

CHALMERS UNIVERSITY OF TECHNOLOGY AND
UNIVERSITY OF GOTHENBURG

Department of Computer Science and Engineering

A Minimal-Input Framework for Cut-In Detection and Pair-Specific Risk Analysis in Highway Trajectory Data

Traj2Rel-SFC: Trajectory-to-Relation Reconstruction and
Context Signatures for Cut-In Interactions

MASTER'S THESIS IN SOFTWARE ENGINEERING

30 credits

Authors: Sandeep Arsule
Shradha Shinde

Supervisor: Junyi Gu, Ph.D.
Examiner: Aris Alissandrakis

Gothenburg, Sweden
15 April 2026

Abstract

Highway cut-ins create new same-lane leader–follower interactions that may require the follower to brake, yet most analysis pipelines depend on dataset-provided lane and neighbour identifiers, limiting reuse across datasets. This thesis presents Traj2Rel-SFC, a minimal-input framework that reconstructs lane assignment and same-lane relations from trajectory geometry (x, y) and lane-marking metadata, detects cut-ins as explicit cutter–follower pairs, computes pair-specific surrogate safety measures (DHW, THW, TTC, DRAC), and encodes surrounding traffic context as reversible 16-bit Hilbert space-filling-curve signatures.

Evaluated on 60 *highD* recordings (with highD lane/neighbour identifiers used only as reference labels, and highD indicator columns used only once to calibrate a fixed geometry convention for pairwise SSM computation), the framework achieves mean reconstruction accuracy >0.9998 , cut-in detection $F_1 > 0.999$, and context-signature agreement of 99.49% over 1.4M stage rows. Multi-indicator analysis shows that THW, TTC, and DRAC capture complementary severity aspects; only 0.08% of events exceed the hard-braking DRAC threshold. Decision-stage features predict execution-stage THW risk with ROC-AUC 0.82 under leave-one-recording-out cross-validation, and SFC context features alone achieve AUC 0.62, showing that spatial context signatures carry standalone predictive signal, although they do not significantly improve AUC over the kinematic-only model. A small exploratory extension on 10 *exiD* recordings further shows that the same lane-reconstruction and cut-in mining core can be moved to highly interactive highway entry/exit scenes without redesigning the pipeline, although this add-on is intentionally limited and is not presented as a second benchmark.

From a software-engineering perspective, the result is a portable and auditable analysis pipeline: method inputs are explicitly restricted, intermediate relations are reconstructable and testable, and outputs can be reproduced without relying on dataset-specific derived fields.

Keywords: cut-in detection, minimal-input reconstruction, pair-specific risk analysis, surrogate safety measures, deceleration rate to avoid crash, interaction context signature, space-filling curve, risk prediction, highD, reproducibility.

Acknowledgements

We thank our supervisor for continuous feedback and for pushing us to define a clear system boundary: what information the framework is allowed to use as input, and what information must be treated as evaluation labels. We also thank the course staff and seminar group for discussions about scope, evaluation, and writing. Finally, we are grateful to the authors of the highD dataset for making the dataset available to the research community.

Contents

Abstract	iii
Acknowledgements	v
List of Figures	xi
List of Tables	xiii
Abbreviations	xv
1 Introduction	1
1.1 Motivation	1
1.2 Problem statement and approach	1
1.3 Research questions	2
1.4 Scope and delimitations	3
1.5 Contributions	3
1.6 Thesis outline	4
2 Related Work	5
2.1 The highD dataset	5
2.2 Surrogate safety measures	6
2.3 Lane-changing and cut-in analysis	6
2.4 Context encoding for interaction analysis	7
2.5 Research gap	7
3 Research Method	9
3.1 Methodological approach	9
3.2 Evaluation levels	9
3.3 Ethical considerations	10
3.4 Use of AI tools	10
4 Theory	11
4.1 Lane change vs cut-in	11
4.1.1 Lane change	11
4.1.2 Cut-in	11
4.2 Cut-in onset and stage model	11
4.3 Surrogate safety measures	12
4.3.1 Deceleration Rate to Avoid Crash (DRAC)	13
4.4 Risk label	13
4.5 Interaction context signatures (SFC encoding)	14

5	Methods	17
5.1	Framework overview	17
5.2	Data preparation and relation reconstruction	20
5.2.1	Preliminary processing and data harmonisation	20
5.2.2	Lane inference from lateral position (XY-lane)	20
5.2.3	Same-lane leader/follower reconstruction	21
5.3	Interaction mining	21
5.3.1	Lane-change detection	21
5.3.2	Cut-in extraction	21
5.3.3	Configuration	22
5.4	Safety quantification	23
5.4.1	Stage-wise features	24
5.4.2	Deceleration-based indicator: DRAC computation	24
5.5	Context signatures	24
5.6	Predictive evaluation: decision-stage risk forecasting	25
5.7	Reproducibility	27
6	Results	29
6.1	Event extraction and dataset overview	29
6.2	Reconstruction accuracy (RQ1/RQ2)	29
6.3	Execution-stage indicator distributions	30
6.4	Risk characterisation (RQ2)	30
6.5	Multi-indicator analysis (RQ2)	31
6.6	Context signature maps and interpretation	33
6.6.1	Exploratory decision-stage context archetypes	36
6.7	Predictive evaluation (RQ4)	38
6.8	Reference agreement for context signatures (RQ3)	40
6.9	Exploratory extension on <i>exiD</i>	40
7	Discussion	43
7.1	Answers to the research questions	43
7.2	Interpretation and implications	43
7.3	Threats to validity	46
7.4	Future work	47
8	Conclusion	49
	Bibliography	51
A	Terminology and highD Field Mapping	55
B	Qualitative Inspection Dashboard	57
C	Plain-Language Algorithms	59
C.1	Lane change detection (stable blocks)	59
C.2	Cut-in detection (follower identification)	59
C.3	DHW, THW, and TTC for the interacting pair	59
C.4	Reference 3×3 occupancy grid from highD neighbour ids	60

D Exploratory NGSIM Feasibility Probe	61
D.1 Probe setup and fixed assumptions	61

List of Figures

4.1	Stage model around the cut-in onset t_0 used for stage-wise feature extraction and comparison.	12
4.2	Construction of an interaction context signature	15
5.1	Overview of Traj2Rel-SFC	18
6.1	Decision-stage occupancy maps	34
6.2	Stage-wise risk-minus-non-risk occupancy differences	35
6.3	Seven decision-stage context archetypes	37
6.4	THW risk prevalence by context archetype	38
6.5	Exploratory exiD overview for the 10 selected recordings	41
6.6	Execution-stage TTC summary for the exploratory exiD extension . .	42
B.1	Schematic layout of the interactive inspection dashboard	58
B.2	Example exiD dashboard view used in the exploratory extension . . .	58
D.1	Exploratory NGSIM feasibility probe on the two freeway subsets . . .	63

List of Tables

2.1	Comparison with related lane-change and cut-in studies	8
5.1	Macro-phases and outputs of Traj2Rel-SFC.	19
5.2	Key configuration values for the final framework run.	23
6.1	Event counts for the final run (recordings 01–60).	29
6.2	Duration statistics (highD 01–60, 25 Hz).	29
6.3	Execution-stage indicator minima across all events (10,023 cut-ins).	30
6.4	Execution-stage risk summary. $\Delta v = \bar{v}_{\text{cutter}} - \bar{v}_{\text{follower}}$	30
6.5	Effect sizes between risk groups (execution stage). δ is Cliff’s delta.	31
6.6	THW threshold sensitivity (10,023 events).	31
6.7	Execution-stage DRAC distribution (finite only, $n = 3,393$).	31
6.8	DRAC threshold sensitivity (10,023 events).	32
6.9	Execution-stage TTC by THW risk group (finite TTC only, $n = 3,606$).	32
6.10	THW-risky vs TTC-threshold overlap (infinite TTC counted as not risky).	32
6.11	Top-3 extreme events by minimum execution-stage TTC (finite only).	32
6.12	Risk prediction: LORO-CV summary (stump ensemble, 34 features from the intention and decision stages, 60 held-out folds).	38
6.13	Top-5 features by Gini importance (stump ensemble, LORO-CV).	39
6.14	Feature ablation: LORO-CV (stump ensemble).	39
7.1	Contribution claims and supporting evidence.	45
A.1	Mapping between thesis terminology and highD columns (tracks file).	55
D.1	Summary of the NGSIM feasibility probe	63

Abbreviations

Abbreviation	Meaning
AUC	Area under the ROC curve
DHW	Distance headway (longitudinal gap)
DRAC	Deceleration rate to avoid crash
highD	Drone-based highway trajectory dataset used in this study [1]
LORO-CV	Leave-one-recording-out cross-validation
ROC	Receiver operating characteristic
SFC	Space-filling curve
SSM	Surrogate safety measure
THW	Time headway (gap divided by follower speed)
TTC	Time-to-collision (gap divided by closing speed, if closing)
XY-lane	Lane inference from y and lane markings

Chapter 1

Introduction

1.1 Motivation

Lane changes are among the most common interactions on highways. A *cut-in* occurs when a lane-changing vehicle moves directly in front of another vehicle in the target lane, creating a new same-lane leader–follower interaction that may require the following vehicle to brake or adjust its speed. Understanding these interactions at scale is important for safety assessment, scenario extraction for automated driving validation, and traffic modelling.

Trajectory datasets such as *highD* [1] enable large-scale analysis of cut-in interactions. However, a fundamental limitation of many analysis pipelines is their dependence on dataset-provided *derived fields*, that is, dataset-computed columns such as lane identifiers, neighbour identifiers, and pre-computed surrogate safety measures. When these fields are unavailable, inconsistently defined, or computed for a different vehicle pair than the one under study, the downstream analysis inherits hidden assumptions that are difficult to detect or control. For cut-in analysis specifically, a mismatch between the pair used for indicator computation (e.g., a generic same-lane predecessor) and the actual interacting pair (the cutter and the impacted follower) can lead to internally inconsistent risk assessments even when individual computations are correct.

1.2 Problem statement and approach

This thesis addresses the lack of a reproducible, minimal-input framework that can: (i) reconstruct lane membership and same-lane neighbour relations directly from trajectory geometry, (ii) detect cut-ins as explicit cutter–follower events, (iii) compute surrogate safety indicators for the correct interacting pair, and (iv) represent the surrounding traffic context in a compact, verifiable form.

We propose Traj2Rel-SFC, a framework with three connected layers:

1. a **trajectory-to-relation reconstruction layer** (Traj2Rel) that infers lane assignment and same-lane neighbour relations from x, y trajectories and lane-marking metadata,

2. an **explicit interaction-pair layer** that defines each cut-in as a cutter–follower event with a reproducible onset definition and pair-consistent indicator computation (DHW, THW, TTC, and DRAC), and
3. a **context-signature layer** that encodes the surrounding traffic configuration into a reversible 16-bit integer code using a Hilbert space-filling-curve traversal for compact, direction-normalised context comparison.

A central design decision is the *evaluation boundary*. The framework is evaluated on the *highD* dataset, but highD’s lane and neighbour identifiers are used only as reference labels for validation and never as inputs to the reconstruction method. For pairwise DHW/THW/TTC computation, highD indicator columns are used only once to calibrate a fixed geometry convention; after that calibration, the final cut-in features are computed from geometry alone. This separation allows the framework to be assessed as a method rather than as a dataset-analysis tool, and makes it transferable to other trajectory sources that provide positions and lane markings but not pre-computed relations.

The specific angle of this thesis is to move one step earlier than most cut-in studies. Rather than assuming that lane identifiers, neighbouring vehicles, and interaction pairs are already available, it treats those elements as reconstructable intermediates that must themselves be validated before risk indicators are interpreted. The contribution is therefore not a new risk score in isolation; it is a way to make normally hidden preprocessing assumptions explicit and auditable.

From a software-engineering perspective, this matters because data-driven safety analysis depends on how analysis pipelines are engineered, not only on which indicators are reported. Restricting method inputs, separating them from evaluation labels, and validating intermediate products all support reproducibility, portability, and testability across datasets and implementations.

1.3 Research questions

RQ1: How accurately can lane indices and same-lane neighbour relations be reconstructed from trajectories and lane markings, compared to highD’s provided identifiers?

RQ2: How consistently can cut-in events be detected from reconstructed relations, and what does explicit cutter–follower pairing contribute to risk-feature construction beyond dataset-provided indicators?

RQ3: How can spatial context around cut-ins be encoded using SFC signatures, and does the resulting representation support direction-consistent context comparison and interpretable scene grouping?

RQ4: Do the reconstructed features carry predictive value? More specifically, can features from the intention and decision stages predict execution-stage risk, and do SFC context signatures contribute useful standalone information alongside scalar indicators?

1.4 Scope and delimitations

The framework is evaluated primarily on *highD* highway recordings (drone-based trajectories at 25 Hz) using offline batch processing. To give one concrete example beyond *highD* without expanding the thesis into a second full benchmark, the thesis also includes a small exploratory extension on 10 *exiD* recordings from highway entry/exit scenes [2]. This add-on is meant only to illustrate transferability of the event-mining core to a related but more interactive highway geometry; it is not presented as evidence equivalent to the main highD evaluation. Older public freeway datasets such as NGSIM were not included in the evaluated thesis pipeline. Appendix D records a narrow internal feasibility probe on the NGSIM freeway subsets, but that probe reuses dataset-provided lane and neighbour fields after light harmonisation and is therefore not treated as evidence on the same level as either the main highD evaluation or the exploratory *exiD* extension. Full NGSIM integration remains future work because its noisier kinematics and different neighbour semantics would require a broader harmonisation effort than this thesis can justify. The thesis does not aim to infer driver intent beyond observable trajectories, does not include simulation-based re-enactment, and does not claim that any single risk threshold is universally valid. The emphasis is on reproducible scenario extraction, pair-consistent indicator computation, and reference-label evaluation.

1.5 Contributions

The main contributions are:

1. **Minimal-input relation reconstruction:** lane inference and same-lane predecessor/follower reconstruction from x, y trajectories and lane markings, evaluated frame by frame against highD reference labels (mean accuracy >0.9998).
2. **Explicit cutter–follower pairing and pair-consistent indicators:** cut-ins

are detected as explicit interacting pairs with a reproducible onset definition; DHW, THW, TTC, and DRAC are computed for the actual cutter–follower pair and interpreted stage-wise.

