SUMO User Conference 2026

TIBconferencesession

TIB-OP will set DOI with \TIBdoi

Published: 1970-01-01

Reproducing Weekly Crash Patterns Using a Minimalist Driver Error Model

Can Simple Driver Inattention Models Reproduce Spatio-Temporal Crash Distributions in Microscopic Traffic Simulation?

Andreas Leich[1], Ronald Nippold [1], and Peter Wagner[1,2]

[1]DLR Institute for Transportation Systems, Germany

[2]TU Berlin, Institute for Land- and Sea Traffic, Germany

*Correspondence: Ronald Nippold, ronald.nippold@dlr.de

Abstract: This study compares weekly patterns of real crashes at one of Berlin's most crash-prone intersections ("Schlesisches Tor") with simulated crashes generated using a simple "blackout" driver error model in SUMO. We analyze both temporal trends and spatial distributions of crashes between observed and simulated data.

The crash model induces a 3-second phase of reduced situational awareness during which vehicles continue moving at constant speed, with activation probabilities calibrated for different traffic modes (passenger cars, trucks, bicycles). Simulated crash patterns closely match observed weekly trends with distinct morning and evening peaks on weekdays and flatter curves on weekends.

Spatially, simulated rear-end collisions cluster along main roads, matching observations. The simulation, however, overpredicts turning-area collisions and underpredicts conflicts at pedestrian crossings, revealing limitations in modeling pedestrian behavior (e. g., hesitation, red-light crossing).

Keywords: Spatio-temporal crash analysis, crash pattern reproduction, urban intersection safety, microscopic traffic simulation, SUMO (Simulation of Urban MObility)

1 Introduction

Research in traffic safety has a long tradition spanning several decades, with foundational work establishing the statistical and theoretical underpinnings of road crash analysis [1], [2]. Early attempts to combine traffic simulation with safety assessment were performed as far back as the late 20th century, followed by a long series of studies that have equipped traffic simulations – including mesoscopic and macroscopic ones –

with the ability to assess traffic safety [3], [4]. These efforts reflect an increasing recognition that understanding crash mechanisms requires not only observational data but also experimental approaches that can isolate causal factors.

Since traffic simulation became a handy tool for transportation research, numerous approaches have been developed to study traffic safety and reveal the underlying mechanisms behind road crashes. Microscopic traffic simulations, in particular, offer the ability to model individual vehicle interactions at high temporal resolution, making them well-suited for investigating specific crash scenarios [5]. Safety surrogate measures (SSM) have been integrated into simulation platforms such as SUMO to provide additional metrics beyond traditional crash counts [4], [6]. Nevertheless, the general consensus remains undisputed: “the more objects, the more crashes” is a fundamental relationship in traffic safety research [4].

However, while microscopic traffic simulation models are increasingly used for safety assessment, their ability to reproduce realistic crash patterns remains poorly understood. Despite the availability of many approaches, a reliable method is still missing [4]. This gap exists partly because traffic safety represents one of the more difficult objects of science, where a lot of work has been performed but something like an overarching theory is still lacking – too much reliance on correlation rather than causation persists. Crashes are rare events that lack objective observers and may even include strong history-dependence: a tiny decision made a couple of seconds earlier can determine whether a real crash happens or only a dangerous situation occurs [4].

Up to now, not enough comparison and knowledge about deviations between real-world crash data and simulated crashes exist in the literature. While other studies have explored relationships between crashes and conflicts within simulation environments using built-in SSM devices, these investigations remain largely incomplete [4]. The assignment of crash severity, the role of traffic speed, and spatial distribution patterns represent open research questions that have yet to be systematically addressed through rigorous empirical comparison. This represents a significant limitation given that a reliable microscopic traffic simulation might be capable of modeling realistic traffic crashes.

Building on prior work of the authors in this area, this study reconstructs the weekly patterns of real crashes at one of Berlin’s most crash-prone intersections (“Schlesisches Tor”) with simulated crashes generated using a “blackout” driver error model in SUMO. The intersection, which consists of two sub-intersections and has a total dimension of approximately 170×170 meters, provides an ideal testbed to compare simulated with real crashes. By utilizing averaged weekly traffic demand derived from multi-year municipal count data, this research aims to contribute to filling the existing knowledge gap in validation of simulated crash patterns against empirical observations.

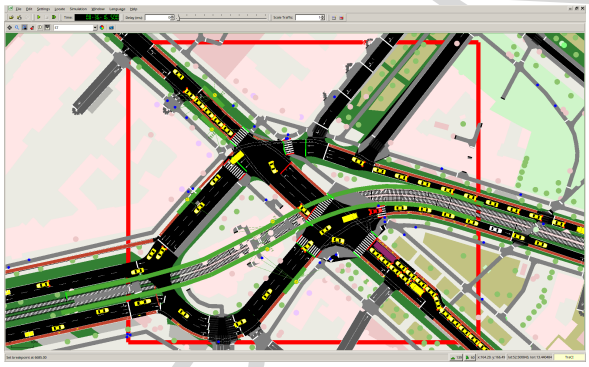
2 Data used, Simulation Setup and Crash Modeling

2.1 Network and Data Basis

The microscopic traffic simulation for this study is based on the intersection “Schlesisches Tor” in the central part of Berlin (Germany), which – measured by absolute crash figures – was identified as one of the most crash-prone locations in the entire city during the period from 2001 to 2024. The intersection features a complex geometry, as well as an elevated subway line and a station situated in the center. Traffic volume

on all streets remains relatively high – for both motorized vehicles and cyclists – even during the fringe hours of the day, as the immediate vicinity of the intersection can be characterized as a nightlife district. Figure 1 shows location, geometry, and immediate surroundings of this intersection.

