are shown in the left figure of Fig. 8. Further, we differentiate these combinations by considering whether the second-passing vehicle is from the left or the right of the first-passing vehicle. For simplification, we use two letters and an arrow to represent these regimes. For example, **P** → **C** indicates that two vehicles were running parallel before the conflict point and crossing afterwards, and the second-passing vehicle moved from the left to the right side of the first-passing vehicle. For merging conflicts, we only annotate which side (**L** or **R**) the turning vehicle merges to the going-straight vehicle. There exist some vague cases between merging, crossing, and lane-changing, as shown in the bottom-right figure of Fig. 8. We annotate these cases as miscellaneous (**M**).

We further distinguish three types of passing order in conflict resolution:

(1) *AV-first*: An AV passing the conflict point before a human road user.
(2) *AV-second*: An AV passing the conflict point before a human road user
(3) *AV-free* or *HV*: the conflict happens between an HV and a human road user.

Fig. 9 presents the statistics of the three types of passing order for vehicle–vehicle interactions. Crossing conflicts show similar regime distributions. Around and more than 90% of crossing conflicts fall into **O**-**C**, **C**-**O**, and **C**-**C**. Unprotected left turns belong to **O**-**C** and **C**-**O**. Other types such as U-turn (**P**-**O**) are relatively rare. The regime distributions of merging conflicts are also similar. We notice that more than 60% cases are right-turn merges.

To summarize, the enhanced dataset contains highly consistent and smooth trajectory data, different types of road users, and diverse conflict regimes. This dataset is suitable for comparing the conflict resolution behaviours in AV-involved and AV-free scenarios. The dataset is publicly available on the 4.TU data platform.[3]

[3] https://github.com/RomainLITUD/conflict_resolution_dataset

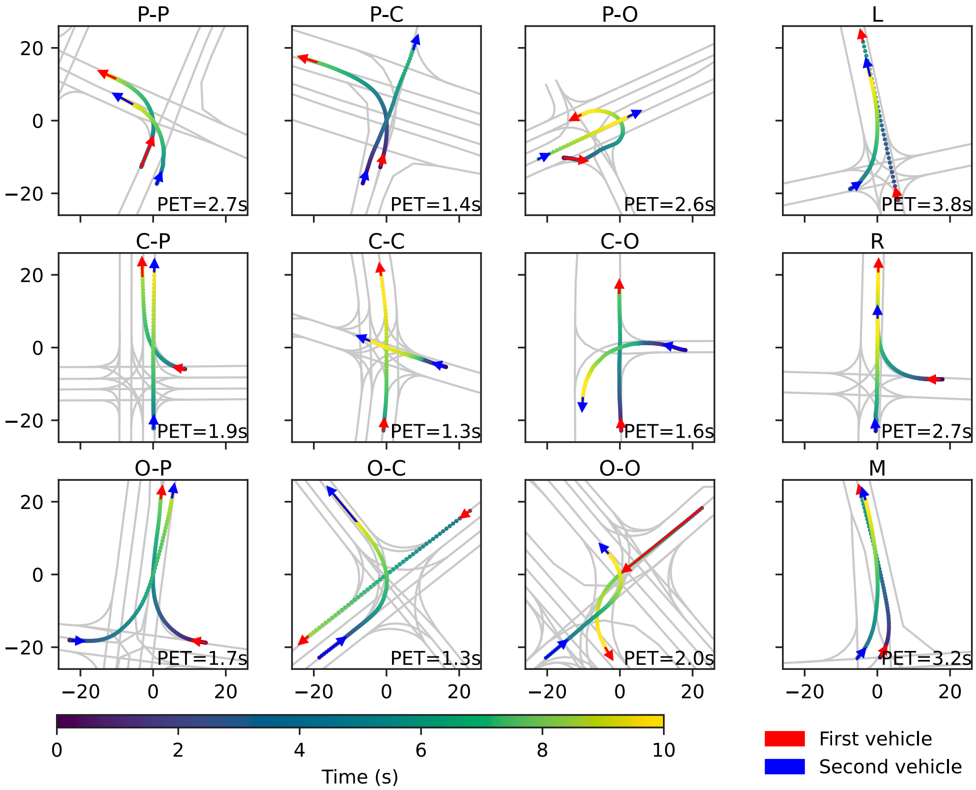

Fig. 8. Left 3 columns are 9 combinations of crossing conflict regimes and the right column shows 3 merging regimes.

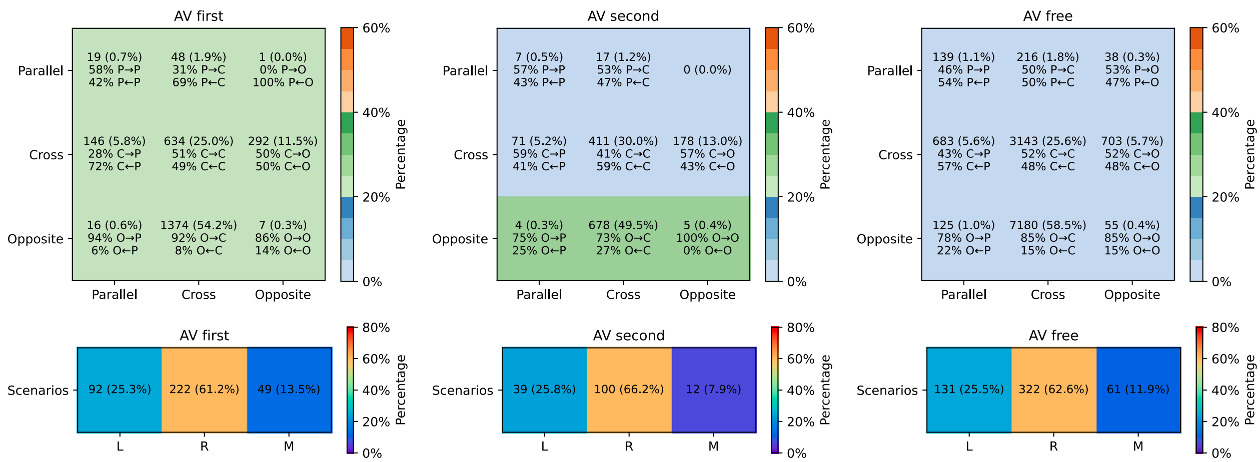

Fig. 9. Statistics of conflict regimes for different passing orders. The first row is for crossing conflicts and the second is for merging conflicts.

5. Impact assessment

This section uses the processed conflict data to conduct a preliminary analysis of the safety and efficiency impact of AVs on human behaviours and mixed traffic at intersections. We first consider several key confounders that may affect the assessment results and check if they are balanced between comparison groups. Next, we give a detailed analysis of the interested safety and efficiency measures.