3. **Reversible context signatures:** a direction-normalised 3×3 occupancy grid encoded into a 16-bit Hilbert SFC signature, validated against highD reference occupancy (99.49% exact row agreement over 1.4 M stage rows).
4. **Predictive evaluation:** a LORO-CV experiment showing that decision-stage features predict execution-stage THW risk with ROC-AUC 0.82; SFC context features alone achieve AUC 0.62, showing that the framework produces predictive features beyond descriptive agreement even though the kinematic-only model remains strongest.
5. **Reproducible two-path evaluation:** the minimal-input reconstruction path and the reference-label path are executed in parallel, producing auditable outputs that can be regenerated from the same settings; exact numerical replication of stochastic substeps additionally requires the same random seeds.

Taken together, these contributions define a validated analysis pipeline rather than a single metric or classifier: the main contribution is the combination of minimal-input reconstruction, explicit interaction pairing, reversible context encoding, and a fixed evaluation boundary between method inputs and reference labels.

1.6 Thesis outline

Chapter 2 reviews the literature and motivates the framework. Chapter 3 describes the evaluation methodology. Chapter 4 defines cut-ins, stages, safety indicators, and the SFC context encoding. Chapter 5 presents the framework, the two-path evaluation design, and the prediction protocol. Chapter 6 reports the main results and the exploratory exiD extension. Chapter 7 covers implications, variability, and validity threats. Chapter 8 concludes the thesis.

Chapter 2

Related Work

This chapter reviews the research context for this thesis: the *highD* trajectory dataset that serves as the evaluation basis, surrogate safety measures and their validation, lane-changing and cut-in analysis, and compact context encodings for interaction representation.

2.1 The highD dataset

The *highD* dataset is a drone-based collection of naturalistic highway trajectories recorded at six locations on German highways [1]. It provides 60 recordings at 25 Hz, each containing per-frame vehicle states (position, velocity, acceleration), lane identifiers, and neighbour identifiers for surrounding vehicles (same-lane preceding/following and adjacent-lane neighbours). These dataset-computed columns are referred to as *derived fields* in this thesis. A mapping between the terminology used in this thesis and the relevant highD columns is provided in Appendix A. A closely related dataset from the same broader collection is *exiD*, which focuses on highly interactive highway entry/exit scenarios rather than straight motorway segments [2]; this thesis uses exiD only for a small exploratory extension, not as a second main evaluation dataset.

In this thesis, highD’s lane and neighbour identifiers are used exclusively as *reference labels* for evaluating the reconstruction method. The framework derives all relations from trajectory geometry and lane-marking metadata; it never uses highD’s derived relation fields as inputs. The only calibration use of highD-derived indicators is a one-time selection of a geometry convention for pairwise DHW/THW/TTC computation, after which the final cut-in features are computed from geometry alone. This separation is important because recent work has identified lane-change inconsistencies in highD that motivate careful validation when reconstructing events from provided identifiers [15].

2.2 Surrogate safety measures

Because crashes are rare in naturalistic data, surrogate safety measures (SSMs) are widely used to assess interaction severity without requiring crash observations. The main indicators used in this thesis are time-to-collision (TTC), time headway (THW), distance headway (DHW), and the deceleration rate to avoid crash (DRAC). Recent reviews and validation studies emphasise that SSMs require careful interpretation and validation because their values depend on modelling assumptions and data quality [3, 4]. Deceleration-based indicators expressed in m/s^2 , such as DRAC, complement time-based measures (e.g., TTC, THW); Archer [14] uses 3.35 m/s^2 as a threshold to distinguish serious from non-serious conflicts, while Wang et al. [3] report DRAC thresholds in the range of $3.0\text{--}3.5 \text{ m/s}^2$ across the literature.

2.3 Lane-changing and cut-in analysis

Lane changes and cut-ins are often described through a combination of incentive, safety, and temporal structure.

Naturalistic cut-in studies show that cut-ins are not just lane changes with a binary flag. Wang et al. develop an extraction algorithm for 5,608 cut-in events and report substantial variation in cut-in duration (0.7–12.4 s) together with gap-acceptance effects from environment, vehicle type, and kinematics [6]. Zhou and Itoh study driver risk perception in two lane-change stages and find that drivers tend to rely on TTC when deciding to change lanes and on both TTC and THW while cutting into the target lane [7]. This directly motivates phase-aware interpretation of surrogate measures in the framework presented in this thesis.

Recent risk-prediction frameworks further reinforce the need for explicit scenario construction. Using the highD dataset, Shangguan et al. integrate lane-change intention recognition with risk prediction and show that interaction characteristics with surrounding vehicles are central to risk estimation [8]. Liu et al. jointly predict lane-change manoeuvre, time-to-lane-change, and risk level, again using highD as the evaluation basis [9]. These studies are predictive rather than reconstructive, but they support the thesis decision to preserve explicit pair identities, stage alignment, and interaction context features.

For cut-in and lane-merging behaviour in automated driving, structured phase models with explicit risk logic are also common. For example, Hwang et al. propose a finite-state-machine approach for lane-merging cut-ins with TTC-based risk handling [5]. While this thesis does not implement a controller, such phase-based framing

motivates clear scenario segmentation and transparent risk definitions for offline analysis.

2.4 Context encoding for interaction analysis

Scalar indicators such as minimum THW compress an interaction to a single number and discard information about the surrounding traffic configuration. A vehicle cutting in with an empty adjacent lane faces a different situation than one merging between closely spaced neighbours, yet both may produce identical THW values. Grid-based representations such as occupancy grids address this limitation by encoding the presence or absence of surrounding vehicles in a structured spatial layout [11]. In the trajectory-prediction literature, Deo and Trivedi [13] demonstrate that a spatial grid populated with encoded vehicle states (convolutional social pooling) captures interaction structure that improves manoeuvre prediction on highway data, which supports the premise that grid-based context carries information beyond scalar gap measures. Occupancy grids have been adopted widely in robotics and automated driving perception, but their use as an analysis-level context descriptor for offline interaction studies is less common.

When the grid is small (e.g., 3×3 cells), the resulting binary pattern can be compressed into a compact integer signature using a space-filling curve (SFC). Among the available SFC orderings, the Hilbert curve preserves two-dimensional spatial locality better than simpler alternatives such as the Morton (Z-order) curve, meaning that grid patterns that differ by a single neighbouring cell tend to map to numerically close codes [12]. This property is useful when clustering or comparing context patterns, because small spatial changes produce small code changes. This thesis uses a cutter-centred 3×3 occupancy grid encoded into a 16-bit Hilbert SFC signature, yielding a reversible integer-code representation of local context that supports efficient storage, retrieval, and grouping.

2.5 Research gap

The reviewed literature establishes surrogate safety measures, phase-aware lane-change models, and predictive risk frameworks, but three gaps motivate this thesis. First, cut-in analysis requires explicit pairing between the cutter and the impacted follower, yet dataset-provided SSMS may correspond to a different vehicle pair. Second, existing highD-based risk models assume that scenario samples and interaction partners are already available; they do not address the minimal-input reconstruction

problem when lane and neighbour identifiers are unavailable or inconsistent. Third, compact context representations (e.g., occupancy grids with SFC encoding) have not been combined with end-to-end reference-label verification for cut-in interactions. This thesis addresses these gaps through a framework that reconstructs relations from minimal inputs, computes pair-consistent and phase-aware indicators, and produces reference-validated SFC-based context signatures. Most prior studies therefore begin after scenario extraction has already happened. The specific contribution here is to treat scenario extraction, pair construction, and context encoding as first-class methodological objects that are validated end to end instead of assumed as given preprocessing.

Positioning summary. Table 2.1 compares the framework presented in this thesis to the most closely related studies across five dimensions that together define the research gap.

Table 2.1: Comparison of the framework presented in this thesis with related lane-change and cut-in studies. ✓ = supported, – = not addressed.

Study	Minimal input	Explicit pair	Multi-indicator	Context encoding	Reference validation
Wang et al. [6]	–	✓	–	–	–
Zhou & Itoh [7]	–	–	✓	–	–
Shangguan et al. [8]	–	–	–	–	–
Liu et al. [9]	–	–	✓	–	–
Hwang et al. [5]	–	✓	–	–	–
This thesis	✓	✓	✓	✓	✓

Chapter 3

Research Method

3.1 Methodological approach

This study adopts a design-oriented empirical methodology. We implement a reconstruction-based framework and evaluate its outputs against reference labels provided by the *highD* dataset. The evaluation is structured at four levels, each addressing a different aspect of the framework’s validity.

3.2 Evaluation levels

Frame-level validation (RQ1). For every frame, the predicted lane assignment and same-lane leader/follower identifiers are compared against highD reference labels using exact-match accuracy. We report accuracy both over all frames and conditional on frames where the reference relation exists (non-zero id), to avoid inflated scores from trivial no-neighbour matches.

Event-level validation (RQ2). Cut-in events detected from reconstructed relations are compared against events detected using reference relations, reporting precision, recall, and F_1 with an onset tolerance.

Representation-level validation (RQ3). For cut-in stage frames, 3×3 semantic occupancy grids from the reconstructed and reference paths are compared using cell-wise match ratio and Hamming distance between SFC codes:

$$\text{Match Ratio} = \frac{1}{9} \sum_{i=1}^9 \mathbf{1}(G_r[i] = G_{ref}[i]),$$

where G_r is the reconstructed grid and G_{ref} the reference grid.

Predictive evaluation (RQ4). We test whether features from the intention and decision stages, including SFC context signatures, predict execution-stage risk using leave-one-recording-out cross-validation (LORO-CV). An ablation study isolates the contribution of SFC features versus kinematic indicators.

Beyond these four levels, we report stage-based summaries, risk-threshold sensitivity, and indicator-disagreement analysis to characterise the analytical value of explicit pair construction.

3.3 Ethical considerations

The thesis uses an existing trajectory dataset collected from public roads. No new data is collected and no human subjects are recruited.

The intended benefit of the work is methodological: to support more transparent extraction and analysis of highway cut-in interactions for safety research, scenario review, and reproducible benchmarking. Because the framework is developed and evaluated offline, it should be understood as an analysis tool rather than as a real-time decision system.

At the same time, any safety-related model can be misused if its assumptions are ignored. Pairing errors, threshold choices, or dataset-specific biases could lead to misleading conclusions if the outputs were treated as direct operational truth. For that reason, the thesis keeps the reconstruction boundary explicit, reports threshold sensitivity, and treats all risk labels as operational definitions within the limits of the data rather than as ground-truth measures of danger.

3.4 Use of AI tools

Generative AI tools were used in a supportive role for troubleshooting development environment issues, resolving runtime errors, suggesting code refactoring patterns, and improving grammar and language clarity in the written thesis. All research ideas, problem formulations, methodological decisions, evaluation design, and interpretation of results were developed independently by the authors. The authors take full responsibility for the correctness and integrity of the work.

Chapter 4

Theory

4.1 Lane change vs cut-in

4.1.1 Lane change

We define a **lane change event** as a transition from one stable lane block to another stable lane block. A lane block is treated as stable when the vehicle remains in the same valid lane identifier for at least a minimum persistence duration. In the final highD configuration, a transition is accepted only if the source lane remains unchanged for at least $N_{\text{before}} = 25$ consecutive frames before the transition and the target lane remains unchanged for at least $N_{\text{after}} = 25$ consecutive frames after it. At the highD sampling rate of 25 Hz, these thresholds correspond to 1.0 s on each side. Transitions involving invalid lane identifiers (e.g., lane id 0) are ignored.

4.1.2 Cut-in

We define a **cut-in event** as a lane change where the lane-changing vehicle (the *cutter*) becomes the same-lane predecessor of another vehicle (the *follower*) in the target lane. This definition yields an explicit interacting pair (A, B) with cutter A and follower B .

4.2 Cut-in onset and stage model

We define the cut-in onset, t_0 , as the first frame in which the cutter becomes the vehicle directly ahead of the follower in the target lane. This frame marks the start of the follower relationship. We divide the time around t_0 into four stages:

- **Intention:** $[t_0 - 4 \text{ s}, t_0 - 2 \text{ s})$
- **Decision:** $[t_0 - 2 \text{ s}, t_0)$
- **Execution:** $[t_0, t_0 + 2 \text{ s})$
- **Recovery:** $[t_0 + 2 \text{ s}, t_0 + 4 \text{ s}]$

Figure 4.1 illustrates the four-stage alignment around t_0 . Because *highD* is recorded at 25 Hz, one frame corresponds to 0.04 s, and all frame counts in the implementation are converted to seconds using this rate. The recovery stage is retained for alignment completeness and descriptive inspection, but the main inferential analyses in this thesis focus on the intention, decision, and execution stages.

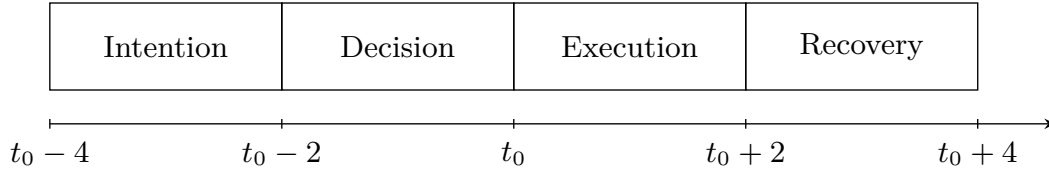


Figure 4.1: Stage model around the cut-in onset t_0 used for stage-wise feature extraction and comparison.

Justification of stage durations. The ± 4 s window and 2 s stage width are informed by three empirical observations. First, Liu et al. [9] report recall above 94% for lane-change prediction at horizons of 3.2–3.6 s before the vehicle crosses lane boundaries, supporting a multi-second pre-onset window. Second, Zhou and Itoh [7] distinguish two phases of driver risk perception: an earlier phase dominated by TTC and a later insertion phase where both TTC and THW are relevant; the 2 s split separates these phases. Third, Wang et al. [6] report that cut-in durations range from 0.7 to 12.4 s (mean ≈ 3.9 s), so a 2 s execution window captures the initial post-onset interaction without assuming the event is instantaneous. The sensitivity of this choice is addressed through the threshold analysis in the results.

4.3 Surrogate safety measures

Let A be the leader (cutter) and B the follower. We use:

- **DHW** (distance headway): longitudinal gap between leader rear and follower front.
- **THW** (time headway): $\text{THW} = \text{DHW}/v_B$ (with safe handling of small v_B).
- **TTC** (time-to-collision): under constant-velocity assumptions,

$$\text{TTC} = \begin{cases} \text{DHW}/(v_B - v_A), & \text{if } v_B > v_A \\ \infty, & \text{otherwise.} \end{cases}$$

Because cut-ins change the relevant interacting pair, all three measures are computed for the explicit cutter–follower pair rather than reusing dataset-provided values for a possibly different neighbour pair. Following Zhou and Itoh [7], who report that TTC is salient during the merging decision while THW remains informative during the insertion phase, we compute all three measures throughout the stage-aligned window and interpret them as complementary indicators.

