

TRANSPORT FINDINGS

Lane Change Inconsistencies in the highD Dataset

Zhaohan Wang¹, Yang Gao²¹ University of Sydney, ² City University of Hong Kong

Keywords: highD dataset, lane change, vehicle trajectories, data evaluation

<https://doi.org/10.32866/001c.132341>

Findings

The highD dataset is renowned for its large-scale, high-frequency data points and smooth, precise vehicle trajectories, making it a valuable resource for various traffic research applications. However, its suitability for studying lane-changing behavior is questionable due to the lack of upstream and downstream context at the recording sites. Further analysis shows inconsistency in speed gains during lane changes in opposite directions, suggesting potential limitations on its application in lane change studies.

1. Questions

The highD dataset is a large-scale naturalistic dataset capturing human-driven vehicle trajectories. It is frequently praised for its precise, high-frequency data points, which result in smooth and accurate vehicle trajectories. The mean positional errors of vehicle midpoints in both longitudinal and lateral directions are less than 3 cm (Krajewski et al. 2018). Consequently, it has been widely used in numerous research projects, particularly in microscopic traffic studies, including car-following and lane-changing studies.

Despite its high quality, the dataset's suitability for lane-changing research is questionable. Lane changes can be broadly categorized into discretionary lane changes (DLC) and mandatory lane changes (MLC). DLC mainly occurs when drivers perceive that the target lane offers better driving conditions, such as higher speeds, improved safety, or compliance with driving rules. In contrast, MLC occurs when drivers are forced to change lanes to avoid downstream obstructions or to position themselves in the appropriate lane for an upcoming maneuver (Wang, Ramezani, and Levinson 2024).

The highD dataset, which does not feature ramps or lane drops in its recordings, is primarily used for studying DLC (C. Zhang et al. 2022; Ji, Ramezani, and Levinson 2023; Y. Zhang et al. 2023). However, its recording segments are relatively short (approximately 420 m), and the specific locations of the recording sites are not disclosed. As a result, the dataset lacks upstream and downstream context, making it challenging to infer the broader trajectory conditions.

In this paper, we examine the suitability of the highD dataset for studying DLC by analyzing drivers' lane-changing actions, as well as their behaviors before and after lane changes.

2. Methods

The highD dataset consists of 60 drone recordings that are on average 17 minutes in duration. These recordings, referred to as tracks, were recorded at 6 different locations near Cologne, Germany. Location 1 is the most represented, with 37 out of the 60 tracks. As a result, our study primarily focuses on tracks from this location.

The highD dataset is largely clean and requires minimal pre-processing. The only adjustments made were shifting vehicle locations from the upper-left corner to their center points and converting the speeds in the x and y directions to a Euclidean speed. We then extracted the two traffic directions: lanes 6, 7, and 8 for the rightward (positive velocity) direction, and lanes 2, 3, and 4 for the leftward (negative velocity) direction. Lane changes were identified as the moments when a vehicle's lane value changed. We extracted a total of 10,127 lane changes from location 1, where 4,763 are from lanes 6, 7, 8, and 5,364 are from lanes 2, 3, 4.

Since discretionary lane changes (DLC) are typically assumed to occur when drivers seek better driving conditions, often for speed gains, we evaluated whether the recorded lane changes resulted in increased speed. This study focuses on short-term speed gains, for which we defined two measures: the 1-second speed gain and the 5-second speed gain. The 1-second gain is calculated as the difference between the average speed 1 second after the lane change and the average speed 1 second before the lane change. Similarly, the 5-second gain is the difference between the average speed 5 seconds after the lane change and the average speed 1 second before the lane change. These speed gains were calculated separately for the two traffic directions across all tracks at location 1.

3. Findings

The speed gains from all tracks at location 1 were aggregated for comparison. The distributions of speed gains following a lane change differ significantly between the two traffic directions, as shown in [Figure 1](#). A summary of the distribution statistics is provided in [Table 1](#).

