Analysing the relevance of road marking quality on road safety with the application of geospatial techniques and statistical and machine learning models

[1]Cafiso, S., [1]Dimauro, A., [1]*Pappalardo, G., [2]Mauriello, F., & [2]Fucito, S.
*lead presenter
[1]giuseppina.pappalardo1@unict.it, University of Catania, Italy
[2] University of Naples, Italy

Road markings are indispensable elements of road infrastructure, serving as visual cues that convey essential information to drivers. They play a pivotal role in guiding drivers along the road, delineating lanes, indicating traffic regulations, and highlighting potential hazards. Unlike verbal signs or signals, which may require interpretation, road markings communicate their message through universally understood symbols, designs, and colours, making them accessible to drivers of diverse backgrounds and languages.

In adverse weather conditions, such as heavy rain, fog, or snow, and during night-time driving when natural light is limited, road markings become even more critical for maintaining road safety. Visibility challenges in these conditions heighten the importance of clear, well-maintained markings that stand out against the road surface. Properly designed and maintained markings ensure that drivers can navigate safely, anticipate upcoming roadway configurations, and make informed driving decisions.

The significance of road marking visibility is underscored by its influence on driver behaviour and decision-making. Research has shown that drivers' responses to road markings are influenced by their visibility and legibility. Visible markings provide drivers with confidence and reassurance, enabling them to navigate with greater precision and predictability. Conversely, faded, worn, or obscured markings can lead to uncertainty, hesitancy, and potentially hazardous manoeuvres, increasing the risk of accidents, particularly in high-speed or complex traffic environments.

Understanding the relationship between road marking visibility and road safety requires comprehensive analysis and consideration of various factors. Previous studies have examined the impact of factors such as retro-reflectivity, colour contrast, and maintenance practices on marking visibility and safety outcomes. While some studies have found a direct correlation between visible markings and reduced accident rates, others have yielded mixed or inconclusive results.

To support the understanding of the correlation between road marking conditions and accident occurrence, the presented research aims to conduct a detailed analysis using a combination of spatial data analysis with geographic information systems (GIS), and statistical and machine learning for modelling.

The aim of the present study is the development of crash modification functions (CMFs) for the retro-reflectivity coefficient for different crash types (total, run-off-the-road, and others), pavement conditions (dry and wet), and lighting conditions (daytime and nighttime) based on data from two-lane rural roads in Italy.

To accomplish this task, data was gathered regarding the retro-reflectivity coefficient (RL) [mcd/lx/m2], which serves as an indicator of the of road markings quality. Subsequently, these road segments were spatially linked at the road network, utilizing Geographic Information System (GIS) software, specifically QGIS.

The road network is split into homogeneous segments of constant length. Accident, traffic, geometry, and weather data were also collected for the study network from national and local databases. The accident data are the micro-data provided by the Italian National Institute of Statistics (ISTAT) from 2019 to 2023. ISTAT data, relative to injury and fatal accidents, are further classified according to the crash type and the location of crash.

Detailed traffic flow data were provided by ANAS agency, and traffic volume is categorized also into heavy and light vehicles. Climate and rain duration data were derived from the regional Functional Hydrogeological Centre. The geometric data is distinguished by cross-section and horizontal alignment variables. The cross-section data includes lane width and shoulder width. For the horizontal alignment of the road, eight geometric design exploratory variables are extracted: Curvature Ratio; Curvature Change Rate; Average value of 1/R (AV(1/R)); Standard deviation of 1/R (SD1/R); Coefficient of variation of 1/R (CV(1/R)); Ratio between the maximum radius and the minimum radius of section (RRmax-min); Ratio between average radius of the section and minimum radius of the section (RRavg-min).

RL values are derived by the ANAS survey that collect, with the DELFHI advanced mobile laboratory, the average, minimum, maximum and standard deviation of the road marking at a continuous step of 50 m. The road marking condition are referred to the average value of RL parameter in the road segment in the reference year by using decay curves derived from the ANAS data.

This study uses safety performance functions (SPFs), fitted with generalized linear modelling techniques and a negative binomial distribution error structure, to develop CMFs. The SPFs quantify the effect of a specific variable on crash occurrence, and CMFs are then derived from the model coefficients.

SPFs will be carried out using two analytic methods, generalized linear models with Negative Binomial distribution (NB) and Geographically Weighted Negative Binomial Regression Models (GWNBR). The NB is commonly employed for SPF development since the state of the art (e.g. Highway Safety Manual) recommends it due to its ability to handle the dispersion in the crash data. The NB model is chosen also due to its more straightforward interpretation of results, such as the dispersion parameter of the model as in the case of estimating accidents using the Empirical Bayes method.

GWNBR is an extension of the traditional NB regression, and it can test whether a relationship is stable or varies substantially over space. Some studies showed that GWNBR models significantly reduced the spatial dependence of model residuals and highlighted that they are appropriate for spatially analyzing accident data while accounting for their over-dispersion.

By examining data on retro-reflectivity coefficient, traffic volumes, road geometry, environmental factors, and accident records over multiple years, the study seeks to identify patterns and trends that may elucidate the relationship between road marking visibility and road safety outcomes.

In addition to the analysis of the link between the dependent and independent variables carried out by NB and GWNBR, the random forest (RF) model makes it possible to compute the relative importance of the input variables as predictors of the crash outcome by using the random forests score of importance.

The analysis and subsequent processing of the results are still in progress and will be ready for presentation during the ICTCT conference in October 2024.