(a) *"Real-world" geometry (source: Geoportal Berlin, Data licence Germany - Zero - Version 2.0, captured: May 5, 2026).*



(b) *Corresponding microscopic traffic simulation network of the intersection "Schlesisches Tor", Berlin in Eclipse SUMO.*

Figure 1. *Representation of the analyzed intersection "Schlesisches Tor".*

The network geometry as displayed in Figure 1 b was imported from OpenStreetMap (OSM) and corrected in terms of the number of lanes, their connections, assignments to turning lanes, and cycle path layouts in order to match current infrastructure conditions. Since signal timing data were unavailable and varied over time, the traffic light phases were manually adjusted based on field observations [4].

For validation purposes, the simulation utilizes historical police-reported crash data from the time period 2001 – 2024 covering all modes of transport (i. e., passenger cars, trucks, bicycles). While flow data is available for a subset of the years 2016 – 2024 via road detectors [7], the traffic demand profile was reconstructed based on aggregated weekly courses to ensure temporal consistency. To gain an impression of the average traffic load at this intersection – as well as the accuracy with which traffic volumes are represented in the simulation – the following Figure 2 compares actual measured traffic

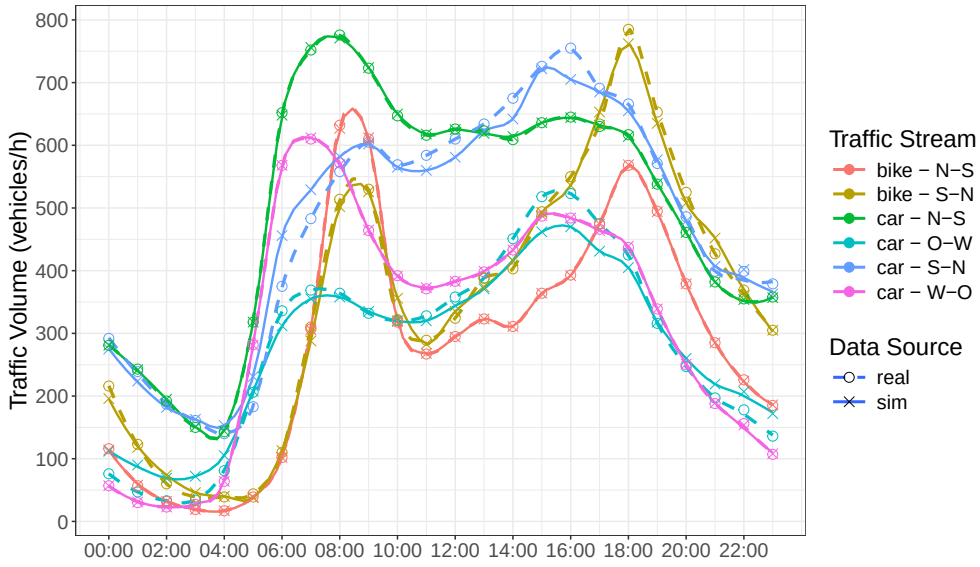


Figure 2. *Measured and simulated traffic volumes at "Schlesisches Tor" for an average working day.*

volumes with simulated ones for each direction (or for the available detectors) for both motorized traffic and bicycles, based on an average working day. The share of heavy goods traffic has been omitted for the clarity of the overview.

Unlike the initial study [4], which aimed to reproduce crash counts within a single day simulation, this work reproduced the traffic situation during an average week. Given the absence of direct empirical pedestrian counting data for the target scenario, route assignment for pedestrians crossing this intersection was implemented via stochastic path generation calibrated against on-site observational proxies. Daily demand distribution was modeled through five discrete temporal intervals derived from visual surveys and estimated site capacity utilization. Consequently, pedestrian flows are presented as a baseline scenario within the range of estimated demand fluctuations to account for potential variance in absolute flow figures. To maximize statistical robustness, 100 independent simulation runs were executed.

2.2 Crash Mechanism: The Blackout Model

To model traffic crashes in SUMO, we used an approach that is not based on pure collision physics but on driving behavior and attention deficits of the drivers [4]. Vehicles are assigned a probability for transiting into a “blackout” state when entering the critical intersection zone (cf. the red-marked area in Figure 1 b). During this blackout phase, vehicles do not react to environmental changes – such as other road users or traffic signals – for a defined duration of three seconds.

We defined the activation probabilities aligned to the crash rates observed in the 24-year dataset:

- Passenger Cars: $p_{\text{car}} = 0.047$ (4.7 %);
- Trucks: $p_{\text{truck}} = 0.016$ (1.6 %);
- Bicycle: $p_{\text{bicycle}} = 0.005$ (0.5 %).

Upon selection, a delay period is applied between the vehicle entering the red geofence and the onset of the blackout to randomly distribute the positions of the blackout occurrence across the intersection area – also taking into account varying traffic volumes in different traffic situations. This delay is sampled from a uniform distribution within [12 s, 35 s]. Once activated, the blackout state persists for exactly three seconds, after which normal driving behavior resumes. This can be observed graphically in Figure 1 b: The red passenger car, traversing the crosswalk, is currently in blackout state ($actionStepLength = 1$ s), whereas the white vehicle downstream (which has just exited the red geofence) has already completed a blackout phase and returned to the “normal” driving state ($actionStepLength = 3$ s).

Unlike e. g. VISSIM’s “Fehlverhalten” module [8], which stochastically modifies reaction times and signal compliance, or SUMO’s IDM-based car-following models (which lack built-in distraction parameters), the blackout approach decouples error simulation from vehicle dynamics. Implemented via TraCI, it offers a lightweight, mode-specific proxy for driver inattention that can be rapidly parameterized and scaled. The model does not replace detailed car-following or distraction frameworks but provides a pragmatic baseline for spatial-temporal pattern reproduction.

