ANALYSIS OF THE SAFETY CHARACTERISTICS OF UNSIGNALIZED INTERSECTIONS

Mohamed Abdel-Aty, Ph.D., P.E.

Kirolos Haleem, Ph.D.

Department of Civil, Environmental & Construction Engineering

University of Central Florida

Orlando, FL 32816-2450

Phone: (407) 823-5657

Fax: (407) 823-3315

Email: mabdel@mail.ucf.edu

ABSTRACT

This study is concerned with using multiple approaches for analyzing safety at unsignalized intersections. This was investigated by analyzing total crashes, one of the most frequent crash types at unsignalized intersections (angle crashes) and crash injury severity. Some of the developed methodological techniques in this study are considered recent and have not been extensively applied. Massive data collection effort was conducted, and resulted in collecting around 2500 unsignalized intersections from six counties in Florida.

The first analysis dealt with applying the Bayesian updating concept for better crash prediction. This was investigated by updating the coefficients of the probabilistic negative binomial (NB) models. Different non-informative and informative prior structures using the NB and log-gamma distributions were attempted. The log-gamma distribution showed the best prediction capability.

The second analysis dealt with analyzing crash injury severity using the binary probit framework. The important factors found in the fitted probit models were the logarithm of the annual average daily traffic "AADT" on the major road, and the speed limit on the major road. It was found that higher severity probability is always associated with a reduction in AADT, as well as an increase in speed limits.

Finally, a recently-developed data mining technique (the multivariate adaptive regression splines or MARS) was used, which is capable of yielding high prediction accuracy. MARS was used to analyze angle crashes. MARS yielded the best prediction performance while dealing with continuous responses. Additionally, screening the covariates using random forest technique before fitting a MARS model was very encouraging.

1. INTRODUCTION

According to documented statistics (e.g., FDOT, 2006), intersections are among the most hazardous locations on roadway systems. Many studies have extensively analyzed safety of signalized intersections, but did not put their major focus on the most frequent type of intersections; unsignalized intersections. One important reason is the inadequacy and difficulty to obtain data at these intersections, as well as the limited crash counts. Unsignalized intersections can be differentiated from their signalized counterparts in that their operation takes place without the presence of a traffic signal, and they include intersections with stop control, yield control and no traffic control.

This study gives broad comprehensive analysis of crash frequency and severity at unsignalized intersections; hence, aiming at achieving four main objectives. The first one is improving crash prediction through using a reliability process in terms of the Bayesian updating concept. This concept was used for updating the coefficients of the fitted negative binomial (NB) models, as well as for reducing uncertainty (or standard error) estimates associated with the parameters. The second objective aims at identifying those geometric, traffic and driver-related factors contributing to injury severity at unsignalized intersections using the binary probit framework.

The third objective aims at improving crash prediction as well, but using a data mining technique that could yield high prediction performance, which is the multivariate adaptive regression splines (MARS) technique. MARS could efficiently accommodate nonlinearities usually found in crash data. For this study's scope, angle crashes were used in the analysis. Finally, the fourth objective aims at introducing some research applications and countermeasures as a remedy for alleviating those safety deficiencies identified in this study. Hence, the main idea was to present to the readers different ways to deal with crash frequency (i.e., total and angle crash frequencies), as well as crash severity analysis on locations that were not extensively addressed; unsignalized intersections. For this, the study is attempting to provide beneficial information to the readers; both methodology-wise and application-wise.

2. BACKGROUND

Traditional NB models are widely used in the prediction of crash frequencies at intersections and have been applied extensively in different highway safety studies (Harwood et al., 2000; Lord and Bonneson, 2006). The NB model could efficiently accommodate over-dispersion observed in crash count data (i.e., variance is greater than the mean); hence, it is more preferable than the Poisson model.

Another approach well accepted by safety researchers is the Bayesian framework. A Bayesian formulation combines prior and current information to reach a final estimate for the expected safety performance of the site being evaluated (Persaud et al., 2009). Empirical Bayes (EB) and full Bayes are the two types of Bayesian approaches. The full Bayesian approach has been suggested lately as a useful (however complex) alternative to the EB approach that can better account for uncertainty in the data. The full Bayesian approach provides more flexibility in selecting crash count distributions. The EB concept was extensively used in crash analysis, and was originally developed to control the regression-to-the-mean effect in before-and-after studies evaluating the effects of specific traffic safety countermeasures. An example is using the EB approach for testing the reduction of crash rates at a certain location when introducing a specific countermeasure (Powers and Carson, 2004). Also, the EB estimates were used for identifying black spot locations (Saccomanno et al., 2001) by computing the difference between the results of some predictive models and the EB estimate.

For analyzing injury severity, researchers have employed various statistical techniques, and those techniques have been used extensively in traffic safety analysis. Examples of those techniques are the multinomial logit, nested logit, and ordered probit models. Abdel-Aty (2003) used the multinomial logit, nested logit and ordered probit frameworks to identify those factors affecting injury severity at toll plazas. He concluded that the multinomial logit model produced poor results when compared to the ordered probit model. Also, the author found that the ordered probit model is better than the nested logit model due to its simplicity. In addition to toll plazas, he used the ordered probit model to compare those factors that affect injury severity at other locations, including roadway sections and signalized intersections.

For the ordered probit framework, Quddus et al. (2002) analyzed motorcycle's injury severity using a 9-year crash data in Singapore. An unexpected result found is that a higher road design standard increases the probability of severe injuries and fatalities. Kockelman and Kweon (2002) used the ordered probit formulation to investigate the risk of different injury levels for single and two-vehicle crashes. They concluded that pickups and SUVs are less safe than passenger cars for single-vehicle crashes. However, in two-vehicle crashes, they found them to be safer for their drivers and more dangerous for the passengers. Duncan et al. (1998) used the ordered probit framework to examine occupant characteristics as well as roadway and environmental conditions that influence injury severity in rear-end crashes involving truck-passenger car crashes. Two models were developed, the first with the basic exogeneous variables, and the second with interactions among those exogeneous variables. They found that there is an increased severity risk for high speed crashes, those occurring at night, for women, and when alcohol is involved.

Recently, new pioneering statistical methods for modeling and predicting crashes were proposed that are very comparable to NB and Poisson models. Examples of those methods are neural networks (Mussone et al., 1999; Abdelwahab and Abdel-Aty, 2002), Bayesian neural networks (Xie et al., 2007; Riviere et al., 2006) and support vector machine "SVM" (Li et al., 2008). However, neural networks models suffer from their interpretation complexity. For this, Bayesian neural networks were introduced, and as an example, Xie et al. (2007) applied the Bayesian neural networks in predicting crashes, and found that they are more efficient than NB models. Li et al. (2008) applied a simpler technique than the Bayesian neural networks, which is SVM, to data collected on rural frontage roads in Texas. They fitted several models using different sample sizes, and compared the prediction performance of those models with the NB and Bayesian neural networks models. They found that SVM models are more efficient predictors than both NB and Bayesian neural networks.

A promising data mining (machine learning) technique explored in this study is the MARS technique that was introduced by Friedman (1991). MARS is considered a nonparametric technique since it does not require priori assumption regarding the relationship between both dependent and independent variables. This technique is effective while analyzing complex structures in the data such as nonlinearities and interactions. Crash data are those types of data that are characterized by a nonlinear relationship between the predictors and the dependent variable. Also, MARS is a regression-based technique, not suffering from the "black-box" limitation.