4.3.1 Deceleration Rate to Avoid Crash (DRAC)

DHW, THW, and TTC describe proximity and gap acceptance, but they do not directly measure the *braking demand* a cut-in imposes on the follower. To complement these indicators, we include the Deceleration Rate to Avoid Crash:

$$\text{DRAC} = \begin{cases} \frac{(v_B - v_A)^2}{2 \cdot \text{DHW}}, & \text{if } v_B > v_A \text{ and } \text{DHW} > 0, \\ \text{undefined}, & \text{otherwise.} \end{cases}$$

DRAC represents the minimum constant deceleration the follower must apply to avoid a collision under constant-speed assumptions. Following Archer [14], who uses 3.35 m/s^2 as a conflict-severity threshold, values above this level are interpreted as a hard-braking situation in this thesis. Unlike TTC, which describes *time until* a potential collision, DRAC describes the *deceleration required* to prevent one, providing a complementary perspective on interaction severity. DRAC is only defined when the follower is closing, the same condition as TTC. We report DRAC alongside TTC and THW in the execution stage to see how often closing events require non-trivial braking demand.

4.4 Risk label

We define a binary risk label for descriptive comparison:

$$\text{risky} := \min(\text{THW}) \text{ in execution stage} < 0.7 \text{ s.}$$

The threshold of 0.7 s represents a deliberately strict cut-off corresponding to a very short accepted gap. Tscharn et al. [10] evaluate 0.7 s as a short time headway in their perceived-criticality experiments, while guidance values for safety-critical situations are typically in the range 1.8–2.0 s. We use execution-stage THW because it remains defined after the follower relation is established, even when TTC becomes infinite under non-closing conditions. Threshold sensitivity is reported in the results to make

the operational choice explicit [3].

4.5 Interaction context signatures (SFC encoding)

Scalar indicators like THW or TTC reduce each interaction to a single number and discard information about the surrounding traffic layout. Two cut-ins with identical THW may differ considerably in how constrained the cutter’s manoeuvre space is: one may occur in light traffic with empty adjacent lanes, while the other may involve a dense corridor with vehicles on both sides. To capture this spatial context in a form that is compact enough for storage and comparison yet rich enough to distinguish different scene configurations, we encode the local traffic layout as a short integer code.

For each frame around a cut-in event, we define a cutter-centred **3×3 semantic occupancy grid**:

- columns: {left adjacent lane, cutter’s lane, right adjacent lane},
- rows: {ahead of the cutter, alongside the cutter, behind the cutter}.

Each cell is treated as binary occupancy (1 if at least one vehicle occupies the cell, else 0). The purpose of the SFC step is not to create a new risk score; it is to turn the spatial occupancy pattern for one frame into a compact, reversible context identifier. The 3×3 grid is first mirror-normalised to a common cut-in direction, then embedded in a fixed 4×4 grid and encoded as a 16-bit integer using a locality-preserving space-filling-curve order (Hilbert order by default). Because the mapping is deterministic and reversible, we can decode a code back into the same occupancy grid. Figure 4.2 summarises this conversion from surrounding-vehicle layout to context signature.

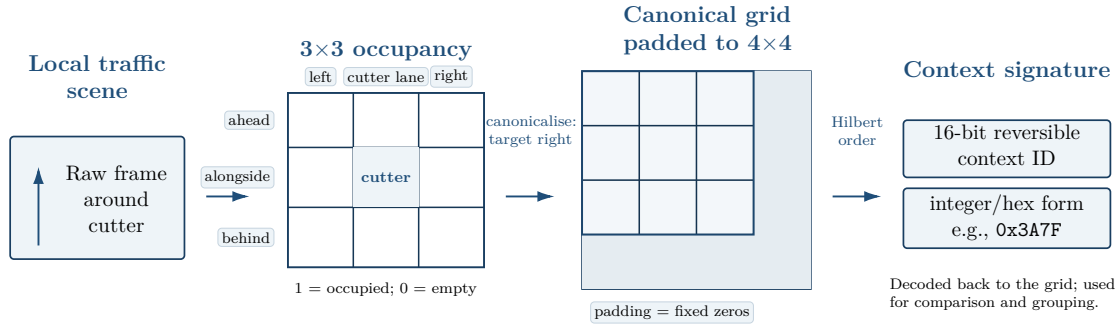


Figure 4.2: Construction of an interaction context signature. The first panel is the cutter-centred local scene for one frame; the vertical arrow defines the forward/ahead direction. The scene is discretised into 3×3 occupancy cells, mirrored to a common target-right orientation, padded with fixed zero cells to a 4×4 grid, and read in Hilbert order as a reversible 16-bit context code. The code identifies the local traffic layout; it is not a risk score.

We perform an exhaustive round-trip test across all 2^{16} possible bit patterns because the later code-level agreement checks are only meaningful if the encoding is lossless. The result confirms reversibility: encoding followed by decoding returns the original grid for both Hilbert and Morton (Z-order) traversals.

Chapter 5

Methods

5.1 Framework overview

Traj2Rel-SFC starts from trajectories, vehicle geometry, and lane-marking metadata. From these inputs it reconstructs lane assignments, neighbour relations, cut-in events, pair-specific safety indicators, and context signatures. For evaluation, the same downstream analysis is run in two parallel paths:

- **Reference-label path:** uses highD-provided lane and neighbour identifiers to produce reference relations, events, and signatures.
- **Reconstruction path:** uses only trajectories and lane markings to produce the same outputs without using highD’s derived relational fields as inputs.

This two-path design enables like-for-like agreement measurements at the frame, event, and representation levels. Figure 5.1 summarises the processing flow and highlights the evaluation boundary. Table 5.1 presents the same pipeline as a compact phase/output summary so that the main artefacts can be located quickly in the later results chapter.

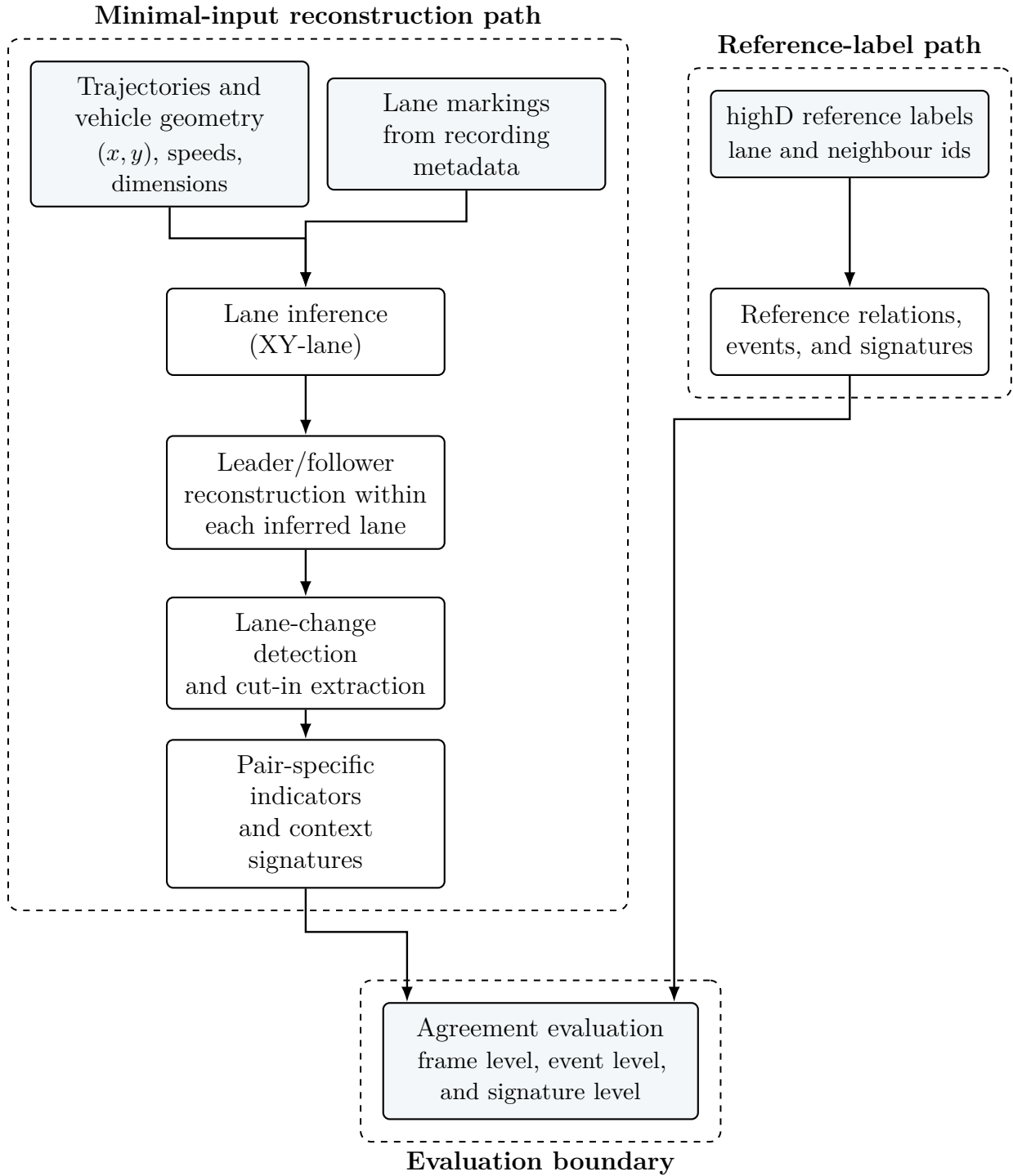


Figure 5.1: Overview of Traj2Rel-SFC. The minimal-input reconstruction path starts from trajectories, vehicle geometry, and lane markings, then reconstructs lanes, same-lane relations, cut-in events, pair-specific indicators, and context signatures. In parallel, a reference-label path derived from highD lane and neighbour identifiers generates comparison outputs that are used only for agreement evaluation.

Table 5.1: Macro-phases and outputs of Traj2Rel-SFC.

Phase	Component	Main outputs
Preliminary processing	Data harmonisation	Consolidated per-frame dataset; sanity checks.
Core reconstruction	XY-lane inference (Traj2Rel)	Predicted lane index $\hat{L}$ per frame from y and lane markings.
Core reconstruction	Neighbour reconstruction (Traj2Rel)	Predicted same-lane leader and follower identifiers per frame via within-lane longitudinal ordering.
Interaction mining	Cut-in event extraction	Cut-in table with explicit cutter–follower pairs and onset t_0 .
Safety quantification	Pair-consistent indicators & stages	DHW/THW/TTC for the interacting pair; stage-wise aggregates and risk summaries.
Context signatures	Occupancy & SFC encoding	Mirror-normalised occupancy matrices; reversible 16-bit SFC context signatures (hex + bits); stage-level context features.
Reference-label evaluation	Agreement reporting	Frame-level accuracy, event-level precision/recall/ F_1 , and representation-level match/Hamming-distance diagnostics.

The remainder of this chapter describes each component. Qualitative inspection was additionally performed using an interactive inspection dashboard (Appendix B). The dashboard was an internal validation aid used to inspect onset alignment, decoded occupancy, and extreme events; it was not a separate user study or a manual relabelling step.

5.2 Data preparation and relation reconstruction

5.2.1 Preliminary processing and data harmonisation

Each highD recording is distributed as three data tables: (i) per-frame trajectories, (ii) per-vehicle metadata, and (iii) recording-level metadata. In preliminary processing we load these tables, perform basic consistency checks (e.g., frame rate and track bounds), and combine them into a consolidated per-frame dataset where each row corresponds to one vehicle at one time step together with the metadata needed in later phases. This consolidated representation is used by both the reference-label path and the reconstruction path so that subsequent comparisons are like-for-like.

We also extract lane-marking positions from the recording metadata because they provide the only map-related input used by the reconstruction method. Dataset-provided lane and neighbour identifiers are retained only for evaluation and are never used as inputs in the reconstruction path.

The next step is the reconstruction layer (Traj2Rel), which derives two pieces of relational information from minimal inputs: (i) lane membership for each vehicle and time step, and (ii) the identity of the closest vehicle ahead and behind in the same lane. Together these relations are sufficient to define leader–follower interactions and to express cut-in events as explicit cutter–follower pairs.

5.2.2 Lane inference from lateral position (XY-lane)

Lane membership is inferred by mapping lateral position to the interval between two consecutive lane markings provided in the recording metadata. Values outside the roadway are mapped to lane id 0 and are excluded from interaction mining. Algorithm 1 summarises the procedure.

Algorithm 1: Lane inference from lateral position and lane markings

Input: Vehicle lateral position $y(t)$ and lane boundaries $\{b_0, \dots, b_K\}$ from recording metadata.

Output: Inferred lane index $\ell(t) \in \{1, \dots, K\}$ (or 0 if outside the roadway).

foreach *frame* t **do**

if $y(t) \notin [b_0, b_K)$ **then**

$\ell(t) \leftarrow 0$ // outside valid lanes

else

 find k such that $b_{k-1} \leq y(t) < b_k$

$\ell(t) \leftarrow k$

5.2.3 Same-lane leader/follower reconstruction

Within each frame and inferred lane, vehicles are ordered by longitudinal position in the travel direction. The immediate leader (ahead) and follower (behind) are given by adjacent vehicles in this ordering. Algorithm 2 summarises the procedure.

Algorithm 2: Same-lane leader/follower reconstruction from geometry

Input: Frame-wise vehicle states with longitudinal position $x(t)$ and inferred lane index $\ell(t)$.

Output: For each vehicle: leaderId and followerId in the same lane (or 0 if none).

foreach *frame* t **do**

 group vehicles by lane ℓ

foreach *lane group* $\mathcal{V}_{\ell,t}$ **do**

 sort $\mathcal{V}_{\ell,t}$ by $x(t)$ in travel direction

foreach *vehicle* v_i *in sorted order* **do**

 leaderId(v_i) $\leftarrow$ closest vehicle ahead in the list (else 0)

 followerId(v_i) $\leftarrow$ closest vehicle behind in the list (else 0)

These two components form the core of the **system under evaluation**: starting from trajectories and lane markings, we can reconstruct the relations needed for downstream interaction mining, indicator computation, and context-signature generation without using dataset-provided neighbour identifiers as inputs.

5.3 Interaction mining

5.3.1 Lane-change detection

A lane change is accepted when a vehicle transitions from one stable lane block to another. In the final highD configuration, stability means that the source lane remains unchanged for at least $N_{\text{before}} = 25$ frames before the transition and the target lane remains unchanged for at least $N_{\text{after}} = 25$ frames after it, corresponding to 1.0 s on each side at 25 Hz. Transitions involving invalid lane identifiers (e.g., lane id 0) are ignored.

