

Impact of Road Geometry and Temporal Factors on Number of Vehicles in Accidents: A GIS Study

Gavrilov Igor^{1,#} Hruběš Pavel²

¹ Department of Transport Telematics, Czech Technical University in Prague, Prague, 101 00, Czechia, gavriigo@cvut.cz

² Department of Transport Telematics, Czech Technical University in Prague, Prague, 101 00, Czechia, hrubes@k620.fd.cvut.cz

Abstract

Road safety studies often rely on predefined road segments of varying lengths, which can introduce spatial inconsistency and bias in statistical modelling. This paper proposes a GIS-based framework that uses dynamically generated fixed-length road segments before and after each accident location to evaluate the influence of road geometry and temporal factors on multi-vehicle crash occurrence. The approach ensures a spatially consistent and localised assessment of highway conditions at each crash site. The methodology is demonstrated using detailed police-reported accident data from the D1 highway in the Czech Republic, covering January to December 2025, combined with a national road network dataset. For each accident, local geometric attributes, including road width, number of lanes, and curvature, were derived from a 300 m dynamically extracted highway segment centred on the projected crash location. A binomial generalised linear model with a logit link function was applied to estimate the probability of multi-vehicle accidents as a function of these geometric factors and temporal variables, including time of day, month, and day of week. Results show that wider road segments and daytime periods, particularly morning and afternoon hours, significantly increase the likelihood of multi-vehicle crashes, whereas higher lane counts and weekend periods are associated with lower risk. This framework provides a transferable, spatially robust tool for localised road safety assessment and infrastructure planning.

Keywords: Road Traffic Accidents, Geographic Information System, Poisson Regression, Dynamic Road Segments, Spatial-Temporal Analysis

Introduction

The World Health Organisation (WHO, 2026) at the end of 2023 reported that every year approximately 1.19 million people die as a result of road accidents. Also, between 20 and 50 million people have non-fatal injuries, which can be permanent or temporary disabilities. One of the risks of road crashes is unsafe road infrastructure, which may include dangerous curves and insufficient width or number of lanes for growing road traffic.

In previous studies, specified parts (length segment level) (Mahmud et al., 2021) with varying lengths of rural roads (Asif and Gayah, 2021; Liu et al., 2025; Anas, 2021) or highways were examined (Li et al., 2024), and as a result of road crashes, the equivalent property damage only (EDPO) or severity (fatality, heavy, soft, or non-injured) (Fangrong et al., 2021; Ji and Levinson, 2020; Chen et al., 2021) was used. Authors used different descriptive or predictive techniques, the most popular: Poisson regression (Lord, Mannering, 2010), Negative Binomial regression, generalised linear models, machine learning approaches (Yasir et al., 2024; Ji et al., 2020), Classification (Theofilatos, 2017), and neural networks. Using the different lengths of road segments and normalising them adds more variability to the research results.

This paper investigates the use of dynamically generated road segments of fixed length before and after the location of a road accident, a topic that has not been addressed in previous studies. The proposed approach enables a localised and spatially consistent assessment of how road geometry and temporal factors influence the number of vehicles involved in traffic accidents,

providing a GIS-based framework for improved road safety analysis and infrastructure planning.

Data and Method

For the research, two data sources were used: detailed information on road accidents in the Czech Republic along the D1 highway from January 2025 to December 2025 reported to the Police (Centrum dopravního výzkumu, 2024), and the Czech Republic's road network as of March 2025 (ČÚZK, 2025).

Under Czech law, road accidents must be reported to the police in cases involving fatalities or property damage exceeding 4000 euros. Therefore, the dataset represents a sample of the total population of road accidents that occurred on the highway D1 during the study period.

The initial dataset was cleaned by removing records not related to the studied highway. The attribute “ref” contains a textual reference to the corresponding rural road or highway name for each accident record. Records with missing values in the “ref” attribute, as well as records with missing day or time information, were also excluded from the dataset. This attribute was used to filter the dataset, retaining only accident records associated with the analysed highway section while excluding records related to other road segments. Also, records of road accidents involving drivers under the influence of alcohol or drugs were excluded as non-relevant for this study. The dataset contains 635 road accidents, and the available attributes are: total cars, date, time, and the accident locations' geographical coordinates in WGS84 format.

