Accident Analysis and Prevention 43 (2011) 1991–1998

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Modeling animal–vehicle collisions considering animal–vehicle interactions

Yunteng Lao[a], Guohui Zhang[b,1], Yao-Jan Wu[a,2], Yinhai Wang[a,*]

[a] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA
[b] Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Austin, TX 78705, USA

ARTICLE INFO

Article history:
Received 18 October 2010
Received in revised form 2 May 2011
Accepted 17 May 2011

Keywords:
Accident modeling
Animal–vehicle collision
Carcass removal
Roadway Safety
Vehicle–animal interaction-based probability model
Negative binomial (NB) regression

ABSTRACT

Animal–Vehicle Collisions (AVCs) have been a major safety problem in the United States over the past decades. Counter measures against AVCs are urgently needed for traffic safety and wildlife conservation. To better understand the AVCs, a variety of data analysis and statistical modeling techniques have been developed. However, these existing models seldom take human factors and animal attributes into account. This paper presents a new probability model which explicitly formulates the interactions between animals and drivers to better capture the relationship among drivers' and animals' attributes, roadway and environmental factors, and AVCs. Findings of this study show that speed limit, rural versus urban, and presence of white-tailed deer habitat have an increasing effect on AVC risk, whereas male animals, high truck percentage, and large number of lanes put a decreasing effect on AVC probability.

Published by Elsevier Ltd.

1. Introduction

Over the past decade, the number of Animal–vehicle collisions (AVCs) has been rising with the continued increase of motor vehicle traffic (Curtis and Hedlund, 2005). Romin and Bissonette (1996) reported that at least 1.5 million deer-vehicle collisions occurred annually nationwide. In Washington State, approximately 3000 collisions occur annually with deer and elk on state highways (Wagner and Carey, 2006). These increasing AVCs have caused significant damage to human safety, property, and wildlife in the past decades. These collisions caused about 200 human fatalities, and 20,000 human injuries annually in the United States (Huijser et al., 2007). Property damage related with AVCs exceeds one billion dollars each year. In most AVCs, the animal dies immediately or soon after (Allen and McCullough, 1976). AVCs may also affect the population level of some precious species (e.g., Van der Zee et al., 1992; Huijser and Bergers, 2000) or even lead to a serious decrease in the probability of population survival (Proctor, 2003). Thus, a better understanding of the factors contributing to AVCs is critical for indentifying the high risk locations and prioritizing potential countermeasures, such as signs, fences, wildlife underpasses and overpasses, roadside reflectors, whistles, and diversionary feeding areas (Danielson and Hubbard, 1998).

To identify the contributing factors in general traffic accidents, a number of statistical modeling techniques have been developed based on the diverse characteristics of collisions in different circumstances. Poisson regression (e.g., Jovanis and Chang, 1986; Miaou and Lum, 1993; Miaou, 1994), negative binomial (NB) regression (or Poisson-gamma regression) (Miaou, 1994; Maher and Summersgill, 1996; Milton and Mannering, 1998; Chin and Quddus, 2003; Wang et al., 2003; Wang and Nihan, 2004; El-Basyouny and Sayed, 2006; Donnell and Mason, 2006; Kim et al., 2007; Malyshkina and Mannering, 2010; Daniels et al., 2010), and Poisson-lognormal models (Miaou et al., 2005; Lord and Miranda-Moreno, 2008) have been commonly used in accident modeling. Recently, some other innovative accident models, including finite-mixture/Markov switching models, random parameter models, Bayesian neural networks, neural networks, and support vector machines, have been used in the collision analysis studies. A detail review of these recent accident models was elaborated in (Lord and Mannering, 2010).

These regression models have been used for modeling vehicle–vehicle collisions and are able to provide insight into the contributing factors of accidents. For most AVC research, Poisson regression and negative binomial regression models are used for modeling deer-vehicle collisions to investigate the factors that influence the frequency and severity of deer-vehicle crashes

* Corresponding author. Tel.: +1 206 616 2696; fax: +1 206 543 5965.
E-mail addresses: laoy@u.washington.edu (Y. Lao), guohui@mail.utexas.edu, yaojan@slu.edu (G. Zhang), yaojan@u.washington.edu, guohui08@gmail.com (Y.-J. Wu), yinhai@u.washington.edu (Y. Wang).
[1] Address: Department of Civil Engineering, University of New Mexico, Albuquerque, NM 87106, USA.
[2] Address: Department of Civil Engineering, Parks College of Engineering, Aviation and Technology, 3450 Lindell Boulevard, McDonnell Douglas Hall Room 2051, Saint Louis University, St. Louis, MO 63103, USA. Tel.: +1 314 977 8249; fax: +1 314 977 8 388.

0001-4575/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.aap.2011.05.017

(Gkritza et al., 2010). However, most previous accident modeling studies did not reflect human factors, despite their critical roles in the crash mechanism (Wang et al., 2003). Even though Wang (1998) implemented a microscopic probability (MP) model to include drivers' responses as part of the collision model, the MP model is only designed to model vehicle collisions. In order to further investigate animal-related factors, such as animal population distribution and vehicle–animal interactions for AVCs, we are motivated to propose a vehicle–animal interaction-based probability (VAIP) model to characterize the responses of drivers and animals and the unique impacts of animal habitats on collisions to better understand AVCs and their associated contributing factors.

The remainder of this paper is organized as follows. The MP and VAIP models are introduced in Section 2 before describing the details of the test data in Section 3. Section 4 shows the model estimation results, and model interpretation and discussion are detailed in Section 5. The model spatial and temporal transferability tests are described in Section 6. In the end, this research effort will be concluded with findings and recommendations.