5.1. Confounder distribution check

As discussed in Section 3, checking the distribution of confounders is imperative for giving credible results. Here two confounders are considered: the *passing direction* (the second passing agent comes from the left or right side of the first passing vehicle) that may

Table 4
Passing direction of pedestrians (from which side of vehicles)

Regime	HV-first		AV-first	
Direction	left	right	left	right
Percentage (%)	45.3	54.7	41.9	58.1

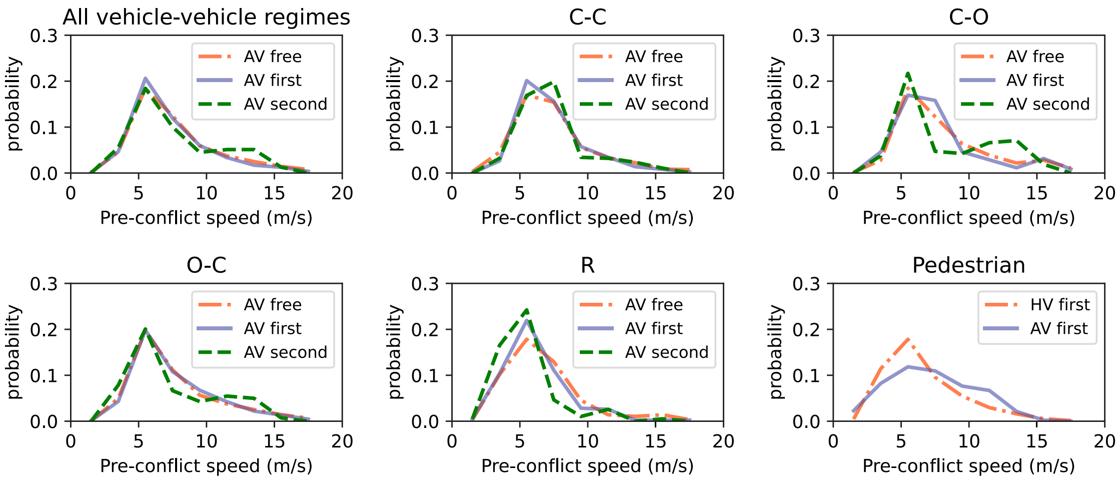

Fig. 10. Distributions of pre-conflict speed for different passing orders.

imply rule-based passing priority, and the *speed distribution* representing traffic states. Speed distribution is particularly important for traffic efficiency estimation.

For vehicle–vehicle interaction, Fig. 9 in the previous section shows that the passing direction (arrows) distributions are similar between the three passing orders. For example, for **C**-**C**, the directions of **C** → **C** and **C** ← **C** are around 50%. For merging, the percentages of **L**, **R**, and **M** are also close. For pedestrian–vehicle interactions, we only consider vehicle-first cases because we are interested in pedestrians' behaviours. Table 4 shows that the passing direction of pedestrians is also well-balanced between the two comparison groups.

The distributions of pre-conflict speed are compared in Fig. 10. Here we consider the maximum speed of the first-passing vehicle before passing the conflict point. The top-left sub-figure shows that the overall traffic state distributions are almost the same for the three passing orders in vehicle–vehicle interactions, concentrated around 5–8 $\mathrm{m\,s^{-1}}$. For the 4 vehicle–vehicle regimes with sufficient samples, i.e., C–C, C-O, O-C, and R, their speed distributions are also similar. For pedestrian–vehicle regimes, the vehicle speed distribution does not differ a lot in the two comparison groups (e.g. modes are at the speed of 5 $\mathrm{m\,s^{-1}}$).

In summary, both passing direction and traffic state are roughly balanced. Therefore, in the subsequent assessment, these confounders will be considered to not bias the comparison between AV-involved and AV-free cases.

5.2. Safety impact assessment

This section focuses on analysing the safety impact of AVs on HVs and pedestrians. Because the perception range and decision-making mechanism of AVs are different from human drivers, we do not assess the safety performance of AVs. SSMs and deceleration profiles are compared between AV-free and AV-involved cases in different conflict regimes. In these comparisons, AV-free cases serve as references. We employ the Kolmogorov–Smirnov (K–S) hypothesis test to assess if the measure distribution of AV-involved cases can be considered to follow the same distribution as the AV-free reference.

PET comparison. For vehicle–vehicle conflicts, we focus on the regimes shown in Fig. 10. Since cases with longer PET may entail no resolution processes and the distribution may be unstable around the selection threshold (5 s), we restrict the hypothesis test to PET < 4 s. For vehicle–pedestrian conflicts, we do not restrict PET ranges (because pedestrians' behaviours are more flexible) to assess whether PET samples of AV-involved cases stem from the same reference distribution as AV-free cases. All statistical outcomes are summarized in Table 5. μ and σ represent mean and standard deviation respectively.

For vehicle–vehicle conflicts, AV-first cases have almost the same PET distribution as AV-free cases in all evaluated conflict regimes. They have close mean and standard deviation, and all p-values are higher than 0.05. In contrast, AV-second cases have significantly higher average PET. All p-values are smaller than 0.01, falsifying the null hypothesis of no significant difference. Fig. 11 presents PET distributions for two conflict regimes, **O**-**C**, **C**-**C**. We do observe that the distributions of PET < 4 s for AV-free and AV-first cases are very similar, containing low-PET samples (< 2 s). But in AV-second cases, PET is always larger than 2 s. Considering

Table 5
PET distribution assessment (unit: second).

Vehicle–vehicle conflicts

Type		Crossing			Merging
Direction		**C-C**	**C-O**	**O-C**	**R**
AV free	μ	3.45	3.71	3.25	3.30
	σ	1.01	0.88	0.97	1.17
AV first	μ	3.87	3.87	3.17	3.16
	σ	1.19	0.91	1.06	1.12
	p-value	0.80	0.89	0.06	0.07
AV second	μ	4.06	4.00	4.15	3.63
	σ	0.85	0.75	0.97	0.93
	p-value	<0.01	0.01	<0.01	<0.01

Vehicle–pedestrian conflicts

HV first	μ	3.42
	σ	1.01
AV first	μ	3.88
	σ	1.48
	p-value	<0.01
HV second	μ	3.89
	σ	0.81
AV second	μ	5.55
	σ	1.22
	p-value	<0.01

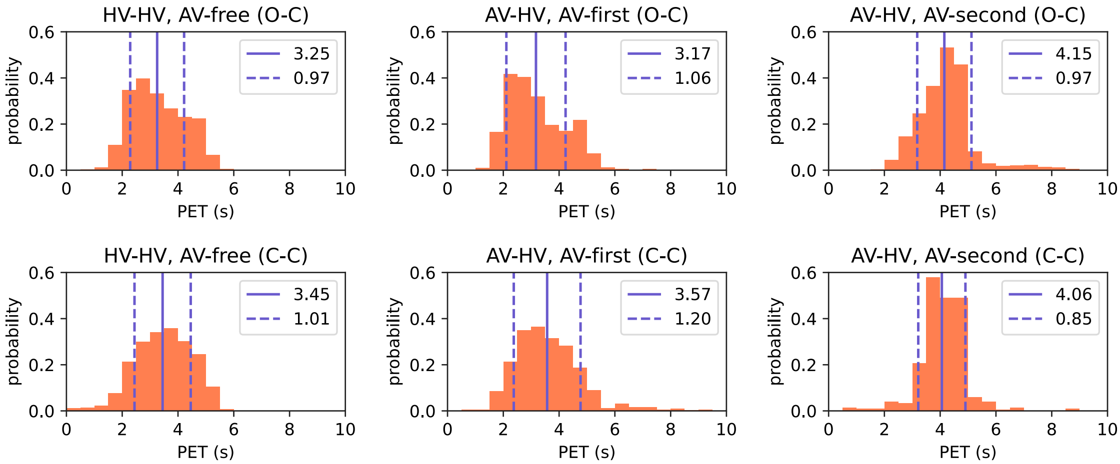

Fig. 11. Distributions of PETs for different conflict regimes. The blue solid lines mark the mean values (μ) and the dotted lines mark the standard deviation widths ($\mu \pm \sigma$).