5.3.2 Cut-in extraction

Given a lane change into target lane L by vehicle A :

1. starting from the lane change start, scan frames forward within a fixed search window,
2. in each frame, identify a vehicle B in lane L for which A is the immediate leader in that frame (according to the chosen source of neighbour relations: reconstructed relations or reference labels),
3. if the relationship is established within the allowed delay and holds for at least a minimum number of frames, record a cut-in event with cutter A , follower B , and onset t_0 at the first matching frame.

Algorithm 3 summarises the rule-based cut-in identification used in the reconstruction and reference-label paths.

Algorithm 3: Cut-in identification as a cutter–follower pair

Input: Lane-change candidate for vehicle A with target lane L ,
reconstructed relations over time.

Output: Cut-in event (A, B, t_0) or no event.

initialise $t_0 \leftarrow \text{undefined}$

foreach frame t in a search window after the lane change begins **do**

if A is the leader of some vehicle B in lane L **then**

set t_0 to the first such frame and keep B as follower

if the relation $A \rightarrow B$ persists for at least $d_{\min}$ frames **then**

return (A, B, t_0)

return no event

Design rationale. The framework separates event extraction from risk labelling. Lane changes are detected between stable segments, which reduces sensitivity to label jitter, and follower pairing requires a minimum persistence duration to avoid one-frame oscillations. This is consistent with naturalistic evidence that cut-ins have non-trivial duration rather than occurring instantaneously [6].

5.3.3 Configuration

Table 5.2 summarises the key parameters used in the final run (highD 01–60, 25 Hz). These parameters are informed by empirical cut-in duration distributions [6] and the two-phase risk perception model [7], and are reported to support reproducibility.

Table 5.2: Key configuration values for the final framework run.

Component	Parameter	Value
Lane-change detection	Min stable before/after	25 frames (≈ 1.0 s)
Lane-change detection	Ignored lane ids	{0}
Cut-in detection	Search window	50 frames (≈ 2.0 s)
Cut-in detection	Min relation duration	10 frames (≈ 0.4 s)
Cut-in detection	Max relation delay	15 frames (≈ 0.6 s)
Indicator computation	Min speed for THW	0.1 m/s
Indicator computation	Closing speed epsilon	10^{-6}
Indicator computation	Window (pre/post) for diagnostics	50/75 frames (≈ 2 s/3 s)
Stage model	Total analysis window	± 4 s around t_0
Risk label	Execution THW threshold	0.7 s
SFC encoding	Ahead/behind range	Disabled (no fixed cutoff)
SFC encoding	Alongside threshold	4.0 m (final Step 14 thesis run)
SFC encoding	Order	Hilbert

5.4 Safety quantification

All safety indicators are computed for the explicit cutter–follower pair rather than a generic predecessor, consistent with recent risk models that use interaction-specific features [8, 9]. Following Zhou and Itoh [7], we interpret TTC and THW as complementary indicators across the stage-aligned window.

We compute DHW/THW/TTC for arbitrary pairs using geometry and speeds. Because trajectory datasets may define position references differently (e.g., centre point vs bounding-box corner), we *select a geometry model* by comparing our computed indicators against highD-provided indicators for same-lane (ego, predecessor) pairs on a large random sample. This calibration step is separate from the reconstruction boundary: highD indicators are used here only as external reference values to choose a consistent position convention, not as hidden inputs to the final cut-in features. We test multiple plausible reference conventions and select the one that best matches dataset-provided DHW/THW/TTC. In the final run, the best match was obtained with a bounding-box–based interpretation of the position variables (a top-left reference point), combined with vehicle length to compute front and rear positions.

To avoid negative gaps due to noise or reference mismatches, reported minima are clipped to be non-negative for DHW/THW. TTC is treated as infinite when the follower is not closing ($v_B \leq v_A$).

5.4.1 Stage-wise features

For each detected cut-in, we align a fixed time window around onset t_0 and compute stage-wise summaries for the four stages in Chapter 4. The feature table includes stage minima of DHW/THW/TTC and kinematic aggregates for cutter and follower. These stage-wise summaries are later used not only for descriptive risk analysis, but also for operational-usefulness evaluation through threshold sensitivity and TTC-versus-THW disagreement analysis.

5.4.2 Deceleration-based indicator: DRAC computation

We compute DRAC for the explicit cutter-follower pair in the execution stage using the formula defined in Section 4.3.1. The implementation reuses the same pair-consistent longitudinal states (positions, speeds, and DHW) already computed for TTC and THW, so all indicators describe the same interaction pair. DRAC is reported alongside the existing indicators to assess braking demand alongside gap acceptance (THW) and collision timing (TTC).

5.5 Context signatures

To capture local traffic context around the cutter compactly, we encode neighbourhood occupancy into the 3×3 semantic grid defined in Chapter 4. Each frame is mapped to a binary occupancy pattern, and the 3×3 cells are embedded into a 4×4 grid and linearised using a Hilbert-type space-filling-curve order. To remove direction-induced asymmetry, events are mirror-normalised to a canonical orientation prior to aggregation.

Minimal-input SFC signatures. In the minimal-input path, occupancy is computed from inferred lanes (XY-lane) and lane-snapshot neighbour queries (geometry). This means SFC signatures can be generated without accessing highD neighbour columns.

Reference SFC signatures. For comparison, we also construct a reference occupancy grid from the highD surrounding-vehicle identifiers (same-lane preceding/following plus left/right neighbours and alongside vehicles). Mapping each non-zero neighbour id to occupancy yields a reference 3×3 grid per frame, which can be encoded into a reference SFC signature.

Signature agreement. We assess the SFC branch by comparing the reference and minimal-input codes for matching (event, frame) rows. We report both exact code match rate and the distribution of Hamming distances (bitwise differences), which localises whether mismatches affect a single cell or multiple cells.

Exploratory clustering of decision-stage signatures. As an additional analysis of representational usefulness, we cluster event-level decision-stage context signatures after the core agreement check. For each validated cut-in event, the mirror-normalised decision-stage 3×3 occupancy grids are averaged over the frames in that stage, yielding a 9-dimensional vector of cell-wise occupancy probabilities. We then apply k -means clustering for $K = 3, \dots, 8$, assess internal cohesion with the silhouette score, and assess solution stability with the mean adjusted Rand index (ARI) across repeated random initialisations. This clustering is not used as a primary correctness criterion; it is reported only as an exploratory check of whether the validated SFC representation supports interpretable grouping of decision-stage interaction contexts.

Algorithm 4 summarises how the context signatures are constructed for each stage-aligned frame.

5.6 Predictive evaluation: decision-stage risk forecasting

Beyond descriptive agreement, we evaluate the *predictive value* of the framework’s reconstructed features. We test whether features available *before* the cut-in is established (intention and decision stages) can predict execution-stage THW risk ($\text{THW}_{\min} < 0.7 \text{ s}$).

Algorithm 4: Cutter-centred occupancy grid and space-filling-curve context signature

Input: Cut-in event (A, B, t_0) , reconstructed lane/relations, neighbourhood range parameters.

Output: A stage-aligned sequence of binary grids and a 16-bit context signature per frame.

foreach frame t in the analysis window around t_0 **do**

- build a 3×3 cutter-centred occupancy grid around the cutter A
- (optional) mirror-normalise the grid to a canonical direction
- embed the 3×3 grid into a 4×4 grid with fixed padding
- traverse cells in Hilbert order and write occupancy bits into a 16-bit code
- store the code and (optionally) the decoded grid for inspection

Feature set. The feature vector used for prediction consists of 34 features: 18 kinematic features from the intention and decision stages ($DHW_{\min}$, $THW_{\min}$, $TTC_{\min}$, cutter/follower speeds, accelerations, and lateral velocity for each stage) and 16 SFC context features (mean occupancy probabilities for each bit position in the decision-stage mirrored Hilbert code).

Evaluation protocol. We use Leave-One-Recording-Out Cross-Validation (LORO-CV): for each of the 60 recordings, the model is trained on the remaining 59 recordings and tested on the held-out recording. This ensures that all test events come from a traffic environment not seen during training. The classification threshold is optimised on the training fold by maximising F_1 .

Model. We train a bagged decision-stump ensemble (50 trees with random feature subsampling), which provides both predictions and feature-importance estimates via Gini gain. The choice of a shallow ensemble rather than a deeper model (e.g., gradient boosting or neural network) is deliberate: with 34 features, 10,023 events, and a 20% minority class, a high-capacity model risks overfitting to recording-specific artefacts. A decision-stump ensemble offers stable feature-importance rankings for interpretability and has low variance under the 60-fold LORO-CV protocol. The goal of RQ4 is to test whether the reconstructed features carry predictive value at all, not to maximise classification accuracy; a simple model that clearly exceeds the baseline provides stronger evidence for this claim than a tuned model whose performance is harder to attribute to the features alone.

Baseline. The majority-class baseline assigns the most frequent training-set label to all test events. Because the minority class (risky, 20.1%) is never the majority, the baseline predicts “non-risky” for all events and achieves $F_1 = 0$ for the risky class.

Ablation. To isolate the contribution of SFC context features versus kinematic indicators, we run the same LORO-CV protocol on three feature subsets: all features (34), kinematic features only (18), and SFC features only (16).

Unless otherwise noted, uncertainty intervals on reconstruction metrics are computed over the 60 recording-level scores rather than over individual frames. We report these as 95% normal-approximation intervals for the mean ($\bar{x} \pm 1.96 \text{ SE}$), so that recordings rather than frames act as the unit of aggregation.

5.7 Reproducibility

All results reported in this thesis can be regenerated from the released pipeline given the configuration in Table 5.2 and the *highD* dataset. Some substeps use stochastic procedures (random sampling for geometry calibration, repeated k -means initialisations, and random feature subsampling in the ensemble model), so exact numerical replication additionally requires the same random seeds. The complete pipeline, including scripts for all tables and figures, is publicly available at <https://github.com/arsulesandy/cutin-risk-analysis>.

Chapter 6

Results

6.1 Event extraction and dataset overview

The final run processes all 60 highD recordings. From 11,856 detected lane changes, the framework extracts 10,026 cut-in events (84.6%), of which 10,023 remain after stage-window validity checks. The three excluded events are those for which the full stage-aligned feature window falls outside the available track boundaries. Table 6.1 summarises the event yield, and Table 6.2 reports duration statistics.

Table 6.1: Event counts for the final run (recordings 01–60).

Artefact	Lane changes	Cut-ins	Notes
Interaction mining	11,856	10,026	Extracted from reconstructed relations.
Stage feature table	–	10,023	After removing 3 events whose full stage window falls outside the available track range.
Context signatures	–	1,421,980	Frame-level signatures for 10,023 events.

Table 6.2: Duration statistics (highD 01–60, 25 Hz).

Event type	Min (s)	P25 (s)	Median (s)	Mean (s)	P75 (s)	Max (s)
Lane change duration	0.960	3.680	6.240	6.551	8.960	69.120
Observed cut-in relation duration within post-onset window	0.360	1.960	1.960	1.870	1.960	1.960

Note: The cut-in relation duration is right-censored by the fixed post-onset extraction window, so the repeated upper quantiles at 1.96 s reflect the event-extraction design rather than an unconstrained natural duration distribution.

6.2 Reconstruction accuracy (RQ1/RQ2)

Geometry-based same-lane neighbour reconstruction matches highD reference relations with mean accuracy 0.999997. When lane membership is also inferred from geometry (XY-lane), agreement remains high: mean preceding/following accuracy

0.999810/0.999811 (mismatch rate $\approx 0.02\%$). Cut-in detection from reconstructed relations matches reference events with $F_1 = 0.998938$, with recording-level 95% intervals of $\pm 2.3 \times 10^{-5}$ for neighbour accuracy and $\pm 7.3 \times 10^{-4}$ for cut-in F_1 .

These results establish that the minimal-input path reproduces the reference-label path with high accuracy, validating the reconstruction as a basis for downstream analysis.

6.3 Execution-stage indicator distributions

Table 6.3 summarises execution-stage minima across all 10,023 events. TTC statistics are shown only for events with finite TTC (i.e., closing speed).

Table 6.3: Execution-stage indicator minima across all events (10,023 cut-ins).

Indicator	Min	P25	Median	Mean	P75	Max
DHW _{min} (m)	1.37	20.69	32.21	51.28	59.82	373.98
THW _{min} (s)	0.051	0.769	1.194	1.700	2.028	12.383
TTC _{min} (s), finite only (n=3,606)	0.220	10.769	19.726	74.161	42.025	14428.50

The TTC distribution is strongly right-skewed, so the median and quartiles are more informative than the mean for interpreting the typical severity of finite-closing events.

6.4 Risk characterisation (RQ2)

Using the risk label defined in Chapter 4 (execution-stage THW_{min} < 0.7 s), 2,018 of 10,023 events (20.1%) are classified as risky. Table 6.4 compares the two groups, and Table 6.5 reports effect sizes using Cliff’s δ .

Table 6.4: Execution-stage risk summary. $\Delta v = \bar{v}_{\text{cutter}} - \bar{v}_{\text{follower}}$.

Group	Events	THW _{min} median (s)	DHW _{min} median (m)	Δv median (m/s)
Non-risky	8,005	1.427	39.680	2.304
Risky	2,018	0.539	14.945	2.853

Table 6.5: Effect sizes between risk groups (execution stage). δ is Cliff’s delta.

Metric	Median non-risky	Median risky	Cliff’s δ
$DHW_{\min}$ (m)	39.68	14.95	−0.950
Δv (m/s)	2.304	2.853	0.032
$ v_{y,\text{cutter}} _{\max}$ (m/s)	0.88	0.93	0.097

The large effect size for $DHW_{\min}$ ($\delta = -0.950$) is expected given the direct relationship $THW = DHW/v_B$. The negligible effect for lateral velocity ($\delta = 0.097$) indicates that the THW-based label captures longitudinal gap acceptance rather than lateral kinematics.

Threshold sensitivity. Table 6.6 shows how risk prevalence varies with the THW threshold, ranging from 7.58% at 0.50 s to 62.21% at 1.50 s. Across the 60 recordings, the risk ratio varies between 11.1% and 34.1% (median 20.0%).

Table 6.6: THW threshold sensitivity (10,023 events).

THW threshold (s)	Risky events	Prevalence
0.50	760	7.58%
0.70	2,018	20.13%
1.00	4,004	39.95%
1.50	6,235	62.21%

6.5 Multi-indicator analysis (RQ2)

Indicator complementarity: THW, TTC, and DRAC. The three indicators capture fundamentally different aspects of cut-in severity. TTC is finite for only 3,606 of 10,023 events (35.98%), meaning that a TTC-only analysis would miss the majority of established interactions. DRAC is finite for a similar share (3,393 events, 33.9%), but its median value is 0.071 m/s^2 , which is well below any critical threshold. Only 8 events (0.08%) exceed the hard-braking threshold of 3.35 m/s^2 . In contrast, THW flags 20.1% of events as risky. Tables 6.7–6.10 detail these distributions.