2. Methodology

2.1. Extended application of the MP model

2.1.1. MP model structure

This study is based on the MP model proposed by Wang (1998). An overview of the MP model and its association with the AVC model are summarized in this section. The MP model describes the relationship between the presence of a leading vehicle and the ineffective response of a driver in the following vehicle. An important advantage of this approach is its capability of considering the mechanism of accident occurrence in risk modeling. This approach has been successfully applied in many subsequent studies of accident risks (see for example Siddique, 2000, Wang et al., 2003 and Kim et al., 2007) and achieved favorable results. Although animals' behavior exhibit different patterns than drivers', we will apply Wang' MP model to formulate the VAC before the new VAIP model is developed and investigated. Their performance will be examined and analyzed as follows.

In the MP model, the probability for a randomly selected vehicle to have an accident on a certain roadway section is the probability of the driver's ineffective response P_{vf} to the obstacles conditioned on the probability of an obstacle presenting on the road P_o. In other words, the probability for a driver to have an AVC(P_{AVC}) can be expressed as the product of P_o and P_{vf} (Wang, 1998):

$$P_{AVC} = P_o \cdot P_{vf} \tag{1}$$

However, P_o and P_{vf} are not directly observable, and require further estimation.

2.1.2. P_o formulation

An animal becomes an obstacle for vehicles if its highway-crossing movement interrupts the smooth movement of vehicles. When an animal highway-crossing movement occurs within a certain period, the animal may become an obstacle to the arriving vehicle. This period is called "effective time." As the arrival of an obstacle is discrete, nonnegative, and random, it is assumed to be a Poisson arrival process. In such a process, intervals between arrivals are independent and follow the same exponential distribution (Pitman, 1993). Assuming the arrival rate is η_j in the time interval j, the average number of Poisson disturbances during the entire time period is $\sum_j \eta_j t_j$ then probability of encountering an obstacle animal, P_o, is equivalent to the probability that at least one disturbance occurs within the effective period. Therefore, P_o can be formulated as:

$$P_o = 1 - e^{-\sum_j \eta_j t_j} \tag{2}$$

In Eq. (2), $\sum_j \eta_j t_j$ should always be positive and dependent on a set of variables. Thus, an exponential link function can be employed to reflect the effects of the explanatory factors as shown as (Wang, 1998; Kim et al., 2007):

$$\sum_j \eta_{dj} t_{dj} = e^{\boldsymbol{\beta_o x_o}} \tag{3}$$

P_o then becomes:

$$P_o = 1 - e^{-e^{\boldsymbol{\beta_o x_o}}} \tag{4}$$

where $\boldsymbol{\beta_o}$ and $\boldsymbol{x_o}$ are vectors of unknown parameters and explanatory variables of disturbance frequency, respectively. $\boldsymbol{\beta_o}$ does not change with location, while $\boldsymbol{x_o}$ does. Animal habitat integrity, habitat size, and animal population are very likely contribution variables to $\boldsymbol{x_o}$.

2.1.3. P_{vf} formulations

It is assumed that a driver cannot avoid a collision if their Necessary Perception Reaction Time (NPRT) is longer than the Available Perception Reaction Time (APRT). The APRT refers to the time a driver has for completing their perception and response under a given condition. The NPRT is the ability-oriented minimum required perception reaction time and typically varies from person to person. Both the APRT and the NPRT are random variables and are assumed to follow normal distributions. Since a normal distribution does not have a closed form for cumulative probability calculation, the Weibull distribution is used instead. The NPRT is assumed to follow the Weibull (α, λ) distribution, and the APRT is assumed to follow the Weibull (α, γ) distribution. Here, λ and γ are the scale parameters. The Weibull distribution shape parameter α is chosen to be 3.25 in this study because it has been empirically verified that when α = 3.25, the Weibull distribution is a very good approximation to the normal distribution (Kao, 1960; Plait, 1962). Using the assumed distributions for the APRT and the NPRT, P_{vf} can be calculated as:

$$\begin{aligned} P_{vf} &= \int_0^\infty \int_{t_{av}}^\infty f(\lambda, t) f(\gamma, t_{av}) dt dt_{av} \\ &= \int_0^\infty e^{-\lambda t_{av}^\alpha} \alpha \gamma t_{av}^{\alpha-1} e^{-\gamma t_{av}^\alpha} dt_{av} = \frac{1}{1+\lambda/\gamma} \end{aligned} \tag{5}$$

where t_{av} is the variable used to represent the APRT. Eq. (5) shows that P_{vf} is only dependent on λ/γ, and has no relationship to α. Since the parameters λ and γ are positive variables, λ/γ can be related to various factors by using an exponential link function as shown in Eq. (5) (Wang, 1998; Kim et al., 2007). Correspondingly, P_{vf} can be written as.

$$\frac{\lambda}{\gamma} = e^{-\boldsymbol{\beta_{vh} x_{vh}}} \tag{6}$$

$$P_{vf} = \frac{1}{1 + e^{-\boldsymbol{\beta_{vh} x_{vh}}}} \tag{7}$$

where $\boldsymbol{\beta_{vh}}$ and $\boldsymbol{x_{vh}}$ are vectors of unknown parameters and explanatory variables, respectively, related to P_{vf}. Variables affecting drivers' task load and action complexity need to be included in $\boldsymbol{x_{vh}}$.