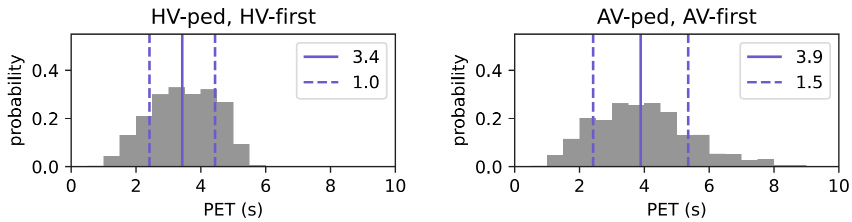

Fig. 12. Distributions of PETs for HV-first and AV-first vehicle–pedestrian crossing conflicts.

that PET is mainly determined by the second-passing vehicle, the result suggests that ArgoAI AVs drive more conservatively than HVs.

For vehicle–pedestrian conflicts (all are crossing conflicts), whether an AV is involved makes a significant difference. Both AV-first and AV-second cases have p-values much smaller than 0.01, which means statistically pedestrians behave differently in AV-involved conflicts. Fig. 12 demonstrates that compared to HV-first, the pedestrians in AV-first cases show a long-tail PET distribution at the high-value end. Considering the same critical distance threshold and traffic state distribution, this observation implies that some pedestrians become more conservative when interacting with an AV than with an HV. They tend to move slower.

Deceleration profile comparison. PSD is another applicable SSM for lateral conflicts. PSD is defined as the ratio between the remaining distance to the potential collision point and the minimum acceptable stopping distance (Allen et al., 1978):

$$\mathrm{PSD} = \frac{d_t}{v_t^2/(2|a_{max}|)} \tag{3}$$

where d_t is the longitudinal distance (along the lanes or the trajectory) to the conflict point and a_{max} is the maximum comfortable deceleration, set as $-3.35\,\mathrm{m\,s^{-2}}$ (Astarita et al., 2012). For each case, we compute how the PSD of the second-passing agent changes in time until the first-passing agent passes the conflict point. The minimum value in this process is selected. Pedestrians are not

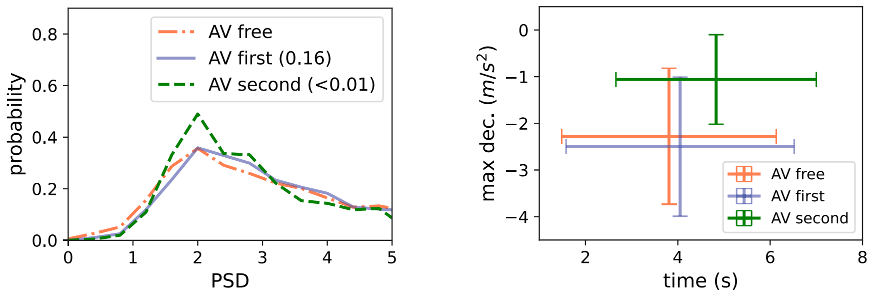

Fig. 13. *Left*: PSD distributions and p-values. *Right*: Mean value and standard deviation of both the maximum deceleration and the time gap between passing the conflict point and reaching the maximum deceleration.

Table 6
List of notations used for computing MRCT.

Notation	Meaning
t	time
t_i	moment of the vehicle i passing the conflict point
s	curvilinear distance to the conflict point, negative when approaching the conflict point
v	speed
$d_g(v)$	critical gap acceptance
$d_h(v)$	critical distance headway

subject to an acceptable deceleration. Therefore, we only consider the PSD of vehicles in all conflict regimes. The left plot in Fig. 13 presents the PSD distribution for three passing orders and the p-values of AV-first and AV-second cases (compared to the AV-free reference). Again, there is no statistically significant difference in PSD between AV-free and AV-first cases. The right plot in Fig. 13 further shows the mean and the standard deviation of the maximum deceleration before reaching the conflict point (how hard the vehicle brakes), and the time interval between the moment of reaching the maximum deceleration and the moment of passing the conflict point for the second-passing vehicle (how early the vehicle starts braking). We see that AVs tend to brake earlier and more softly than HVs. As for HVs, their deceleration profiles are highly similar whether passing through the conflict point after an AV or another HV.

In summary, by comparing PET, PSD, and deceleration profiles, we do not observe a statistically significant impact of AVs on the safety measures of conflicting human drivers. In contrast, some pedestrians demonstrate more conservative behaviours when interacting with AVs. Therefore, subject to the Argoverse-2 dataset, AVs do not cause a higher risk for human road users when resolving conflicts at intersections.

5.3. *Efficiency impact assessment*

This subsection measures the efficiency impact of AVs. Considering the limited number of merging conflicts, we only consider crossing conflicts.

5.3.1. *Minimum recurrent clearance time (MRCT)*

A conflict resolution process can be divided into two stages, before and after one agent passes the conflict point (for convenience, we call them the *pre-conflict* stage and *post-conflict* stage). In this subsection, we propose a novel measure, the so-called **Minimum Recurrent Clearance Time** (MRCT), to measure conflict resolution efficiency in these two stages. The basic idea of MRCT is, assuming that the same interaction process repeatedly happens at the same location, what is the minimum required time to clear the conflict area for the next interaction? The measure is based on the following assumptions:

- A-1 Vehicles from two conflicting streams pass the conflict point alternately.
- A-2 Each interaction between two conflicting vehicles is not influenced by the previous interaction.
- A-3 The same conflict resolution process happens repeatedly in a uniform time interval around the conflict point.

Based on the three assumptions, we can derive the mathematical form of MRCT. For convenience, we use subscripts 1 and 2 to represent the 1st and 2nd passing vehicles in one conflicting vehicle pair, respectively. We use superscript _' to represent the current pair of vehicles and without _' means the previous pair of vehicles. The 4 vehicles in the adjacent two interactions are denoted as V1, V2, V1', and V2'. The meaning of the used notations are listed in Table 6.