The application of the MARS technique can be found in Put et al. (2004) and Attoh-Okine et al. (2003). As an example, Put et al. (2004) concluded that MARS has some advantages compared to the more traditionally complicated techniques such as neural networks. Attoh-Okine et al. (2003) used the MARS technique to develop a flexible pavement roughness prediction model, and concluded that MARS allows easy interpretation of the pavement, environmental and traffic predictors detected in the model.

From the abovementioned studies, in spite of the fact that the Bayesian concept (particularly empirical Bayes) was extensively used in traffic safety analysis, a reliability method based on full Bayesian updating to reduce the uncertainties from the predictive models has not been extensively applied to date. Hence, the NB and log-gamma likelihood functions were examined in this study in the Bayesian updating procedure using informative and non-informative priors.

Also, to the authors' knowledge, almost no study addressed injury severity at unsignalized intersections, where most studies analyzing safety at unsignalized intersections (e.g., Kulmala, 1997; Vogt and Bared, 1998) have focused on exploring those factors affecting crash frequency. Therefore, an essential objective of this study is to investigate injury severity at unsignalized intersections for exploring the effect of traffic and roadway covariates on crash injury severity. The MARS technique is worth investigation as it has the advantage of accounting for nonlinearities in crash data, and for improving prediction.

3. METHODOLOGICAL APPROACHES

3.1 Bayesian Updating Framework

The Bayesian updating used both the NB and log-gamma likelihood functions while updating the coefficients of the NB models. For applying the Bayesian updating framework with the log-gamma likelihood function, the log-gamma Bayesian model of crash frequency is:

$$C_i = \exp(X_i^T \Theta)\exp(\varepsilon_i) = \hat{\lambda}\, h_i \qquad (1)$$

where: C_i is the number of crashes, X_i is the vector of variables, Θ is the vector of coefficients for these parameters, $\hat{\lambda}$ is the best estimate of the crash prediction model, and h_i = exp(ε_i) is the error term that has the one parameter gamma distribution with mean = 1, and variance σ^2 equals the over-dispersion parameter.

The NB Bayesian model uses the same functional form as shown in Equation (1); however, the error term is described by the NB distribution. The prior distribution of Θ is updated using the following formula:

$$f(\Theta) = c\, L(\Theta)\, p(\Theta) \qquad (2)$$

where: $L(\Theta)$ is the likelihood function that contains observations regarding the model that are used to update the prior joint distribution of parameters $p(\Theta)$. The resulting posterior distribution of the parameters is obtained after determination of the normalization constant *c*.

The formula used to describe the log-gamma likelihood function in this study is:

$$L(\Theta) \propto \prod p[C_i = \hat{\lambda}(\Theta)\, \exp(\varepsilon_i)] \qquad (3)$$

In this study, both non-informative and informative priors were explored. For both priors, two likelihood distributions were examined, the NB and log-gamma distributions. The informative prior uses known information and often result in lower uncertainty in the posterior distributions for each of the parameters being updated. The non-informative prior reflects a lack of information at the beginning of the analysis and can be used to estimate the joint distribution of the parameters $f(\Theta)$. As the parameters Θ considered in this study are diffuse, the non-informative prior is a constant and absorbed by the normalization constant *c*, except for the parameters describing the one-parameter log-gamma distribution (θ_g) and negative binomial distribution (α). These two parameters are limited to the positive domain; therefore the non-informative prior takes the form of $1/\theta_g$ and $1/\alpha$, respectively.

As the informative prior distribution need not be exact to obtain accurate posterior results, it is assumed in this paper that the parameters follow a multinormal distribution. This was applied for the NB and log-gamma likelihood functions with informative priors. The multinormal prior distribution is specified according to:

$$p(\Theta) = \frac{1}{(2\pi)^{n/2}\left|\Sigma_{\Theta\Theta}\right|^{1/2}} \exp[-\frac{1}{2}(\Theta - M_{\Theta})^{T}\, \Sigma_{\Theta\Theta}{}^{-1}(\Theta - M_{\Theta})] \tag{4}$$

where: M_{Θ} is the mean parameter vector, $\Sigma_{\Theta\Theta}$ is the covariance matrix determined so as to have a desirable confidence interval for the updated parameters, and *n* is the total number of parameters being estimated.

For the case of the NB likelihood function updated using an informative prior in Equation (4), the authors and the research team selected the mean parameter values based on expert traffic engineering judgment and opinion. To illustrate this point, for example, it is expected that the logarithm of annual average daily traffic (AADT) increases crash frequency at intersections, as shown in Wang and Abdel-Aty (2006), hence, it was assigned a high positive sign (e.g., +1). Other new variables such as the presence of right and left turn lanes on the major approach were based on the engineering assessment. The presence of a right turn lane on each major approach is expected to reduce crash more than the existence on one approach only. Hence, the presence of one right turn lane on each approach was assigned a value of -1, and on one approach only was assigned a value of -0.5. The covariance matrix was assigned values that could yield a 70% confidence interval, by assuming the standard deviation is equal to or greater than the parameter mean estimate (i.e., a coefficient of variation of one).

For the log-gamma likelihood with informative prior, M_{Θ} is the mean parameter vector determined from the posterior estimates of the log-gamma likelihood with non-informative prior, and $\Sigma_{\Theta\Theta}$ is the covariance matrix determined from the same posterior estimates. As the unsignalized intersection data were used to estimate these posterior statistics, a set of additional 162 three and four-legged intersection data was collected by the research team from a neighboring county (Seminole County) and used to populate the log-gamma likelihood function for the second updating to avoid using data twice. As an illustration, for the 3-legged model, the values 0, 1, 1 and 10.5 correspond to an intersection having no stop sign on the minor approach, one right turn lane on each major approach, one left turn lane on each major approach and a natural logarithm of AADT of 10.5.

The normalization constant in Equation (2) is computed according to:

$$c = \left[\int L(\Theta)p(\Theta)d\Theta\right]^{-1} \tag{5}$$

where the integral is over as many dimensions as the order of Θ. There are many numerical approaches for computing the posterior statistics, including crude Monte Carlo simulation, importance sampling, and Markov chain Monte Carlo (MCMC) simulation.

An importance sampling method of computing the Bayesian integrals is adopted in this study based on the approach shown by Gardoni et. al. (2002). To make sure that the importance sampling method yielded robust results, an established MCMC algorithm was utilized for comparison purposes. Under a sufficient number of simulations for both the importance sampling and MCMC methods, there was no significant difference in the posterior statistics. Therefore, the importance sampling approach was adopted for the remainder of the study.

The joint sampling distribution is calculated based on a Nataf distribution (Nataf, 1962). The mean is obtained from the maximum likelihood estimate of Θ and the covariance is obtained from the negative inverse of the Hessian of the log-likelihood function, evaluated at the maximum likelihood estimate. For clarification purposes, the functional form of the NB prediction model is estimated as:

$$p_i = \exp(X_i^T \Theta) \tag{6}$$

where: p_i is the predicted crash count at intersection *i*; X_i *and* Θ are previously explained in Equation (1).