2.1.4. Integrated MP model

The application of Wang's (1998) MP model in AVC only has the terms of the probability of an animal being present on the road (P_o) and the probability of an ineffective response by the driver (P_{vf}). Substituting Eqs. (4) and (7) into Eq. (1), the probability of an individual vehicle being involved in an AVC is formulated as:

$$P_{AVC} = P_o P_{vf} = \frac{1 - e^{-e^{\boldsymbol{\beta_o x_o}}}}{1 + e^{-\boldsymbol{\beta_{vf} x_{vf}}}} \tag{8}$$

2.2. Vehicle–animal interaction-based probability (VAIP) model

As discussed in Section 1, the AVC process is difficult to accurately model and interpret because many subjective and objective factors, such as human and animal factors, cannot be properly reflected in the model. It is needed to have a modeling process that considers two significant AVC contributors: insufficient responses from drivers, such as a lack of deceleration, swerving and late responses from animals, such as freezing, running in the wrong direction. These two contributors interact with each other so that an AVC may be caused by either one or both. Since the MP model was originally developed for vehicle-to-vehicle collisions, the responses of animals were not considered in the modeling structure. An AVC could be avoided if drivers can react early and quickly to the obstacle or if the animals can notice to oncoming vehicles in a timely manner. Therefore, a third item addressing animal's response is desired in the MP model to enhance model rationality and applicability on AVCs. Thus, we propose a vehicle–animal interaction-based probability (VAIP) model as an extension of the MP model.

2.2.1. VAIP model structure

In this study, we consider that the occurrence of an AVC is conditioned on the presence of an animal in the roadway, ineffective response of the arriving vehicle driver, and the animal's failure to escape. Therefore, the vehicle–animal interaction probability can be formulated as

$$P_{AVC} = P_o \cdot P_{vf} \cdot P_{af} \tag{9}$$

where P_o is the probability of a hazardous crossing presence of an animal when vehicles travel along roadways, P_{vf} is the probability of ineffective response of the driver, and P_{af} is the probability of the animal failing to escape being hit. Thus the probability for a randomly selected vehicle to have an AVC on a certain roadway section is the product of P_o, P_{vf}, and P_{af}: In this VAIP model, P_o and P_{vf} are defined according to the MP model and P_{af} describes animals' responses in a collision. In this study we simplify animals' responses by following a similar model structure of P_{vf} by comparing the animals' necessary perception reaction time with the available perception reaction time. Here, the available preception reaction time refers to the time an animal has for noticing and escaping from the approaching vehicle. The necessary perception reaction time is the minimum required perception reaction time depending on factors such as animal species and characteristics. Both are random variables and are assumed to follow normal distributions. By following the same modeling process with P_{vf} in Section 2.1.3 "P_{vf} Formulation", P_{af} can be written as:

$$P_{af} = \frac{1}{1 + e^{-\boldsymbol{\beta_{ah}} \boldsymbol{x_{ah}}}} \tag{10}$$

where $\boldsymbol{\beta_{ah}}$ and $\boldsymbol{x_{ah}}$ are vectors of unknown parameters and explanatory variables, respectively, related to P_{af}. Variables affecting animal' action need to be included in $\boldsymbol{x_{ah}}$.

2.2.2. Integrated VAIP model

By substituting Eqs. (4), (7), and (10) into Eq. (9), the integrated VAIP risk model for each roadway section can be rewritten as:

$$P_{AVC} = P_o P_{af} P_{vf} = \frac{1 - e^{-e^{\boldsymbol{\beta_o x_o}}}}{(1 + e^{-\boldsymbol{\beta_{af} x_{af}}})(1 + e^{-\boldsymbol{\beta_{vf} x_{vf}}})} \tag{11}$$

where P_o is the probability of an animal being present on the road, P_{af} is the failure probability by the animal to escape from being hit, and P_{vf} is the probability of an ineffective response by the driver. One can see that the model contains not only road environment related factors, but also factors related to human and animal behaviors. The inclusion of human and animal factors is one of the major distinctions between the proposed model and most existing AVC models. Note that if the animals' reactions are dispensable as stationary objects, the probability, $P_{af} = 1$, and the VAIP model reduces to the MP model.

2.3. P_{AVC} formulation

It is assumed that vehicles within a traffic flow have an average AVC risk, P_{AVCi}. Since AVCs are very rare, P_{AVCi} should be very small while traffic volume f_i is very large for the given span of time. Thus, the Poisson distribution is good approximations to the binomial distribution (Pitman, 1993):

$$P(n_i) = \frac{{m_i}^{n_i} \cdot e^{-m_i}}{n_i!} \tag{12}$$

with Poisson distribution parameter:

$$m_i = E(n_i) = f_i \cdot P_{AVCi} \tag{13}$$

where f_i is the annual traffic volume that can be calculated from the annual average daily traffic (AADT) for roadway section i, and n_i is the number of AVC occurred within f_i.

The mean and variance in a Poisson distribution need to be the same. However, in most cases, accident data are over-dispersed. An easy way to overcome this difficulty is to add an independently distributed error term, ε_i, to the log transformation of Eq. (13). That is:

$$\ln m_i = \ln(f_i P_{AVCi}) + \varepsilon_i \tag{14}$$

We assume $\exp(\varepsilon_i)$ is a Gamma distributed variable with mean 1 and variance δ. Substituting Eq. (14) into Eq. (12) yields:

$$P(n_i|\varepsilon_i) = \frac{e^{(-f_i P_{AVCi} \exp(\varepsilon_i))} \cdot (f_i P_{AVCi} \exp(\varepsilon_i))^{n_i}}{n_i!} \tag{15}$$

Integrating ε_i out of Eq. (15), we can directly derive a negative binomial distribution model as the following:

$$P(n_i) = \frac{\Gamma(n_i + \theta)}{\Gamma(n_i + 1)\Gamma(\theta)} \left(\frac{\theta}{f_i \cdot P_{AVCi} + \theta} \right)^{\theta} \left(\frac{f_i P_{AVCi}}{f_i \cdot P_{AVCi} + \theta} \right)^{n_i} \tag{16}$$

where $\theta = 1/\delta$. The expectation of this negative binomial distribution equals to the expectation of the Poisson distribution shown in Eq. (14). The variance is now:

$$V(n_{ik}) = E(n_{ik})[1 + \delta E(n_{ik})] \tag{17}$$

Note that the Poisson regression model is regarded as a limiting NB regression model when δ approaches zero (Washington et al., 2003).