Firstly, assumption A-1 implies that V1' and V2' are subjected to the trajectories of V1 and V2. The relationship between the vehicles on the same stream is car-following so we define a critical distance headway $d_h(v)$. From the perspective of one vehicle, it is also subjected to the vehicles from another stream. So, we also define a critical gap acceptance $d_g(v)$. If the distance headway and the gap are both longer than the predefined critical values, we say that the current interaction is not influenced by the

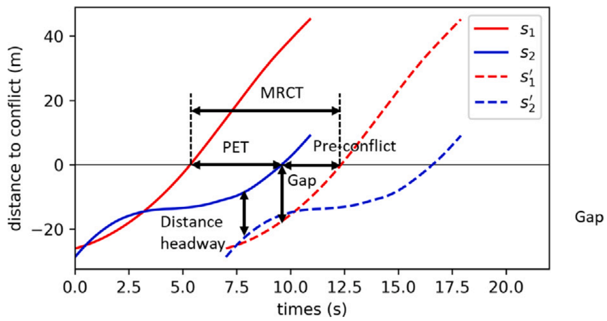

Fig. 14. Illustration of concepts for computing MRCT.

previous interaction (assumption A-2). Further, according to assumption A-3, if the time interval between two consecutive interaction processes is Δt, then the distance and speed of V1' and V2' can be expressed by the motion of V1 and V2 as follows:

$$\begin{aligned} s_i'(t) &= s_i(t - \Delta t) \\ v_i'(t) &= v_i(t - \Delta t) \end{aligned} \tag{4}$$

For vehicle V1', the following conditions must be met:

$$\begin{aligned} s_1(t) - s_1'(t) &\geq d_h(v_1'(t)), \quad \forall t \leq t_1 \\ s_1'(t_2) &\geq d_g(v_1'(t_2)) \end{aligned} \tag{5}$$

For vehicle V2', we have the constraints:

$$s_2(t) - s_2'(t) \geq d_h(v_2'(t)), \quad \forall t \leq t_2 \tag{6}$$

By using Eq. (4), we can write:

$$\begin{aligned} s_1(t) - s_1(t - \Delta t) &\geq d(v_1(t - \Delta t)), \quad \forall t \leq t_1 \\ s_1(t_2 - \Delta t) &\geq d(v_1(t_2 - \Delta t)) \\ s_2(t) - s_2(t - \Delta t) &\geq d(v_2(t - \Delta t)), \quad \forall t \leq t_2 \end{aligned} \tag{7}$$

The minimum value of Δt that satisfies Eq. (7) gives MRCT:

$$\text{MRCT} = \min \Delta t \quad \text{s.t. Eq.(7) hold} \tag{8}$$

These concepts are illustrated in Fig. 14 for a more intuitive understanding. In this way, we convert the efficiency evaluation problem into an optimization problem. If there is no conflict at all, this model degrades to the basic car-following model with a critical distance headway. It is straightforward that MRCT $\geq$ PET. If PET is interpreted as the post-conflict passing time that is mainly decided by the second-passing vehicle, MRCT−PET can be understood as the negotiation time of the first vehicle and thus the conflict resolution efficiency. We call it *pre-conflict duration*. If the negotiation takes too long (for example, both vehicles decelerate and hesitate before determining the passing order), or if the first-passing vehicle does not accelerate to pass the conflict point faster, the pre-conflict duration will be longer.

The form of $d_h(v)$ and $d_g(v)$ is calibrated from field test data or simulators. Here we use an empirical and simplified linear relationship adopted from a car-following model (Bose and Ioannou, 1999):

$$\begin{aligned} d_h(v) &= 2 \cdot v + 8 \\ d_g(v) &= \max(2 \cdot v, 8) \end{aligned} \tag{9}$$

Note that problem (8) does not have a solution if a segment of the trajectories is missing or the pre-conflict trajectory of the second-passing vehicle is too short. We omit these invalid cases.

To demonstrate the validity of the MRCT, we show three examples in Fig. 15. The first example has a long PET, but a shorter pre-conflict duration. The passing order is determined quickly because we see that the red vehicle brakes immediately without hesitation and the blue vehicle accelerates to pass the conflicting point faster. However, the red HV stands still and waits for several seconds, which causes a long PET. The second example has a long pre-conflict duration (5.7 s) and a relatively shorter PET (3.3 s). From the trajectory profiles, we see that there is a long negotiation process. Both vehicles decelerate and finally, the left-turning AV passes first instead of waiting for the straight-ahead HV (which has the priority to pass according to traffic rules). The third example is a highly efficient case of human interaction. Both PET and pre-conflict duration are short. Only the blue vehicle slightly decelerates (due to turning or conflicts) and the two vehicles pass the conflict point quickly. Therefore, MRCT can effectively capture the efficiency of conflict resolution based on experience and domain knowledge.

Fig. 16 shows the distributions of pre-conflict time duration and the associated p-values (compared to the AV-free reference). For all scenarios, the pre-conflict duration concentrates around 2.5 s. However, AV-first cases do not have any pre-conflict duration

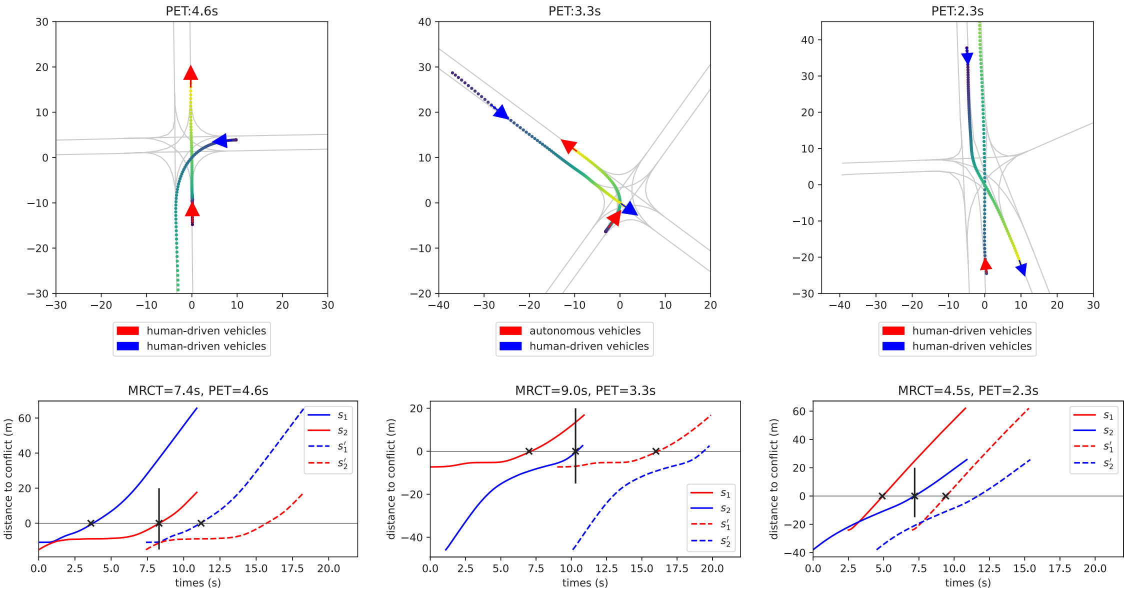

Fig. 15. Three examples of conflict resolution. The black lines separate PET from pre-conflict duration.

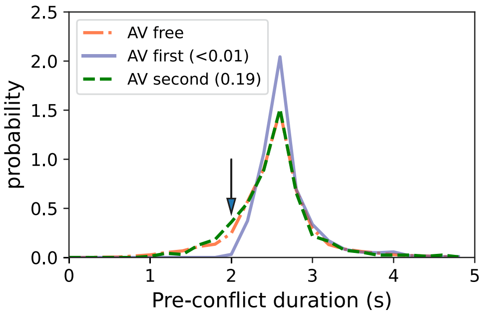

Fig. 16. Distributions of pre-conflict time duration for different passing orders. P-values to the AV-free distribution are also given.

shorter than 2 s. We do not observe a statistically significant difference in pre-conflict duration between AV-free and AV-second scenarios (p-value is 0.19). Recall that pre-conflict duration depends on the decision-making of the first-passing vehicle but PET is mainly decided by the second-passing vehicle. Thus, both shreds of evidence suggest that in this dataset, AVs drive more conservatively and their conflict resolution capability is less efficient than human drivers. When HV passes first, the pre-conflict duration is the same, whether it interacts with an AV or another HV.