The road network is a spatial dataset; the structure is based on ISO 19131 Geographic information - Data production standard. The dataset contains information on multiple types of transport infrastructure, including road, rail, air, and water transport; however, for the purpose of this study, the longest and most heavily trafficked corridor - the D1 highway - was selected. The highway connected the capital of the Czech Republic with the second largest city, Brno, and ended at the Polish border. The first segment, 54.1km of the highway, was opened in 1977, and at the time of writing, the highway has not been completed. The first part, Prague – Rukovice, has a length of 266632.45 meters, and the second part, Prerov-Polish border, has a length of 94122.43 meters. The highway is direction-separated (one-way), so collisions involving a vehicle in the opposite direction (MVOD) or intersection-related (IR) collisions are impossible.

The main complexity of the highway D1 is its long construction period, which began around 50 years ago, during which the Czech Republic applied different construction standards, from those of the Czechoslovak Federal Republic to the current European Union standards for highways. And as a result, the highway with 2 or 3 lanes has an absolutely different width of the road: 7, 7.2, 7.3, 7.4, 7.5, 7.6, 7.9, 8.2, 8.4, 8.5, 9, 9.1, 9.3, 9.4, 9.6, 9.7, 9.8, 9.9, 10, 10.1, 10.2, 10.4, 10.5, 10.7, 10.8, 10.9, 11, 11.1, 11.2, 11.3, 11.4, 11.5, 11.6, 11.7, 12, 12.7, 12.8, 13, 13.8, 14, 14.3, 14.5, 14.6, 14.7, 14.8, 14.9, 15, 15.1, 15.2, or 15.4 meters.

For analysis, the author's previous studies mainly used software such as ArcGIS (Esri, 2026) and QGIS (QGIS, 2026), as well as commercial or free software, which have limited functionality. For the examination, a relational database, PostgreSQL 15.7 (PostgreSQL, 2026), with the PostGIS extension 3.4 (PostGIS, 2026), was used. That combination provides a more robust solution to manipulate geographical data.

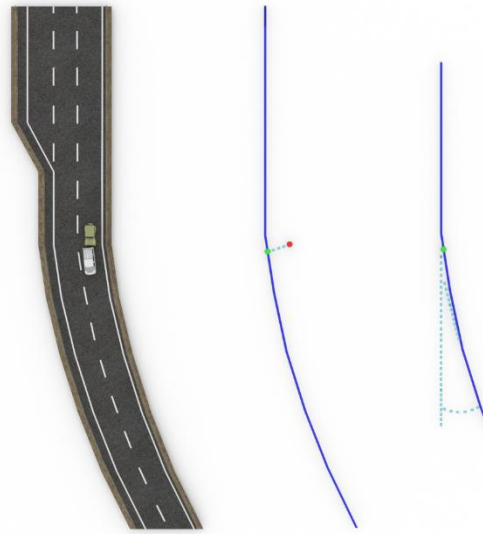


Fig. 1 Data transformation process: (left) original accident location, (middle) projection onto the highway vector, and (right) extraction of a 150 m road segment before and after the projection for angle measurement.

For each recorded accident, the original geographic location was first projected onto the corresponding highway vector. A fixed-length road segment extending 150 meters before and after the projected point was then extracted from the highway vector. This dynamically generated segment was used to compute the local road curvature by measuring the change in angle along the extracted segment.

A binary response variable was constructed to distinguish between single-vehicle (1 vehicle) and multi-vehicle (more than one vehicle) accidents. A binomial generalised linear model with a logit link function was applied to estimate the influence of road geometry and temporal factors on the probability of multi-vehicle crashes.

$$\mu_i = \log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \sum_{k=1}^K \beta_k X_{ik} \quad (1)$$

Where p_i denotes probability of the outcome for the i -th observation ($i=1, \dots, n$), β_0 is the intercept, β_k represents the coefficient associated with the k -th predictor ($k=1, \dots, K$), and X_{ik} is the value of the k -th explanatory variable for the i -th observation.

From the original dataset, the date and time attributes were extracted, and each month was added as an explanatory variable. The time attribute was aggregated into 6-hour intervals and categorised as night (0–6), morning (6–12), afternoon (12–18), and evening (18–24).