3. Data description

Three major data sources are used in this study:

- Carcass removal data by Washington State Department of Transportation (WSDOT) stores the information of animal carcass being collected. The information includes location (by milepost), date, weather, animal type, sex, age, etc. Carcass removal data have been commonly used in AVC research (Reilley and Green, 1974; Allen and McCullough, 1976; Knapp and Yi, 2004; Lao et al., submitted for publication). This study used two years (2005–2006) of carcass removal data from ten highway routes (US 2, SR 8, US 12, SR 20, I-90, US 97, US 101, US 395, SR 525 and

Table 1
Description of explanatory variables in the models.

	Variable	Min	Max	Mean	S.D.
Y[a]	Number of carcasses per segment[b]	0	16	0.095	0.564
z1	Annual average daily traffic (in thousands)	0.31	148.8	15.11	21.07
z2	Restrictive access control (Yes: 1; No: 0)			0.24	
z3	Speed level (>50 mph: 1; otherwise: 0)			0.68	
z4	Truck percentage level (>5%: 1; otherwise: 0)			0.78	
z5	Median width (>6 feet: 1; others: 0)			0.33	
z6	Total number of lanes (in both directions)	2	9	2.96	1.32
z7	Roadway length (mile)	0.01	6.99	0.22	0.4
z8	Terrain type (Rolling: 1; Otherwise: 0)			0.72	
z9	Terrain type (Mountainous: 1; Otherwise: 0)			0.095	
z10	Lane width (feet)	10	20	12.5	1.88
z11	Left shoulder width (feet)	0	18	2.44	2.04
z12	Right shoulder width (feet)	0	20	4.03	3.52
z13	Rural area (Rural: 1; Urban: 0)			0.76	
z14	White-tailed deer habitat (Yes: 1; No: 0)			0.31	
z15	Mule deer habitat (Yes: 1; No: 0)			0.51	
z16	Elk habitat (Yes: 1; No: 0)			0.31	
z17	Sex of animal (Male: 1; Female: 0)			0.328	
z18	Horizontal curve (Curve degree > 3: 1; otherwise: 0)			0.16	
z19	Vertical curve (Grade percentage > 3%: 1; otherwise: 0)			0.22	

[a]Specific to carcass removal data only; [b]dependent variable, number of carcasses within two years (2005–2006); Min: Minimum; Max: Maximum; S.D.: standard deviation.

SR 970) as the study routes following the recommendation from WSDOT experts.
- Deer distribution data by Washington Department of Fish Wildlife (WDFW) is in the form of GIS-based maps for mule deer, white-tailed deer, and elks.
- Roadlog data by Highway Safety Information System (HSIS) provides geometric information for the roadway, such as median width, number of lanes and shoulder width.

Table 1 lists all explanatory variables used in the modeling process. Most of the quantitative and dummy variables were directly selected from the combined dataset. Several variables were created based on the observed data. For example, the variable "Speed Level" was created based on posted speed limits. This variable is a dummy variable. The variable is set to 1 when the posted speed limit is greater than 50 mph and 0 otherwise. This is because a dramatic increase in AVCs was found when the speed limit > 50 mph. Other examples, such as variables $z14$, $z15$, and $z16$, were created for representing habitats of different types of animal.

The minimum, maximum, mean, and standard deviation (S.D.) of each variable are shown in Table 1. One can find that the reported collision data is over-dispersed as indicated by the variance being higher than the mean.

4. Model estimation

For the purpose of comparison, both a Poisson regression model (Eq. (12), when δ approaches zero in Eq. (16)) and a negative binomial regression model (Eq. (16)) were produced for the MP and VAIP model estimation using the carcass removal data. An open source statistical analysis package, R (http://www.r-project.org/, 2010), was used for model estimation in this research.

In order to evaluate the explanatory and predictive power of the model, two measures of goodness-of-fit (GOF) are adopted here for model comparisons: Adjusted ρ^2 (Ben-Akiva and Lerman, 1985), and Akaike's Information Criterion (AIC) (Akaike, 1974). Adjusted ρ^2 (rho-squared) is the log-likelihood ratio index, and is used to evaluate model's GOF for random, discrete, and sporadic count data (Ben-Akiva and Lerman, 1985; Chin and Quddus, 2003; Washington et al., 2003). The index is formulated as:

$$\rho^2 = 1 - \frac{\ln L(\hat{\boldsymbol{\beta}}) - K}{\ln L(0)} \tag{18}$$

where $L(\hat{\boldsymbol{\beta}})$ is the maximum likelihood estimation of the compared model, $L(0)$ is the initial maximum likelihood estimation of the same model with only the constant term, and K is the number of parameters estimated in the model.

AIC is another measure of GOF for a statistical model (Akaike, 1974). AIC is often used for model selection. The model with the lowest AIC is considered the best model. In general, AIC is formulated as:

$$\mathrm{AIC} = 2K - 2\ln(L) \tag{19}$$

where L is the maximum likelihood estimation of the model.

The variables were firstly assigned based on preliminary analysis, then the assignment of variables was adjusted based on their significance. The final model is selected based on the AIC value. Table 2 shows the coefficients of explanatory variables and statistical test results of the convergence MP model, estimated by negative binomial regression. Variables significantly associated with the probability of a hazardous crossing of an animal, P_o and the probability of the driver's ineffective response, P_{vf}, are shown as the explanatory variables in the models.