5.3.2. MRCT-based traffic flow and delay time

MRCT is the time interval between passing two succeeding vehicles on the same stream. Therefore, 1/MRCT can be regarded as an indication of traffic flow (number of vehicles passing the conflict point per lane in a unit of time). For convenience, we use 3600/MRCT so the unit is the widely-used vehicle/lane/h. Table 7 summarizes statistics of this MRCT-based traffic flows for different crossing conflict regimes and also the average value over the enhanced dataset. Compared to AV-free cases, the efficiency of AV-first cases slightly decreases by around 3.5%–5.0%. Notably, the efficiency of AV-second cases is significantly lower. The average flow of AV-second unprotected left turns decreases by 16.6% in **O-C** regime and by 10.3% in **C-C** regime. This observation supports the argument that AVs and HVs have the largest efficiency performance gaps in unprotected left turns (Zhao et al., 2023), which is the most complicated conflict in automated driving.

Notice that the concept of MRCT and the traffic flow in Table 7 are calculated under the equilibrium assumption, which means the departure rate is the same as the arrival rate. If one wants to further derive macroscopic traffic efficiency measures, such as average delay time or intersection capacity, then more information (and assumptions) beyond trajectory data must be provided. For example, the penetration rate of AVs, the service time of conflicting traffic streams, the ratio of AV-second conflicts, traffic demand (determining the arrival rate), and the average discharging rate are all required to estimate the delay time. Appendix gives a toy-model example about how to estimate average delay time on a left-turning lane using MRCT.

Table 7
Mean (μ) and standard variance (σ) of MRCT-based flow (veh/lane/hour).

	AV free		AV first		AV second	
	μ	σ	μ ($\Delta\mu$%)	σ	μ ($\Delta\mu$%)	σ
C-C	643	130	611(−5.0%)	104	577(−10.3%)	75
C-O	617	106	593(−3.9%)	102	572(−7.3%)	89
O-C	675	135	651(−3.5%)	113	563(−16.6%)	80
Average over dataset	663	134	634(−4.3%)	112	573(−13.5%)	89

Using the proposed concept of MRCT, again, we do not observe statistically significant changes in pre-conflict duration for HVs when interacting with AVs. Nevertheless, the analysis of MRCT-based traffic flow demonstrates that the conservative driving style of AVs can decrease traffic efficiency. AVs cannot efficiently handle unprotected left turns like human drivers yet.

6. Main findings and discussion

Coming back to the research questions, the answers are stated as follows:

- From the in-depth analysis and comparison of surrogate safety measures, deceleration profiles, pre-conflict duration, and efficiency measures, subject to these measures and the Argoverse-2 dataset, *human drivers do not demonstrate statistically significant behaviour changes to the presence of AVs when resolving lateral conflicts at intersections*.
- Regarding surrogate safety measures, *some pedestrians behave more conservatively when interacting with AVs* compared to interacting with HVs.
- Subject to the Argoverse-2 dataset, *AVs do not increase risks for HVs and pedestrians*.
- On average, AVs in Argoverse-2 drive more conservatively than HVs, which causes *decreased traffic efficiency*, especially for unprotected left turns.

One possible explanation for human drivers' unchanged behaviours is that they only interact with AVs once at intersections. Therefore, different from car-following, human drivers do not have enough time to observe AVs' driving style and adjust their behaviours correspondingly. Some pedestrians spend more time waiting for an AV to pass first. This may be out of cautiousness or curiosity, which requires further behavioural research. With the increasing penetration rate of AVs, human road users' reactions may change in the future.

Limitations. We recognize that the conclusiveness of the above findings is restricted by the quality, type and coverage of used data, behaviour measures, considered confounders, and statistical methods. This study provides a "tangential" perspective. One can consider more potential confounders like driving variability. Another limitation is that the confounder distributions are roughly compared. Using finer confounder balancing methods in causal inference, such as sample reweighting (Desai and Franklin, 2019), can further improve the reliability of impact assessment. However, readers should note that a fine balance of confounding factors requires complete trajectories before and after the conflict to infer, for example, pre-conflict traffic states. Otherwise, short trajectory durations may cause bias.

Table 8 lists the results of research that investigates the behaviour change of human road users when interacting with AVs. Note that there are some studies utilizing micro-simulation with varying driving models and assumptions of interaction behaviour, which are not included in this table because those results cannot be compared with real-world data analyses. We see that, even for experiments in the same scenario, the results depend on used measures and data sources, and are not necessarily conclusive. This table also shows that this study is one of the first attempts to analyse AV impacts at intersections from road test data.

7. Conclusion

In this paper, we derive a high-quality lateral conflict resolution dataset from the open Argoverse-2 data. This well-labelled dataset comprises diverse AV-involved and AV-free conflict resolution regimes at intersections. Using this dataset, the safety and efficiency impacts of autonomous vehicles are assessed. The results suggest that HVs do not behave differently when interacting with an HV or an AV during conflict resolution. The conservative driving strategy of Argoverse AVs decreases traffic efficiency at intersections as measured by MRCT-based flow, particularly when they pass the conflict area after human drivers in unprotected left turns.

Based on these findings, we further propose several research directions. First, at intersections, vulnerable road users, such as pedestrians, are sensitive to the presence of AVs and will adjust their behaviours. Therefore, AV–pedestrian interaction cannot be completely learned from the HV–pedestrian interaction data. How to resolve this behaviour shift needs more attention in the following studies. Second, the assessment of the efficiency impact of AVs has not drawn as much attention as the functional safety of AVs. This topic requires novel methods and standards to comprehensively evaluate the traffic efficiency of mixed traffic flow from field test data, which is critical for the smooth integration of AV into human traffic. Third, the reliability of the comparative results can be further strengthened by using advanced causal reasoning methods to exclude spurious correlations caused by confounding factors. We believe that this presented dataset can help address these concerns.

Table 8
Summary of human road user behaviours in interactions with AVs (compared to with HVs).

Road user	Data source	Scenario	Used measure	Analysis result
Driver	Field experiment	Car-Following	Distance gap	Rahmati et al. (2019)[a], Mahdinia et al. (2021)[+], Zhang and Talebpour (2023)[+], Soni et al. (2022)[−]
		Merging	Critical gap	Soni et al. (2022)[a]
	Driving simulator	Car-following	Time-headway	Rad et al. (2021)[−], Stange et al. (2022)[a]
		Merging	Decision	Trende et al. (2019)[a]
		Merging	TTC	Sultana and Hassan (2024)[a]
		Merging	Passing time	Miller et al. (2022)[d]
		T-junction	Critical gap	Reddy et al. (2022)[a]
	AV road test	Car-following	Time-headway	Wen et al. (2022)[a], Hu et al. (2023)[−], Jiao et al. (2024)[d]
Pedestrian	Field experiment	Crossing	Perceived vulnerability	Hulse (2023)[c]
		Crossing	Decision	Dey et al. (2017)[d]
	AV road test	Crossing	Estimated crash risk	Lanzaro et al. (2023)[d]
	VR experiment	Crossing	Decision	Zhao et al. (2024a)[b]

[a] significant difference
[b] minor difference
[c] insignificant difference
[d] inconclusive/depending on external factors

Table A.1
List of notations for estimating delay time.