3.2 Probit Framework

Ordered probit, binary probit and nested logit frameworks were adopted in this study for analyzing crash injury severity (only the binary probit models are shown to economize on space). Florida's officers classify severity as any of the five levels; property damage only (PDO), possible injury, non-incapacitating injury, incapacitating injury and fatal (within 30 days after the accident). For applying the ordered probit model, the response has the five possible levels; whereas for applying the binary probit model, the response is either severe or non-severe. Severe injury includes incapacitating injury and fatal injury, and non-severe injury included PDO, possible injury and non-incapacitating injury. The functional form of the binary probit model is:

$$Log\frac{\pi_0}{\pi_1} = Logit\{p(y \le 0)\} = \alpha + \beta' x \tag{7}$$

where: π_0 and π_1 (=1-π_0) are the probability of non-severe and severe injuries, respectively; y is the response (either "0" for non-severe injury or "1" for severe injury); α is the intercept; β is the vector of non-intercept coefficients; and x is the vector of all variables except the intercept.

Details about the probit method can be found in Abdel-Aty (2003). As for the nested logit formulation, please refer to Ben-Akiva and Lerman (1985) and Abdel-Aty and Abdelwahab (2004).

3.3 Multivariate Adaptive Regression Splines "MARS" Technique

According to Friedman (1991), the MARS method is "a local regression method that uses a series of basis functions to model complex (such as nonlinear) relationships". The global MARS model is defined according to Put et al. (2004) as shown in Equation (8).

$$\hat{y} = a_0 + \sum_{m=1}^{M} a_m B_m(x) \tag{8}$$

where: $\hat{y}$ is the predicted response; a_0 is the coefficient of the constant basis function;

$B_m(x)$ is the mth basis function, which can be a single spline function or an interaction of two (or more) spline functions; a_m is the coefficient of the mth basis function; and

M is the number of basis functions included in the MARS model.

According to Put et al. (2004), there are three main steps to fit a MARS model. The first step is a constructive phase, in which basis functions are introduced in several regions of the predictors. The second step is the pruning phase, in which some basis functions are deleted. In the third step, the optimal MARS model is selected from a sequence of smaller models.

As indicated in Put et al. (2004), the first step for describing the three MARS steps is created by continually adding basis functions to the model. Basis functions in MARS consist of a single spline function or a product (interaction) of two (or more) spline functions for different predictors (Put et al., 2004). Those basis functions are added using forward stepwise procedure. Basis functions could have either one left-sided or one right-sided truncated function defined by a given knot location, as shown in Equations (9) and (10), respectively.

$$[-(x-t)_+^q] = \begin{cases} (t-x)^q \,;\; x < t \\ 0\text{; otherwise} \end{cases} \tag{9}$$

$$[+(x-t)_+^q] = \begin{cases} (x-t)^q \,;\; x > t \\ 0\text{; otherwise} \end{cases} \tag{10}$$

From Put et al. (2004), the predictor as well as knot location having the most contribution to the model, are selected first. Also, at the end of each iteration, the introduction of an interaction is checked for model improvement. As shown by Put et al. (2004), the order of any fitted MARS model indicates the maximum number of basis functions that interact (i.e., in a third-order MARS model, the interaction order of the splines is no more than three).

The second step is the pruning step, where the backward deletion procedure is applied and the basis functions with the least contribution to the model are eliminated. The pruning is based on the generalized cross-validation (GCV) criterion (Friedman, 1991). The *GCV* criterion is used to find the overall best model from a sequence of fitted models, and to avoid the model's over-fitting. Finally, the third step deals with selecting the optimal MARS model based on the prediction characteristics of various fitted MARS models.

3.4 Random Forest Technique

Random forest is one of the promising machine learning techniques proposed by Breiman (2001) for selecting important variables from a set of variables. This technique showed promising results for variable selection (as opposed to classification and regression trees) as indicated in Harb et al. (2009). For more details about random forest, readers are encouraged to refer to Kuhn et al. (2008) and Harb et al. (2009).

4. DATA COLLECTION AND PREPARATION

The analysis conducted in this study was performed on around 2500 unsignalized intersections (specifically 2498) collected from six counties in Florida. The county selection was based on its geographic location to represent the Northern, Southern, Central, Eastern and Western parts. These data are considered the most comprehensive data collection effort in Florida.

Randomly-selected state (major) roads (SRs) were selected in each of the six counties, then the 2498 unsignalized intersections were then identified along these SRs using "Google Earth" and "Video Log Viewer Application". This application is an advanced tool developed by FDOT, and possesses two video perspectives, the "right view" and the "front view". The "right view" provides the opportunity of identifying whether a stop sign and a stop line exist or not. The "front view" identifies the median type as well as the number of lanes per direction clearly. The Geometric, traffic and control fields of the collected intersections were merged with the Crash Analysis Reporting System "CAR" database for the 4 years "2003-2006".

From the 2498 unsignalized intersections, 1955 were 3-legged intersections, and 543 were 4-legged. For the crash summary, 6088 crashes occurred in "2003-2004", of which 3900 occurred at 3-legged intersections, and 2188 occurred at 4-legged intersections. Moreover, 6446 crashes occurred in "2005-2006" (4169 and 2277 crashes at 3 and 4-legged intersections, respectively).

New important roadway and traffic covariates were explored in this study that were not extensively examined before. Examples of those new roadway covariates are the existence of crosswalks on the minor and major approaches, effect of various minor approach control types (e.g., stop sign, no control and yield sign), various sizes of intersections, intersection type (whether it is a regular unsignalized intersection, access point or ramp junction), various median types on the major approach (open, closed, two-way left turn lane, etc.), distance between unsignalized intersections and signalized ones (from both the upstream and downstream aspects), distance between successive unsignalized intersections, and left (or median) shoulder width. Figure 1 illustrates the upstream distance measure between unsignalized and signalized intersections. An important traffic covariate explored in this study is the surrogate measure for AADT on the minor approach, which is represented by the number of through lanes on this approach. The AADT on the minor approaches was not available for most of the cases, since they are mostly non-state roads.

Figure 1: Upstream Distance Measure between Unsignalized and Signalized Intersections

Three and four-legged intersections were analyzed separately in this study since both intersection types have different maneuvers and operating characteristics (Jonsson et al., 2009). For the Bayesian updating analysis, the use of the NB framework was appropriate since data over-dispersion was detected. As for the severity analysis, for either 3 or 4-legged dataset, the percentage of non-severe injury was around 90%.

5. RESULTS AND DISCUSSIONS

5.1 Bayesian Updating Analysis

Before applying the Bayesian updating analysis, two NB crash frequency models at 3 and 4-legged intersections were fitted. The two years “2003-2004” were used for calibrating the models, and the two years “2005-2006” were used for prediction and for examining the updating performance. Of the identified factors affecting total crash frequency at unsignalized intersections were AADT on the major road, existence of stop signs, configuration of the intersection, number of right and/or left turn lanes, median type on the major road, and left and right shoulder widths. Afterwards, the full Bayesian updating approach was used to update the coefficients of the models using the log-gamma and NB likelihood functions for each of the non-informative and informative priors. Hence, four Bayesian-structure models were examined; non-informative prior with a log-gamma likelihood function, non-informative prior with an NB likelihood function, informative prior with an NB likelihood function, and informative prior with a log-gamma likelihood function.

Mathematica (Wolfram Mathematica 6) was used to perform the Bayesian updating procedure for both the 3 and 4-legged NB models. There is no built-in code for executing the Bayesian updating concept; hence this was done by writing a code for estimating the posterior estimates of the parameters. The main objective was to update the distribution of the parameters in the NB model for more accurate prediction of crashes and to reduce uncertainty of the associated predicted crashes. The before and after updating mean parameter estimates for the NB and the four Bayesian updating structures for the 4-legged model are shown in Table 1.

