Accident Analysis and Prevention 55 (2013) 12–21

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

Using multivariate adaptive regression splines (MARS) to develop crash modification factors for urban freeway interchange influence areas

Kirolos Haleem*, Albert Gan[1], Jinyan Lu[2]

Department of Civil and Environmental Engineering, Florida International University, Miami, FL 33174, United States

ARTICLE INFO

Article history:
Received 12 June 2012
Received in revised form
28 December 2012
Accepted 12 February 2013

Keywords:
Crash modification factors
Multivariate adaptive regression splines
Urban freeway interchange influence areas
Outside shoulder width
Inside shoulder width
Median width

ABSTRACT

Crash modification factors (CMFs) are used to measure the safety impacts of changes in specific geometric characteristics. Their development has gained much interest following the adoption of CMFs by the recently released Highway Safety Manual (HSM) and *SafetyAnalyst* tool in the United States. This paper describes a study to develop CMFs for interchange influence areas on urban freeways in the state of Florida. Despite the very different traffic and geometric conditions that exist in interchange influence areas, most previous studies have not separated them from the rest of the freeway system in their analyses. In this study, a promising data mining method known as multivariate adaptive regression splines (MARS) was applied to develop CMFs for median width and inside and outside shoulder widths for "total" and "fatal and injury" (FI) crashes. In addition, CMFs were also developed for the two most frequent crash types, i.e., rear-end and sideswipe. MARS is characterized by its ability to accommodate the nonlinearity in crash predictors and to allow the impact of more than one geometric variable to be simultaneously considered. The methodology further implements crash predictions from the model to identify changes in geometric design features. Four years of crashes from 2007 to 2010 were used in the analysis and the results showed that MARS's prediction capability and goodness-of-fit statistics outperformed those of the negative binomial model. The influential variables identified included the outside and inside shoulder widths, median width, lane width, traffic volume, and shoulder type. It was deduced that a 2-ft increase in the outside and inside shoulders (from 10 ft to 12 ft) reduces FI crashes by 10% and 33%, respectively. Further, a 42-ft reduction in the median width (from 64 ft to 22 ft) increases the rear-end, total, and FI crashes by 473%, 263%, and 223%, respectively.

1. Introduction

Crash modification factors (CMFs) are multiplicative factors that are used to quantify the impact on crash occurrences as a result of changes in specific geometric and traffic characteristics. The changes, which may increase or decrease, are measured in relative to a baseline value that is assigned a CMF of 1.0. Accordingly, if a CMF for "lane width = 11 feet" is 1.1 and the base lane width is 12 ft, a roadway segment with 11-ft lanes is expected to increase crashes by 10% compared to one with 12-ft lanes. A similar term that has traditionally been used to measure such changes is the crash reduction factor (CRF), which is the complement of CMF, or CRF = 1 − CMF. Both CMFs and CRFs are needed in the economic analysis of safety improvement projects. Specifically, they help in the selection of improvement projects by estimating the benefit from potential crash reduction associated with each project (AASHTO, 2010a).

The development of CMFs has gained much interest over the last decade (Shen and Gan, 2003; Lord and Bonneson, 2007; Bahar et al., 2007; Harkey et al., 2008; Bahar et al., 2009; Bahar, 2010; Gross et al., 2010; Stamatiadis et al., 2011; Li et al., 2011). The newly released Highway Safety Manual (HSM) (AASHTO, 2010a) has made exclusive use of CMFs for measuring the safety impacts of geometric features such as lane width, shoulder width, and presence of left turn lanes at intersections. In the HSM, CMFs for these geometric features were developed for different roadway facilities such as rural two-lane roadways, rural multilane highways, and urban and suburban arterials.

As pointed out by Stamatiadis et al. (2011), most studies have evaluated individual design features, with no consideration for the combined effect of multiple features. In fact, CMFs are often dependent. For example, Li et al. (2011) emphasized the

* Corresponding author. Tel.: +1 321 276 7889; fax: +1 305 348 2802.
E-mail addresses: khaleemm@fiu.edu (K. Haleem), gana@fiu.edu (A. Gan), jlu002@fiu.edu (J. Lu).
[1] Tel.: +1 305 348 3116; fax: +1 305 348 2802.
[2] Tel.: +1 305 495 9871; fax: +1 305 348 2802.

0001-4575/$ – see front matter
http://dx.doi.org/10.1016/j.aap.2013.02.018

simultaneous impact of lane and shoulder width changes on traffic safety. Studies including Bonneson et al. (2007) and Gross et al. (2009) have also suggested the need to account for the impact of more than one design feature at once in CMF development. Moreover, some researchers have observed that design features such as lane width and shoulder width follow a nonlinear relationship with crash frequency, e.g., the U-shaped in Li et al. (2008, 2011) and Xie et al. (2007).

To account for the simultaneous effect of multiple geometric design features, as well as the nonlinearity issue as aforementioned in CMF development, this study explores the potential use of a data mining modeling technique called the multivariate adaptive regression splines (MARS). MARS is known for its ability to consider a mixture of continuous and categorical variables and their nonlinear relationships to the dependent variable, as well as its potentially high prediction accuracy (Briand et al., 2004). Additionally, MARS is a non-black-box model, making it advantageous over neural networks (Veaux et al., 1993).