Similarly, the coefficients of the explanatory variables and their significance are shown for the VAIP model in Table 3. The traditional NB regression model with the standard structure was also estimated in Table 3. Compared with the traditional NB regression model, the VAIP model has made (19,484 − 17,177)/19,484 = 12% improvement on the AIC value. In addition to the probabilities, P_o and P_{vf}, the probability of the animal's failure to escape from being hit, P_{af}, is explicitly formulated. One variable, the sex of animal, is identified significant by P_{af}. Additionally, to fully understand the marginal effects of each independent variable, their elasticity values are calculated as (Shankar et al., 1995; Abdel-Aty and Radwan, 2000; Washington et al., 2003):

$$E_{x_{ik}}^{\lambda_i} = \frac{\partial \lambda_i}{\partial x_{ik}} \frac{x_{ik}}{\lambda_i} = \beta_k x_{ik} \tag{20}$$

where λ_i is the expected number of accidents for roadway segment i, x_{ik} is the k-th variable in the vector of explanatory variables for roadway segment i, and β_k is the corresponding coefficient of the k-th variable. The elasticity in Eq. (20) applies when the explanatory variable x_{ik} is continuous. In case of an indicator variable, pseudo-elasticity is estimated as an approximate elasticity of this variable (Washington et al., 2003):

$$E_{x_{ik}}^{\lambda_i} = \frac{exp(\beta_k) - 1}{\exp(\beta_k)} \tag{21}$$

5. Model interpretation

The estimated coefficients, their t-values, and GOF for the MP model and the VAIP model are shown in Tables 2 and 3 respectively. Comparing the estimation results from Tables 2 and 3, one can find

Table 2
Description of explanatory variables in the MP model.

Explanatory variables	Coeff[a]	st. err[b]	*t*-value
Variables affecting the probability of a hazardous crossing of an animal (P_o)			
Constant	−16.359	0.268	−60.945
Median width (>6 feet: 1; others: 0)	−1.016	0.137	−7.444
Total number of lanes	−0.290	0.057	−5.119
Terrain type (Rolling: 1; Otherwise: 0)	0.248	0.070	3.525
Rural area (Rural: 1; Urban: 0)	1.890	0.133	14.197
White-tailed deer habitat (Yes: 1; No: 0)	1.516	0.056	26.963
Animal sex (Male: 1; Female: 0)	−0.720	0.056	−12.876
Variables affecting the probability of ineffective response of the driver (P_{vf})			
Speed level (>50 mph: 1; otherwise: 0)	1.954	0.277	7.042
Truck percentage Level (>5%: 1; otherwise: 0)	−1.219	0.183	−6.646
Model evaluation			
AIC at base model[#]			26,861
AIC at convergence with Poisson regression			19,653
AIC at convergence with NB regression (δ = 1.66)			17,177
ρ^2			0.36

[a]Coefficients in the model; [b]standard error; ρ^2 was calculated by comparing the log-likelihood with the base model; base model[#]; δ approaches zero and β = 0.

that the GOF of these two models are almost the same: both the adjusted ρ^2 values are 0.36, and the AIC values are undistinguished. Based on the AIC values within Table 2 or 3, the negative binomial regression outperformed the corresponding Poisson regression. The estimate results show that the δ value is 1.66 in both MP and VAIP models and their p value is 0.00, which verifies that δ is significantly greater than 0, and the carcass removal data are over-dispersed. In this case, the model estimated with Poisson regression should not be used because it requires the mean and variance of the carcass removal data to be the same. Model estimated with the NB regression is a better choice for this study.

For both the MP and VAIP models estimated by the NB regression, a total of eight variables are identified as significant, including the number of lanes, terrain type, rural area, white-tailed deer habitat, median width, sex of animals, “truck percentage level”, and “speed level”. Among them, two variables, “truck percentage level” and “speed level”, have significant impacts on P_{vf}, the probability of drivers' ineffective response, and the other six variables play significant roles in determining P_o, the probability of encountering a disturbance animal in the MP model. However, in the VAIP model, one variable, sex of animal, is explicitly identified as significant by P_{af}, the probability of the animal's failure to escape from being hit, instead of P_o in the MP model. Although both models show the similar GOF, further analyses show that the VAIP model demonstrates more capability of interpreting the AVC process and the impacts of explanatory variables. Therefore, the detailed explanations and discussions regarding the VAIP model follows.

5.1. Interpretation of estimation results for P_o

The five significant variables affecting the probability of an animal's presence reflects both roadway geometric characteristics and animal distribution features as shown in Table 3. Compared with the level terrain type, rolling terrain tends to have an increasing effect on the possibility of the presence of an animal on the road P_o (Coef. = 0.248, t = 3.525, E = 0.220). This may be because rolling terrain has a higher animal population than that of level terrain. The elasticity value here shows that an incremental change of 0.22% to the AVC accident risk is caused by the changes from level terrain to rolling terrain. Similarly, compared to the highways in urban areas, those in rural areas also tend to have a higher P_o (Coef. = 1.890, t = 14.197, E = 0.849). This may also be due to the higher animal population and activity levels in rural areas. The elasticity value here shows that an incremental change of 0.849% to the AVC accident risk is caused by the changes from urban area to rural area.

Table 3
Description of explanatory variables in the VAIP model.

Explanatory variables	Coeff[a]	st. err[b]	*t*-value	E^c
Variables affecting the probability of a hazardous crossing of an animal (P_o)				
Constant	−15.666	0.268	−58.363	
Median width (>6 feet: 1; others: 0)	−1.016	0.137	−7.444	−1.762
Number of lanes	−0.290	0.057	−5.119	−0.336
Terrain type (Rolling: 1; otherwise: 0)	0.248	0.070	3.525	0.220
Rural area (Rural: 1; Urban: 0)	1.890	0.133	14.197	0.849
White-tailed deer habitat (Yes: 1; No: 0)	1.516	0.056	26.963	0.780
Variables affecting the probability of the animal failure to escape from being hit (P_{af})				
Animal sex (Male: 1; Female: 0)	−1.134	0.074	−15.347	−2.108
Variables affecting the probability of ineffective response of the driver (P_{vf})				
Speed level (>50 mph: 1; otherwise: 0)	1.954	0.277	7.042	0.858
Truck percentage level (>5%: 1; otherwise: 0)	−1.219	0.183	−6.646	−2.384
Model evaluation				
AIC at base model[e]			26,861	
AIC at convergence with Poisson regression			19,653	
AIC at convergence with standard NB regression[d]			19,484	
AIC at convergence with NB regression (δ = 1.66)			17,177	
ρ^2			0.36	

[a]Coefficients in the model; [b]standard error; [c]average elasticity value; δ is referred to as the overdispersion parameter; ρ^2 was calculated by comparing the log-likelihood with the base model; base model[e]; δ approaches zero and β = 0; standard NB regression[d]; the traditional NB regression model with the standard structure.

