Naturalistic Spatial Road Safety Analysis: The SmartMaps Project

Dimitrios Nikolaou[1], Armira Kontaxi[1*],
Apostolos Ziakopoulos[1], George Yannis[1],
Petros Fortsakis[2], Eleni Konstantina Frantzola[2],
Konstantinos Sigalos[3], George Kouridakis[3]

[1] National Technical University of Athens, 5 Heroon Polytechniou St., GR-15773, Athens, Greece
[2] OSeven Single Member Private Company, 27B Chaimanta Str., GR-15234, Athens, Greece
[3] Global Link, 87 Amarousiou Str., GR-15125, Athens, Greece
* akontaxi@mail.ntua.gr

Abstract. This paper aims to leverage large-scale spatio-temporal data from smartphone sensors and geometric design characteristics for spatial analysis of telematics-based surrogate safety measures across various road environments. Two distinct statistical models have been developed: a non-spatial log-linear model and a spatial error model. The dependent variable in both models is the logarithm of the number of harsh braking events observed in each considered segment. The study area is located within the Western Greece Region and encompasses 9,355 road segments. The findings reveal a positive correlation between harsh braking events and the length of road segments as well as the number of recorded trips per segment. Furthermore, variables associated with speeding, mobile phone usage, and road segment linearity exhibit positive correlations with the number of harsh braking events in the examined segments. Moreover, harsh braking events on motorways are found to be lower when compared to other road types. In conclusion, the results of this study indicate that the spatial error model outperforms the non-spatial model in terms of data fit and yields more reliable outcomes.

Keywords: Spatial Analysis, Surrogate Safety Measures, Harsh Braking.

1 Introduction

In recent years, there has been growing research focus on Surrogate Safety Measures (SSMs). Several factors have contributed to this research direction, primarily driven by technological advancements that have made data acquisition and analysis significantly more feasible and cost-effective (Nikolaou et al., 2023c). SSMs offer several comparative advantages over crash data. They serve as proactive road safety indicators, enabling the analysis of road safety conditions before crashes occur, or even when events do not necessarily lead to crashes. Moreover, the crash data collection processes remain non-automated and may carry inherent limitations and biases, such as unavailable or inaccurate data (Imprialou & Quddus, 2019), or variations in under-reporting rates among countries (Yannis et al., 2014). These are issues that can be mitigated by leveraging automatically collected SSMs. Tarko (2018) emphasized that SSMs assist in identifying excessive crash risks on the road, enhancing the understanding of crash conditions, and providing a more robust evaluation of the effectiveness of experimental and existing countermeasures.

In a study of 668 road sections of the Olympia Odos motorway in Greece, a Negative Binomial regression model was developed with the number of road crashes (property damage and injury) as the dependent variable and two telematics-based SSMs (harsh accelerations and harsh brakings), the Annual Average Daily Traffic (AADT) and the segment length as independent variables (Nikolaou et al., 2023a). The results of the study indicated a statistically significant and positive correlation between the two harsh driving behaviour metrics and road crash frequency. However, as a follow-up to this research, it was found that harsh brakings contribute significantly to predicting the crash risk level of the road segments under consideration, which is not the case for harsh accelerations (Nikolaou et al., 2023b). It is therefore concluded that harsh brakings are a plausible SSM that can be used in various proactive road safety analyses.

Another notable study is that of Stipancic et al. (2018). The authors modelled road crash frequencies using a Full Bayes model with SSMs as independent variables. The study used large-scale GPS data from smartphones and obtained several SSMs such as harsh braking and traffic flow parameters. The authors used a Latent Gaussian spatial model to model crash frequencies and reported that incorporating spatial correlations provided the greatest improvement in model fit.

Based on the aforementioned, the objective of this research is to use large-scale spatio-temporal data from smartphone sensors and geometric design characteristics for spatial analysis of telematics-based surrogate safety measures across different road environments in the Region of Western Greece.

2 Methodology

2.1 Data Collection

In the framework of the SmartMaps research project, data collection from various sources has been carried out to develop maps of driver behaviour with online information on safety conditions and eco-driving. The ultimate goal is to create a complete and comprehensive tool to promote driving behaviour in order to make it safe and environmentally friendly, while making the overall traffic more efficient and manageable, with application in Greece and around the world.

The dataset examined for the region of Western Greece comprises 9,355 road sections, with an average length of 223 meters. In terms of road types, the distribution is as follows: residential roads (74%), tertiary roads (7%), primary roads (6%), secondary roads (5%), motorways (4%), and the remaining 4% comprises other road types. These data were derived by utilizing OpenStreetMap data, which were processed using appropriate packages in the R programming language (Padgham et al., 2017).