Notation	Meaning
α	penetration rate of AVs
A	arrival rate, how many vehicles arrive in a time unit
D	discharge rate, how many vehicles pass the conflict point after queuing in a time unit
β	total frequency of conflict resolution
m_a	MRCT for AVs
m_h	MRCT for HVs
T	total delay time
T_n	average delay time (per vehicle)

CRediT authorship contribution statement

Guopeng Li: Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Yiru Jiao:** Writing – original draft, Visualization, Software, Formal analysis, Data curation, Conceptualization. **Simeon C. Calvert:** Writing – review & editing, Supervision, Resources, Funding acquisition. **J.W.C. (Hans) van Lint:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Data availability

Data will be made available on request.

Acknowledgements

This research is sponsored by the NWO/TTW project MiRRORS with grant agreement number 16270 and the TU Delft AI Labs programme. We thank them for supporting this study.

Appendix. A toy model for estimating average delay time from MRCT

In this appendix, we consider a typical unprotected left-turn lane at unsignalized intersections. The notations of key parameters are summarized in Table A.1. For simplicity, we assume that left-turning vehicles always yield to the conflicting vehicles based on traffic rules.

The conflict with vehicles from the other direction is considered a fictive traffic light and thus causes delays. MRCT is the time duration for one vehicle passing such a fictive traffic light so 1/MRCT is the departure rate during conflict resolution. All relationships between cumulative vehicles and time are linear in this toy model. Fig. A.1 shows how to calculate the total delay time (the area of the shadowed region) for two sequential AV-involved and AV-free conflicts. It is easy to get the total delay time for both types of conflicts:

$$T_{\text{av}} = \frac{(Am_a - 1)(Dm_a - 1)}{2(D - A)} \tag{A.1}$$

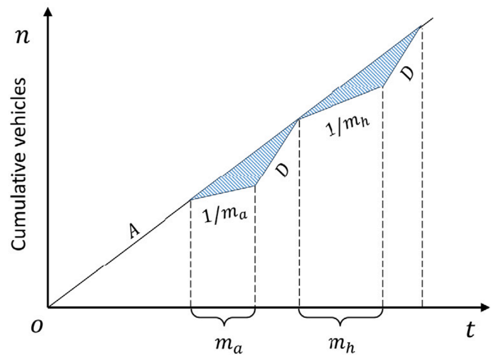

Fig. A.1. Illustration of the calculation of total delay time. The slope of each segment of the line is marked.

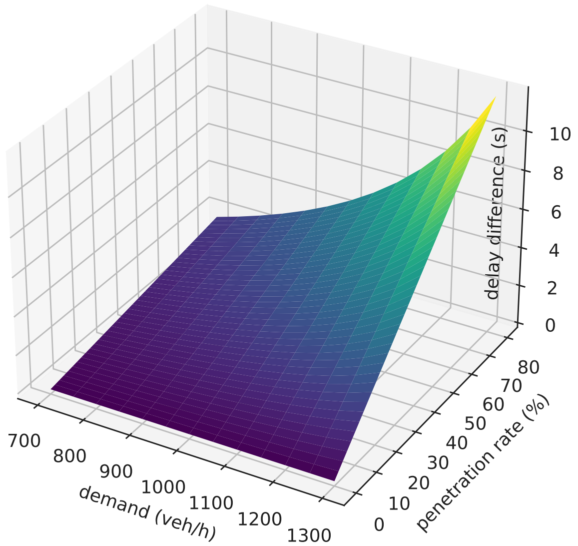

Fig. A.2. Relationship between demand (arrival rate), penetrate rate, and increased average delay time compared to AV-free cases.

$$T_{\mathrm{hv}} = \frac{(Am_h - 1)(Dm_h - 1)}{2(D - A)} \tag{A.2}$$

Therefore, the total delay in unit time is:

$$T_{\mathrm{total}} = \alpha\beta T_{\mathrm{av}} + (1-\alpha)\beta T_{\mathrm{hv}} \tag{A.3}$$

The average delay time is thus:

$$T_n = \frac{T_{\mathrm{total}}}{A} = \frac{\beta}{2(D-A)}[\alpha(Am_a - 1)(Dm_a - 1) + (1-\alpha)(Am_h - 1)(Dm_h - 1)] \tag{A.4}$$

If $m_a > m_h$, we see that the average delay time increases with the peneration rate of AVs α.

Next, we showcase a numerical result. Here we choose the average MRCT of AV-free and AV-second cases in **O**-**C** conflict regimes, $m_h = 5.53\,\mathrm{s}$ and $m_a = 6.50\,\mathrm{s}$. According to the intersection design handbook (Wilson, 2014), we further assume that $D = 1700\mathrm{veh/lane/h}$, $\beta = 100/\mathrm{h}$ (relatively rare), and different conflicts are independent of each other. The relationship between penetration rate α, arrival rate (demand) A, and the difference in average delay time for mixed traffic and AV-free traffic is shown in Fig. A.2. We see that, for the particular left-turning lane, the average delay time increases with higher penetration rates and higher traffic demand due to the conservative driving style of ArogAI AVs.

This toy model shows that macroscopic traffic efficiency measures can be derived from the microscopic MRCT. More complexity can be added, for example, considering the different discharging rates of mixed traffic flow (depending on the acceleration behaviours of AVs) or over-saturated traffic states. Further, if the structure of an intersection is given, we can employ techniques like fictive traffic lights (Chevallier and Leclercq, 2007) to estimate intersection capacity. We will not go into more depth here.

References

Allen, Brian L., Shin, B. Tom, Cooper, Peter J., 1978. Analysis of traffic conflicts and collisions. Transp. Res. Rec. 67–74.

Althoff, Matthias, Mergel, Alexander, 2011. Comparison of Markov chain abstraction and Monte Carlo simulation for the safety assessment of autonomous cars. IEEE Trans. Intell. Transp. Syst. 12 (4), 1237–1247.

Astarita, Vittorio, Guido, Giuseppe, Vitale, Alessandro, Giofré, Vincenzo, 2012. A new microsimulation model for the evaluation of traffic safety performances.

Barmpounakis, Emmanouil, Geroliminis, Nikolas, 2020. On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment. Transp. Res. C: Emerg. Technol. 111, 50–71.

Bock, Julian, Krajewski, Robert, Moers, Tobias, Runde, Steffen, Vater, Lennart, Eckstein, Lutz, 2020. The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections. In: 2020 IEEE Intelligent Vehicles Symposium. IV, IEEE, pp. 1929–1934.

Bose, Arnab, Ioannou, Petros, 1999. Analysis of traffic flow with mixed manual and semi-automated vehicles. In: Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251). Vol. 3, IEEE, pp. 2173–2177.

Brilon, Werner, Wu, Ning, 2001. Capacity at unsignalized intersections derived by conflict technique. Transp. Res. Rec. 1776 (1), 82–90.

Briz-Redón, Álvaro, Martínez-Ruiz, Francisco, Montes, Francisco, 2019. Spatial analysis of traffic accidents near and between road intersections in a directed linear network. Accid. Anal. Prev. 132, 105252.