Table 6.7: Execution-stage DRAC distribution (finite only, $n = 3,393$).

Indicator	Min	P25	Median	Mean	P75	P95
DRAC (m/s^2)	<0.001	0.016	0.071	0.193	0.202	0.693

Table 6.8: DRAC threshold sensitivity (10,023 events).

DRAC threshold (m/s ²)	Flagged events	Ratio
1.0	80	0.80%
2.0	18	0.18%
3.0	9	0.09%
3.35	8	0.08%
5.0	6	0.06%

Table 6.9: Execution-stage TTC by THW risk group (finite TTC only, $n = 3,606$).

Group (by THW label)	Events	TTC median (s)	TTC mean (s)
Non-risky	3,177	22.30	80.27
Risky	429	8.98	28.94

Table 6.10: THW-risky vs TTC-threshold overlap (infinite TTC counted as not risky).

TTC thr. (s)	TTC-flagged	TP	FP	Prec.	Rec.
1.5	10	10	0	1.000	0.005
3.0	43	37	6	0.860	0.018
5.0	156	99	57	0.635	0.049
10.0	788	237	551	0.301	0.117

These results show that no single indicator captures all aspects of cut-in severity. THW identifies short accepted gaps, TTC captures closing-speed conflicts, and DRAC quantifies braking demand. The framework’s pair-consistent computation makes this multi-indicator perspective possible because all measures describe the same cutter–follower interaction [7, 3].

Extreme events. Table 6.11 lists the three events with the lowest finite TTC, useful for qualitative inspection (Appendix B).

Table 6.11: Top-3 extreme events by minimum execution-stage TTC (finite only).

Rec.	Cutter	Follower	From	To	TTC _{min} (s)	THW _{min} (s)
40	682	687	1	2	0.220	0.051
26	496	522	1	2	0.235	0.122
30	839	852	1	2	0.288	0.067

6.6 Context signature maps and interpretation

This section decodes the mirrored SFC signatures from Section 5.5 back into the underlying semantic 3×3 occupancy grid so that the spatial meaning of the codes becomes visible. For each stage-aligned frame, each cell is binary (1 if at least one vehicle occupies that semantic region, else 0); the plotted cell values are then simple group means:

$$\bar{G}_{r,c} = \frac{1}{N} \sum_{i=1}^N G_{r,c}^{(i)}$$

where $G_{r,c}^{(i)} \in \{0, 1\}$ is the occupancy of cell (r, c) in one mirrored stage frame. Each number in Figures 6.1–6.3 should therefore be read as a *mean occupancy probability*. For example, a value of 60% in one cell means that the corresponding semantic region is occupied in 60% of the relevant stage frames.

After mirror normalisation, the left column corresponds to the lane *away from the manoeuvre*, the middle column to the cutter’s current lane, and the right column to the *target lane*. The rows should be read top-to-bottom as *ahead*, *alongside*, and *behind* the cutter. The centre cell is not analytically informative because the cutter itself occupies it by construction; it is therefore shown as fixed rather than interpreted as evidence about surrounding traffic.

We present the decision stage as a representative example because it captures the traffic configuration immediately before the cut-in interaction becomes established. Figure 6.1 therefore shows both absolute occupancy maps (non-risk and risky) and the risk-minus-non-risk difference map for the decision stage. The absolute maps provide context; the difference map is the main analytical view and is computed cell-wise as

$$\Delta_{r,c}^{\text{pp}} = 100 \left(\bar{G}_{r,c}^{\text{risk}} - \bar{G}_{r,c}^{\text{non-risk}} \right),$$

measured in *percentage points* (pp). Thus, a label of +3.2 pp means that the cell is occupied 3.2 percentage points more often in risky events, while -5.6 pp means that the same cell is occupied 5.6 percentage points more often in non-risk events.

To highlight how group differences evolve over time, Figure 6.2 shows risk-minus-non-risk difference maps for the intention, decision, and execution stages using a consistent zero-centred colour scale. The key point is not that risky and non-risky cut-ins form completely different scene types; rather, they share a common cut-in geometry and differ through comparatively small occupancy shifts in a few neighbouring cells. This is why the absolute maps look broadly similar and why the difference maps are more informative.

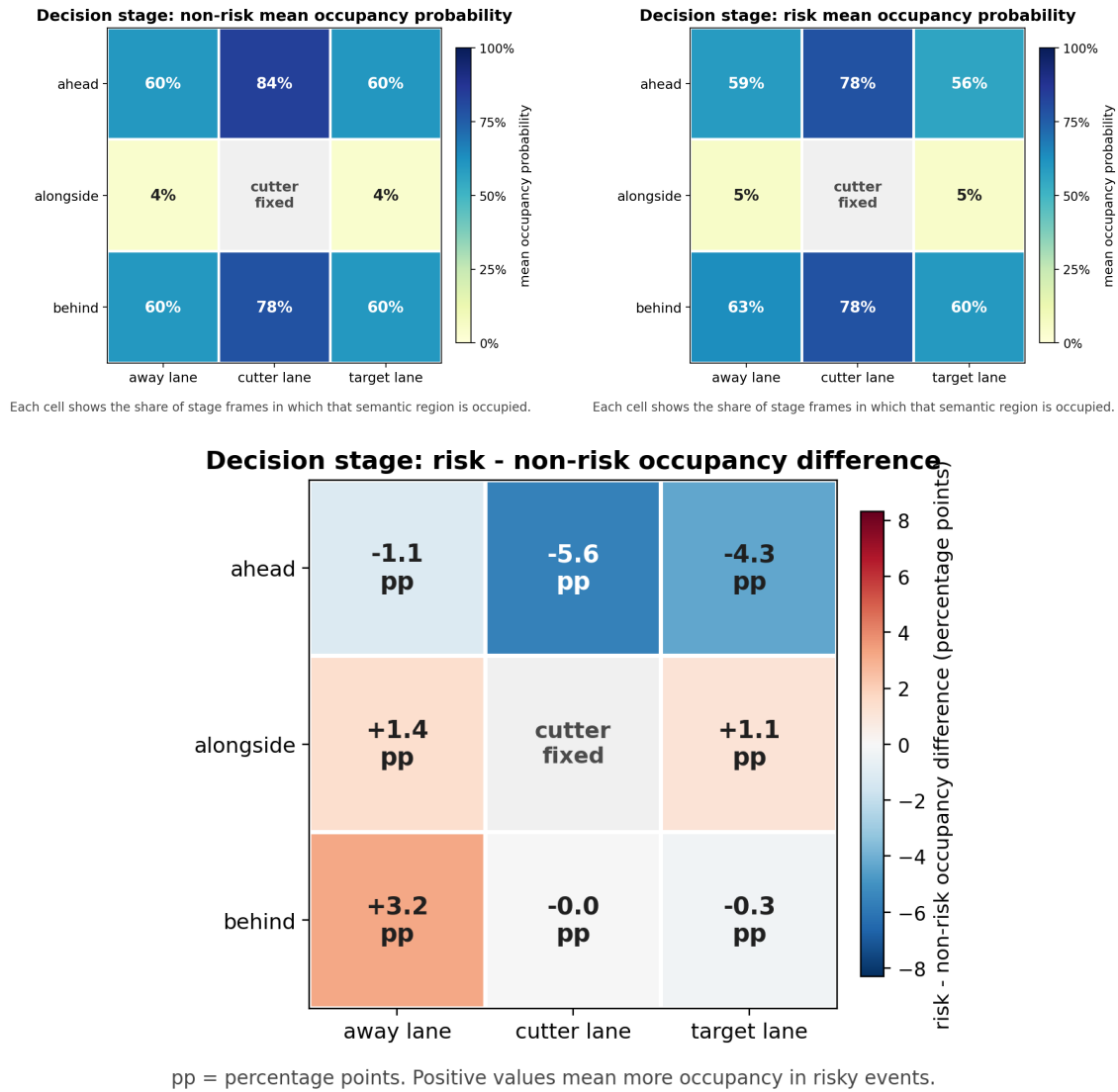


Figure 6.1: Decision-stage occupancy maps. The top panels show cell-wise mean occupancy probabilities for non-risk and risky events after mirror normalisation, so each label is the share of decision-stage frames in which that semantic cell is occupied. The bottom panel shows the cell-wise difference (risk minus non-risk) in percentage points.

In practical terms, these plots should be read as subtle but structured differences in local traffic context. They suggest that the risky group is associated not with an entirely separate scene family, but with small changes in how tightly neighbouring cells around the cutter are occupied during the lead-up to the cut-in. A compact reading is that risky events shift occupancy away from cells *ahead* of the cutter and toward cells *alongside* or *behind* it during the intention and decision stages; by the execution stage, the remaining differences are smaller and are concentrated mainly in the cells behind the manoeuvre. In the decision stage, the most meaningful pattern is therefore a modest tightening around the target-lane gap rather than a wholesale change of scene type, which is consistent with risky cut-ins being executed in more

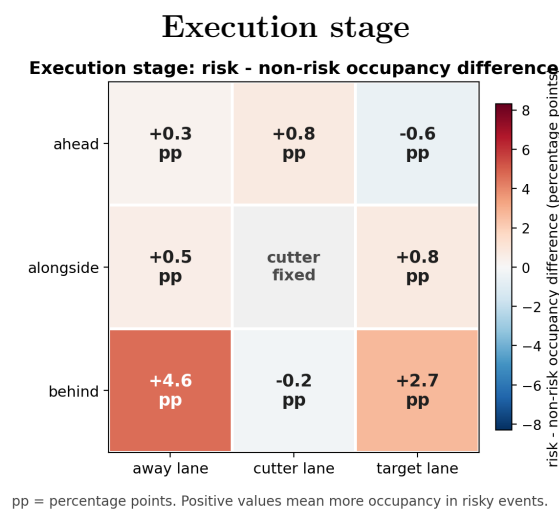
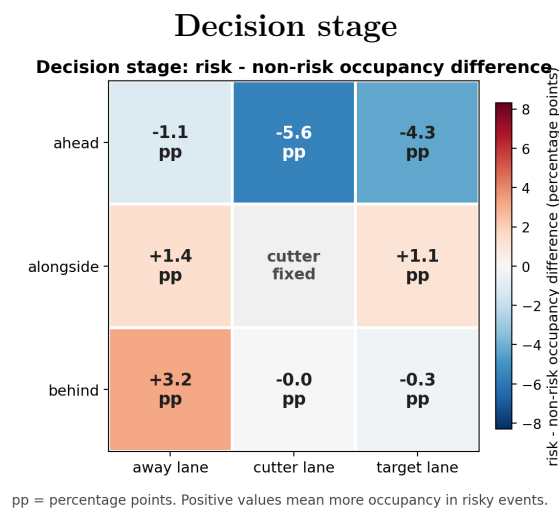
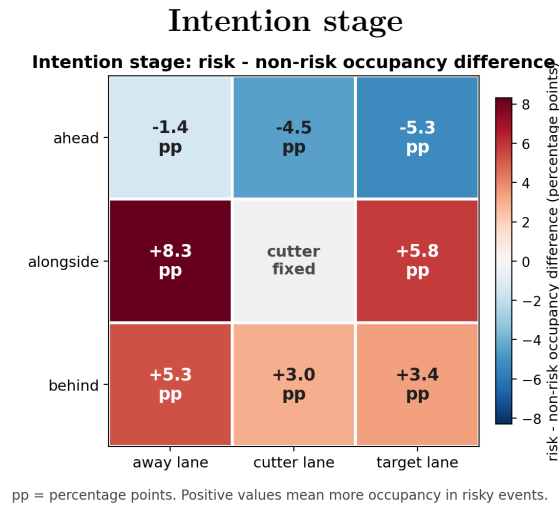


Figure 6.2: Stage-wise risk-minus-non-risk occupancy differences for the intention, decision, and execution stages, shown separately to make the small cell-wise shifts easier to read. Red indicates cells occupied more often in risky events, blue indicates the opposite, and each signed label is a percentage-point difference (pp). The centre cell is neutral because the cutter occupies it by construction.

constrained surrounding traffic.

6.6.1 Exploratory decision-stage context archetypes

Mean occupancy maps show group-level differences, but they do not directly answer whether the validated SFC representation can recover a small set of reusable scene families. As an additional representational-usefulness check, we therefore decode the mirrored binary SFC signatures back into the semantic 3×3 grid, average the decision-stage grids within each event, and cluster the resulting event-level context vectors. We use the *decision* stage rather than the execution stage so that clustering is based on decision-stage context rather than on the same time window used to define the THW risk label.

Searching $K = 3, \dots, 8$ with k -means, the best solution is $K = 7$, with silhouette score 0.498 and mean stability ARI 0.976 across repeated initialisations. The resulting archetypes remain reasonably balanced, spanning 6.0%–30.7% of all 10,023 validated cut-ins. The labels A1–A7 in Figure 6.3 are therefore *cluster identifiers*, not manually defined classes; they are relabelled by cluster size so that A1 is the largest archetype and A7 the smallest. Each panel reports four things at once: the mean occupancy probabilities for that cluster, the cluster’s share of all events, its execution-stage THW-risk prevalence, and its risk ratio relative to the overall 20.13% risk prevalence.

Figure 6.3 shows the seven archetypes, and Figure 6.4 compares their execution-stage THW risk prevalence. Read qualitatively, A1 and A2 are leader-dominated patterns with strong occupancy ahead of the cutter, A3 combines strong leaders ahead with a frequent target-lane follower, A4 represents the tightest target-lane corridor (high occupancy both ahead of and behind the target gap), A5 is follower-dominated in the cutter/target lanes, A6 concentrates traffic mainly ahead across all lanes, and A7 is dominated by followers behind the cutter across all three lanes. The most risk-enriched archetype (A4, 10.27% of events) reaches 26.63% risky outcomes versus 20.13% overall (risk ratio 1.32) with median execution THW 0.975 s. Lower-risk archetypes such as A2 and A6 remain below the overall prevalence (17.55% and 16.29%, respectively). The SFC code therefore supports interpretable unsupervised grouping of cut-in scenes based on decision-stage context.



Figure 6.3: Seven decision-stage context archetypes derived from mirrored binary SFC signatures. A1–A7 are unsupervised cluster IDs relabelled by cluster size (A1 largest). In each panel, cell labels show mean occupancy percentages, while the text above the grid reports the archetype’s event share, risk prevalence, and risk ratio relative to the overall dataset. The grey hatched centre cell is neutralised because the cutter occupies it by construction.