Among all the variables, white-tailed deer habitat was found to be the most significant explanatory variable affecting AVCs (Coef. = 1.516, t = 26.963, E = 0.780). This may be due to the higher animal population in the white-tailed deer habitat, contributing to the increased probability of animal crossing P_o. If a highway section segments a white-tailed deer habitat area, a driver using this section will have a higher probability of encountering an animal. Compared with white-tailed deer habitats, the variable of elk habitats is not significant at 95% significance level. This can be explained by the fact that the total number of collisions with elk only contributes a small part of the whole AVC records for the study period. Mule deer habitat also was not significant in the model. The reason for this may be because the mule deer habitat distribution is relatively uniformly in Washington State and covers a large portion of the study routes. The elasticity value for the white-tailed deer habitat indicates an incremental change of 0.780% on the AVC accident risk caused by the changes from other areas to white-tailed deer habitat areas. The finding is consistent with another AVC study (Lao et al., 2011).

The number of lanes is the significant factor having a negative effect on the presence of animals, P_o (Coef. = −0.290, t = −5.119, E = −0.336). With an increase in the total number of lanes, the probability of animals present on the road tends to be lower. This is understandable because roadway sections with more travel lanes are typically wider, which might increase the crossing difficulty for animals. Therefore, animals would be reluctant to cross a wider segment and thus the P_o is lower. The elasticity value here shows that a 1% increase in the number of lane decreases the AVC accident risk by 0.336%.

The variable of median width is related to roadway geometric design elements. A median width of greater than 6 feet was found to have a significant decreasing effect on P_o (Coef. = −1.016, t = −7.444, E = −1.762). This variable is similar to the number of lanes in that a wider median will increase the crossing hesitation for animals, and hence reduce the likelihood of AVCs. The elasticity value here shows a decrement change of 1.762% on the AVC accident risk caused by the changes from median width less than 6 feet to median width more than 6 feet.

5.2. Interpretation of P_{vf}

Among the factors affecting the probability of the driver's ineffective response, P_{vf}, two explanatory variables, "Speedlevel" and "Truck percentage level", were found to be significant. The speed limit level has a positive estimated coefficient (Coef. = 1.954, t = 7.042, E = 0.858). This implies that when a highway segment had a speed limit greater than 50 mph, the probability of a driver's ineffective response would increase. A vehicle running at a higher speed requires a longer stopping distance. Hence, when an animal is perceived, the reaction time for a faster vehicle is shorter. This explains why speed limit has an increasing effect on P_{vf}. This finding is consistent with many previous AVC related studies, e.g. Rolley and Lehman (1992) and Allen and McCullough (1976). The elasticity value here indicates an incremental change of 0.858% to the AVC accident risk is caused by the changes from the highways of speed limit lower than 50mph to the highways of speed limit higher than 50mph.

The truck percentage level was found to have an increasing impact on the probability of driver's effective response (decreased failure to avoid collision, Coef. = −1.219, t = −6.646, E = −2.384). This is presumably because truck drivers drive at relatively lower speeds. High-profile trucks have taller profiles, providing the drivers with longer sight distances and most truck drivers are professionally trained and well experienced. This result is supported by the motor vehicle accident research (Milton and Mannering, 1998) in which the increase in the percentage of trucks may decrease the accident probabilities. The elasticity value here indicates a decrement change of 2.384% to the AVC accident risk is caused by the changes from the areas with lower truck percentage to the areas with higher truck percentage. Additionally, it could be because truck-animal collisions were underreported due to the possibly smaller risks of property damage and people injury. Further studies are desirable to consolidate this finding.

5.3. Interpretation of P_{af}

Turning to the factors affecting the probability of animal's response, P_{af}, one variable, sex of animal, were found to affect P_{af} significantly. Compared with female animals, male animals tend to have lower collision risk (Coef. = −1.025, t = −12.877, E = −1.787). This may be because male animals require less response time than female animals. However, further study is still needed for this argument. The elasticity value here indicates a decrement change of 1.787% on the AVC accident risk caused by the change from female animals to male animals. The modeling capability of the MP model is extended by the item, P_{af} to explicitly explain unique animal response behavior with different attributes. For instance, animal species and gender may play significant roles in determining their reactions when a vehicle is approaching. Some animal species may detect the approaching vehicles much earlier than the others. Male animals may respond and run faster than females. The proposed VAIP model is capable of capturing specific animal responses in an AVC and enhances the MP model's ability in data interpretation. Due to data constraints, animal species data are not available for model calibration and estimation. Table 3 shows that sex of animal is considered significant in describing animal response behavior.

5.4. Findings and practical implications

The model estimation results indicate the proposed VAIP model extends the MP modeling capability and enables better formulation of an AVC process and identification of its significant contributing factors.

Among all the significant variables, rural area type, speed limit, white-tailed deer habitat, and rolling terrain type have positive effects on AVC risk. The remaining four variables, median width, sex of animal, truck percentage level, and number of lanes may reduce the probability of AVC risk when the values of these variables increase. Results from this model are useful to compile countermeasures against AVCs. For example, in the areas where the highway crosses the habitat of non-domestic animals, such as white-tailed deer, transportation agencies should further examine some key associated variables, such as speed limit, and develop suitable countermeasures.