The p-value from the two-sample t-test is less than 0.001, indicating that the difference between the two samples is statistically significant. Notably, vehicles traveling in the rightward direction tend to exhibit positive speed gains, while those traveling leftward often experience negative speed gains. DLCs are typically associated with positive speed gains, as they are assumed to occur when drivers seek better driving conditions. Therefore, the negative speed gains observed for leftward vehicles may indicate that some of these lane changes are MLCs. This interpretation is further supported by the one-sample p-values of less than 0.001, which confirm that both the positive and negative speed gains differ significantly from 0.

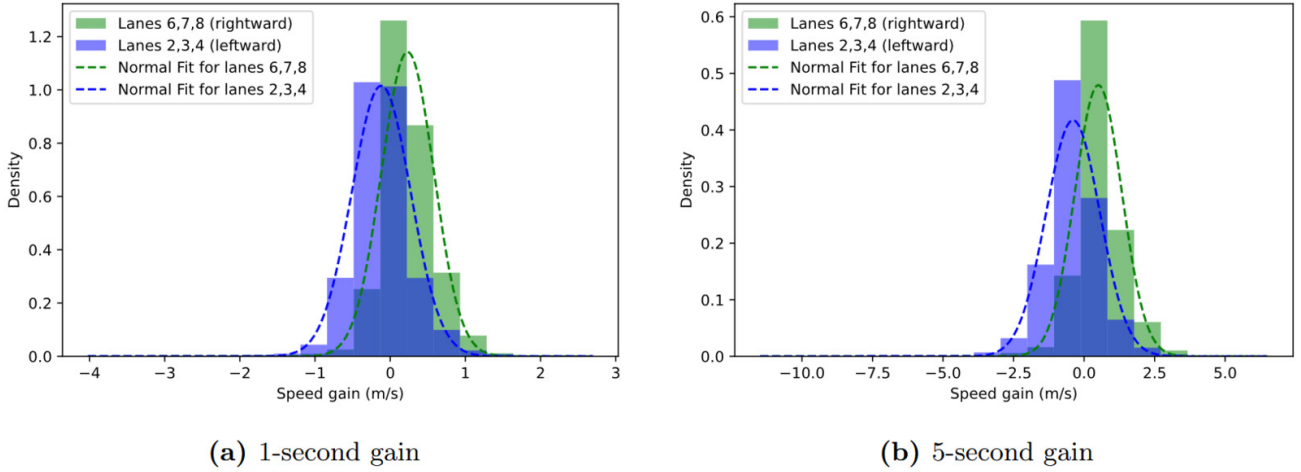


Fig. 1. Distributions of speed gains for the two directions at location 1.

Table 1. Statistics for distributions of speed gains on the two directions.

	1-second gain		5-second gain	
	Lanes 6, 7, 8	Lanes 2, 3, 4	Lanes 6, 7, 8	Lanes 2, 3, 4
2-3 (lr)4-5 Mean (m/s)	0.23	-0.12	0.50	-0.39
Std (m/s)	0.35	0.39	0.83	0.96
Sample size	4,763	5,364	4,763	5,364
p-value (null: $\mu = 0$)	<0.001	<0.001	<0.001	<0.001
2-3 (lr)4-5 p-value (null: $\mu_1 = \mu_2$)	<0.001		<0.001	

A similar pattern is observed for the 5-second speed gain, as shown in [Figure 1b](#) and summarized in [Table 1](#). With a p-value less than 0.001, the difference between the two samples is once again statistically significant. Additionally, the one-sample p-values confirm that the means are significantly different from 0. Some studies have highlighted that unnecessary DLCs can contribute to congestion by creating oscillation in the traffic (Ahn and Cassidy 2007; Gao and Levinson 2023). However, the speed drop from the leftward direction should not be attributed to this, as both would have experienced the same speed reduction. Furthermore, over a short-term horizon (1 or 5 seconds), the resulting speed reduction should not be statistically significant.

We, therefore, hypothesize that the recordings from location 1 are influenced by the presence of ramps as illustrated in the red box in [Figure 2](#). Specifically, the rightward traffic likely includes vehicles that have recently merged from an on-ramp, with the lane changes primarily being DLCs aimed at balancing the traffic across lanes. In contrast, a portion of the lane changes in the leftward traffic are likely MLCs, performed with the intention of exiting the freeway via an off-ramp. This distinction could explain the observed discrepancy between the two directions and the negative speed gains experienced by the leftward vehicles.