2.3 TraCI-Based Implementation

The crash mechanism is activated and deactivated dynamically via the Traffic Control Interface (TraCI). A custom Python script controls the simulation loop in real-time. The implementation relies on modifying the *actionStepLength* attribute of selected vehicles. By setting the actionStepLength to 3.0 seconds (matching the blackout duration), the vehicle's decision-making processes are paused for that interval, effectively simulating a driver who is unaware of changing traffic conditions [9]. The script performs the following steps every simulation step:

- Detection: Monitor all active vehicles within the geofence (in local intersection coordinates: $x \in [183, 343], y \in [124, 261]$);
- Selection: Apply type-specific probabilities after a time delay from the interval [12 s, 35 s] to select vehicles for blackout state;
- Activation: Update actionStepLength and visualize the state (red color) via TraCI commands;
- Recovery: After expiration of the blackout duration, revert actionStepLength to the standard value ($actionStepLength = 1\,\text{s}$) to let the vehicle return to normal operation. Additionally, convert vehicle color to white to mark former "blackout" vehicles.

Vehicles that have already experienced a blackout are added to a blacklist and will not undergo any further adjustment of the actionStepLength. This approach enables precise temporal and spatial control of the crash cause parameters. By decoupling the logic from SUMO's internal configuration files, this configuration allows for multiple runs without manual intervention.

2.4 Experimental Design

While traffic volume is a primary predictor of crash frequency, the blackout model does not merely replicate exposure relationships. It enables mechanistic exploration of how spatial and temporal heterogeneity in traffic state variables interacts with stochastic driver error. Unlike correlation-based safety performance functions, the simulation allows controlled perturbation of intersection geometry or signal timing while holding exposure constant, isolating the effect of operational or geometric changes on crash distribution. This capability is not accessible through observational data alone.

Absolute crash counts are not directly comparable across the used datasets due to temporal scaling. The simulation uses a weekly average demand profile derived from detector data (2016 – 2024) and accelerates the temporal domain to reproduce crash rates. Blackout probabilities were calibrated to match the empirical crash frequency from the time period 2001 – 2024 per vehicle-hour entering the intersection zone. To account for stochastic variance, 100 independent simulation runs of an entire week were executed and averaged. This multi-run approach minimizes stochastic fluctuations that can occur when modeling rare events such as crashes. However, a direct count comparison is therefore inappropriate; validation relies on rate-normalized temporal patterns and spatial distribution metrics.

3 Methodology

3.1 Temporal Validation

To address the discrepancy between absolute crash counts, it is important to note that the simulation does not reproduce empirical counts directly. Instead, it compresses a representative weekly traffic demand cycle to generate sufficient crash events for spatial-temporal pattern analysis. The blackout activation probabilities were calibrated to match empirical crash rates per vehicle-hour entering the intersection zone, ensuring exposure equivalence at the macroscopic level.

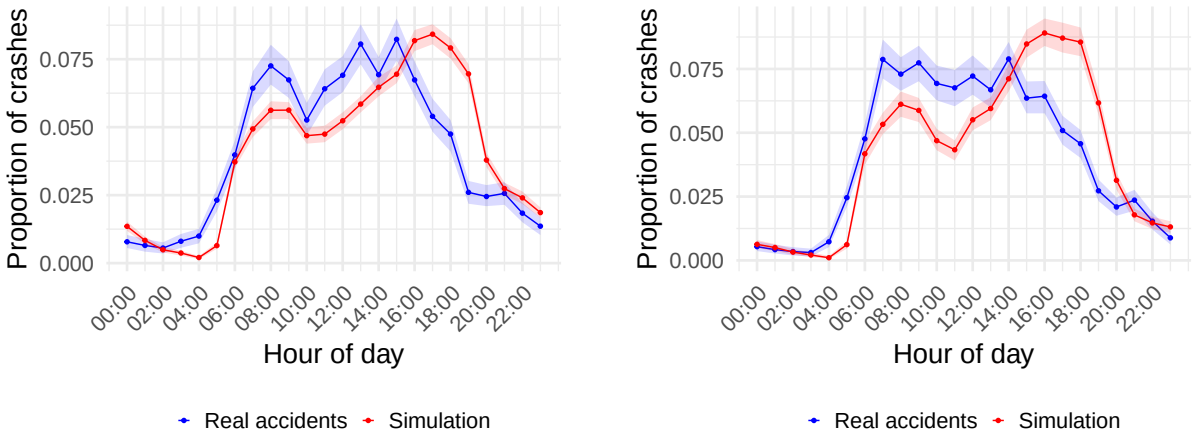
Table 1 summarizes the empirical and simulated crash counts alongside two-sample Cramér–von Mises (CVM) tests comparing hourly and daily crash distributions. Both

Table 1. *Empirical (real-world) versus simulated crash counts and macroscopic temporal distribution tests for "collision" (rear-end) and "junction" (crossing) crashes.*

Measure	**"Collision"**	**"Junction"**
Real crash count (n)	5,233	5,210
Simulated crash count (n)	18,849	8,880
Hourly CVM statistic	508.31	549.72
Hourly permutation p-value	0.001	0.001
Daily CVM statistic	14.94	2.52
Daily permutation p-value	0.001	0.090

Note: CVM = Cramér–von Mises statistic. Simulated counts are higher because blackout probabilities were calibrated to reproduce the 24-year empirical crash rate per vehicle-hour, applied to a single representative weekly demand cycle. Temporal tests assess distributional alignment under equivalent traffic exposure.