6. Spatial and temporal transferability test

The relationship between AVCs and their associated factors may change temporally and spatially. Thus, a concern with the model is whether its estimated coefficients are transferable spatially or temporally. When testing spatial and temporal transferability, the following likelihood ratio test can be conducted (Washington et al., 2003):

$$X^2 = -2[LL(\beta_T) - LL(\beta_a) - LL(\beta_b)] \tag{22}$$

where $LL(\beta_T)$ is the log likelihood at convergence of the model using the data from both regions (or time periods), $LL(\beta_a)$ is the log likelihood at convergence of the model using the data from a region (or time period a), $LL(\beta_b)$ is the log likelihood at convergence of the model using the data from b region (or time period b). This X^2 statistic is a χ^2 distribution with degrees of freedom equal to the

Table 4
Spatial and temporal transferability test results for AI model 3.

	# of segments	# of accidents	log-likelihood
Spatial transferability test			
*First five routes	10,415	1,290	−3,369
#Second five routes	9,993	2,607	−5,132
Overall data	20,408	3,897	−8,585
X^2	$= -2[LL(\beta_T) - LL(\beta_a) - LL(\beta_b)] = -2[-8585 + 5132 + 3369] = 168$		
Temporal transferability test			
2005	9,942	2,110	−4,572
2006	10,466	1,787	−3,992
Overall data	20,408	3,897	−8,585
X^2	$= -2[LL(\beta_T) - LL(\beta_a) - LL(\beta_b)] = -2[-8585 + 4572 + 3992] = 42$		

*First five routes: SR 8, US 12, I-90, US 101, and SR970; #Second five routes: US 2, SR 20, US 97, US 395, and SR 525.

summation of the number of coefficients in region a and region b minus the number of coefficients in the overall model.

We statistically tested spatial and temporal transferability for the model. Table 4 shows the transferability test results. The number of segments, number of accidents, and the log-likelihoods for different data sets are also show in Table 4.

For both transferability tests, the null hypothesis was that the coefficients are transferable. For the spatial test, the first data set was routes SR 8, US 12, I-90, US 101, and SR 970, with the second data set having the remaining five routes. For the temporal test, the first data set was the year 2005, and the second the year 2006. Following Eq. (22), the data sets were estimated separately and then together. For the spatial test, X^2 was 168 with 9 degrees of freedom, which is greater than 16.92 at a 95% confidence level. For the temporal test, X^2 was 42 with 9 degrees of freedom, which is greater than 16.92 at a 95% confidence level. Thus, the coefficients were found to not be transferable between routes or years.

Although the estimated coefficients could not transfer from year to year or from location or location, the significant explanatory variables and their sign (positive or negative) converged from these different data sets are basically exact the same. The poor transferability may be attributed to the unsatisfied data quality and availability or may be a reflection of the performance and characteristic differences among drivers and animals in different time period or location. Thus, if we want to estimate a more accurate elasticity for different explanatory variables, we need to recalibrate the model using the data set in a particular time and location. However, the impacts from those variables, either being with a decreasing or an increasing factor on the AVCs, remain the same in different time periods and different locations. This implies that the model can still be applied to develop AVC countermeasures in practice.

7. Conclusions

A series of count data models have been used in AVC analysis in many previous studies. However, most of these models used in vehicle collisions seldom include human factors or animal characteristics in their analysis process, although these attributes are critical to the occurrence of AVCs. Thus, the previous models could not be directly used in the AVC analysis process.

This paper presents the MP and VAIP models and their estimation results. Both models consider the probability of drivers' ineffective response and animals' presence. As an improvement, the VAIP model includes a third term, the probability of an animals' response failure to escape, to capture animals' reaction characteristics in AVCs. The test results show that this model can provide a reasonably good explanation of the relationship among human factors, animal distributions, roadway design factors, and AVCs. Key research findings are summarized as follow:

- Compared with urban areas, the probability for a vehicle to encounter an animal is high in rural areas. This is likely due to the animal population difference between the two areas.
- The probability for a vehicle to hit a deer is much higher when driving on a highway through a white-tailed deer habitat.
- The probability of a driver's ineffective response will increase with speed limit. It goes up significantly when speed limit is greater than 50 mph.
- Compared with female animals, male animals are more alert and have a better chance to escape from potential AVCs.

Results from this model are helpful to transportation agencies for determining countermeasures against AVCs. The authors recommend that transportation agencies should further examine some key associated variables, such as speed limit, and develop suitable countermeasures in the areas where the highway crosses the habitat of non-domestic animals, such as white-tailed deer. For better application of this method, the spatial and temporal transferability tests are still needed in future study.

Acknowledgements

The authors are grateful for the financial support to this project from the Washington State Department of Transportation (WSDOT) and Transportation Northwest (TRANSNow). The authors wish to express sincere appreciation to WSDOT's Environmental Services Office and Research Office personnel, specifically Kelly McAllister and Rhonda Brooks, for their help on the data collection. The authors also want to acknowledge Highway Safety Information System (HSIS) staff member Yusuf Mohamedshah for his help with data collection. Special thanks also go to Barrett Welford Taylor at the University of Washington for his editing work, and Heather Turner from University of Warwick for her suggestions on model calculation.