Caesar, Holger, Bankiti, Varun, Lang, Alex H., Vora, Sourabh, Liong, Venice Erin, Xu, Qiang, Krishnan, Anush, Pan, Yu, Baldan, Giancarlo, Beijbom, Oscar, 2020. nuscenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11621–11631.

Calvert, Simeon C., Schakel, Wouter J., Van Lint, J.W.C., 2017. Will automated vehicles negatively impact traffic flow? J. Adv. Transp. 2017.

Chang, Ming-Fang, Lambert, John, Sangkloy, Patsorn, Singh, Jagjeet, Bak, Slawomir, Hartnett, Andrew, Wang, De, Carr, Peter, Lucey, Simon, Ramanan, Deva, et al., 2019. Argoverse: 3d tracking and forecasting with rich maps. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8748–8757.

Cheng, Cheng, Du, Yuchuan, Sun, Lijun, Ji, Yuxiong, 2016. Review on theoretical delay estimation model for signalized intersections. Transp. Rev. 36 (4), 479–499.

Chevallier, Estelle, Leclercq, Ludovic, 2007. A macroscopic theory for unsignalized intersections. Transp. Res. B 41 (10), 1139–1150.

Curtis, Carey, McLeod, Sam, Hultén, John, Pettersson-Lofstedt, Fredrik, Paulsson, Alexander, Hedegaard Sørensen, Claus, 2021. Knowledge for policy-making in times of uncertainty: the case of autonomous vehicle model results. Transp. Rev. 41 (4), 478–498.

Desai, Rishi J., Franklin, Jessica M., 2019. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. bmj 367.

Dey, Debargha, Martens, Marieke, Eggen, Berry, Terken, Jacques, 2017. The impact of vehicle appearance and vehicle behavior on pedestrian interaction with autonomous vehicles. In: Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct. pp. 158–162.

Dosovitskiy, Alexey, Ros, German, Codevilla, Felipe, Lopez, Antonio, Koltun, Vladlen, 2017. CARLA: An open urban driving simulator. In: Conference on Robot Learning. PMLR, pp. 1–16.

Duarte, Fábio, Ratti, Carlo, 2018. The impact of autonomous vehicles on cities: A review. J. Urban Technol. 25 (4), 3–18.

Filon, Louis Napoleon George, 1930. III.—On a quadrature formula for trigonometric integrals. Proc. Roy. Soc. Edinburgh 49, 38–47.

Hankey, Jonathan M., Perez, Miguel A., McClafferty, Julie A., 2016. Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets. Technical report, Virginia Tech Transportation Institute.

Houston, John, Zuidhof, Guido, Bergamini, Luca, Ye, Yawei, Chen, Long, Jain, Ashesh, Omari, Sammy, Iglovikov, Vladimir, Ondruska, Peter, 2021. One thousand and one hours: Self-driving motion prediction dataset. In: Conference on Robot Learning. PMLR, pp. 409–418.

Hu, Xiangwang, Zheng, Zuduo, Chen, Danjue, Sun, Jian, 2023. Autonomous vehicle's impact on traffic: empirical evidence from Waymo Open Dataset and implications from modelling. IEEE Trans. Intell. Transp. Syst..

Hu, Xiangwang, Zheng, Zuduo, Chen, Danjue, Zhang, Xi, Sun, Jian, 2022. Processing, assessing, and enhancing the waymo autonomous vehicle open dataset for driving behavior research. Transp. Res. C 134, 103490.

Huang, Yan, Ye, Yingjun, Sun, Jian, Tian, Ye, 2023. Characterizing the impact of autonomous vehicles on macroscopic fundamental diagrams. IEEE Trans. Intell. Transp. Syst..

Hulse, Lynn M., 2023. Pedestrians' perceived vulnerability and observed behaviours relating to crossing and passing interactions with autonomous vehicles. Transp. Res. F: Traffic Psychol. Behav. 93, 34–54.

Jiao, Yiru, Calvert, Simeon C., van Cranenburgh, Sander, van Lint, Hans, 2023. Inferring vehicle spacing in urban traffic from trajectory data. Transp. Res. C 155, 104289.

Jiao, Yiru, Li, Guopeng, Calvert, Simeon C., van Cranenburgh, Sander, van Lint, Hans, 2024. Beyond behavioural change: Investigating alternative explanations for shorter time headways when human drivers follow automated vehicles. Transp. Res. C 164, 104673.

Lanzaro, Gabriel, Sayed, Tarek, Fu, Chuanyun, 2023. A comparison of pedestrian behavior in interactions with autonomous and human-driven vehicles: an extreme value theory approach. Transp. Res. F: Traffic Psychol. Behav. 99, 1–18.

Li, Guopeng, Jiao, Yiru, Knoop, Victor L., Calvert, Simeon C., van Lint, J.W.C., 2023. Large car-following data based on lyft level-5 open dataset: Following autonomous vehicles vs. Human-driven vehicles. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 5818–5823.

Li, Guopeng, Li, Zirui, Knoop, Victor Lambert, van Lint, J.W.C., 2024. How far ahead should autonomous vehicles start resolving predicted conflicts? Exploring uncertainty-based safety-efficiency trade-off. IEEE Trans. Intell. Transp. Syst..

Loder, Allister, Ambühl, Lukas, Menendez, Monica, Axhausen, Kay W., 2019. Understanding traffic capacity of urban networks. Sci. Rep. 9 (1), 16283.

Lu, Qiong, Tettamanti, Tamás, Hörcher, Dániel, Varga, István, 2020. The impact of autonomous vehicles on urban traffic network capacity: an experimental analysis by microscopic traffic simulation. Transp. Lett. 12 (8), 540–549.

Mahdinia, Iman, Mohammadnazar, Amin, Arvin, Ramin, Khattak, Asad J., 2021. Integration of automated vehicles in mixed traffic: Evaluating changes in performance of following human-driven vehicles. Accid. Anal. Prev. 152, 106006.

Makridis, Michail, Mattas, Konstantinos, Anesiadou, Aikaterini, Ciuffo, Biagio, 2021. OpenACC. An open database of car-following experiments to study the properties of commercial ACC systems. Transp. Res. C: Emerg. Technol. 125, 103047.

Malinin, Andrey, Band, Neil, Chesnokov, German, Gal, Yarin, Gales, Mark J.F., Noskov, Alexey, Ploskonosov, Andrey, Prokhorenkova, Liudmila, Provilkov, Ivan, Raina, Vatsal, et al., 2021. Shifts: A dataset of real distributional shift across multiple large-scale tasks. arXiv preprint arXiv:2107.07455.

Matin, Ali, Dia, Hussein, 2022. Impacts of connected and automated vehicles on road safety and efficiency: A systematic literature review. IEEE Trans. Intell. Transp. Syst. 24 (3), 2705–2736.

Miller, Linda, Koniakowsky, Ina Marie, Kraus, Johannes, Baumann, Martin, 2022. The impact of expectations about automated and manual vehicles on drivers' behavior: Insights from a mixed traffic driving simulator study. In: Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. pp. 150–161.

Pan, Quan, Zhang, Lei, Dai, Guanzhong, Zhang, Hongai, 1999. Two denoising methods by wavelet transform. IEEE Trans. Signal Process. 47 (12), 3401–3406.