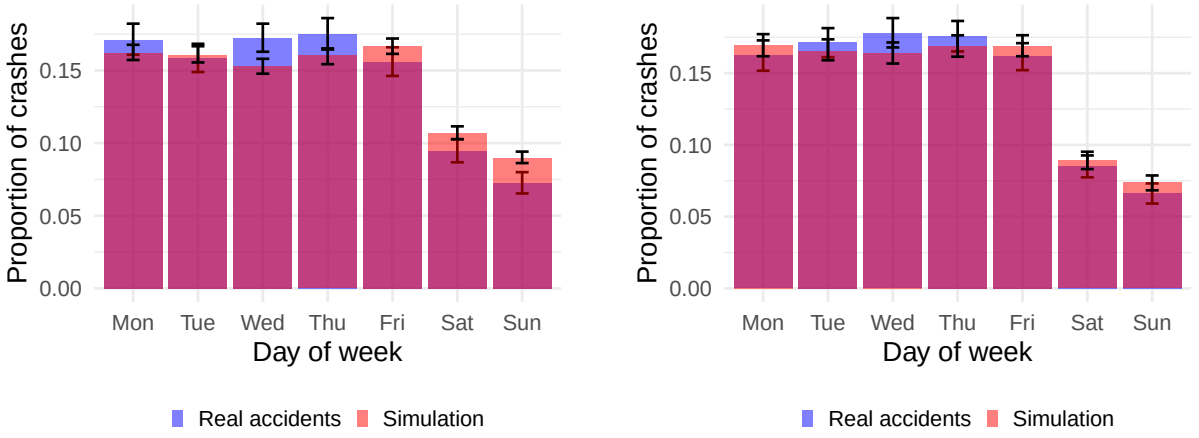
crash types show statistically significant alignment in hourly patterns ($p = 0.001$), reproducing the characteristic morning and afternoon peaks. This degree of correspondence is also evident from Figure 3. Daily patterns also align closely for rear-end collisions ($p = 0.001$), while junction crashes show a marginally non-significant deviation



(a) *"Collision" crashes.* ***(b)*** *"Junction" crashes.*

Figure 3. *Hourly crash proportion comparison with bootstrapped 95 % confidence intervals for an averaged working day.*

($p = 0.090$), likely attributable to stochastic fluctuations in pedestrian-vehicle interactions not fully captured by the current model. Figure 4 displays the level of correspondence for all week days graphically. Please note the flatter weekend profile, indicating that the blackout model scales appropriately with time-dependent traffic demand. This temporal match confirms that the stochastic error trigger correctly interacts with the modeled traffic state fluctuations so that the blackout approach successfully reproduces macroscopic temporal trends despite differences in absolute event counts.



(a) *"Collision" crashes.*

(b) *"Junction" crashes.*

Figure 4. *Daily crash proportion comparison with bootstrapped 95 % confidence intervals for an all day of week.*

3.2 Data Acquisition and Classification

The aim of this study is to compare two different datasets representing crash events: (1) real-world crash records, which serve as the ground truth, and (2) simulated crash events generated using SUMO with the blackout model described in subsection 2.2. To ensure a meaningful comparison in terms of spatial distribution, both datasets were divided into specific crash types. In a first step, the dataset of the real-world crash records was limited to crashes classified as "collisions" (VUtype = 6 in the real-world dataset). These crashes correspond to rear-end collisions in longitudinal traffic. Similarly, the simulated dataset was filtered by these events, with this crash type also being referred to as a "collision" in SUMO. In a second step, intersection or crossing crashes were examined, i. e., crashes at intersections when turning or merging (VUtype = 2 and 3 in the real-world dataset). In SUMO, this crash type is referred to as a "junction" crash.

3.3 Spatial Preprocessing and Framework

Traffic crashes are inherently spatial phenomena; therefore, point pattern analysis is employed rather than standard univariate statistical tests (e. g., t-tests), which do not account for geographic coordinates. The preprocessing pipeline involved three critical steps: coordinate transformation, window definition, and data cleaning.

3.3.1 Coordinate System Alignment

Both datasets were projected into the European Terrestrial Reference System 89 (ETRS89) / UTM Zone 33N (EPSG:25833). This projection ensures that Euclidean distance calculations reflect physical distances in meters, which is a prerequisite for spatial statistical analysis.

3.3.2 Study Area Windowing

To analyze the interaction between the two point clouds within a comparable region, we defined a bounded study area (window). The window was constructed based on the extent of the real-world crash distribution to ensure coverage of all empirical events, with an additional buffer zone added to accommodate potential simulated outliers. Mathematically, the spatial window W is defined as: $W = [x_{\text{min}} - \delta, x_{\text{max}} + \delta] \times [y_{\text{min}} - \delta, y_{\text{max}} + \delta]$, where $(x_{\text{min}}, y_{\text{min}})$ and $(x_{\text{max}}, y_{\text{max}})$ represent the bounding box of the real crash coordinates, and δ represents the buffer size.

3.3.3 Duplicate Handling

In point pattern analysis, exact duplicate coordinates can bias density estimations and segregation tests. We identified and removed duplicate points within each dataset prior to statistical testing using a deterministic exclusion method rather than stochastic jittering. This ensures that the resulting Point Pattern Objects represent distinct spatial events without introducing artificial noise.

3.4 Statistical Analysis: Point Pattern Comparison

The primary objective was to investigate whether the real and simulated crash samples originated from the same underlying probability distribution. We utilized a multivariate approach based on spatial point process theory, implemented via the “spatstat” package in R [10]. Three complementary statistical tests were performed that are described in the following.

3.4.1 Monte Carlo Segregation Test

To test the null hypothesis that real and simulated crashes are spatially independent (i. e., they do not segregate from one another), we employed a Monte Carlo-based segregation test (“segregation.test” in R). This permutation test randomly reassigns the “marks” (labels “real” vs. “simulated”) to the combined set of points n_{total} times. The observed segregation statistic is compared against the distribution generated by these permutations.

- Null Hypothesis (H_0): The real and simulated crash locations are drawn from the same spatial intensity function (i. e., they are intermingled).
- Alternative Hypothesis (H_1): There is a significant preference for segregation or clustering between the two types within the defined window.