References

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev. 32 (5), 633–642.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723.
Allen, R.E., McCullough, D.R., 1976. Deer-car accidents in Southern Michigan. J. Wildl. Manage. 40, 317–325.
Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press Series in Transportation Studies, 9. MIT Press, Cambridge, MA, pp. 167–168.
Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accid. Anal. Prev. 35, 253–259.
Curtis, P.D., Hedlund, J.H., 2005. Reducing Deer-Vehicle Crashes. Wildlife Damage Management Fact Sheet Series. Cornell Cooperative Extension, Ithaca, NY.
Danielson, B.J., Hubbard, M.W., 1998. A Literature Review for Assessing the Status of Current Methods of Reducing Deer-Vehicle Collisions. A report prepared for The Task Force on Animal Collisions, Iowa Department of Transportation and Iowa Department of Natural Resources, 25.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining variation in safety performance of roundabouts. Accid. Anal. Prev. 42 (2), 292–402.
Donnell, E.T., Mason, J.M., 2006. Predicting the frequency of median barrier crashes on Pennsylvania interstate highways. Accid. Anal. Prev. 38 (3), 590–599.
El-Basyouny, K., Sayed, T., 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. Transp. Res. Rec. 1950, 9–16.
Gkritza, K., Baird, M., Hans, Z.N., 2010. Deer-vehicle collisions, deer density, and land use in Iowa's urban deer herd management zones. Accid. Anal. Prev. 42 (6), 1916–1925.
http://www.r-project.org/. Accessed July 5, 2010.
Huijser, M.P., Bergers, P.J.M., 2000. The effect of roads and traffic on hedgehog (Erinaceus europaeus) populations. Biol. Conserv. 95, 111–116.
Huijser, M.P., Fuller, J., Wagner, M.E., Hardy, A., Clevenger, A.P., 2007. Animal-vehicle Collision Data Collection: A Synthesis of Highway Practice. National Cooperative Highway Research Board Program: Synthesis 370. Transportation Research Board, Washington, DC.
Jovanis, P.P., Chang, H., 1986. Modeling the relationship of accidents to miles traveled. Transp. Res. Rec. 1068, 42–51.
Kao, J.H.K., 1960. A summary of some new techniques in failure analysis. In: Proceedings of Sixth National Symposium on Reliability and Quality Control , Washington, DC, pp. 190–201.
Kim, J., Wang, Y., Ulfarsson, G., 2007. Modeling the probability of freeway rear-end crash occurrence. ASCE J. Transp. Eng. 133 (1), 11–19.
Knapp, K.K., Yi, X., 2004. Deer-Vehicle Crash Patterns and Proposed Warning Sign Installation Guidelines. Transportation Research Board, Washington, DC.
Lao, Y., Wu, Y., Corey, J., Wang, Y., 2011. Modeling animal-vehicle collisions using diagonal inflated bivariate Poisson regression. Accid. Anal. Prev. 43 (1), 220–227.
Lao, Y., Wu, Y., Wang, Y., McAllister, K. submitted for publication. Fuzzy Logic-based Mapping Algorithm for Improving Animal-Vehicle Collision Data. Working paper, Department of Civil and Environmental Engineering, University of Washington, submitted for publication.
Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Saf. Sci. 46 (5), 751–770.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Accid. Anal. Prev. 44 (5), 291–305.
Maher, M.J., Summersgill, I.A., 1996. Comprehensive methodology for the fitting of predictive accident models. Accid. Anal. Prev. 28 (3), 281–296.
Malyshkina, N., Mannering, F., 2010. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents. Accid. Anal. Prev. 42.
Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accid. Anal. Prev. 25 (6), 689–709.
Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accid. Anal. Prev. 26 (4), 471–482.
Miaou, S.-P., Bligh, R.P., Lord, D., 2005. Developing median barrier installation guidelines: a benefit/cost analysis using Texas data. Transp. Res. Rec. 1904, 3–19.
Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation 25 (4), 395–413.
Pitman, J., 1993. Probability. Springer-Verlag, New York, Inc.
Plait, A., 1962. The Weibull distribution – with the tables. Ind. Qual. Control (19), 19–26.
Proctor, M.F., 2003. Genetic Analysis of Movement, Dispersal, and Population Fragmentation of Grizzly Bears in Southwestern Canada, Ph.D. dissertation, The University of Calgary, Calgary, AB, Canada.
Reilley, R.E., Green, H.E., 1974. Deer mortality on a Michigan Interstate Highway. J. Wildl. Manage. 38, 16–19.
Rolley, R.E., Lehman, L.E., 1992. Relationships among raccoon road – kill surveys, harvests, and traffic. Wildl. Soc. Bull. 20, 313–318.
Romin, L.A., Bissonette, J.A., 1996. Deer–vehicle collisions: status of state monitoring activities and mitigation efforts. Wildl. Soc. Bull. 24 (2), 276–283.
Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometric and environment factors on rural freeway accident frequencies. Accid. Anal. Prev. 27 (3), 371–389.
Siddique, Z.Q., 2000. Accident Risk Modeling of Vehicle-to-Bicycle and Vehicle-to-Pedestrian Accidents at Four-Legged Signalized Intersections, Master Thesis, Asian Institute of Technology.
Van der Zee, F.F., Wiertz, J., ter Braak, C.J.F., van Apeldoorn, R.C., Vink, J., 1992. Landscape change as a possible cause of the Badger Meles meles L. decline in The Netherlands. Biol. Conserv. 61, 17–22.
Wagner, P. and M. Carey. 2006. Connecting Habitats and Improving Safety. WSDOT 2007-09 Transportation Research Project Problem Statements.
Wang, Y., 1998. Modeling Vehicle-to-Vehicle Accident Risks Considering the Occurrence Mechanism at Four-Legged Signalized Intersections. Ph.D. Dissertation, University of Tokyo.
Wang, Y., Ieda, H., Mannering, F.L., 2003. Estimating rear-end accident probabilities at signalized intersections: an occurrence-mechanism approach. J. Transp. Eng. 129 (4), 1–8.
Wang, Y., Nihan, N.L., 2004. Estimating the risk of collisions between bicycles and automobiles at signalized intersections. Accid. Anal. Prev. 36 (3), 313–321.
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman & Hall/carcass removal C, Boca Raton, Florida, pp. 242–243.