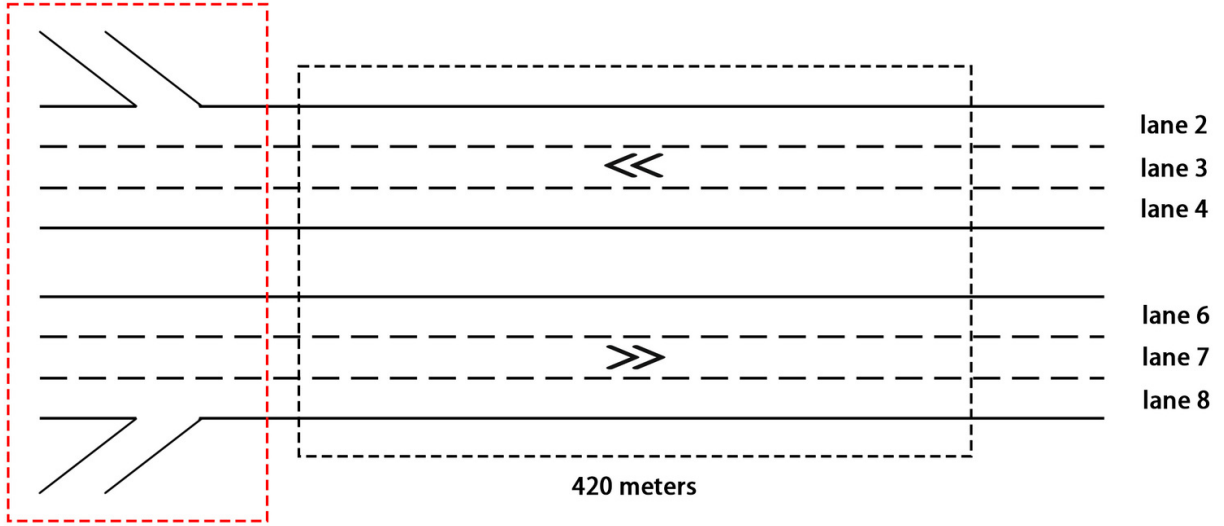


Fig. 2. The hypothesized context of location 1 from the highD dataset.

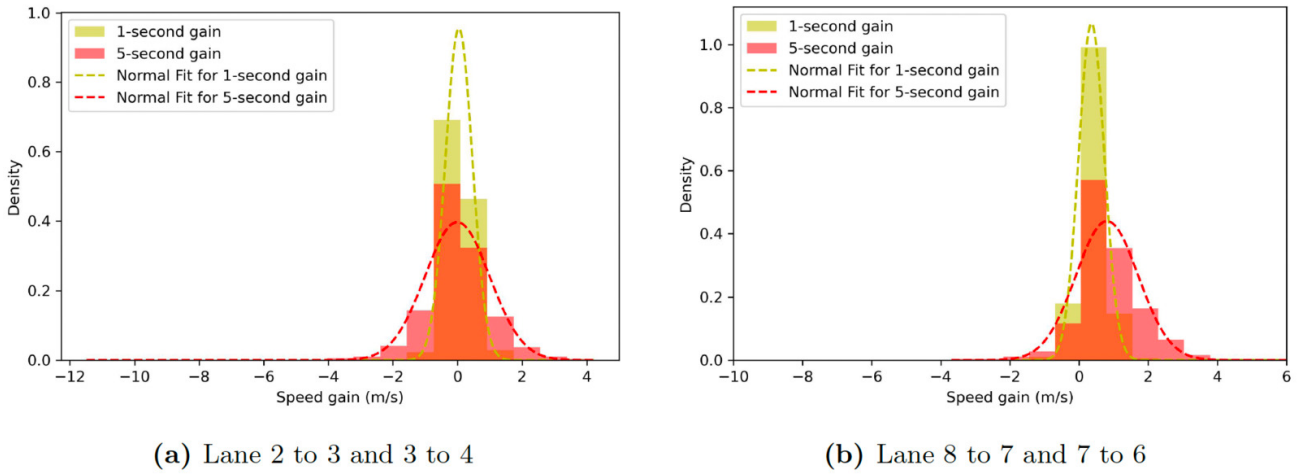


Fig. 3. Distributions of speed gains for lane changes separating the two directions.