3.4.2 Cross-K Function Analysis

To analyze the interaction at specific distance scales, we calculated the bivariate cross-K function ($K_{\text{cross}}(r)$) [11]. This statistic measures the expected number of simulated crashes found within a distance r of an arbitrary real crash. The cross-K function was evaluated across a sequence of radii (r) ranging from 0 meters to k_{max}, with 100 discrete steps. This allowed for the assessment of spatial dependence at multiple scales, ensuring that the simulation captures both local clustering and broader regional crash patterns. To determine if deviations from spatial independence were significant, we generated simulation envelopes using the "envelope" function with global correction ("global = TRUE"). If the observed K_{cross} curve falls outside the 95% confidence envelope of the simulated random patterns, it indicates a statistically significant attraction or repulsion between real and simulated locations.

3.4.3 Cross-Nearest Neighbor Distance (NND)

In addition to aggregate functions, we analyzed the local spatial relationship by computing the nearest neighbor distances from each real crash point to its closest simulated counterpart. This metric provides insight into the "closeness" of the simulation to reality at a micro-scale. The minimum Euclidean distance d_{min} for a real point i is calculated as: $d_i = \min_j \left(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\right)$, where j indexes all simulated crash points. Descriptive statistics (mean, median, min/ max) of this distribution were used to quantify the average spatial discrepancy between the two datasets.

3.5 Software Implementation

All statistical analyses and data visualizations were performed using R version 4.5.2 [12]. Spatial operations relied on the "spatstat" package for point pattern manipulation, while data ingestion was handled via standard CSV parsing tools.

4 Results

The multi-metric approach described in section 3 allows for an evaluation of both global distribution patterns and local clustering tendencies within the intersection geometry. The primary metric for evaluating spatial overlap between real and simulated crash locations was the Segregation Test Statistic (T_1), evaluated using a Monte Carlo permutation test. The results for the two analyzed crash types ("collision" and "junction") are summarized in Table 2. Both crash categories utilized a selected bandwidth (σ) of 20 meters, consistent with the spatial resolution required to capture intersection-level interactions.

4.1 Interpretation of Monte Carlo Significance

The Monte Carlo segregation test yielded a p-value of exactly 0.05 for both crash types. In classical hypothesis testing frameworks, a significance level (α) of 0.05 is the conventional threshold for rejecting the null hypothesis. Under strict formal interpretation, $p \leq 0.05$ warrants rejection of the null hypothesis that real and simulated crash locations are spatially indistinguishable (i. e., "there is a significant difference").

Table 2. *Summary of segregation test statistics for "collision" (rear-end) and "junction" (crossing) crashes.*

Measure	**"Collision"**	**"Junction"**
Number of Real Crashes	5,233	5,210
Number of Simulated Crashes	18,849	8,880
Test Method	Segregation + Cross-K + Nearest Neighbor	
Selected Bandwidth (σ)	20	20
Segregation Test Statistic T	158.5685	164.1628
p-Value (Monte Carlo) - Segregation	0.05	0.05
Mean NND Distance	2.4	1.8
Permutation Test p-Value	1	0.001

However, in the context of Monte Carlo simulations with $N_{\text{sim}} = 999$, the precision of p-value estimation is constrained by the discrete nature of the permutation distribution. The smallest non-zero resolvable probability increment is approximately $1/(N_{sim} + 1) \approx 0.001$. A reported value of exactly 0.05 suggests a "precision landing" at the threshold rather than a robust margin away from it. Consequently, relying solely on binary significance (reject/fail to reject) may be misleading for model validation purposes.

Following contemporary statistical recommendations regarding borderline p-values, it is prudent to characterize this result as demonstrating marginal significance or a trend toward segregation. This indicates that while the probability of observing such extreme spatial divergence under random permutation is low (approximately $5\,\%$), the evidence is not sufficiently strong to rule out stochastic variation entirely. The simulation and reality are therefore statistically very similar, yet distinguishable at the chosen confidence level. This is also reflected by the average distance between a real crash location and its nearest simulated counterpart. The magnitude of less than 3 meters indicates that simulated crashes are generally located within the same micro-location as real ones supporting good spatial fidelity despite global segregation.

4.2 Spatial Implications of High Test Statistics

While the p-value indicates marginal significance, the magnitude of the Segregation Test Statistic ($T_1 \approx 158 - 164$) provides critical insight into the nature of the model discrepancy. A high T_1 value implies that the spatial separation between real and simulated points is greater than expected under the null hypothesis of perfect overlap.

This suggests a systematic misalignment rather than random noise. Specifically:

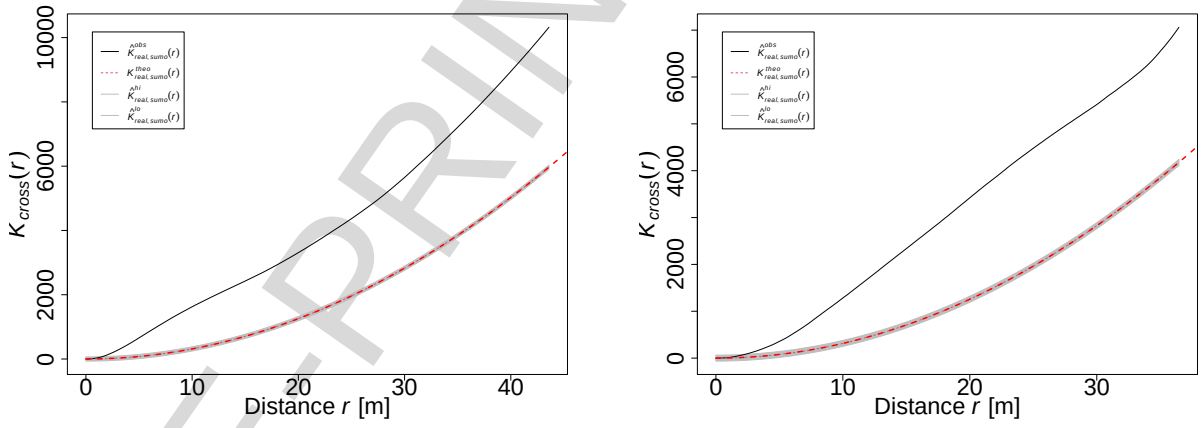
1. Global vs. Local Fidelity: The simulation appears to capture the macroscopic distribution of crashes at the intersection (large-scale density), as indicated by the non-rejection in broader cross-K analyses (implied by the low NND distance of around $2\,\text{m}$). However, the high T_1 reveals deficiencies in local density matching.