From Table 1, the coefficient estimates using the non-informative and informative priors with NB as the likelihood function are close to those before updating. On the other hand, using the non-informative prior with log-gamma as the likelihood function, as well as the second Bayesian iteration structure led to different coefficients for some parameters. Also, it is noticed that structure 4 led to the least standard errors (bolded values) for the parameters, compared to the other three structures. The Bayesian model with structure 4 (with log-gamma as the likelihood function) is the best Bayesian-structure model, since it has the lowest AIC (and DIC) value.

Comparing the standard errors (a surrogate measure for uncertainty) for those parameters from the fitted NB model before updating and the best Bayesian-structure model (structure 4) after updating shows that there is always an uncertainty reduction after updating the parameters. The highest uncertainty reduction is 61.39% for the “log_AADT” variable. Also, to assess the prediction performance of the four Bayesian models after updating and the NB model before updating, three measures of effectiveness (MOE) were used; mean absolute deviance (MAD), mean square prediction error (MSPE) and the overall prediction accuracy, as shown in Table 2.

Table 1: Parameter Estimates for the 4-legged Model Before and After the Bayesian Updating Framework (for 4 Different Structures)

		NB Model Before Updating		Non-informative Prior (Log-gamma *) (Structure 1)		Non-informative Prior (NB *) (Structure 2)		Informative Prior (NB *) (Structure 3)		Second Bayesian Updating Iteration Using Posterior from Structure 1 (Log-gamma *) (Structure 4)	
Parameter	**Variable Description**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
Intercept		-13.4765	5.1569	-5.3639	2.6324	-14.0285	5.6094	-11.4059	3.3946	-9.8359	**2.0104**
stop_sign_mnr 2	One stop sign exists on each minor approach	0.5636	0.2606	0.5567	0.2225	0.5799	0.2748	0.5012	0.2143	0.3355	**0.1984**
stop_sign_mnr 1	One stop sign exists on one of the minor approaches	--- [a]	---	--- [a]	---	--- [a]		--- [a]		--- [a]	
major_RT 2	One right turn lane exists on each major road direction	0.6775	0.3673	0.4828	0.3193	0.6839	0.4379	1.083	0.2358	0.5388	**0.2755**
major_RT 1	One right turn lane exists on only one major road direction	-0.6274	0.3725	-0.2075	0.3027	-0.6384	0.391	-1.028	0.2774	-0.286	**0.2473**
major_RT 0	No right turn lane exists	--- [a]	---	--- [a]	---	--- [a]		--- [a]		--- [a]	
minor_through 2	Two through movements exist on both minor approaches (one on each minor approach)	-1.0664	0.3851	-1.2026	0.3475	-1.0578	0.4379	-0.602	0.2025	-1.1567	**0.2807**
minor_through 1	One through movement exists on one minor approach only	--- [a]	---	--- [a]	---	--- [a]		--- [a]		--- [a]	
major_MDT 4	A two-way left turn lane median on the major road	0.4737	0.2412	0.5101	0.2049	0.4874	0.2486	0.2217	0.1217	0.3732	**0.1464**
major_MDT 1	An open median on the major road	---[a]	---	---[a]	---	---[a]		---[a]		--- [a]	
log_AADT	Natural logarithm of the section annual average daily traffic on the major road	1.3501	0.4761	0.6270	0.2412	1.3995	0.5127	1.1693	0.3118	1.0465	**0.1838**

		NB Model Before Updating		Non-informative Prior (Log-gamma *) (Structure 1)		Non-informative Prior (NB *) (Structure 2)		Informative Prior (NB *) (Structure 3)		Second Bayesian Updating Iteration Using Posterior from Structure 1 (Log-gamma *) (Structure 4)	
Parameter	**Variable Description**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
SLDWIDTH_num	Right shoulder width on the major road (in feet)	0.0818	0.0503	0.0802	0.0402	0.0826	0.0567	-0.0029	0.0092	0.0936	**0.0325**
ISLDWDTH_num	Left shoulder width near the median on the major road (in feet)	-0.1443	0.1023	-0.0657	0.0763	-0.1728	0.1145	-0.1618	0.072	-0.0912	**0.0631**
Dispersion		0.2889	0.1321	0.3521	0.066	0.4218	0.1779	0.3717	0.1582	0.2696	0.0345
AIC [b]		283.65		111.26		184.15		190.1		**42.96**	
DIC [c]		---		112.67		184.02		180.37		**30.52**	

[*] Used likelihood function
[a] Baseline
[b] Akaike Information Criterion
[c] Deviance Information Criterion

Table 2: MOE Values for the Five 4-legged Models (Before and After Bayesian Updating)

	Before Bayesian Updating	After Bayesian Updating			
MOE	**NB Model Before Updating**	**Non-informative Prior (Log-gamma) (Structure 1)**	**Non-informative Prior (NB) (Structure 2)**	**Informative Prior (NB) (Structure 3)**	**Second Bayesian Updating Iteration Using Posterior from Structure 1 (Log-gamma) (Structure 4)**
MAD [1]	**1.79**	**1.79**	**1.80**	**1.80**	**1.71**
MSPE [2]	**5.55**	**4.98**	**5.52**	**6.33**	**4.98**
Overall Prediction Accuracy [3]	**0.68**	**0.92**	**0.71**	**0.81**	**0.84**

[1] $MAD = \frac{1}{Sample\ size}\sum\left|Actual\ crash\ frequency - \Pr edicted\ crash\ frequency\right|$ [2] $MSPE = \frac{1}{Sample\ size}\sum(Actual\ crash\ frequency - \Pr edicted\ crash\ frequency)^2$

[3] Overall prediction accuracy is estimated by dividing the total predicted crashes by the total observed crashes at the collected intersections in the prediction dataset "2005-2006"

From Table 2, the two best models are structures 1 and 4 (for the log-gamma likelihood function). It can be noted that structure 1 has the highest overall prediction accuracy (0.92), followed by structure 4 (0.84); however, structure 4 was deemed the best Bayesian-structure model, as it has a lower MAD value and there is little difference between both prediction accuracies. This indeed demonstrates the importance of using the log-gamma likelihood function as a valid distribution for updating the parameters. Thus, in conjunction with previous findings, these results show the significant effect of applying the Bayesian updating approach to increase the prediction accuracy, with structure 4 being the best Bayesian model.

5.2 Crash Injury Severity Analysis

To achieve the second objective for exploring those geometric, traffic and driver-related factors affecting crash severity, the binary probit framework was used, as shown in Table 3. From Table 3, regular unsignalized intersections are those intersections having distant stretches on the minor approaches; whereas access points include parking lots at plazas and driveways that are feeding to the major approach. For the size of the intersection (e.g., 2x4), the first number represents the total number of approach lanes for the minor approach, while the second represents the total number of through lanes for the major approach.