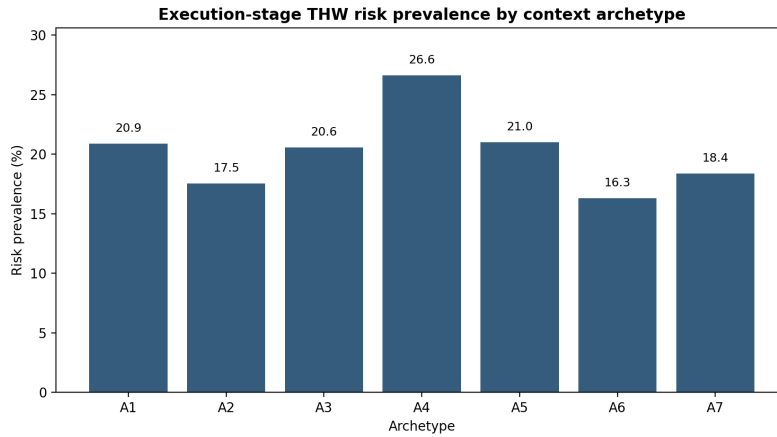


Figure 6.4: Execution-stage THW risk prevalence by decision-stage context archetype.

6.7 Predictive evaluation (RQ4)

Table 6.12 summarises the LORO-CV results for the decision-stump ensemble on the full feature set.

Table 6.12: Risk prediction: LORO-CV summary (stump ensemble, 34 features from the intention and decision stages, 60 held-out folds).

Metric	Mean	$\pm$ Std
ROC-AUC	0.822	± 0.058
F_1 (risky class)	0.506	± 0.088
Precision	0.357	± 0.083
Recall	0.907	± 0.060
Baseline F_1	0.000	± 0.000
ΔF_1 vs baseline	+0.506	± 0.088

The model achieves ROC-AUC of 0.82, indicating good discriminative ability between risky and non-risky events based on features available before the execution stage. The high recall (0.91) means the model identifies most risky events before the execution stage, while the lower precision (0.36) reflects false positives: events that appear risky based on decision-stage features but resolve safely. This asymmetry is expected because not all tight decision-stage gaps lead to risky outcomes; follower adaptation and cutter acceleration can widen the gap during execution.

Feature importance. Table 6.13 lists the five most important features by Gini importance.

Table 6.13: Top-5 features by Gini importance (stump ensemble, LORO-CV).

Feature	Importance
Decision-stage THW _{min}	0.268
Decision-stage DHW _{min}	0.247
Intention-stage DHW _{min}	0.119
Intention-stage THW _{min}	0.110
Decision-stage TTC _{min}	0.055

Decision-stage DHW and THW minima together account for about 52% of the total importance, which confirms that gap acceptance in the decision stage is the strongest predictor of execution-stage risk. Among the SFC features, the two most informative bits (SFC positions 13 and 8) contribute modestly, suggesting that specific occupancy configurations carry additional predictive information.

Ablation study. Table 6.14 compares models trained on different feature subsets.

Table 6.14: Feature ablation: LORO-CV (stump ensemble).

Feature set	n	Mean AUC ($\pm$ std)	Mean F_1 ($\pm$ std)
All (kinematic + SFC)	34	0.822 (± 0.058)	0.506 (± 0.088)
Kinematic only	18	0.832 (± 0.061)	0.519 (± 0.086)
SFC only	16	0.621 (± 0.059)	0.349 (± 0.056)

Kinematic features alone achieve a slightly higher AUC (0.832) than the full model (0.822), while the SFC-only model still reaches an AUC of 0.621, which is clearly above chance (0.50). To test whether these differences are meaningful, we first used a Friedman test across the three ablation models over the 60 held-out folds. The result was significant ($\chi^2(2) = 81.43$, $p = 2.07 \times 10^{-18}$), so we then carried out Bonferroni-corrected Wilcoxon signed-rank tests for the follow-up comparisons. After correction, the difference between the full model and the kinematic-only model was not significant ($p = 0.267$). In contrast, both the full model and the kinematic-only model performed significantly better than the SFC-only model ($p = 6.52 \times 10^{-11}$ and $p = 9.28 \times 10^{-11}$, respectively). The SFC-only model also remained significantly above chance ($p = 4.64 \times 10^{-11}$). Overall, this suggests that the SFC features do contain useful predictive information, but most of the predictive signal is already captured by the kinematic features.

Point-biserial correlations. The strongest individual predictors are decision-stage THW_{min} ($r = -0.34$) and decision-stage DHW_{min} ($r = -0.31$), both negatively

correlated with risk: smaller pre-onset gaps predict riskier outcomes.

6.8 Reference agreement for context signatures (RQ3)

We assess the SFC encoding in two complementary ways:

- **Algorithmic round-trip:** exhaustive decode $\leftrightarrow$ encode testing across all 2^{16} codes confirms that the 16-bit SFC mapping is reversible for both Hilbert and Morton orders.
- **Reference agreement:** for cut-in stage frames, we compute a reference occupancy grid from highD surrounding-vehicle ids and compare its SFC code to the minimal-input code computed from inferred lanes and geometry. Exact matches indicate identical neighbour occupancy under the two definitions; non-matches localise differences to specific cells via Hamming distance.

Using the final thesis configuration, with an alongside threshold of 4.0 m and no fixed ahead/behind cutoff, we compare 1,421,980 stage rows and obtain 1,414,762 exact row matches (99.49%). Stage-wise exact-match rates are 99.35% (intention), 99.52% (decision), and 99.60% (execution). In the thesis event view, stage-window ANY-frame agreement is 100.00%, while stage-window ALL-frame agreement is 79.82%. The raw unmirrored match rate is only 57.56%; this is best read as a sanity check, because opposite travel directions swap left/right occupancy by construction. Mirror normalisation is therefore required before any like-for-like comparison of left/right context across recordings.

6.9 Exploratory extension on *exiD*

To make the transferability claim concrete without turning the thesis into a second benchmark, we ran a deliberately limited exploratory extension on *exiD*, a related drone dataset of highly interactive highway entry/exit scenarios in Germany [2]. We restricted this add-on to 10 recordings (07, 18, 28, 29, 42, 50, 60, 70, 77, 83). These files were chosen pragmatically rather than statistically: the aim was to cover all seven *exiD* locations, keep visible lane-change activity at each site, and prefer shorter recordings where possible so that extraction, visual review, and figure generation remained tractable.

The question here is narrower than in the main *highD* evaluation: can the same explicit cut-in event definition, cutter–follower pairing logic, and pairwise risk computation be moved to a related but more interactive highway geometry with only limited adaptation work? For that reason, the *exiD* results are not pooled with the *highD* statistics, and we do not repeat the full *highD*-style reference agreement or prediction study.

Figure 6.5 summarises the event yield across the selected recordings. The exported cut-in counts vary strongly across the 10 files, from low tens to more than one hundred events, with recording 77 contributing the largest set. Median execution-stage THW also varies substantially across recordings (roughly 1.0–3.1 s), which is expected because *exiD* mixes mainline and ramp interactions rather than the straighter motorway segments represented in *highD*. The point of this figure is therefore not to compare *highD* and *exiD* statistically, but to show that the same event-mining core still produces coherent outputs on a related dataset family.

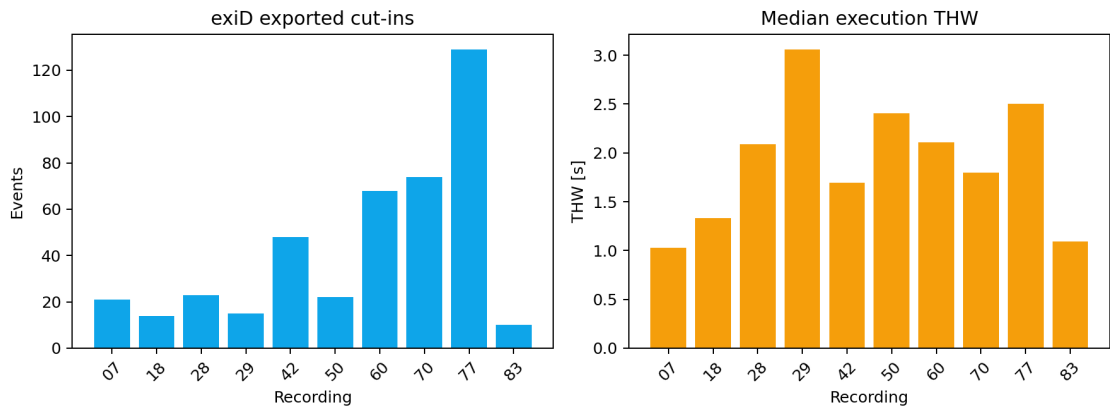


Figure 6.5: Exploratory *exiD* overview for the 10 selected recordings. Left: number of exported cut-ins per recording. Right: median execution-stage THW per recording. The variation across files reflects the heterogeneous interaction density of entry/exit scenes and reinforces that this add-on should be read as a limited transferability illustration rather than as a second benchmark.

Because TTC is a standard cut-in severity cue in naturalistic studies [6] but is only defined for closing interactions, Figure 6.6 reports two TTC-aware summaries: the finite-TTC share and the median over finite TTC events only. This follows the general surrogate-safety guidance that indicator interpretation depends on explicit operational assumptions rather than on a single pooled value [3]. Across the 424 exploratory *exiD* events, finite TTC is available for 46.5% of cases overall, with substantial variation across recordings (14.3%–68.2%). Median finite TTC ranges from about 17.6 s to 75.8 s. The dispersion is expected in heterogeneous entry/exit scenes; the methodological point is that the same explicit cutter–follower event

definition yields coherent TTC summaries on *exiD* without redesigning the pipeline.

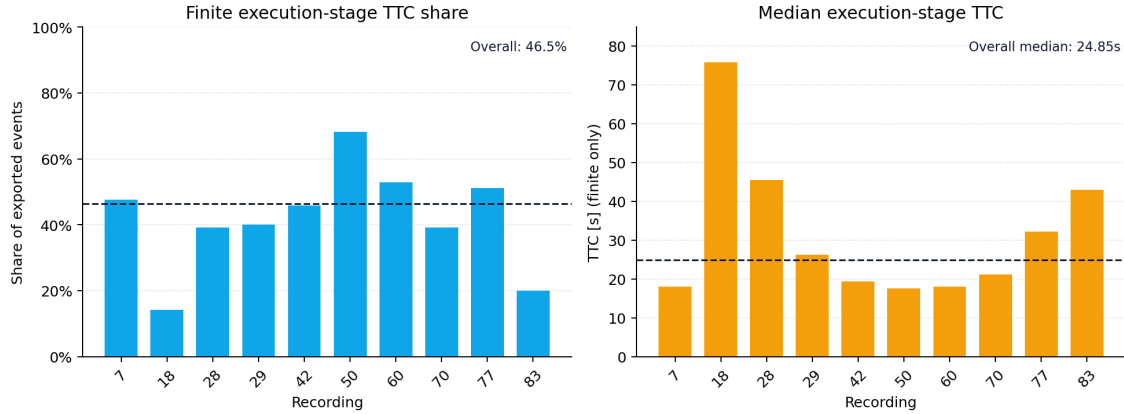


Figure 6.6: Execution-stage TTC summary for the exploratory *exiD* extension. Left: share of exported cut-ins with finite TTC in each recording. Right: median execution-stage TTC among finite-TTC events only. Dashed lines show the pooled overall values (46.5% finite share; median 24.85 s). Reporting both panels avoids conflating TTC severity with TTC availability, since TTC is undefined for non-closing interactions.

Taken together, the *exiD* results show a limited but useful form of transfer. The framework is not tied to one exact highD road geometry, and the same minimal-input event definition still yields interpretable cut-in sets together with coherent pairwise THW/TTC summaries in a related dataset family. At the same time, the extension remains intentionally lightweight: only 10 recordings are used, the full reference-label agreement study is not repeated, and no claim is made that *exiD* provides evidence on the same level as the main highD evaluation.

Chapter 7

Discussion

7.1 Answers to the research questions

RQ1 (reconstruction accuracy): Same-lane neighbour reconstruction from geometry reaches mean accuracy >0.9998 against highD reference labels. Cut-in detection achieves $F_1 = 0.999$, which confirms that the minimal-input path reproduces the reference structure.

RQ2 (cut-in analysis): The framework extracts 10,023 validated cut-ins with explicit cutter–follower pairing. Multi-indicator analysis shows that THW, TTC, and DRAC capture different severity aspects: THW flags 20.1% of events as risky, while DRAC flags only 0.08% as hard-braking.

RQ3 (context signatures): The 16-bit SFC signature achieves 99.49% exact row agreement with reference occupancy over 1.4 M stage rows. Exploratory clustering yields seven decision-stage archetypes, with the most constrained archetype showing $1.32\times$ higher risk prevalence.

RQ4 (predictive value): Decision-stage features predict execution-stage risk with AUC 0.82. SFC context features alone achieve AUC 0.62, which shows that spatial context signatures carry useful standalone predictive information, although they do not significantly improve AUC over the kinematic-only model.

7.2 Interpretation and implications

The principal contribution is methodological: the framework separates observation, derived relation, event definition, and risk label into auditable layers. By treating lane and neighbour identifiers as reconstructable intermediate products rather than immutable dataset fields, the pipeline’s assumptions become explicit and measurable against reference labels.

At first glance the framework may seem straightforward: infer lanes, recover leader–follower relations, and then compute indicators. The non-obvious contribution is the enforced evaluation boundary linking these steps. Many related studies operate downstream of scenario extraction and therefore inherit dataset-specific pairing or neighbour assumptions implicitly. This thesis makes those prerequisites explicit,

reconstructs them from minimal inputs, and validates them before drawing risk conclusions.

This design has three practical implications. First, explicit cutter–follower pairing ensures that safety indicators describe the intended interaction rather than a generic predecessor relation. Second, the multi-indicator results (THW, TTC, DRAC) show that risk interpretation is indicator-dependent and phase-dependent [7, 3], emphasizing the need for transparent indicator choice. Third, the LORO-CV prediction experiment shows that the framework produces predictive features: achieving AUC 0.82 from features available before the execution stage means the outputs support early risk identification, not merely reference-label reproduction. The ablation showing AUC 0.62 for SFC features alone further shows that the context signatures encode risk-relevant spatial information as a standalone signal, even though most of the predictive strength is already captured by the kinematic features.