With regard to the naturalistic driving data, this study utilizes data obtained from an existing smartphone application developed by OSeven Telematics (www.oseven.io). A total of 14,161 trips in the study area within the year 2021 were examined, and a spatial matching of the naturalistic driving data and the road segments examined was performed. Table 1 demonstrates some key descriptive statistics of the collected data by road segment.

Table 1. Key descriptive statistics of geometry and driving behaviour data per road segment.

Variable	Min.	Mean	Max.
trip [count]	0.0	61.3	1,293.0
length [m]	1.3	222.7	18,029.1
linearity index [0-1]	0.03	0.95	1.00
harsh braking [count]	0.0	1.3	221.0
speeding [sec]	0.0	30.5	27,279.0
mobile phone usage [sec]	0.0	35.2	8,561.0

2.2 Statistical Analysis

The data presented in section 2.1. including road geometric features and naturalistic driving behaviour data via smart phone sensors were analysed by applying two distinct statistical models: a non-spatial log-linear model and a spatial error model. Log-linear regression is a widely known and simple technique and as such the mathematics behind it are omitted. The spatial error model handles the spatial autocorrelation in the residuals. The idea is that residuals from regression are autocorrelated in that the error from one spatial feature can be modeled as a weighted average of the errors of its neighbors. This model can be expressed as:

$$y = X\beta + u, \quad u = \lambda_{Err} W u + \varepsilon \quad (1)$$

where y is a (N×1) vector of observations on a dependent variable taken at each of N locations, X is a (N×k) matrix of covariates, β is a (k×1) vector of parameters, u is a (N×1) spatially autocorrelated disturbance vector, ε is a (N×1) vector of independent and identically distributed disturbances and λ_{Err} is a scalar spatial parameter.

The non-spatial log-linear regression model and the spatial error model can be compared using the Akaike Information Criterion (AIC), where lower values of this criterion indicate better statistical model quality (Akaike, 1970).

3 Results and Discussion

As mentioned in the Introduction, harsh brakings are SSMs that can be used in analyses before road crashes occur. To this end, harsh brakings were used as a dependent variable in both the non-spatial log-linear model and the spatial error model of this study. The results of the two developed statistical models are presented in Table 2.

Table 2. Non-spatial log-linear regression and Spatial Error Model results.

Dependent variable: log (harsh brakings + 1)				
	Log-linear Model		**Spatial Error Model**	
Parameters	**Estimate**	**p-value**	**Estimate**	**p-value**
(Intercept)	-0.756	<0.001	-0.756	<0.001
trip count	0.003	<0.001	0.003	<0.001
log(1+length)	0.099	<0.001	0.099	<0.001
log(1+speeding)	0.115	<0.001	0.115	<0.001
log(1+linearity index)	0.468	<0.001	0.467	<0.001
Mob. phone use/trips	0.012	<0.001	0.012	<0.001
Motorway	-0.169	<0.001	-0.167	<0.001
Lamda	-	-	0.016	0.041
R-squared	0.556	-	-	-
AIC	11,826	-	11,824	-

Based on Table 2 results, the values and the signs of the independent variables' coefficients remain consistent between the two developed models. Specifically, both the length of the road segment under consideration and the number of trips per road segment can be considered as indicators of risk exposure and, as expected, are positively correlated with the number of harsh brakings. In addition, the positive sign of the independent variable of the road segment linearity index suggests that road segments with fewer curves have a higher number of harsh braking events. Moreover, the variables related to speeding and mobile phone use are positively correlated with the number of harsh brakings on the road segments considered. Lastly, harsh braking events on motorways are found to be lower when compared to other types of roads.

Regarding the spatial error model, the Lambda value of 0.016 is statistically significant, indicating that the error term is spatially autoregressive. Moreover, based on AIC values, it can also be observed that the spatial error model outperforms the non-spatial log-linear model. The spatial error model's predictions for harsh braking events in the Western Greece Region's road network are illustrated in Figure 1. Figure 2 offers a close-up look at the predictions specifically for the city of Patras, Greece.

Fig. 1. Spatial error model results for the examined road network of Western Greece.