2. Specific Spatial Discrepancies: The deviation likely stems from specific interaction mechanisms not fully parameterized in the simulation logic. Potential drivers for this segregation include:

- Pedestrian-Vehicle Interactions: The model may underrepresent conflict points involving high pedestrian volumes at specific corners of the intersection, leading to simulated crashes clustering differently than real-world data.
- Visibility Constraints: Obstructed sightlines (e. g., due to infrastructure or parked vehicles) at specific quadrants might not be accurately modeled, causing agents to react differently than human drivers in reality.

4.3 Visual Interpretation

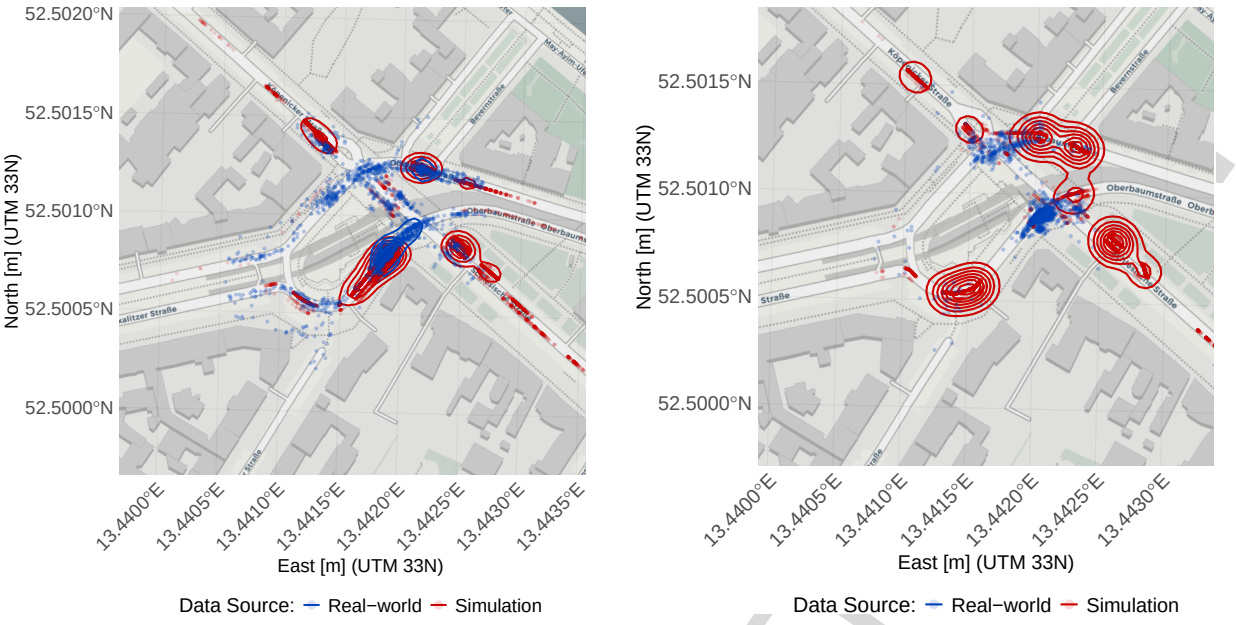
Figure 5 presents the results of the bivariate cross-K function analysis for rear-end collisions. The solid black line represents the observed estimator, $\hat{K}^{obs}_{real,sumo}(r)$, which quantifies the expected number of simulated crash locations within a radius r of each real crash location. The dashed red line indicates the theoretical expectation under complete spatial randomness (CSR), $K^{theo}_{real,sumo}(r)$. The gray shaded region denotes the $95\,\%$ Monte Carlo simulation envelope generated under the null hypothesis of spatial independence between real and simulated crash points. Throughout the analyzed distance range ($r \leq 45$ m), the observed curve remains consistently above the upper bound of the simulation envelope. This indicates a statistically significant spatial attraction between real and simulated crash locations, leading to the rejection of the null hypothesis of independence. The result confirms that the simulation reproduces the macroscopic spatial clustering of crashes at conflict-prone areas of the intersection. The deviation from the envelope aligns with the segregation test results in sub-section 4.1, suggesting that while the model captures general crash-prone zones, local intensity distributions differ. These discrepancies are attributed to simplified pedestrian behavior modeling and the absence of visibility constraints in the current implementation.