Table 3: Binary Probit Estimates for 3 and 4-legged Unsignalized Intersections

	Three-Legged Model		**Four-Legged Model**	
Variable Description	**Estimate [a]**	**P-value**	**Estimate [a]**	**P-value**
Intercept	-0.5872 (0.8890)	0.5089	0.6682 (0.6980)	0.3384
Natural logarithm of AADT on the major road	-0.1015 (0.0592)	0.0866	-0.1643 (0.0651)	0.0117
Natural logarithm of the upstream distance to the nearest signalized intersection	0.0528 (0.0255)	0.0383	N/S [b]	
Natural logarithm of the downstream distance to the nearest signalized intersection	0.0639 (0.0265)	0.0161	N/S	
No stop line exists on the minor approach	0.1133 (0.0629)	0.0718	N/S	
A stop line exists on the minor approach	--- [c]		N/S	
Posted speed limit on major road < 45 mph (72.4 km/hr)	-0.1252 (0.0633)	0.0481	-0.2547 (0.0722)	0.0004
Posted speed limit on major road >= 45 mph (72.4 km/hr)	--- [c]		--- [c]	
Skewness angle <= 75 degrees	N/S		0.3183 (0.1178)	0.0069
Skewness angle > 75 degrees	N/S		--- [c]	
No right turn lane exists on the major approach	-0.2139 (0.1413)	0.1302	-0.1964 (0.1106)	0.0758
One right turn lane exists on only 1 major road direction	-0.2363 (0.1464)	0.1066	0.0133 (0.1236)	0.9142
One right turn lane exists on each major road direction	--- [c]		--- [c]	
No left turn lane exists on the major approach	0.0036 (0.0751)	0.9613	N/S	
One left turn lane exists on only 1 major road direction	0.1124 (0.0607)	0.0641	N/S	
One left turn lane exists on each major road direction	--- [c]		N/S	
15 <= At-fault driver's age <= 19 (very young)	-0.2720 (0.1496)	0.0692	N/S	
20 <= At-fault driver's age <= 24 (young)	-0.2360 (0.1480)	0.1109	N/S	
25 <= At-fault driver's age <= 64 (middle)	-0.1837 (0.1391)	0.1867	N/S	
65 <= At-fault driver's age <= 79 (old)	-0.1401 (0.1591)	0.3785	N/S	
At-fault driver's age >= 80 (very old)	--- [c]		N/S	
Right shoulder width on the major road	0.0209 (0.0113)	0.0651	N/S	
Daylight lighting condition	-0.4425 (0.0864)	<0.0001	N/S	
Dusk lighting condition	-0.6063 (0.1696)	0.0004	N/S	
Dawn lighting condition	-0.3626 (0.2316)	0.1175	N/S	
Dark (street light) lighting condition	-0.2314 (0.0971)	0.0172	N/S	
Dark (no street light) lighting condition	--- [c]		N/S	
Access point unsignalized intersections	0.4426 (0.2853)	0.1209	N/S	
Ramp junctions	-4.1439 (0.1987)	<0.0001	N/A [d]	
Regular unsignalized intersections	0.4640 (0.2798)	0.0972	N/S	
Unsignalized intersections close to railroad crossings	--- [c]		N/S	
"1x2", "1x3" and "1x4" intersections	4.8632 (0.1987)	<0.0001	N/A	
"2x2" and "2x3" intersections	-0.1546 (0.2140)	0.4701	N/S	
"2x4", "2x5" and "2x6" intersections	0.0419 (0.2064)	0.8391	N/S	
"2x7" and "2x8" intersections	0.1258 (0.2489)	0.6132	N/S	
"3x2", "3x3", "3x4", "3x5", "3x6" and "3x8" intersections	0.0174 (0.2199)	0.9367	N/S	
"4x2", "4x4", "4x6" and "4x8" intersections	--- [c]		N/S	
Intersection is in Brevard County	-0.1314 (0.1216)	0.2798	0.1706 (0.1460)	0.6467

	Three-Legged Model		Four-Legged Model	
Variable Description	**Estimate [a]**	**P-value**	**Estimate [a]**	**P-value**
Intersection is in Hillsborough County	-0.1444 (0.1018)	0.1562	-0.0534 (0.1166)	0.0975
Intersection is in Leon County	-0.6443 (0.1109)	<0.0001	-0.2390 (0.1442)	0.0109
Intersection is in Miami-Dade County	-0.4746 (0.1070)	<0.0001	-0.3263 (0.1281)	0.6467
Intersection is in Orange County	-0.2244 (0.1041)	0.0312	-0.0477 (0.1331)	0.7198
Intersection is in Seminole County	--- [c]		--- [c]	
Percentage of trucks on the major road	-0.0096 (0.0085)	0.2612	N/S	
Log-likelihood at convergence	-1869		-1039	
Log-likelihood at zero	-1971.1		-1095.7	
AIC	3804		2100	

[a] Standard error in parentheses [b] N/S means not significant [c] Baseline [d] N/A means not applicable

5.2.1 Three-legged Model Interpretation

From Table 3, as anticipated, increasing the natural logarithm of AADT on the major road (which inherently means increasing AADT) significantly reduces severe injury probability. As the AADT increases, speed decreases, and hence fatal/severe crashes decrease as well, whereas crashes occurring at higher AADT (like rear-end and sideswipe crashes) are not generally severe. This is consistent with the finding of Klop and Khattak (1999).

There is a significant increase in severity probability for an increase in the natural logarithm of the upstream and downstream distances to the nearest signalized intersection. As the distance between intersections increases, drivers tend to drive at (or above) the speed limit on that stretch (which is mostly high), and thus accident severity increases as well, which is an expected outcome.

Having no stop lines on the minor approach significantly increases severity probability, when compared to having stop lines. This is a reasonable outcome, emphasizing the importance of marking stop lines at unsignalized intersections for reducing severity.

Lower speed limits (less than 45 mph “72.4 km/hr”) significantly reduce severe injury probability, when compared to speed limits greater than 45 mph “72.4 km/hr”. This conforms to the study done by Malyshkina and Mannering (2008) and Renski et al. (1998).

An interesting finding is that having 1 left turn lane on one of the major approaches significantly increases severe injury probability, when compared to having 2 left turn lanes.

The highest reduction in the probability of having severe injury occurs in young and very young at-fault drivers, when compared to very old drivers. This result is consistent with the study by Abdel-Aty et al. (1998), who concluded that young and very young drivers are associated with fatal injury reduction as well. Although very old drivers tend to drive slowly and carefully, their weak physical condition, as well as their higher reaction time could explain the higher severity risk.

Increasing the right shoulder width significantly increases the severity probability. This can be attributed to the fact that wide shoulders encourage to inappropriately using the shoulder; hence, there is a high sideswipe and rear-end crash risk, which might be severe at relatively high speeds. This conforms to that of Noland and Oh (2004), who found that increasing the right shoulder width increases severity.

The highest significant reduction in the probability of having a severe injury occurs at dusk, when compared to dark with no street lights. This might be attributed to the relatively lower conflict risk.

Although ramp junctions are usually controlled by a yield sign, and merging maneuvers are more dominant, those intersection types reduce severe injury than intersections nearby railroad crossings.

The highest increase in the probability of severe injury occurs at “1x2”, “1x3” and “1x4” intersections, when compared to intersections with four lanes on the minor approach. Intersection’s configurations (“1x2”, “1x3” and “1x4”) could exist at ramp junctions with yield signs, where merging and diverging maneuvers are frequent; hence traffic conflicts and serious injuries are more likely, especially at higher speeds.

The second highest reduction in the probability of severe injury occurs at Miami-Dade County, when compared to Seminole County. Miami-Dade County is the heaviest-populated and most urbanized county used in this study (U.S. Census, 2000); thus, more crash frequency is expected to occur. However, less severe/fatal injuries could happen due to high-dense roadways (relatively high AADT).

Increasing the percentage of trucks on the major road reduces the probability of severe injury. This could be interpreted as drivers are very attentive while overtaking or driving behind trucks. However, the probit estimate is not statistically significant.

5.2.2 Four-legged Model Interpretation

From Table 3, a skewness angle less than or equal to 75 degrees increases severity probability, when compared to skewness angle greater than 75 degrees, since sight distance is a problem. This illustrates the significant importance of designing intersections with skewness angle around 90 degrees, to reduce severe crashes.