From a software-engineering perspective, the work contributes an auditable data-processing and validation pipeline rather than only a traffic-analysis result. The central engineering properties are explicit input contracts, reconstructable intermediate artefacts, reference-based testing, and reproducible outputs. These properties matter because downstream models, dashboards, and safety assessments are only as reliable as the scenario extraction logic that feeds them. In a safety-critical software context, making preprocessing assumptions explicit is valuable not only for reproducibility but also because silent pairing or reconstruction errors can otherwise propagate into misleading validation evidence.

A smaller supporting contribution is the adapted inspection dashboard. It was not a user-facing evaluated system, but it made manual sanity checks of onset alignment, interaction pairing, and decoded occupancy practical across 60 recordings, which in turn supported the trustworthiness of the final pipeline.

Transfer beyond highD. The exploratory exiD extension in Section 6.9 makes the transfer question more concrete. Because exiD belongs to the same broader dataset family but focuses on highly interactive highway entry/exit scenes rather than straight motorway segments, it is a useful stress test of whether the framework depends on one particular road geometry. Reusing the same cut-in event definition, cutter–follower pairing logic, and pairwise risk-summary path on 10 exiD recordings still produces coherent event sets and interpretable qualitative cases, which supports the claim that transfer is not blocked by the event definition itself. At the same time, the extension is deliberately limited: we do not repeat the full highD-style reference agreement or predictive evaluation, so the exiD results should be read as an

illustrative probe of transferability rather than as full cross-dataset validation. Older public freeway datasets such as NGSIM remain relevant future targets. Appendix D records a narrow scoping probe showing that the downstream event-mining path can be exercised after a thin NGSIM adapter, but because that probe depends on dataset-provided relation fields and inherits known NGSIM data-quality issues, it does not change the conclusion that full NGSIM transfer requires broader harmonisation.

Per-recording variability. The cross-recording perspective also reveals notable variability. Risk prevalence ranges from 11.1% to 34.1% across the 60 recordings (median 20.0%), which likely reflects differences in traffic density, speed distributions, and lane configurations across the six recording sites. LORO-CV fold-level AUC ranges from approximately 0.65 to 0.95, with lower-performing folds concentrated in recordings with fewer events or atypical speed distributions. This variability underscores the importance of the leave-one-recording-out protocol: pooled cross-validation would overestimate performance by allowing the model to learn site-specific patterns. The limited exiD extension supports this interpretation: once the framework moves to a different highway geometry, transfer remains possible, but the surrounding traffic mix and event yield become more heterogeneous immediately.

Table 7.1 maps each contribution claim to the supporting evidence.

Table 7.1: Contribution claims and supporting evidence.

Contribution	Evidence	Significance
Minimal-input reconstruction	Accuracy >0.9998; $F_1 = 0.999$	Enables cut-in mining without dataset-specific relation columns.
Pair-consistent indicators	10,023 events; THW/TTC/DRAC disagreement analysis	Ensures indicators describe the correct interaction pair.
Validated context signatures	99.49% row agreement; 7 archetypes; peak risk ratio 1.32	Compact, reversible descriptors supporting scene retrieval and grouping.
Predictive evaluation	AUC 0.82 (all); AUC 0.62 (SFC alone)	Features carry predictive value; SFC provides a useful standalone signal, but not a significant gain over kinematics alone.
Reproducible evaluation	Two-path design; parameter table; released pipeline	Separates method inputs from reference labels and supports auditable re-runs.

7.3 Threats to validity

Construct validity: Risk is operationalised via a THW threshold; alternative definitions (TTC, DRAC, post-encroachment measures) would label different event subsets. The threshold sensitivity analysis in Section 6.5 makes this dependence explicit. The 0.7 s threshold is anchored in a single perceived-criticality study [10]; guidance values for safety-critical situations are typically higher (1.8–2.0 s), so the operational risk label should be interpreted as identifying very short gap-acceptance events rather than universally “dangerous” interactions. Additionally, highD lacks wider upstream/downstream context that may influence lane-change decisions [15].

Internal validity: Reconstruction depends on lane-marking accuracy and coordinate conventions. Errors in these inputs can shift the onset t_0 and bias indicators. This is mitigated through geometry model selection against dataset-provided indicators and independent reconstruction variants that agree closely with the reference path. A further consideration is that highD positions are derived from overhead drone video via computer-vision tracking; the resulting measurement noise differs from on-vehicle sensing (e.g., radar or lidar), so the absolute indicator values reported here should not be directly compared to values from instrumented-vehicle studies without accounting for positional uncertainty.

External validity: The main quantitative results are based on *highD* highway recordings collected at six German motorway sites under free-flow to moderately congested conditions. The framework’s assumptions (straight road segments, clearly marked lanes, and moderate traffic density) may not hold for urban driving, ramp merges, construction zones, or different sensing setups. The small *exiD* extension shows that the event-mining core can be moved to a related highway dataset with entry/exit interactions, but it remains limited to 10 recordings and does not establish full cross-dataset validity because the full reference-agreement and prediction stack is not repeated there.

Conclusion validity: The thesis reports descriptive differences, sensitivity analyses, and cross-validated predictions rather than causal claims about driver behaviour. Adding the Friedman test and the Bonferroni-corrected Wilcoxon follow-up tests makes the statistical argument stronger, especially for the claim that the main gap is between the SFC-only model and the two models that include kinematic information. At the same time, the 60-fold LORO-CV setup still produces correlated folds because the training sets overlap heavily, so these p -values should be treated as supportive evidence rather than as fully independent significance tests. A broader repeated-measures comparison across more model variants would strengthen this

part further.

7.4 Future work

Five directions for future work emerge from this thesis:

1. **Full cross-dataset validation:** extending the current exploratory exiD extension into a calibrated multi-dataset study on more exiD recordings and additional datasets such as inD, rounD, or harmonised NGSIM subsets, including coordinate harmonisation, kinematic smoothing, and representation-level validation beyond event mining.
2. **Parameter sensitivity:** systematic ablation over relevance-distance, pair-persistence, and stage-boundary criteria, with manual review of disagreement cases.
3. **Archetype transferability:** testing whether decision-stage context archetypes generalise across datasets, road classes, and prediction tasks.
4. **Richer predictive models:** sequence models over full stage windows (rather than stage-level aggregates) to exploit the temporal structure of reconstructed features.
5. **Per-frame DRAC profiles:** computing frame-level DRAC time series for more granular deceleration-based risk assessment.

Chapter 8

Conclusion

This thesis presented Traj2Rel-SFC, a minimal-input framework that reconstructs lane and neighbour relations from trajectory geometry, detects cut-ins as explicit cutter–follower events, computes pair-specific safety indicators (DHW, THW, TTC, DRAC), and encodes surrounding traffic context as reversible 16-bit SFC signatures.

Evaluated on 60 *highD* recordings with *highD* identifiers used only as reference labels, the framework achieves mean reconstruction accuracy >0.9998 , cut-in detection $F_1 = 0.999$, and context-signature agreement of 99.49% over 1.4 M stage rows.

Three lines of evidence demonstrate that the contribution extends beyond reconstruction accuracy. First, multi-indicator analysis shows that THW, TTC, and DRAC capture different severity aspects, with risk interpretation depending on indicator choice and interaction phase. Second, decision-stage features predict execution-stage risk with AUC 0.82, and SFC features alone achieve AUC 0.62, showing that the framework produces features with predictive value even though the kinematic-only ablation remains strongest. Third, exploratory context-archetype analysis shows that the SFC signatures support interpretable scene grouping linked to differential risk outcomes.

The thesis contribution is therefore not that cut-ins or surrogate indicators were previously unknown. Rather, it surfaces usually hidden preprocessing assumptions – lane assignment, neighbour reconstruction, interaction pairing, and local-context encoding – and validates them within one reproducible pipeline. That is the main reason the approach matters despite being conceptually simple at the level of its individual parts.

This is also where the software-engineering relevance lies. The framework defines explicit input boundaries, reconstructable intermediate products, and reproducible outputs that can be checked against reference labels. As a result, it is relevant not only for analysing *highD* but also as a basis for portable tooling for scenario mining, safety analysis, and validation workflows in automated-driving software engineering.

A small exploratory *exiD* extension further supports this claim: reusing the same lane-reconstruction and cut-in mining logic on 10 entry/exit recordings still produces coherent event sets and interpretable qualitative cases without redesigning the pipeline. However, the limitation is equally important. The *exiD* add-on is

intentionally lightweight and does not repeat the full highD-style reference agreement or predictive evaluation. It therefore illustrates basic transferability, but full multi-dataset validation remains future work.

The result is a reproducible, auditable pipeline for cut-in analysis that serves as a basis for cross-dataset transfer, richer predictive models, and downstream applications requiring pair-consistent interaction features.

Bibliography

- [1] R. Krajewski, J. Bock, L. Kloecker, and L. Eckstein. The highD dataset: A drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems. In *Proceedings of the IEEE 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018. DOI: <https://doi.org/10.1109/ITSC.2018.8569552>.
- [2] T. Moers, L. Vater, R. Krajewski, J. Bock, A. Zlocki, and L. Eckstein. The exiD dataset: A real-world trajectory dataset of highly interactive highway scenarios in Germany. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, pages 958–964, 2022. DOI: <https://doi.org/10.1109/IV51971.2022.9827305>.
- [3] C. Wang, Y. Xie, H. Huang, and P. Liu. A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. *Accident Analysis & Prevention*, 157:106157, 2021. DOI: <https://doi.org/10.1016/j.aap.2021.106157>.
- [4] C. Johnsson, A. Laureshyn, and C. D’Agostino. A relative approach to the validation of surrogate measures of safety. *Accident Analysis & Prevention*, 161:106350, 2021. DOI: <https://doi.org/10.1016/j.aap.2021.106350>.
- [5] S. Hwang, K. Lee, H. Jeon, and D. Kum. Autonomous vehicle cut-in algorithm for lane-merging scenarios via policy-based reinforcement learning nested within finite-state machine. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):17594–17606, 2022. DOI: <https://doi.org/10.1109/TITS.2022.3153848>.
- [6] X. Wang, M. Yang, and D. Hurwitz. Analysis of cut-in behavior based on naturalistic driving data. *Accident Analysis & Prevention*, 124:127–137, 2019. DOI: <https://doi.org/10.1016/j.aap.2019.01.006>.
- [7] H. Zhou and M. Itoh. How does a driver perceive risk when making decision of lane-changing? *IFAC-PapersOnLine*, 49(19):60–65, 2016. DOI: <https://doi.org/10.1016/j.ifacol.2016.10.462>.
- [8] Q. Shangguan, T. Fu, J. Wang, S. Fang, and L. Fu. A proactive lane-changing risk prediction framework considering driving intention recognition and different

- lane-changing patterns. *Accident Analysis & Prevention*, 164:106500, 2022. DOI: <https://doi.org/10.1016/j.aap.2021.106500>.
- [9] Y. Liu, J. Zhang, N. Lyu, and Q. Zhao. Predicting lane change maneuver and associated collision risks based on multi-task learning. *Accident Analysis & Prevention*, 209:107830, 2025. DOI: <https://doi.org/10.1016/j.aap.2024.107830>.
- [10] R. Tscharn, F. Naujoks, and A. Neukum. The perceived criticality of different time headways is depending on velocity. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58:1043–1052, 2018. DOI: <https://doi.org/10.1016/j.trf.2018.08.001>.
- [11] A. Elfes. Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6):46–57, 1989. DOI: <https://doi.org/10.1109/2.30720>.
- [12] B. Moon, H. V. Jagadish, C. Faloutsos, and J. H. Saltz. Analysis of the clustering properties of the Hilbert space-filling curve. *IEEE Transactions on Knowledge and Data Engineering*, 13(1):124–141, 2001. DOI: <https://doi.org/10.1109/TKDE.2001.859372>.
- [13] N. Deo and M. M. Trivedi. Convolutional social pooling for vehicle trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. DOI: <https://doi.org/10.1109/CVPRW.2018.00196>.
- [14] J. Archer. Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: A study of urban and suburban intersections. PhD thesis, Royal Institute of Technology (KTH), Stockholm, 2005.
- [15] Z. Wang and Y. Gao. Lane change inconsistencies in the highD dataset. *Findings*, 2025. DOI: <https://doi.org/10.32866/001c.132341>.
- [16] U.S. Federal Highway Administration. Interstate 80 freeway dataset. FHWA-HRT-06-137, 2006. <https://www.fhwa.dot.gov/publications/research/operations/06137/index.cfm>.
- [17] U.S. Federal Highway Administration. US Highway 101 dataset. FHWA-HRT-07-030, 2007. <https://www.fhwa.dot.gov/publications/research/operations/07030/index.cfm>.

- [18] V. Punzo, M. T. Borzacchiello, and B. Ciuffo. On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. *Transportation Research Part C: Emerging Technologies*, 19(6):1243–1262, 2011. DOI: <https://doi.org/10.1016/j.trc.2010.12.007>.
- [19] M. Montanino and V. Punzo. Making NGSIM data usable for studies on traffic flow theory: Multistep method for vehicle trajectory reconstruction. *Transportation Research Record*, 2390(1):99–111, 2013. DOI: <https://doi.org/10.3141/2390-11>.
- [20] B. Coifman and L. Li. A critical evaluation of the Next Generation Simulation (NGSIM) vehicle trajectory dataset. *Transportation Research Part B: Methodological*, 105:362–377, 2017. DOI: <https://doi.org/10.1016/j.trb.2017.09.018>.

Chapter A

Terminology and highD Field Mapping

This appendix maps the descriptive lane and neighbour terms used in the thesis to the corresponding highD column names.

Table A.1: Mapping between thesis terminology and highD columns (tracks file).

Concept in this thesis	highD column	Meaning
Lane index (reference labels)	<code>laneId</code>	Discrete lane identifier per frame.
Same-lane leader (ahead)	<code>precedingId</code>	Track id of closest vehicle ahead in the same lane (0 if none).
Same-lane follower (behind)	<code>followingId</code>	Track id of closest vehicle behind in the same lane (0 if none).
Left-lane leader	<code>leftPrecedingId</code>	Closest vehicle ahead in the left adjacent lane (0 if none).
Left-lane alongside	<code>leftAlongsideId</code>	Vehicle alongside in the left adjacent lane (0 if none).
Left-lane follower	<code>leftFollowingId</code>	Closest vehicle behind in the left adjacent lane (0 if none).
Right-lane leader	<code>rightPrecedingId</code>	Closest vehicle ahead in the right adjacent lane (0 if none).
Right-lane alongside	<code>rightAlongsideId</code>	Vehicle alongside in the right adjacent lane (0 if none).
Right-lane follower	<code>rightFollowingId</code>	Closest vehicle behind in the right adjacent lane (0 if none).

HighD distributes each recording as three CSV files: `XX_tracks.csv` (per-frame trajectories), `XX_tracksMeta.csv` (per-track metadata), and `XX_recordingMeta.csv` (recording metadata with lane markings).