To test this hypothesis, we computed the distributions after excluding lane changes from lane 4 to 3 and from lane 3 to 2. We believe that some of these lane changes are intended for exiting the freeway. As shown in [Figure 3a](#) and [Table 2](#), there is insufficient evidence to conclude that the means of the distributions are less than 0. In other words, after removing the potential MLCs, the speed gains are no longer statistically significantly below 0.

Similarly, for lanes 6, 7, 8, we explore the speed gains for vehicles changing lanes to the left. Both [Figure 3b](#) and [Table 3](#) suggest that the speed gains for the left-only lane changes are higher than left and right combined. This again reinforces our assumption that a portion of the vehicles have entered the freeway from an upstream on-ramp which prompts them to change lanes to increase speed. The p-values indicate the speed gains are statistically greater than 0.

Table 2. Statistics for distributions of speed gains for lane changes from lane 2 to 3 and 3 to 4.

	Mean (m/s)	Std (m/s)	Sample size	p-value (null: $\mu \geq 0$)
2-5 1-second gain	0.04	0.42	2,310	1.00
5-second gain	-0.01	1.01	2,310	0.37

Table 3. Statistics for distributions of speed gains for lane changes from lane 8 to 7 and 7 to 6

	Mean (m/s)	Std (m/s)	Sample size	p-value (null: $\mu < 0$)
2-5 1-second gain	0.36	0.37	2,198	<0.001
5-second gain	0.81	0.91	2,198	<0.001

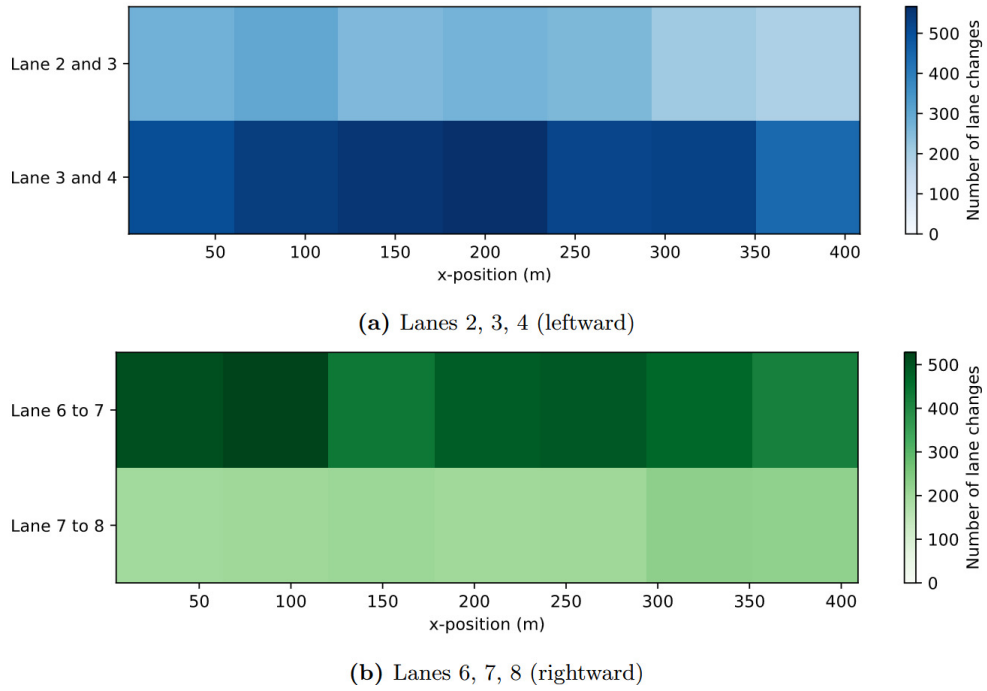


Fig. 4. Heatmap for lane change frequencies.