Noticeably, safety studies focusing on interchange influence areas have been relatively limited. The current version of the HSM has not included freeway facilities (AASHTO, 2010a). *SafetyAnalyst* (AASHTO, 2010b) defines roadway types that explicitly separate interchange influence areas from basic freeway segments. Making use of a comprehensive statewide dataset from Florida comprising four years of crashes from 2007 to 2010, this study applies both the MARS and traditional negative binomial (NB) models to develop CMFs for geometric design elements at urban freeway interchange influence areas. It also identifies those geometric and traffic design features affecting safety at the same facilities to aid in countermeasure selection.

2. Literature review

This section discusses methods for CMF development, studies that investigated geometric changes in key design features (e.g., shoulder width and lane width), as well as the application of data mining techniques in CMF estimation and safety analysis. Noticeably, studies that developed CMFs have focused mostly on two-lane highways (Gross et al., 2009), multi-lane highways (Fitzpatrick et al., 2010; Stamatiadis et al., 2011), and frontage roads (Li et al., 2011). Two studies that developed CMFs for freeways were conducted by Fitzpatrick et al. (2008) and Bonneson and Pratt (2009). Fitzpatrick et al. (2008) recommended CMF functions for freeways with and without median barriers using coefficients of the regression model. Bonneson and Pratt (2009) developed CMF functions for geometric design features along freeways, e.g., outside shoulder width, inside shoulder width, median width, and lane width.

Major study reports that discuss CMF development include Bonneson and Zimmerman (2007), Harkey et al. (2008), and Gross et al. (2010). Bonneson and Zimmerman (2007) developed safety design guidelines and evaluation tools for use by Texas Department of Transportation (TxDOT) engineers by incorporating CMFs in the highway design process. Harkey et al. (2008) discussed the lack of reliability and accuracy in published CMFs and developed CMFs of high or medium-high quality for treatments that did not have a CMF using an expert panel and reanalysis of crash data. Gross et al. (2010) set guidelines on developing quality CMFs and documented various study designs for CMF development in detail. Other studies that researched CMF estimation can be found in Lord and Bonneson (2007), Fitzpatrick et al. (2010), and Elvik (2009).

Specific to geometric features related to this study, Gross et al. (2009) used the case–control approach to identify whether it was safer to increase both lane and shoulder width combinations on two-lane roads given a constant total pavement width. In their study, data from roadways in Pennsylvania and Washington were obtained and analyzed separately for each state. Instead of estimating CMFs, the authors developed an odds ratio (percent change in crashes) and found a slight benefit to increasing the lane width. Similar studies can be found in Gross and Jovanis (2007) and Jovanis and Gross (2007). Similarly, Stamatiadis et al. (2011) developed crash prediction models and CMFs for changes in shoulder and median widths on four-lane roadways with 12-ft lanes. Separate models were developed for divided and undivided medians, as well as for single- and multi-vehicle crashes. The authors used 12 years of crash data from California, Minnesota, and Kentucky. They recommended CMFs for different shoulder widths on divided and undivided roadways, as well as for different median widths on divided roadways. Shoulder width was defined as the average of right shoulders for undivided roadways and the average of left and right shoulders in the same direction for divided roadways.

As discussed in Gross et al. (2010), there are different methods to develop CMFs. A non-exhaustive list would include the before-and-after, cross-sectional, case–control, expert panel, and cohort methods. The before-and-after method computes CMFs based on crash reduction before and after a geometric change. Gross et al. (2010) considered the before-and-after method to be preferable over the cross-sectional method, which develops CMFs by quantifying the differences in the crash experience associated with roadway locations of different design standards (e.g., 12-ft vs. 11-ft lanes). In the case–control method, the odds ratio is a direct estimate of CMF and is expressed as the expected change in the outcome as a result of a treatment. The expert panel is designed to evaluate the findings of published and unpublished research, and then estimate CMFs based on consensus from the panel. The cohort method estimates the relative risk (a direct representation of CMF) and indicates the expected change in the probability of an outcome for a unit change in the treatment.

Use of data mining techniques in CMF development has been relatively few. For example, Li et al. (2011) applied the generalized additive model (GAM) using five years of crash data on rural frontage roads in Texas. They recommended CMFs for combinations of lane and shoulder widths and found their values to be mostly similar to those from Lord and Bonneson (2007). On the other hand, data mining techniques have been more widely applied for crash prediction (e.g., Xie et al., 2007; Li et al., 2008) and variable selection (Qin and Han, 2008). For example, Xie et al. (2007) and Li et al. (2008) applied the Bayesian neural networks and support vector machines, respectively, to data collected on frontage roads in Texas. They concluded that both techniques provided better crash predictions compared to the NB model. Because data mining techniques are nonparametric in nature, they do not require any assumption about the relationship between the response and independent variables, thus, offers flexibility in their usage as opposed to regression models such as the generalized linear models.

While the above studies suggest that there has been no lack of effort in CMF development, the use of data mining techniques, other than the GAM, for CMF development remains in its infancy. A promising data mining technique proposed more recently for transportation applications, including safety, is MARS. For examples, Haleem et al. (2010) applied MARS to analyze rear-end crashes on unsignalized intersections and found the method to have high prediction performance; and Harb et al. (2010) used MARS to analyze factors affecting toll-lane processing time and found the method to be superior to the traditional proportional odds model. First introduced by Friedman (1991), MARS is known for its ability to accommodate nonlinear, complex data structure (such as crash data), as well as variables interactions (Briand et al., 2004). In this study, MARS is being explored as a new method for developing CMFs.