(a) *"Collision" crashes.* ***(b)*** *"Junction" crashes.*

Figure 5. *Bivariate cross-K function for rear-end collisions and crossing crashes. Observed curves (solid black) vs. theoretical CSR expectation (dashed red) and $95\,\%$ Monte Carlo envelope under spatial independence (gray) for both types of collisions.*

Figure 6 displays the spatial density of rear-end and junction crashes for both real-world and simulated datasets, estimated using kernel density estimation (KDE) with a bandwidth of 20 m, consistent with the parameters in Table 2. The normalized density maps indicate substantial overlap at primary conflict zones, particularly along the main through-lanes and central turning paths. However, the simulated distribution (red) exhibits pronounced linear clustering aligned with lane geometries, whereas the empirical



(a) *"Collision" crashes.* ***(b)*** *"Junction" crashes.*

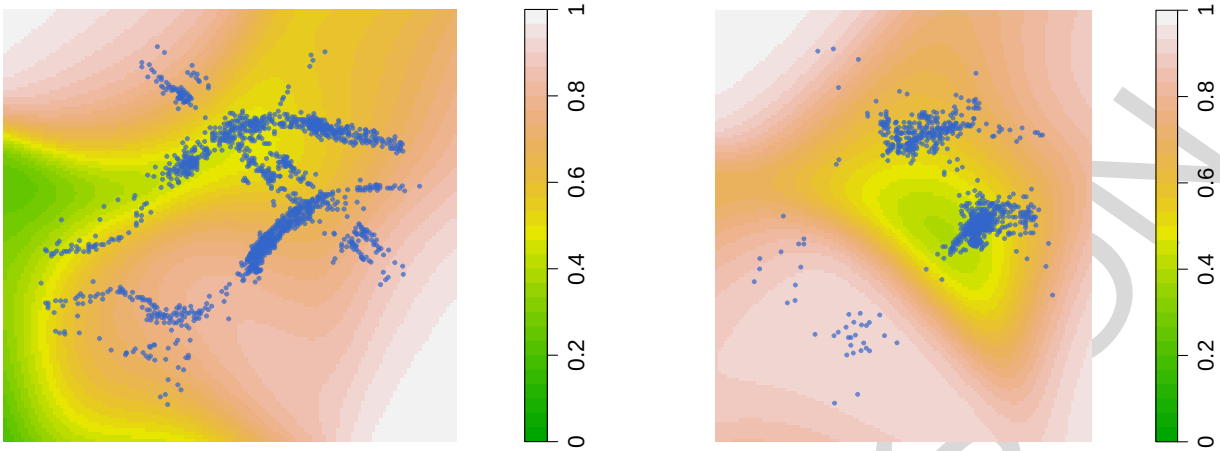
Figure 6. *Spatial density distributions of rear-end and crossing crashes at "Schlesisches Tor". Densities are estimated using kernel density estimation (KDE) with a fixed bandwidth of $\sigma = 20$ m. Blue denotes real-world crash locations aggregated from police records (2001 – 2024); red denotes simulated crash locations averaged over 100 independent runs. Both datasets are projected in ETRS89 / UTM Zone 33N (EPSG:25833).*

distribution (blue) shows a more diffuse spatial pattern. This difference arises from the stochastic nature of real-world crash reporting, unmodeled micro-maneuvers, and heterogeneity in driver behavior that is not fully captured by the uniform blackout probability. The structural divergence in spatial intensity contributes to the high segregation statistic ($T_1 \approx 160$) observed in sub-section 4.1, despite the proximity of density centroids.

Figure 7 presents a local intensity proportion map, defined as $P(x, y) = \lambda_{\text{sim}}(x, y)/[\lambda_{\text{sim}}(x, y) + \lambda_{\text{real}}(x, y)]$, where λ denotes the kernel density estimate of the spatial intensity at coordinates (x, y). The metric is bounded between 0 and 1, with $P(x, y) = 0.5$ indicating perfect spatial agreement between simulated and observed crash intensities. Values significantly above 0.5 indicate localized over-prediction by the simulation, while values below 0.5 indicate under-prediction. The color scale maps the proportion range $[0, 1]$ such that green regions correspond to $P(x, y) < 0.5$ (under-prediction), yellow to near-light areas correspond to $P(x, y) \approx 0.5$ (balanced performance), and red to white regions correspond to $P(x, y) > 0.5$ (over-prediction). The map reveals systematic over-prediction ($P > 0.55$, red to white/light zones) along the western approach, where the blackout model generates excess turning conflicts due to simplified intersection traversal logic. Conversely, under-prediction ($P < 0.45$, green zones) occurs in the eastern quadrant, primarily at pedestrian crossings and corner conflict points. These localized discrepancies align with the marginal segregation result ($p = 0.05$) and indicate that the current model underrepresents vulnerable road user (VRU) interactions and context-dependent risk adaptation in high-conflict quadrants.

4.4 Interpretation of Spatial Segregation

The Monte Carlo segregation test yields a borderline significant result ($p = 0.05$) with a high segregation statistic ($T_1 \approx 158 – 164$). Given the permutation-based nature of the



(a) *"Collision" crashes.* ***(b)*** *"Junction" crashes.*

Figure 7. *Local intensity proportion map comparing simulated to real-world crash intensity at "Schlesisches Tor". The proportion $P(x, y) = \lambda_{sim}(x, y)/[\lambda_{sim}(x, y) + \lambda_{real}(x, y)]$ is computed from kernel density estimation (KDE) with $\sigma = 20$ m for each crash type. Values near 0.5 indicate balanced predictive performance ($\lambda_{sim} \approx \lambda_{real}$), while deviations quantify local model bias. Green regions denote under-prediction ($P < 0.5$), yellow regions denote balanced performance ($P \approx 0.5$), and red to white regions denote over-prediction ($P > 0.5$). The map is projected in ETRS89 / UTM Zone 33N (EPSG:25833).*

test and the discrete resolution of the null distribution ($\Delta p \approx 0.001$), this p-value should be interpreted as indicating a systematic, albeit minor, spatial divergence rather than a complete model failure. The high T_1 statistic reflects a mismatch in local crash intensity distributions, despite strong macroscopic alignment. This divergence is primarily attributed to limitations inherent in the blackout model: The uniform activation probability does not account for context-dependent risk adaptation, such as reduced error rates in complex turning maneuvers or varying sightlines. Furthermore, the simplified pedestrian crossing logic and absence of visibility constraints lead to underpredicted conflicts at corner zones and overpredicted turning-area collisions. Consequently, while the model reliably captures exposure-driven temporal trends and broad spatial clustering, its spatial fidelity is constrained by the lack of micro-level behavioral parameters and environmental context. These limitations define the boundary conditions for applying the model to localized safety evaluations.