As found in the 3-legged model, the highest reduction in the probability of severe injury occurs at Miami-Dade County, compared to Seminole County.

As expected, increasing the natural logarithm of AADT on the major road significantly reduces severe injury probability. Accordingly, speed limits less than 45 mph “72.4 km/hr” significantly reduce severe injury probability, when compared to speed limits greater than 45 mph “72.4 km/hr".

5.3 Angle Crash Analysis Using the MARS Technique

Angle crashes were specifically used in this analysis since they were the most frequent crash type at the collected intersections (also conformed by Summersgill and Kennedy, 1996; Layfield, 1996). Two NB models were initially fitted for highlighting those geometric and traffic factors affecting angle crash occurrence at 3 and 4-legged unsignalized intersections. The significant variables were traffic volume on the major road, the upstream distance to the nearest signalized intersection, the distance between successive unsignalized intersections, median type on the major approach, percentage of trucks on the major approach and size of the intersection.

The MARS technique has three main applications. The first one dealt with a comparison (in terms of prediction capability) between the NB and MARS models while treating the response in each of them as a discrete variable (angle crash frequency). The training dataset used for calibration was 70% of the total data, while the remaining 30% was used for prediction.

The second application dealt with treating the response in the MARS models as a continuous one by considering the natural logarithm of crash frequency. This was proposed due to the high prediction capability of the MARS technique while dealing with continuous responses, as shown by Friedman (1991) and indicated in Kim (2000).

The third application dealt with combining MARS with the random forest technique for screening the variables before fitting a MARS model. Thus, important covariates were identified using random forest, then fitted in a MARS model, and a comparison between the MARS models (with the covariates initially screened using random forest) and MARS models (with the covariates initially screened from the NB model) was held. The angle crash MARS model at 4-legged intersections with different basis functions' coefficients using the R package is presented in Table 4. From Table 4, there is an increase in angle crashes with the increase in the logarithm of AADT (which inherently means an increase in traffic volume). As AADT relatively increases, vehicles coming from the minor approach find it difficult to cross the major road due to congestion; hence angle crash risk might increase. However, the increase is not significant.

Table 4: Angle Crash Frequency Model at 4-Legged Unsignalized Intersections Using MARS

Basis Function	Basis Function Description	Estimate *	P-value
Intercept	Intercept	2.1314 (5.3912)	0.6928
Log_AADT	Natural logarithm of AADT on the major road	0.6831 (0.5134)	0.1840
Hills_County	Hillsborough County	-5.5343 (1.9559)	0.0049
Orange_County	Orange County	-1.4406 (0.4560)	0.0017
Lanes_3	"3x2", "3x3", "3x4", "3x5", "3x6" and "3x8" intersections	-6.3123 (2.2146)	0.0046
Acc_Point	Access points	-1.3737 (0.3382)	<0.0001
Hills_County * Lanes_3	An interaction term	7.4050 (1.8259)	<0.0001

* Standard error in parentheses

Also, it is noticed that there is an interaction term. Hence, the two variables forming the interaction term should be interpreted together. The interaction term is between Hillsborough County and unsignalized intersections with three total lanes on the minor approach. The equation representing this interaction term is: "-5.5343 * Hills_County – 6.3123 * Size_Lanes_3 + 7.4050 * Hills_County * Size_Lanes_3".

The interpretation for the formed equation is described as follows: for the case of Hillsborough (i.e., Hills_County = 1), the equation becomes: "(-6.3123 + 7.4050) * Size_Lanes_3 – 5.5343".

Hence, the equation can be simplified as "1.0927 * Size_Lanes_3 – 5.5343". Thus, the individual coefficient of "Size_Lanes_3" is "1.0927". This means that, in Hillsborough County, more angle crash frequency are observed at intersections with three total lanes on the minor approach, when compared to other intersection sizes used in the analysis (e.g., two lanes).

5.3.1 Comparing MARS and NB Models

The prediction comparison between the MARS models and the corresponding NB models, while treating the response as a discrete one (i.e., crash frequency), is shown in Table 5 on the upper part. From this part, it is noticed that the MSPE values for MARS in the 3 and 4-legged models are lower than the corresponding NB models. As for the MAD values, they are lower for the NB models. However, there is still a good potential in applying the MARS technique.

Table 5: Comparison between Various Fitted MARS Models in terms of Prediction

		Three-legged model		Four-legged model	
		MARS	**NB**	**MARS**	**NB**
MARS and NB (discrete response)	**MAD [1]**	1.27	1.07	1.08	0.85
	MSPE [1]	3.08	3.96	2.95	3.30
MARS (continuous response) [2]	**MAD [1]**	1.01	---	0.69	---
	MSPE [1]	0.74	---	0.61	---
MARS with random forest (continuous response) [2]	**MAD [1]**	0.99	---	0.69	---
	MSPE [1]	0.74	---	0.58	---

[1] MAD and MSPE values are normalized by the average of the response variable

[2] Response is the natural logarithm of angle crash frequency

5.3.2 Examining Fitting MARS Model with Continuous Response

To examine the higher prediction capability of MARS while dealing with continuous responses (Friedman, 1991), the two MARS models using the same significant variables from the NB models were fitted while considering the natural logarithm of angle crash frequency as the response. The assessment criteria for the generated MARS models are shown in Table 5 on the middle part.

By comparing the MAD and MSPE values of the fitted MARS models on the upper and middle parts of Table 5, it is noticed that the MAD and MSPE values shown in the middle part are lower. The estimated MSPE values are close to “zero”, indicating a high prediction capability. This demonstrates the higher prediction performance of MARS while dealing with continuous responses.

5.3.3 MARS in Conjunction with Random Forest

Since the MARS technique showed promising prediction performance, especially while dealing with continuous responses, an additional effort to examine screening all possible variables before fitting a MARS model was attempted using the random forest technique via the R package. The random forest technique was performed with 50 trees grown in the two training datasets. This number was found large enough to obtain stable results.

Figure 2 shows the purity values for every covariate in the 4-legged dataset. The highest variable importance ranking is the natural logarithm of AADT, followed by the county location, then the natural logarithm of the distance between two unsignalized intersections, until ending up with the existence of crosswalk on the major approach. The resulted variable importance ranking demonstrates the significant effect of the spatial covariates on angle crashes, with the distance between successive unsignalized intersections being the most significant. To screen the covariates, a cut-off purity value of “10” was used. This leads to selecting seven covariates (labelled from “1” till “7” in Figure 2). Four out of seven variables were previously found significant in the 4-legged NB model (logarithm of AADT, county, logarithm of the distance between successive unsignalized intersections and intersection size). Those seven covariates were then fitted using MARS, with the response being the natural logarithm of crash frequency, as it revealed the most promising prediction capability.