Results

The results of the binomial generalised linear model (Table 1) indicate that both geometric and temporal factors significantly influence the probability of multi-vehicle crashes on the D1 highway. Road width shows a positive and statistically significant effect ($p = 0.002$), suggesting that wider road segments are associated with a higher likelihood of multi-vehicle accidents. Specifically, a one-meter increase in road width increases the odds of a multi-vehicle crash by approximately 30%.

Table 1 Binomial generalised linear model estimating the probability of multi-vehicle crashes on the D1 highway

Variable	Coefficient	Std. Error	z-value	p-value	Odds Ratio (exp(coef))	Significance
Constant	-0.0663	0.746	-0.089	0.929	0.94	
Angle	-0.0223	0.026	-0.85	0.395	0.98	
Lanes	-0.6716	0.349	-1.925	0.054	0.51	†
Width	0.2649	0.086	3.079	0.002	1.3	**
Morning	0.9167	0.299	3.069	0.002	2.5	**
Afternoon	0.8406	0.295	2.853	0.004	2.32	**
Evening	-0.2748	0.322	-0.853	0.394	0.76	
August	-0.3271	0.525	-0.623	0.533	0.72	
December	-0.6903	0.533	-1.295	0.195	0.5	
February	0.0439	0.596	0.074	0.941	1.04	
January	-1.3212	0.588	-2.249	0.025	0.27	*
July	-1.0396	0.529	-1.966	0.049	0.35	*
June	-0.5095	0.525	-0.97	0.332	0.6	
March	-0.0456	0.596	-0.076	0.939	0.96	
May	-0.2156	0.529	-0.407	0.684	0.81	
November	-0.6416	0.541	-1.186	0.236	0.53	
October	-0.046	0.55	-0.084	0.933	0.96	
September	-0.3247	0.513	-0.632	0.527	0.72	
Monday	-0.0559	0.345	-0.162	0.871	0.95	
Saturday	-0.6715	0.343	-1.956	0.05	0.51	*
Sunday	-0.777	0.319	-2.435	0.015	0.46	*
Thursday	-0.5331	0.296	-1.803	0.071	0.59	†
Tuesday	-0.1273	0.319	-0.399	0.69	0.88	
Wednesday	-0.5766	0.301	-1.918	0.055	0.56	†

Temporal effects are also pronounced. Accidents occurring during the morning and afternoon periods are significantly more likely to involve multiple vehicles compared to the reference category, with odds ratios of 2.50 and 2.32, respectively. This pattern is consistent with increased traffic density during peak and daytime hours.

A significant seasonal effect is observed in January, with the odds of multi-vehicle crashes reduced by approximately 73% relative to the baseline month, potentially reflecting lower traffic volumes or altered driving behaviour during winter conditions.

The number of lanes exhibits a marginally significant negative association with the probability of multi-vehicle crashes ($p=0.054$). The corresponding odds ratio of 0.51 suggests that, all else equal, road segments with more lanes are associated with a lower likelihood of multi-vehicle accidents. This effect may indicate that additional lanes provide greater manoeuvring space and reduce vehicle interactions under normal traffic conditions, although the borderline statistical significance suggests that this relationship should be interpreted with caution.

Day-of-week effects further highlight temporal variability in crash occurrence. Weekend days, particularly Saturday and Sunday, demonstrate a statistically significant reduction in the probability of multi-vehicle accidents, with odds ratios of 0.51 and 0.46, respectively. This pattern may reflect differences in traffic composition and flow characteristics between

weekdays and weekends, such as reduced commuter traffic and a higher proportion of leisure-oriented travel. Additionally, marginal effects observed for Thursday and Wednesday suggest a gradual transition in traffic dynamics toward the weekend, although these associations do not reach conventional levels of statistical significance.

Conclusion

This study demonstrates that dynamically generated fixed-length road segments provide a robust, spatially consistent framework for analysing factors influencing multi-vehicle crash occurrence on high-capacity highways. By extracting standardised segments before and after each accident location, the proposed GIS-based methodology reduces variability associated with heterogeneous road-section definitions and enables a more localised assessment of geometric and temporal effects.

The results of the binomial generalised linear model highlight the significant role of both infrastructure and time-related factors in shaping crash dynamics on the D1 highway. Road width and daytime periods, particularly the morning and afternoon intervals, were found to substantially increase the probability of multi-vehicle accidents, reflecting the combined effects of road design characteristics and elevated traffic demand. Conversely, the number of lanes and weekend periods were associated with a reduced likelihood of multi-vehicle crashes, suggesting that additional manoeuvring space and altered traffic composition may mitigate vehicle interactions under certain conditions.