Chapter B

Qualitative Inspection Dashboard

Quantitative agreement metrics provide an overall assessment of reconstruction quality, but they do not on their own reveal whether failures are systematic (e.g., due to direction conventions, onset misalignment, or lane-boundary interpretation). To support qualitative inspection, we use an interactive inspection dashboard that synchronises (i) aerial video playback and track overlays, (ii) the currently selected cut-in event and its cutter–follower pair, and (iii) the derived stage window, indicators, and context signature for the current frame.

The dashboard was not developed from scratch. It builds on the official *highD* tools repository, which provides functions for handling and visualising *highD* data in Matlab and Python¹. For this thesis, we adapted and extended that tooling with cut-in-specific event selection, cutter–follower highlighting, stage-window synchronisation, and context-signature/indicator panels needed for qualitative inspection of the reconstructed interaction pair. For the small exiD extension, the same inspection concept was reused with the official drone-dataset-tools as a base so that *highD* and *exiD* cases could be reviewed through a consistent event-centred viewer.

The dashboard was used to inspect representative cases and extremes (e.g., lowest-TTC events) and to check three concrete aspects of the pipeline qualitatively: (i) whether the onset t_0 aligned with the first visually stable cutter–follower relation, (ii) whether the decoded 3×3 occupancy signature matched the visibly occupied neighbouring regions, and (iii) whether extreme TTC or THW cases reflected genuinely tight insertions rather than obvious indexing or lane-assignment errors. These checks provide qualitative support for the quantitative agreement results, but they are not reported as a separate manual evaluation dataset. No formal user evaluation was conducted. The dashboard served as a support tool for validation and communication rather than as a standalone thesis contribution: it was not used to manually correct events or labels, and all reported results are produced by the framework’s fixed rule-based components plus the released analysis scripts. Figure B.1 shows the interface used during this qualitative inspection.

¹Official repository: <https://github.com/RobertKrajewski/highD-dataset>.

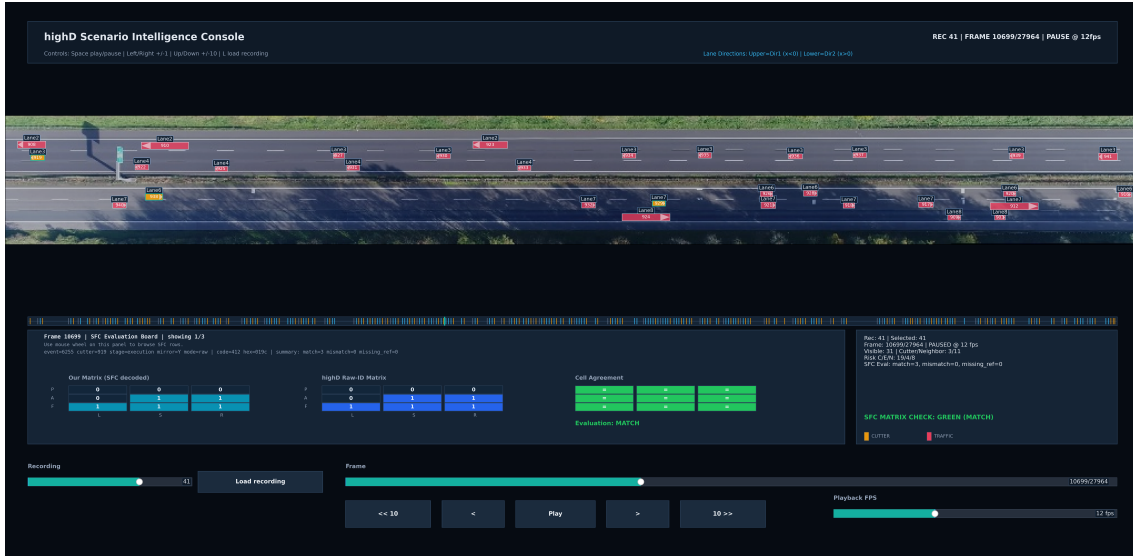


Figure B.1: Schematic layout of the interactive inspection dashboard used for qualitative review. The implemented dashboard is based on a modified version of the official *highD* visualisation tools and was extended with cut-in-specific overlays, event selection, and stage/context displays.

To make the *exiD* reuse more concrete, Figure B.2 shows the same inspection concept applied to an *exiD* recording. The interface logic is unchanged: the viewer still centres on one event, shows the surrounding traffic in the current frame, and keeps the same event-oriented inspection workflow. The main difference is only in the dataset-specific background rendering and geometry.

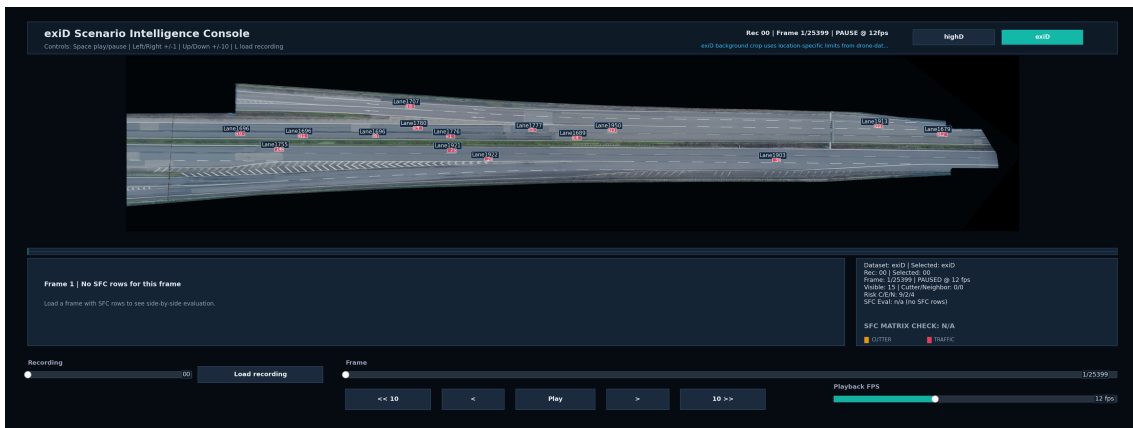


Figure B.2: Example *exiD* dashboard view used in the exploratory extension. The same event-centred inspection workflow was reused for *exiD* scenes, with dataset-specific background rendering but unchanged cut-in review logic.

Chapter C

Plain-Language Algorithms

C.1 Lane change detection (stable blocks)

1. For each vehicle, read its lane id over time (reference) or inferred lane index (XY-lane).
2. When the lane id changes from L_1 to L_2 , check:
 - the vehicle stayed in L_1 for at least N_{before} frames before the change,
 - the vehicle stayed in L_2 for at least N_{after} frames after the change.
3. If both checks pass, emit a lane change event from L_1 to L_2 .

C.2 Cut-in detection (follower identification)

1. Start from a lane change event where the cutter enters a target lane.
2. Move forward frame by frame within a configured search window.
3. In each frame, look for any vehicle in the target lane whose same-lane predecessor equals the cutter id.
4. When this first happens, that vehicle is the follower and the frame is t_0 .
5. Require that the follower relation holds for a minimum number of frames to reduce noise.

C.3 DHW, THW, and TTC for the interacting pair

- DHW is computed as the longitudinal gap between the cutter's rear and the follower's front.
- THW is DHW divided by follower speed (with safe handling of near-zero speeds).

- TTC is DHW divided by closing speed when the follower is gaining on the cutter; otherwise TTC is treated as infinite.

C.4 Reference 3×3 occupancy grid from highD neighbour ids

For a given ego vehicle and frame, highD provides surrounding-vehicle ids for left/right/preceding/following/alongside positions. We map these to the 3×3 grid by setting a cell to 1 if the corresponding neighbour id is non-zero; otherwise 0. The same grid can be produced by geometry-based reconstruction, allowing code-level comparison via SFC signature matching.

Chapter D

Exploratory NGSIM Feasibility Probe

During scoping, we also ran a narrow feasibility probe on the NGSIM freeway subsets US-101 and I-80, both released as video-derived freeway trajectory datasets by the U.S. Federal Highway Administration [17, 16]. The purpose was deliberately narrower than the exiD extension in Chapter 6: we wanted to test whether the downstream lane-change and cut-in mining logic could be exercised on an older public freeway dataset after a thin harmonisation layer. This appendix records that probe so the exclusion of NGSIM from the main evidence base is itself auditable.

The key methodological point is that this probe does *not* reproduce the thesis evaluation boundary. The NGSIM feasibility adapter reads NGSIM’s released `Lane_ID`, `Preceding`, `Following`, `Space_Headway`, and `Time_Headway` columns and feeds them into the downstream event-mining path after light cleaning. It therefore does not reconstruct lane membership or same-lane relations from minimal inputs, does not validate reconstructed relations against reference labels, and does not run the predictive evaluation stack. The appendix should thus be read as a portability probe for the *event-mining core* only, not as additional evidence for RQ1 or as a second benchmark.

This distinction is important because NGSIM trajectory quality has been questioned in the traffic-flow literature. Prior work reported substantial measurement errors and physically implausible differentiated kinematics in the released trajectories, motivating multistep reconstruction and smoothing before theory-heavy use [18, 19, 20]. For the present thesis, that literature supports the conservative decision to keep NGSIM outside the main validation chain unless a broader harmonisation and revalidation effort is performed.

D.1 Probe setup and fixed assumptions

The probe used the following fixed assumptions:

- scope restricted to the two freeway subsets (`us-101`, `i-80`) so the road class remained aligned with the highway cut-in problem,
- process one location at a time to avoid mixing site-specific vehicle ids,

- drop exact duplicate rows on (`Vehicle_ID`, `Frame_ID`, `Global_Time`),
- split repeated source ids when `Global_Time` jumps by more than 1 s,
- remap `Preceding` and `Following` after the split so neighbour links refer to the derived track-instance ids,
- treat `Time_Headway` > 100 s as placeholder/invalid values rather than as genuine car-following measurements,
- adapt the event-persistence parameters to NGSIM’s 10 Hz rate: 10 stable frames before the lane change, 10 stable frames after, a 20-frame follower search window, a 6-frame maximum relation delay, a 5-frame minimum relation duration, and a 10-frame precheck window.

The strongest claim supported by this probe is modest: a thin adapter is sufficient to produce non-empty, qualitatively interpretable cut-in candidate sets and coarse pairwise THW/TTC summaries on both NGSIM freeway sites. That is useful as software-engineering evidence that the downstream event-mining path is not hard-coded to one file format or one exact dataset family.

The probe does *not* show that the thesis’s minimal-input reconstruction layer transfers to NGSIM, because the adapter depends directly on NGSIM’s released lane and neighbour semantics. It also does not establish that the candidate yields are behaviourally meaningful enough for cross-dataset comparison: the very high cut-in-to-lane-change ratio on I-80, the large duplicate removal on US-101, and the non-zero negative-gap share are precisely the kinds of signals that justify a deeper harmonisation pass before any substantive comparison with highD or exiD. Accordingly, NGSIM remains future work at the level of full reconstruction, agreement auditing, and predictive evaluation, even though this appendix records preliminary evidence that a carefully delimited transfer path exists.

Figure D.1 and Table D.1 summarise the resulting artefacts.

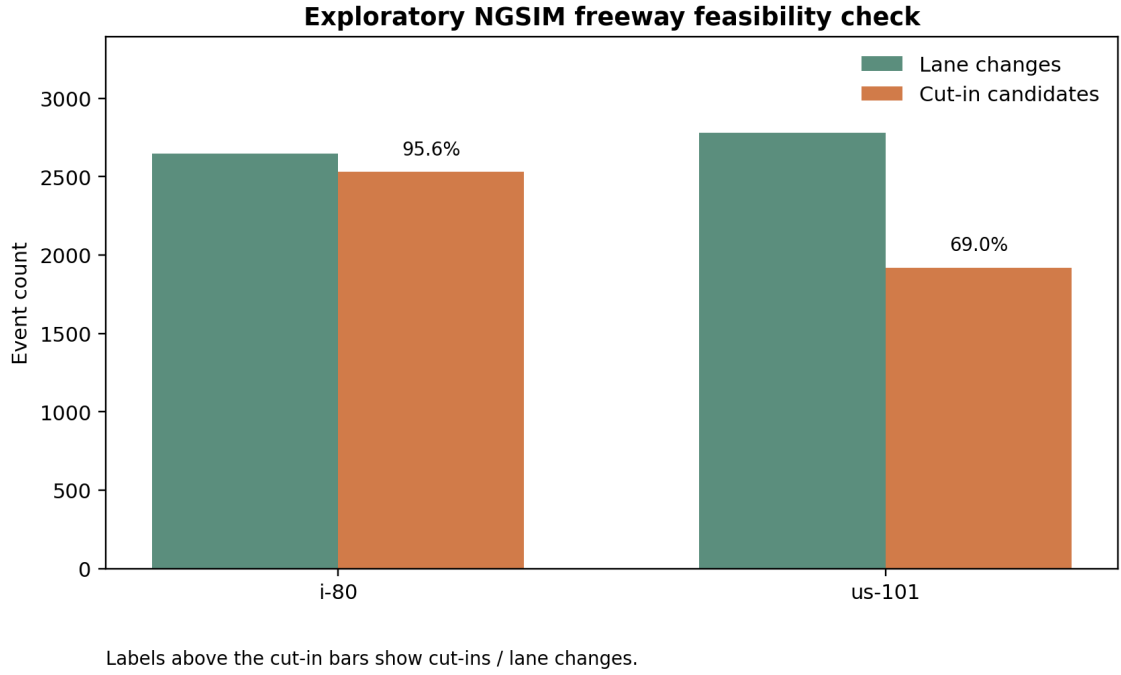


Figure D.1: Exploratory NGSIM feasibility probe on the two freeway subsets. Bars show detected lane changes and cut-in candidates after harmonisation; labels above the cut-in bars show cut-in candidates as a share of detected lane changes. The figure is reported as a scoping artefact only and is not interpreted as cross-dataset validation performance.

Table D.1: Summary of the NGSIM feasibility probe. Negative-gap share denotes candidate events with at least one frame where reported `Space_Headway` minus cutter length became negative; it is used only as a harmonisation stress indicator, not as a behavioural result.

Metric	I-80	US-101
Raw rows	4,566,387	4,802,933
Duplicates removed	0	704,000
Track instances	5,678	6,101
Lane changes	2,647	2,780
Cut-in candidates	2,530	1,919
Cut-in ratio	95.6%	69.0%
Median min THW	2.05 s	1.60 s
Finite TTC share	67.5%	54.3%
Negative-gap share	4.0%	2.4%