Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jtte

Original Research Paper

Investigating the influence of segmentation in estimating safety performance functions for roadway sections

Salvatore Cafiso [a], *Carmelo D'Agostino* [a,*], *Bhagwant Persaud* [b]

[a] *Department of Civil Engineering & Architecture, University of Catania, Catania 95125, Italy*
[b] *Department of Civil Engineering, Ryerson University, Toronto M5B2K3, Canada*

HIGHLIGHTS

- The paper investigates how the approach to segmentation affects the model form and the goodness of fit of safety performance functions for road sections.
- Four different segmentation alternatives are tested in relation to data sets usually available.
- Statistical variability and time trend were addressed using the generalized estimating equation procedure with the negative binomial error distribution.
- The different segmentation approaches are compared based on the goodness of fit of the model.
- Best and worst approaches to segmentation are identified empirically and the results seem to be logically supported.

ARTICLE INFO

Article history:
Received 17 July 2017
Received in revised form
16 October 2017
Accepted 26 October 2017
Available online 30 March 2018

Keywords:
Road safety management
Rural motorways
Safety performance functions
Segmentation
Crash prediction
General estimating equation

ABSTRACT

Safety performance functions (SPFs) are crucial to science-based road safety management. Success in developing and applying SPFs, apart data quality and availability, depends fundamentally on two key factors: the validity of the statistical inferences for the available data and on how well the data can be organized into distinct homogeneous entities. The latter aspect plays a key role in the identification and treatment of road sections or corridors with problems related to safety. Indeed, the segmentation of a road network could be especially critical in the development of SPFs that could be used in safety management for roadway types, such as motorways (freeways in North America), which have a large number of variables that could result in very short segments if these are desired to be homogeneous. This consequence, from an analytical point of view, can be a problem when the location of crashes is not precise and when there is an overabundance of segments with zero crashes. Lengthening the segments for developing and applying SPFs can mitigate this problem, but at a sacrifice of homogeneity. This paper seeks to address this dilemma by investigating four approaches for segmentation for motorways, using sample data from Italy. The best results were obtained for the segmentation based on two curves and two tangents within a segment and with fixed length segments. The segmentation

* *Corresponding author*. Tel.: +39 095 738 2213; fax: +39 095 738 2247.
E-mail addresses: dcafiso@dica.unict.it (S. Cafiso), carmelo.dagostino@dica.unict.it (C. D'Agostino), bpersaud@ryerson.ca (B. Persaud).
Peer review under responsibility of Periodical Offices of Chang'an University.
https://doi.org/10.1016/j.jtte.2017.10.001

characterized by a constant value of all original variables inside each segment was the poorest approach by all measures.

1. Introduction

Safety performance functions (SPFs) are crucial to science-based road safety management using, e.g., the methods prescribed in the Highway Safety Manual (AASHTO and FHWA, 2010). These functions are statistical models used to estimate the expected crash frequency for a facility (Hauer, 1997) based on its characteristics, mainly traffic volume, which accounts for the majority of the variability in crash frequency, and geometric variables (Milton and Mannering, 1998). These functions are developed from data for a number of similar sites. Success in development or application of an SPF for road segments depends strongly on how well the data can be organized into distinct homogeneous entities, i.e., on the approach to segmentation.

Segmentation, when based on multiple variables, may lead to very short homogeneous segments (Resende and Benekohal, 1997). For example, when using the segmentation approach proposed by the Highway Safety Manual (HSM), the presence of very short segments does not allow proper statistical inference for several reasons. The most important are the non-perfect identification of the location of crashes, which is often taken from police reports (Quin and Wellner, 2012), and the fact that crashes are rare events resulting in a great number of segments with zero crash. Lengthening segments to avoid these issues will sacrifice homogeneity.

In the literature there are a number of different approaches to segmentation. Miaou and Lum (1993) suggested that including short segments less than or equal to 80 m in the calibration data could create bias in the estimation of linear models, but not when using Poisson models. Similarly, Ogle et al. (2011) demonstrated that short segment lengths less than 160 m cause uncertain results in crash analysis. Cafiso and Di Silvestro (2011) showed that to increase performance in identifying correct positives as black spots, segment length should be related to AADT with lower AADT values requiring longer segment lengths. Quin and Wellner (2012) studied the relationship between segmentation and safety screening analysis using different lengths of sliding windows to identify hazardous sites, and concluded that short segments as well as those that are too long create a bias in the identification of these sites.

Some studies focused on the relationship between crashes and road geometry in addressing segmentation. For example, Cenek et al. (1997), who investigated this relationship for rural roads data, used a fixed segment length of 200 m. A similar study was done by Cafiso et al. (2010) using homogeneous segments with different lengths on a sample of Italian two lane rural roads, while aggregating variables related to curvature and roadside hazard. They concluded that the model that contained geometry and design consistency variables were more reliable than others. Other studies suggested different ways to aggregate segment data to avoid lengths that are too short. For example, Koorey (2009) proposed the aggregation of curves and tangents when the radius of curves exceeds a predetermined threshold value. D'Agostino (2014) analyzed the performance of two different approach to segmentation using homogeneous segments with different lengths based on HSM or IHSDM approach on a sample of Italian motorways.

The Highway Safety Manual (AASHTO and FHWA, 2010) recommends the use of homogeneous segments with respect to AADT, number of lanes, curvature, presence of ramp at the interchange, lane width, outside and inside shoulder widths, median width and clear zone width. There is no prescribed minimum segment length for application of the predictive models, but there is a suggestion of a segment length not less than 0.10 miles (0.16 km).

Given the variety of concepts and methods, and the fact that there is no apparent preferred one, this paper seeks to investigate alternative approaches for segmentation, including the HSM procedures, using sample data from Italian motorways. The litmus test is how well SPFs can be estimated with calibration datasets that result from each approach. All but one of these approaches aggregate and redefine variables over longer segments while seeking to retain the geometric and exposure characteristics of a segment as best as possible. SPFs calibrated for different segmentations are compared in terms of goodness of fit and the variables captured. At each stage a variable was added or removed and there are several variations on exactly how this can be done. The way followed was to test the p-value of the variable, or a combination of them, and the correlation between the final set of variables.

In addition, for each segmentation concept, two simpler models were estimated and compared, a base model and curvature-based model that is described later.

2. Materials and methods

This section describes the dataset used for the elaboration, the variables used, the segmentation approaches evaluated, and the models investigated, before presenting the results.

2.1. *Data description*

The data used for this investigation pertain to an Italian rural motorway, the "A18" Messina-Catania, which is approximately 76 km (47.2 miles) long. The cross section is made up of 4 lanes, 2 in each direction, divided by a median with barriers