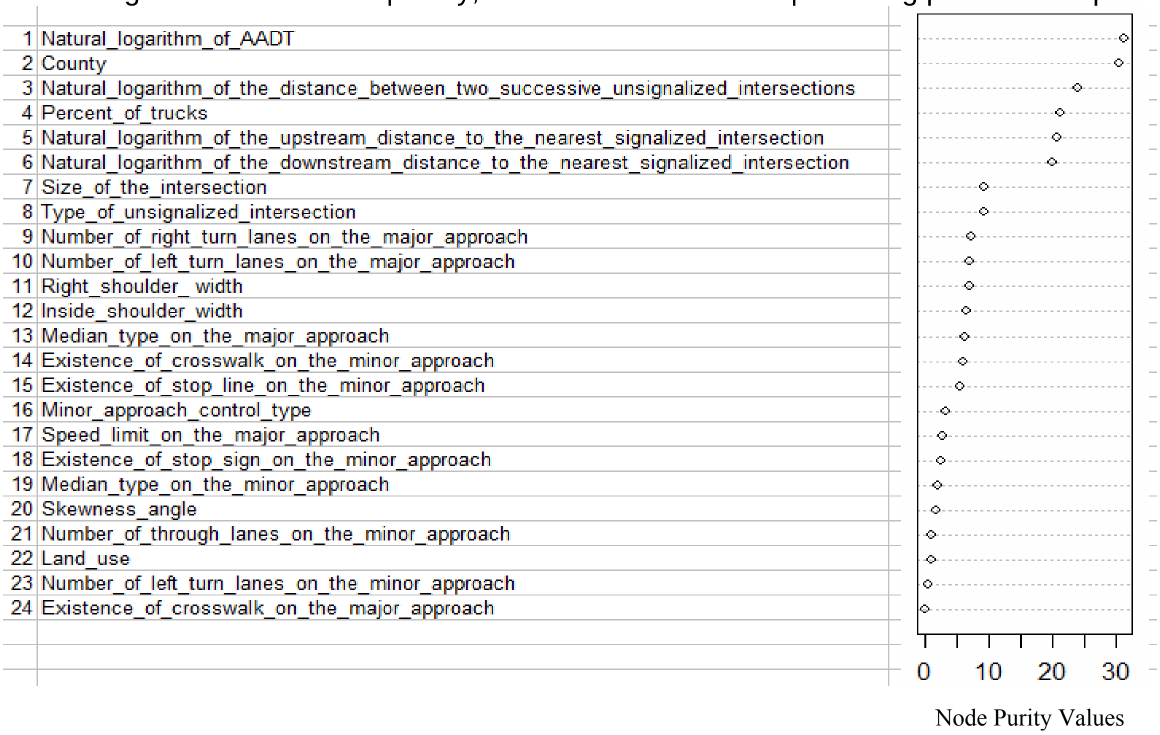

Figure 2: Variable Importance Ranking Using Node Purity Measure

To assess whether there is an improvement over the two generated MARS models using the significant variables from the NB model, the same evaluation criteria are shown on the bottom part of Table 5. Comparing the MAD and MSPE values on the middle and bottom parts of Table 5, it is noticed that there is always a reduction (even if it is small) in the MAD and MSPE values in the bottom part, hence better prediction accuracy. This demonstrates that using MARS after screening the variables using random forest is quite promising.

6. CONCLUSIONS

This study investigated multiple safety applications at unsignalized intersections using the most comprehensive data (around 2500 intersections) collected in the state of Florida. Various crash analyses were investigated through analyzing total crashes, crash injury severity, and one of the most frequent crash types at unsignalized intersections; angle crashes. Broad analysis was conducted and multiple applications were concluded; hence making this study a comprehensive one.

This study also presented predictive models with relatively large number of predictors; however, this was believed to yield high prediction performance, as shown in the study. Since crashes are an interaction between various geometric, roadway and traffic-related factors, models representing those factors are more capable of providing high prediction performance. This was also shown by El-Basyouny and Sayed (2009), as well as Mayora and Rubio (2003). For example, El-Basyouny and Sayed (2009) included also various geometric and traffic variables in their fitted accident prediction models. Also, Mayora and Rubio (2003) found that the combinations of several factors are better predictors of crash rates. The prediction performance mainly relies on the method used, the quality of the data, how the data were prepared, and not specifically on the number of predictors in the models.

A reliability analysis (in terms of the full Bayesian updating framework) was used for reducing uncertainty in predicting crash frequency at unsignalized intersections caused by the probabilistic NB model. A broad exploration of both non-informative and informative priors was conducted using both the NB and the log-gamma likelihood functions. It was concluded that the full Bayesian updating framework for updating the parameter estimates of probabilistic models is promising, and the log-gamma likelihood function is strongly recommended as a robust distribution for updating the parameters of the NB probabilistic models (yielded 84% prediction accuracy for the 4-legged model). However, the use of the estimates from the NB regression models (without updating) still led to acceptable prediction accuracy (around 70% for the 4-legged model).

The second analysis attempted to provide deep insight into factors affecting crash injury severity at 3 and 4-legged unsignalized intersections using the binary probit framework, and it showed several significant traffic, geometric and driver-related factors affecting safety characteristics. Traffic factors include AADT on the major approach, and the number of through lanes on the minor approach (surrogate measure for AADT on the minor approach).

Geometric factors include the upstream and downstream distance to the nearest signalized intersection, existence of stop lines, left and right shoulder width, number of left turn movements on the minor approach, and number of right and left turn lanes on the major approach. As for driver factors, young and very young at-fault drivers were always associated with the least fatal/severe probability compared to other age groups. Also, heavily-populated and highly-urbanized areas experience lower fatal/severe injury. For its simplicity, the binary probit models could be used to model crash injury severity at unsignalized intersections if the objective is to identify the factors contributing to severe injuries in general rather than the specific injury category.

The third analysis investigated multiple applications of a new methodology “MARS” for analyzing angle crashes, which is capable of yielding high prediction accuracy. Those significant factors affecting angle crash frequency deduced using the traditional NB model were traffic volume on the major road, the upstream distance to the nearest signalized intersection, the distance between successive unsignalized intersections, median type on the major approach, percentage of trucks on the major approach and size of the intersection.

MARS yielded the best prediction performance while dealing with continuous responses (the logarithm of angle crash frequency). Additionally, screening the covariates using random forest before fitting MARS model showed the best results. Hence, the MARS technique is recommended as a robust method for effectively predicting crashes at unsignalized intersections if prediction is the sole objective.

7. STUDY APPLICATIONS

For the study applications and countermeasures, which are deemed as the fourth objective, the results from the different methodological approaches can be applicable to diagnose some safety deficiencies identified. From the crash severity analysis, some countermeasures to reduce injury severity at unsignalized intersections could be done by designing safety awareness campaigns encouraging enforcement on speeding. Also, having a 90-degree intersection design is the most appropriate safety design for reducing severity. Moreover, making sure of marking stop lines at unsignalized intersections is quite essential. From analyzing angle crashes, since the increase in the upstream distance to the nearest signalized intersection from the unsignalized intersection of interest decreases angle crashes, it is recommended to have a relatively large spacing between signalized and unsignalized intersections.

For future direction, validating the Bayesian updating procedure could be investigated at other locations rather than unsignalized intersections, such as signalized intersections, toll plazas and roadway stretches. Also, using other techniques for variables’ screening (such as classification and regression trees “CART”) before fitting a MARS model might be interesting.

8. REFRENECES

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using probit models. Journal of Safety Research, Vol. 34, No. 5, pp. 597-603.

Abdel-Aty, M.; Abdelwahab, H. (2004). Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. Accident Analysis and Prevention, 36, pp. 447-456.

Abdel-Aty, M.; Chen, C.; Schott, J. (1998). An assessment of the effect of driver age on traffic accident involvement using log-linear models. Accident Analysis and Prevention, 30(6), pp. 851–861.

Abdelwahab, H.; Abdel-Aty, M. (2002). Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas. Transportation Research Record: Journal of the Transportation Research Board, No. 1784, TRB, National Research Council, Washington, D.C., pp. 115-125.

