ANALYSIS OF SAFETY IMPACT OF FREEWAY DESIGNS USING DATA MINING TECHNIQUES

Juneyoung Park and Mohamed Abdel-Aty
University of Central Florida
4000 Central Florida Blvd, Orlando, FL 32816
Phone: +1-407-823-2733 E-mail: juneyoung.park@ucf.edu

1. INTRODUCTION AND OBJECTIVE
The Highway Safety Manual (HSM) (AASHTO, 2010) was developed by the Transportation Research Board (TRB) and published in 2010 to introduce a science-based technical approach for safety analysis. The HSM provides analytical methods to quantify the safety effects of decisions and treatments in planning, design, operation, and maintenance. One of the main parts in the HSM, Part D, contains crash modification factors (CMFs) for various treatments on roadway segments and at intersections. Generally, a CMF represents the overall safety performance of specific treatment. Although the first edition of HSM provides various CMFs for rural roadways and urban arterials, but there are a lack of CMFs for freeways in the HSM.

According to the HSM, a CMF can be estimated by the observational before-after studies and cross-sectional method. Observational before-after studies are well known approaches for evaluating safety effectiveness and calculating CMFs of specific roadway treatments (Gross et al., 2010). Moreover, the cross-sectional method has been commonly applied to derive CMFs due to the ease with which data can be obtained compared to the before-after approaches (Park et al., 2014). In order to estimate the CMF using the cross-sectional method, development of safety performance functions (SPFs) or crash prediction models (CPMs) is required. Due to its strength of accounting for over-dispersion, the generalized linear regression with negative binomial (NB) distribution has been widely used to develop SPFs. In the cross-sectional method, the CMF can be calculated from the coefficient of the variable associated with specific treatment of the SPF. However, the estimated CMFs from generalized linear regression model (GLM) cannot account for the nonlinear effect of the treatment since the coefficients in the GLM are assumed to be fixed.

Therefore, the objective of this study is to evaluate the safety effects of multiple roadway cross-section design elements on freeways in development of CMFs using data mining approaches. In order to account for both nonlinear effects and interaction impacts between variables, two promising data mining techniques, the multivariate adaptive regression splines (MARS) and the generalized additive models (GAM), were applied in this study.

2. METHOD
2.1. Cross-sectional Method
In this study, the cross-sectional method was used to estimate the CMFs for different roadway design features on freeways. The cross-sectional method is a useful approach to estimate CMFs if there are insufficient crash data before and after a specific treatment that is actually applied. According to the
HSM, the cross-sectional studies can be used to estimate CMFs when the date of the treatment installation is unknown and the data for the period before treatment installation are not available. As stated by Carter et al. (2012), the CMF is calculated by taking the ratio of the average crash frequency of sites with the feature to the average crash frequency of sites without the feature. Thus, the CMFs can be estimated from the coefficient of the variable associated with the treatment as the exponent of the coefficient when the form of the model is log-linear (Lord and Bonneson, 2007).

2.2. Multivariate Adaptive Regression Splines
The MARS analysis can be used to model complex relationships using a series of basis functions. Abraham et al. (2001) described that MARS as a multivariate piecewise regression technique and the splines can be representing the space of predictors broken into number of regions. Piecewise regression, also known as segmented regression, is a useful method when the independent variables, clustered into different groups, exhibit different relationships between the variables in these groups (Snedecor and Cochran, 1980). The independent variable is partitioned into intervals and a separate line segment is fit to each interval. The MARS divides the space of predictors into multiple knots (i.e. the boundary between regions) and then fits a spline functions between these knots (Friedman, 1991).

2.3. Generalized Additive Model
The GAM is a generalized linear regression model in which the linear predictor depends linearly on smooth functions of explanatory variables. The GAM was first introduced by Hastie and Tibshirani (1990). In general, the GAM uses a specified parametric or nonparametric or semi-parametric form (e.g., scatterplot function, polynomial, spline, etc.) as its smoothing function. This flexibility to allow non-parametric fits with relaxed assumptions on the actual relationship between response and predictor, provides the potential for better fits to data than purely parametric models, but arguably with some loss of interpretability.

3. RESULTS
In general, the initial results showed that following countermeasures (or treatments) (i.e., changes of roadway design) are found to be safety effective in reducing crashes:

- Adding shoulder rumble strips
- Adding inside shoulder rumble strips
- Adding a through lane
- Widening shoulder width
- Widening inside shoulder width
- Combination of treatments

4. CONCLUSION
There are very few CMFs in the HSM for the multiple treatments on freeways. Moreover, the CMFs from GLM cannot account for the nonlinear impacts. Therefore, this study analyzed the safety effects of multiple roadway design features on freeways using the cross-sectional method through development and comparison of GLM, MARS, and GAM results for different crash types and severity levels. From the initial preliminary analysis, it was found that MARS models and GAMs generally provide better model fitness than the GLMs due to its strength to account for both nonlinear effects
and interaction impacts between variables. It is expected that the comprehensive evaluation analysis will be completed by September, 2017.

KEYWORDS
Safety Effects, Crash Modification Factors, Safety Performance Function, Multivariate Adaptive Regression Spline, Generalized Additive Model

REFERENCES