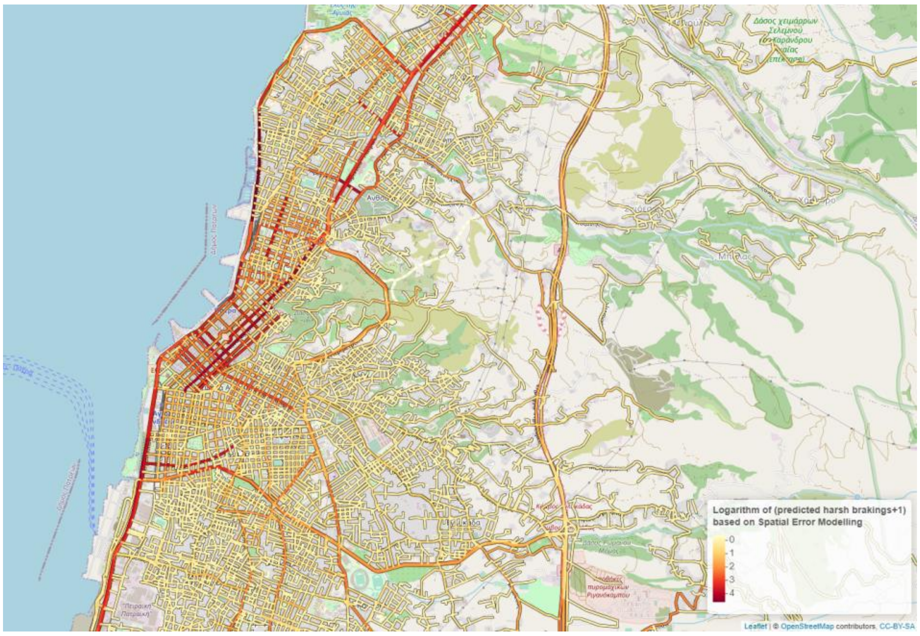

Fig. 2. Spatial error model results for the center of Patras, Greece.

4 Conclusions

This research aims to utilize spatiotemporal data from smartphone sensors on driving behaviour and geometric road features for spatial analysis of SSMs in various road environments in the Western Greece region. Specifically, two different types of statistical models were developed: a non-spatial log-linear model and a spatial error model. The dependent variable in the developed models was the logarithm of the number of harsh braking events per examined segment, as it constitutes a SSM that can be used in road safety analyses either in cases where the exact location of road crashes is not available or as a proactive indicator before road crashes occur.

The results indicated that the number of harsh brakings in the examined road segments is positively correlated with the length and the number of recorded trips per segment. This finding can be considered expected since these variables are risk exposure indicators, and an increase in them leads to more harsh braking events. Furthermore, variables related to speeding and mobile phone use were positively associated with the number of harsh brakings in the examined road segments. Moreover, the results of this research revealed that the spatial error model demonstrates better fit to the data and lead to more reliable results than the non-spatial model.

Finally, it is worth noting that the ultimate goal and the intention of the research team is to create a comprehensive mapping tool covering all regions of Greece, promoting safe and environmentally friendly driving behaviour.

Acknowledgments

This research was performed within the research projects:

- SmartMaps, been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program "Competitiveness, Entrepreneurship and Innovation", under the call "RESEARCH-CREATE-INNOVATE" (project code: T2EDK-04388)
- OptiMo, been financed by the Basic Research Financing (Horizontal support for all Sciences), National Recovery and Resilience Plan (Greece 2.0) (project number: 15629)

References

1. Nikolaou, D., Dragomanovits, A., Ziakopoulos, A., Deliali, A., Handanos, I., Karadimas, C. & Yannis, G. (2023a). Exploiting Surrogate Safety Measures and Road Design Characteristics towards Crash Investigations in Motorway Segments. Infrastructures, 8(3), 40.
2. Nikolaou, D., Ziakopoulos, A., Dragomanovits, A., Roussou, J., & Yannis, G. (2023b). Comparing Machine Learning Techniques for Predictions of Motorway Segment Crash Risk Level. Safety, 9(2), 32.
3. Nikolaou, D., Ziakopoulos, A., & Yannis, G. (2023c). A Review of Surrogate Safety Measures Uses in Historical Crash Investigations. Sustainability, 15(9), 7580.
4. Yannis, G., Papadimitriou, E., Chaziris, A., & Broughton, J. (2014). Modeling road accident injury under-reporting in Europe. European transport research review, 6(4), 425-438.
5. Imprialou, M., & Quddus, M. (2019). Crash data quality for road safety research: current state and future directions. Accident Analysis & Prevention, 130, 84-90.
6. Tarko, A. P. (2018). Surrogate measures of safety. In Safe mobility: challenges, methodology and solutions. Emerald Publishing Limited.
7. Stipancic, J., Miranda-Moreno, L., Saunier, N., & Labbe, A. (2018). Surrogate safety and network screening: Modelling crash frequency using GPS travel data and latent Gaussian Spatial Models. Accident Analysis & Prevention, 120, 174-187.

8. Padgham, M., Lovelace, R., Salmon, M., & Rudis, B. (2017). osmdata. Journal of Open Source Software, 2(14).
9. Akaike, H. (1970). Statistical predictor identification. Annals of the institute of Statistical Mathematics, 22(1), 203-217.