Attoh-Okine, N.; Mensah, S.; Nawaiseh, M. (2003). A new technique for using multivariate adaptive regression splines (MARS) in pavement roughness prediction. Proceedings of the Institution of Civil Engineers, 156 (1), pp. 51-55.

Ben-Akiva, M.; Lerman, S. (1985). Discrete Choice Analysis: Theory Application to 29 Travel Dem. MIT Press, Cambridge, MA.

Breiman, L. (2001). Random Forests, Machine Learning, 45 (1), pp. 5-32.

Crash Analysis Reporting System. Access date: January 20, 2008, http://tlhost01.dot.state.fl.us/bluezone/FDOT_Session/default.htm.

Duncan, C.; Khattak, A.; Council, F. (1998). Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions. Transportation Research Record, Vol. 1635 (pp. 63–71), Washington, D.C.

El-Basyouny, K.; Sayed, T. (2009). Urban Arterial Accident Prediction Models with Spatial Effects. Paper Presented at the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Florida Department of Transportation (2006). Florida Strategic Highway Safety Plan.

Friedman, J. (1991). Multivariate adaptive regression splines. Ann Stat 19:1–141.

Gardoni, P.; Der Kiureghian, A.; Mosalam, K. (2002). Probabilistic Models and Fragility Estimates for Bridge Components and Systems. PEER Report No. 2002/13. Pacific Earthquake Engineering Research Center, Berkeley, CA.

Google Earth. Access date: January 20, 2008, http://earth.google.com/.

Harb, R.; Yan, X., Radwan, E.; Su, X. (2009). Exploring precrash maneuvers using classification trees and random forests. Accident Analysis and Prevention, 41 (1), pp. 98-107.

Harwood, D.; Council, F.; Hauer, E.; Hughes, W.; Vogt, A. (2000). Prediction of the expected safety performance of rural two-lane highways, Federal Highway Administration, Final Report, FHWA-RD-99-207.

Jonsson, T.; Lyon, C.; Ivan, J.; Washington, S.; Schalkwyk, I.; Lord, D. (2009). Investigating Differences in the Performance of Safety Performance Functions Estimated for Total Crash Count and for Crash Count by Crash Type. Paper Presented at the Transportation Research Board 88th Annual Meeting, Washington, D.C.

Kim, J. (2000). MARS modeling for ordinal categorical response data: a case study. The Korean Communications in Statistics 7, pp. 711-720.

Klop, J.; Khattak, A. (1999). Factors Influencing Bicycle Crash Severity on Two-Lane, Undivided Roadways in North Carolina. Transportation Research Record, Vol. 1674, pp. 78–85, Washington, D.C.

Kockelman, K.; Kweon, Y. (2002). Driver injury severity: an application of ordered probit models. Accident Analysis and Prevention, 34(4), pp. 313–321.

Kuhn, S.; Egert, B.; Neumann, S.; Steinbeck, C. (2008). Building blocks for automated elucidation of metabolites: Machine learning methods for NMR prediction. BMC Bioinformatics, 9 (400).

Kulmala, R. (1997). Safety at Highway Junctions Based on Predictive Accident Models, Third International Symposium on Intersections without Traffic signals, Portland, USA, pp. 151-157.

Layfield, R. (1996). Accidents at Urban Priority Crossroads and Staggered Junctions. TRL Report 185, Transport Research Laboratory, Crowthorne, UK.

Li, X.; Lord, D.; Zhang, Y.; Xie, Y. (2008). Predicting motor vehicle crashes using Support Vector Machine models. Accident Analysis and Prevention, 40 (4), pp. 1611-1618.

Lord, D.; Bonneson, J. (2006). Development of Accident Modification Factors for Rural Frontage Road Segments in Texas. Zachry Department of Civil Engineering, Texas A&M University, College Station, TX.

Malyshkina, N.; Mannering, F. (2008). Analysis of the effect of speed limit increases on accident injury severities. Transportation Research Record, Vol. 2083, pp. 122–127, Washington, D.C.

Mayora, J.; Rubio, R. (2003). Relevant Variables for Crash Rate Prediction in Spain's Two Lane Rural Roads. Paper Presented at the 82nd Annual Meeting of the Transportation Research Board, Washington, D.C.

Mussone, L.; Ferrari, A.; Oneta, M. (1999). An Analysis of Urban Collisions Using an Artificial Intelligence Model. Accident Analysis & Prevention, Vol. 31, No. 6, pp. 705-718.

Nataf, A. (1962). Determination des Distribution don't les Marges sont Donnees, Comptes Rendus de l' Academie des Sciences, Paris, 225, pp. 42-43.

Noland, R.; Oh., L. (2004). The Effect of Infrastructure and Demographic Change on Traffic-related Crashes: A Case Study of Illinois County-level Data. Accident Analysis and Prevention, 36, pp. 525–532.

Persaud, B.; Lan, B.; Lyon, C.; Bhim, R. (2009). Comparison of Empirical Bayes and Full Bayes Approaches for Before-after Road Safety Evaluations. Paper Presented at the Transportation Research Board 88th Annual Meeting, No. 2593, Washington, D.C.

Powers, M.; Carson, J. (2004). Before-After Crash Analysis: A Primer for Using the Empirical Bayes Method. Report for the Montana Department of Transportation.

Put, R.; Xu, Q.; Massart, D.; Heyden, Y. (2004). Multivariate adaptive regression splines (MARS) in chromatographic quantitative structure–retention relationship studies. Journal of Chromatography A, 1055, pp. 11–19.

Quddus, M.; Noland, R.; Chin, H. (2002). An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety Research, 33(4), pp. 445–462.

R Software. Access date: April 27, 2009, http://www.r-project.org/.

Renski, H.; Khattak, A.; Council, F. (1998). Impact of Speed Limit Increases on Crash Injury Severity: Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways. Paper Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C.

Riviere, C.; Lauret, P.; Ramsamy, J.; Page, Y. (2006). A Bayesian Neural Network Approach to Estimating the Energy Equivalent Speed. Accident Analysis & Prevention, Vol. 38, No. 2, pp. 248-259.

Saccomanno, F.; Grossi, R.; Greco, D.; Mehmood, A. (2001). Identifying black spots along Highway SS107 in southern Italy using two models. Journal of Transportation Engineering, 127 (6), pp. 515–522.

Summersgill, I.; Kennedy, J. (1996). Accidents at Three-Arm Priority functions on Urban Single Carriageway Roads. TRL Report 184, Transport Research Laboratory, Crowthorne, UK.

U.S. Census Bureau Profile of General Demographic Characteristics 2000. Access date: May 20, 2009, http://www.census.gov/statab/ranks/rank39.html.

Video Log Viewer Application. Access date: February 10, 2008, http://webapp01.dot.state.fl.us/videolog/.

Vogt, A.; Bared, J. (1998). Accident Models for Two-Lane Rural Roads: Segments and Intersections, Report Number FHWA-RD-98-133, No. 3A5A, Final Report, Federal Highway Administration, McLean, USA.

Wang, X.; Abdel-Aty, M. (2006). Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis and Prevention, 38, pp. 1137-1150.

Wolfram Mathematica 6. Access date: May 10, 2008, http://www.wolfram.com/products/mathematica/index.html.

Xie, Y.; Lord, D.; Zhang, Y. (2007). Predicting Motor Vehicle Collisions Using Bayesian Neural Network Models: An Empirical Analysis. Accident Analysis & Prevention, Vol. 39, No. 5, pp. 922-933.