Peesapati, Lakshmi N., Hunter, Michael P., Rodgers, Michael O., 2013. Evaluation of postencroachment time as surrogate for opposing left-turn crashes. Transp. Res. Rec. 2386 (1), 42–51.

Punzo, Vincenzo, Borzacchiello, Maria Teresa, Ciuffo, Biagio, 2011. On the assessment of vehicle trajectory data accuracy and application to the next generation simulation (NGSIM) program data. Transp. Res. C 19 (6), 1243–1262.

Rad, Solmaz Razmi, Farah, Haneen, Taale, Henk, van Arem, Bart, Hoogendoorn, Serge P., 2021. The impact of a dedicated lane for connected and automated vehicles on the behaviour of drivers of manual vehicles. Transp. Res. F: Traffic Psychol. Behav. 82, 141–153.

Rahmati, Yalda, Hosseini, Mohammadreza Khajeh, Talebpour, Alireza, Swain, Benjamin, Nelson, Christopher, 2019. Influence of autonomous vehicles on car-following behavior of human drivers. Transp. Res. Rec. 2673 (12), 367–379.

Reddy, Nagarjun, Hoogendoorn, Serge P., Farah, Haneen, 2022. How do the recognizability and driving styles of automated vehicles affect human drivers' gap acceptance at T-intersections? Transp. Res. F: Traffic Psychol. Behav. 90, 451–465.

Şentürk Berktaş, Ecem, Tanyel, Serhan, 2020. Effect of autonomous vehicles on performance of signalized intersections. J. Transp. Eng. A: Syst. 146 (2), 04019061.

Shekhar Babu, S., Vedagiri, P., 2018. Proactive safety evaluation of a multilane unsignalized intersection using surrogate measures. Transp. Lett. 10 (2), 104–112.

Shirazi, Mohammad Shokrolah, Morris, Brendan Tran, 2016. Looking at intersections: a survey of intersection monitoring, behavior and safety analysis of recent studies. IEEE Trans. Intell. Transp. Syst. 18 (1), 4–24.

Soni, Shubham, Reddy, Nagarjun, Tsapi, Anastasia, van Arem, Bart, Farah, Haneen, 2022. Behavioral adaptations of human drivers interacting with automated vehicles. Transp. Res. F: Traffic Psychol. Behav. 86, 48–64.

Stange, Vanessa, Kühn, Matthias, Vollrath, Mark, 2022. Manual drivers' experience and driving behavior in repeated interactions with automated level 3 vehicles in mixed traffic on the highway. Transp. Res. F: Traffic Psychol. Behav. 87, 426–443.

Sultana, Taniya, Hassan, Hany M., 2024. Does recognizability of connected and automated vehicles (CAVs) platoons affect drivers' behavior and safety? Transp. Res. F: Traffic Psychol. Behav. 103, 368–386.

Sun, Pei, Kretzschmar, Henrik, Dotiwalla, Xerxes, Chouard, Aurelien, Patnaik, Vijaysai, Tsui, Paul, Guo, James, Zhou, Yin, Chai, Yuning, Caine, Benjamin, et al., 2020. Scalability in perception for autonomous driving: Waymo open dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2446–2454.

Trende, Alexander, Unni, Anirudh, Weber, Lars, Rieger, Jochem W., Luedtke, Andreas, 2019. An investigation into human-autonomous vs. human-human vehicle interaction in time-critical situations. In: Proceedings of the 12th ACM International Conference on Pervasive Technologies Related To Assistive Environments. pp. 303–304.

US Department of Transportation – FHWA, 2008. NGSIM – next generation simulation. http://www.ngsim.fhwa.dot.gov/, (Accessed: 2023-11-07).

Virdi, Navreet, Grzybowska, Hanna, Waller, S. Travis, Dixit, Vinayak, 2019. A safety assessment of mixed fleets with connected and autonomous vehicles using the surrogate safety assessment module. Accid. Anal. Prev. 131, 95–111.

Wang, Yiyun, Farah, Haneen, Yu, Rongjie, Qiu, Shuhan, van Arem, Bart, 2023. Characterizing behavioral differences of autonomous vehicles and human-driven vehicles at signalized intersections based on waymo open dataset. Transp. Res. Rec. 03611981231165783.

Wang, Yijing, Liu, Zhengxuan, Zuo, Zhiqiang, Li, Zheng, Wang, Li, Luo, Xiaoyuan, 2019. Trajectory planning and safety assessment of autonomous vehicles based on motion prediction and model predictive control. IEEE Trans. Veh. Technol. 68 (9), 8546–8556.

Wang, Chen, Xie, Yuanchang, Huang, Helai, Liu, Pan, 2021. A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. Accid. Anal. Prev. 157, 106157.

Wen, Xiao, Cui, Zhiyong, Jian, Sisi, 2022. Characterizing car-following behaviors of human drivers when following automated vehicles using the real-world dataset. Accid. Anal. Prev. 172, 106689.

Wilson, A., 2014. Handboek verkeerslichtenregelingen 2014. CROW.

Wilson, Benjamin, Qi, William, Agarwal, Tanmay, Lambert, John, Singh, Jagjeet, Khandelwal, Siddhesh, Pan, Bowen, Kumar, Ratnesh, Hartnett, Andrew, Pontes, Jhony Kaesemodel, et al., 2023. Argoverse 2: Next generation datasets for self-driving perception and forecasting. arXiv preprint arXiv:2301.00493.

Yu, Haiyang, Jiang, Rui, He, Zhengbing, Zheng, Zuduo, Li, Li, Liu, Runkun, Chen, Xiqun, 2021. Automated vehicle-involved traffic flow studies: A survey of assumptions, models, speculations, and perspectives. Transp. Res. C: Emerg. Technol. 127, 103101.

Zhan, Wei, Sun, Liting, Wang, Di, Shi, Haojie, Clausse, Aubrey, Naumann, Maximilian, Kummerle, Julius, Konigshof, Hendrik, Stiller, Christoph, de La Fortelle, Arnaud, et al., 2019. Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. arXiv preprint arXiv:1910.03088.

Zhang, Yanlin, Talebpour, Alireza, 2023. Characterizing human–automated vehicle interactions: An investigation into car-following behavior. Transp. Res. Rec..

Zhao, Jing, Knoop, Victor L., Sun, Jian, Ma, Zian, Wang, Meng, 2023. Unprotected left-turn behavior model capturing path variations at intersections. IEEE Trans. Intell. Transp. Syst..

Zhao, Xiaoyuan, Li, Xiaomeng, Rakotonirainy, Andry, Bourgeois-Bougrine, Samira, Zhu, Zishuo, Delhomme, Patricia, 2024a. Crossing roads in a social context: How behaviors of others shape pedestrian interaction with automated vehicles. Transp. Res. F: Traffic Psychol. Behav. 102, 88–106.

Zhao, Jing, Ma, Ruoming, Wang, Meng, 2024b. A behaviourally underpinned approach for two-dimensional vehicular trajectory reconstruction with constrained optimal control. Transp. Res. C: Emerg. Technol. 159, 104489.

Zhao, Xiangmo, Wang, Zhen, Xu, Zhigang, Wang, Yu, Li, Xiaopeng, Qu, Xiaobo, 2020. Field experiments on longitudinal characteristics of human driver behavior following an autonomous vehicle. Transp. Res. C: Emerg. Technol. 114, 205–224.