From a practical perspective, the proposed approach offers a transferable, scalable tool for road authorities and planners to identify localised risk patterns along long, structurally heterogeneous corridors. The integration of spatial database technology and statistical modelling supports data-driven decision-making for targeted infrastructure interventions and traffic management strategies.

Future research should extend this framework by incorporating traffic volume data, weather conditions, and vehicle type composition to further improve the model's explanatory power. Weather conditions such as precipitation, temperature, fog, or snow can significantly influence road surface conditions, visibility, and driver behaviour, thereby affecting the likelihood and severity of traffic accidents. Including such variables would allow for a more comprehensive understanding of the factors contributing to accident occurrence. Additionally, applying the methodology to other highway networks and urban environments would allow for broader validation and comparative analysis of dynamic fixed-length segmentation in road safety studies.

References

- Anas, A. (2021) A correlated random parameters approach to investigate large truck rollover crashes on mountainous interstates. *Accident Analysis and Prevention*.
- Asif, M., Gayah, V. (2021) Estimation of crash type frequencies on individual collector roadway segments. *Accident Analysis & Prevention*, 161, doi:10.1016/j.aap.2021.106345
- Centrum dopravního výzkumu, v. v. i. Výzkum v dopravě. <https://www.cdv.cz/> (accessed Dec. 5, 2024).
- Chen, F., Yasmin, S., Helai, H., Chan, A. H. S., Hoque, M. M. (2021) Injury severity analysis of motorcycle crashes: A comparison of latent class clustering and latent segmentation based models with unobserved heterogeneity. *Anal. Methods Accid. Res.*, 32, doi:10.1016/j.amar.2021.100188.
- ČÚZK. Geoportál ČÚZK. <https://geoportal.cuzk.cz/> (accessed Oct. 2, 2025).
- Esri. ArcGIS. <https://www.arcgis.com/index.html> (accessed Jan. 18, 2026)

Fangrong, C., Shamsunnahar, Y., Helai, H., H.S. Chan, A., Md. Mazharul, H. (2021) Injury severity analysis of motorcycle crashes: A comparison of latent class clustering and latent segmentation based models with unobserved heterogeneity. *Analytic Methods in Accident Research* 32, doi:10.1016/j.amar.2021.100188

Jean, A. (2010) Long-term trends in weather-related crash risks. *Journal of Transport Geography* 18(2), doi:10.1016/j.jtrangeo.2009.05.002

Ji, A., Levinson D. (2020) Injury Severity Prediction From Two-Vehicle Crash Mechanisms With Machine Learning and Ensemble Models. *IEEE Open Journal of Intelligent Transportation Systems* 1, doi:10.1109/OJITS.2020.3033523

Li, J., Li, C., Zhao X. (2024) Optimizing crash risk models for freeway segments: A focus on the heterogeneous effects of road geometric design features, traffic operation status, and crash units. *Accident Analysis & Prevention*, 205, doi:10.1016/j.aap.2024.107665

Liu, J., Chakraborty, R., Somvanshi, S., Das, S. (2025) Impact of operating speed, roadway curvature, and precipitation on roadway departure risk in rural two-lane roads. *Travel Behaviour and Society*, 41, doi:10.1016/j.tbs.2025.101055

Lord, D., Mannering, F. (2010) The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice* 44(5), doi:10.1016/j.tra.2010.02.001

Mahmud A., V. Gayah V. (2021) Estimation of crash type frequencies on individual collector roadway segments. *Accident Analysis & Prevention*, 161, doi:10.1016/j.aap.2021.106345

PostGIS. <https://postgis.net/> (accessed Jan. 18, 2026).

PostgreSQL. <https://www.postgresql.org/> (accessed Jan. 18, 2026).

QGIS. <https://qgis.org/> (accessed Jan. 18, 2026).

Theofilatos, A. (2017) Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. *Journal of Safety Research*, 61, doi:10.1016/j.jsr.2017.02.003

World Health Organization. Road traffic injuries. <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries> (accessed Jan. 18, 2026).

Yasir, A., Fizza, H., and Md Mazharul, H. (2024) Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. *Accident Analysis & Prevention* 194, doi:10.1016/j.aap.2023.107378