Another piece of evidence that ramps exist (as illustrated in [Figure 2](#)) is that the number of lane changes tends to increase toward the hypothetical ramps. This is illustrated by the heatmap of lane change frequencies in [Figure 4](#). The higher lane change activity may suggest more MLCs for vehicles approaching an off-ramp or DLCs performed by vehicles that have recently merged from an on-ramp. Additionally, the data reveal that lane change frequencies are higher in the inner lanes compared to those between the outer lanes and their adjacent lanes. Several factors could explain this phenomenon. For instance, trucks, which are subject to lower speed limits, often travel in the right-most lanes, while passenger cars may prefer to avoid sharing lanes with slower-moving trucks, leading to fewer lane changes in those areas. The percentages of lane changes for each pair of adjacent lanes are summarized in [Table 4](#).

Table 4. Summary table for lane change frequencies between all adjacent lane pairs

lane change direction	Lanes 6, 7, 8 (rightward)			
	6 to 7	7 to 8	8 to 7	7 to 6
lane change portion	0.37	0.17	0.13	0.33
lane change direction	Lanes 2, 3, 4 (leftward)			
	2 to 3	3 to 4	4 to 3	3 to 2
lane change portion	0.13	0.30	0.38	0.19

In conclusion, based on the evidence and reasoning presented above, we propose that additional pre-processing is necessary before using the highD dataset for lane change studies. Specifically, researchers should extract only those trajectories that align with the study's focus. For instance, studies on DLCs might benefit from concentrating solely on rightward traffic. Although our analysis primarily focuses on location 1, similar inconsistencies are also present at the other locations. This may indicate an inherent bias in the selection of recording locations for the dataset.

Submitted: February 21, 2025 AEDT. Accepted: March 13, 2025 AEDT. Published: March 21, 2025 AEDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-SA-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-sa/4.0> and legal code at <https://creativecommons.org/licenses/by-sa/4.0/legalcode> for more information.

REFERENCES

- Ahn, Soyoung, and Michael J. Cassidy. 2007. “Freeway Traffic Oscillations and Vehicle Lane-Change Maneuvers.” <https://trid.trb.org/View/820158>.
- Gao, Yang, and David Levinson. 2023. “Lane Changing and Congestion Are Mutually Reinforcing?” *Communications in Transportation Research* 3 (December):100101. <https://doi.org/10.1016/j.commtr.2023.100101>.
- Ji, Ang, Mohsen Ramezani, and David Levinson. 2023. “Joint Modelling of Longitudinal and Lateral Dynamics in Lane-Changing Maneuvers.” *Transportmetrica B: Transport Dynamics* 11 (1): 996–1025. <https://doi.org/10.1080/21680566.2022.2154717>.
- Krajewski, Robert, Julian Bock, Laurent Klokner, and Lutz Eckstein. 2018. “The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems.” In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2118–25. Maui, HI: IEEE. <https://doi.org/10.1109/ITSC.2018.8569552>.
- Wang, Zhaohan, Mohsen Ramezani, and David Levinson. 2024. “How Mandatory Are ‘Mandatory’ Lane Changes? An Analytical and Experimental Study on the Costs of Missing Freeway Exits.” *Transportation Research Part B: Methodological* 186 (August):102994. <https://doi.org/10.1016/j.trb.2024.102994>.
- Zhang, Chengyuan, Jiacheng Zhu, Wenshuo Wang, and Junqiang Xi. 2022. “Spatiotemporal Learning of Multivehicle Interaction Patterns in Lane-Change Scenarios.” *IEEE Transactions on Intelligent Transportation Systems* 23 (7): 6446–59. <https://doi.org/10.1109/TITS.2021.3057645>.
- Zhang, Yifan, Qian Xu, Jianping Wang, Kui Wu, Zuduo Zheng, and Kejie Lu. 2023. “A Learning-Based Discretionary Lane-Change Decision-Making Model With Driving Style Awareness.” *IEEE Transactions on Intelligent Transportation Systems* 24 (1): 68–78. <https://doi.org/10.1109/TITS.2022.3217673>.