5 Conclusion

This study evaluated the capability of a minimalist "blackout" driver error model within Eclipse SUMO to reproduce spatio-temporal crash patterns at the "Schlesisches Tor" intersection in Berlin. By integrating empirical police data (2001 – 2024) with aggregated weekly traffic demand and 100 independent simulation runs, the model was assessed against a rigorous statistical framework using point pattern analysis.

Temporally, the simulated crash distribution aligns closely with observed weekly trends, accurately reproducing weekday morning and evening peaks alongside flatter weekend profiles. Spatially, the model successfully clusters rear-end collisions along main road geometries consistent with empirical data. However, the Monte Carlo segregation test indicates a borderline significant spatial divergence ($p = 0.05$, $T_1 \approx 164$), highlighting systematic deviations at the micro-level. Specifically, the simulation overpredicts turning-area conflicts and underpredicts pedestrian-vehicle interactions.

The blackout approach prioritizes computational efficiency and pattern reproduction over behavioral fidelity. It does not capture detailed driver adaptation, risk compensation, or crash severity mechanisms, and is therefore not intended for absolute safety prediction or detailed behavioral validation. Validation is restricted to a single intersection due to data availability and geometric complexity; the spatial segregation metrics and pattern reproduction should be interpreted as a proof-of-concept. Multi-location studies are required to assess the generalizability of the blackout model across different intersection typologies, traffic compositions, and signal control strategies.

In conclusion, the results demonstrate that simple driver error models can effectively capture macroscopic safety characteristics when exposure and traffic composition are controlled. However, accurate micro-level validation requires richer empirical data on pedestrian dynamics and visibility constraints. Future work will focus on refining local hotspot calibration and integrating advanced VRU interaction models to reduce spatial segregation statistics below the marginal threshold identified in this study.

Data availability statement

The crash data are closed source data from the Police of Berlin and cannot be shared. Crash data from all of Germany are open data, but they contain only a small amount of crashes (the ones with injured people) at the “Schlesisches Tor”, about 20 per year. The simulation details, of course, can be shared upon request to the authors.

Author contributions

Ronald Nippold: Software, Simulation, Methodology, and Investigation. **Peter Wagner**: Conceptualization, Methodology **All**: Writing-, Reviewing, and Editing.

Competing interests

The authors declare that they have no competing interests.

Funding

This work used the basic funding of the DLR (German Aerospace Center) and a number of third-party funding projects, most notably the project KI4Safety, funded by the mFUND of the German Ministry of Transport.

Acknowledgements

We are grateful for the data from the police of Berlin. The many persons who have contributed to the development of SUMO are also gratefully acknowledged here.

References

[1] D. Lord and F. Mannering, “The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives,” *Transportation Research Part A: Policy and Practice*, vol. 44, pp. 291–305, Jun. 2010. doi: 10.1016/j.tra.2010.02.001.

[2] F. Mannering, “Cross-sectional modelling,” in *Safe Mobility – Challenges, Methodology, and Solutions*, D. Lord and S. Washington, Eds., Emerald Publishing Limited, 2018, pp. 257–277.

[3] D. Helbing, “Traffic and related self-driven many-particle systems,” Reviews of Modern Physics, Apr. 2001. doi: 10.1103/RevModPhys.73.1067. arXiv: cond-mat/0012229v2 [cond-mat.stat-mech]. [Online]. Available: https://arxiv.org/abs/cond-mat/0012229v2;%20https://arxiv.org/pdf/cond-mat/0012229v2.

[4] A. Leich, R. Nippold, and P. Wagner, “Simulating crashes with sumo for berlin’s most dangerous intersection,” *Transportation Research Arena (Accepted)*, 2026, Forthcoming.

[5] P. A. Lopez, M. Behrisch, L. Bieker-Walz, *et al.*, “Microscopic traffic simulation using sumo,” in *2018 21st international conference on intelligent transportation systems (ITSC)*, Ieee, 2018, pp. 2575–2582.

[6] A. Tarko, *Measuring Road Safety with Surrogate Events*. Elsevier, 2019, isbn: 9780128105047.

[7] “Digitale plattform stadtverkehr berlin - verkehrsdetektion.” German. (2023), [Online]. Available: https://api.viz.berlin.de/daten/verkehrsdetektion (visited on 03/22/2023).

[8] PTV Group, *Driving behavior parameters - error behavior*, https://cgi.ptvgroup.com/vision-help/VISSIM_2025_ENG/Content/4_BasisdatenSim/Fahrverhaltenspar_Fehlverh.htm, Accessed on May 5, 2026, 2025.

[9] Y. Zhu and L. Yue, “Simulation-oriented analysis and modeling of distracted driving,” *Applied Sciences*, vol. 14, no. 13, 2024, issn: 2076-3417. doi: 10.3390/app14135636. [Online]. Available: https://www.mdpi.com/2076-3417/14/13/5636.

[10] A. Baddeley and R. Turner, “Spatstat&58; an r package for analyzing spatial point patterns,” *Journal of Statistical Software*, vol. 12, no. 1, pp. 1–42, 2005. doi: 10.18637/jss.v012.i06. [Online]. Available: https://www.ingentaconnect.com/content/doaj/15487660/2005/00000012/00000001/art00013.

[11] B. D. Ripley, “The second-order analysis of stationary point processes,” *Journal of Applied Probability*, vol. 13, no. 2, pp. 255–266, 1976. doi: 10.2307/3212829.

[12] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2025. [Online]. Available: https://www.R-project.org/.