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

A R T I C L E   I N F O

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

A B S T R A C T

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.

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 Manning, 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; Malyschkina and Manning, 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 Manning, 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.
2.1.2. Estimation.

In Eq. (2), \( \sum \eta_j f_j \) 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):

\[
\sum \eta_j f_j = e^{\beta \eta \gamma}
\]

(3)

\( P_0 \) then becomes:

\[
P_0 = 1 - e^{\beta \eta \gamma}
\]

(4)

where \( \beta \eta \) and \( \gamma \) are vectors of unknown parameters and explanatory variables of disturbance frequency, respectively. \( \beta \eta \) does not change with location, while \( \gamma \) does. Animal habitat integrity, habitat size, and animal population are very likely contribution variables to \( \gamma \).

2.1.3. \( P_{af} \) 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 \( (\alpha, \lambda) \) distribution, and the APRT is assumed to follow the Weibull \( (\alpha, \gamma) \) distribution. Here, \( \alpha \) and \( \gamma \) are the scale parameters. The Weibull distribution shape parameter \( \alpha \) is chosen to be 3.25 in this study because it has been empirically verified that when \( \alpha = 3.25 \), 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, \( P_{af} \) can be calculated as:

\[
P_{af} = \int_{t_{av}}^{\infty} \int_{t_{av}}^{\infty} f(\lambda, \gamma, t_{af})dt_{af}dt_{av}
\]

\[
= \int_{t_{av}}^{\infty} \frac{t_{av}^{\lambda-1}e^{-t_{av}/\gamma}}{(1 + t_{av}/\gamma)^{\lambda+\gamma}} dt_{av}
\]

(5)

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

\[
\frac{\lambda}{\gamma} = e^{\beta_{\lambda} x_{\lambdah}}
\]

(6)

\[
P_{af} = \frac{1}{1 + e^{-\beta_{\lambda} x_{\lambdah}}}
\]

(7)

where \( \beta_{\lambda} \) and \( x_{\lambdah} \) are vectors of unknown parameters and explanatory variables, respectively, related to \( P_{af} \). Variables affecting drivers’ task load and action complexity need to be included in \( x_{\lambdah} \).

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 \( P_{af} \). Substituting Eqs. (4) and (7) into Eq. (1), the probability of an individual vehicle being involved in an AVC is formulated as:

\[
P_{AVC} = P_o P_{af} = \frac{1 - e^{\beta \eta \gamma}}{1 + e^{-P_{af} \gamma}}
\]

(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 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:

\[ P_{AVC} = P_o \cdot P_y \cdot P_y \]  

(9)

where \( P_o \) is the probability of a hazardous crossing presence of an animal when vehicles travel along roadways, \( P_y \) is the probability of ineffective response of the driver, and \( P_y \) 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 \( P_o \), \( P_y \), and \( P_y \): In this VAIP model, \( P_o \) and \( P_y \) are defined according to the MP model and \( P_y \) describes animals’ responses in a collision. In this study we simplify animals’ responses by following a similar model structure of \( P_y \) by comparing the animals’ necessary perception reaction time with the available perception reaction time. Here, the available perception 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 \( P_y \) in Section 2.1.3 “P_y Formulation”, \( P_y \) can be written as:

\[ P_y = \frac{1}{1 + e^{-\beta y \cdot x_y}} \]  

(10)

where \( \beta y \) and \( x_y \) are vectors of unknown parameters and explanatory variables, respectively, related to \( P_y \). Variables affecting animal action need to be included in \( x_y \).

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:

\[ P_{AVC} = P_o P_y P_y = \frac{1 - e^{-\beta y \cdot x_y}}{(1 + e^{-\beta y \cdot x_y})(1 + e^{-\beta y \cdot x_y})} \]  

(11)

where \( P_o \) is the probability of an animal being present on the road, \( P_y \) is the failure probability by the animal to escape from being hit, and \( P_y \) 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, \( P_y = 1 \), and the VAIP model reduces to the MP model.

2.3. AVC formulation

It is assumed that vehicles within a traffic flow have an average AVC risk, \( P_{AVC} \). Since AVCs are very rare, \( P_{AVC} \) 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):

\[ P(n_i) = \frac{m_i^n \cdot e^{-m_i}}{n_i!} \]  

(12)

with Poisson distribution parameter:

\[ m_i = E(n_i) = f_i \cdot P_{AVC} \]  

(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, \( \varepsilon_i \), to the log transformation of Eq. (13). That is:

\[ \ln m_i = \ln(f_i P_{AVC}) + \varepsilon_i \]  

(14)

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

\[ P(n_i|\varepsilon_i) = \frac{e^{-\frac{m_i}{\delta} + f_i P_{AVC} \exp(\varepsilon_i)}}{\varepsilon_i!} \]  

(15)

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

\[ P(n_i) = \frac{\Gamma(n_i + \theta)}{\Gamma(n_i + 1)\Gamma(\theta)} \left( \frac{\theta}{\theta + f_i P_{AVC}} \right)^{\theta} \left( \frac{f_i P_{AVC}}{\theta + f_i P_{AVC}} \right)^{n_i} \]  

(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:

\[ V(n_i) = E(n_i)[1 + \delta E(n_i)] \]  

(17)

Note that the Poisson regression model is regarded as a limiting NB regression model when \( \delta \) 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 (Reiley 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.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y*</td>
<td>0</td>
<td>16</td>
<td>0.095</td>
<td>0.564</td>
</tr>
<tr>
<td>z1</td>
<td>0.31</td>
<td>148.8</td>
<td>15.11</td>
<td>21.07</td>
</tr>
<tr>
<td>z2</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z3</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z4</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z5</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z6</td>
<td>2</td>
<td>9</td>
<td>2.96</td>
<td>1.32</td>
</tr>
<tr>
<td>z7</td>
<td>0.01</td>
<td>6.99</td>
<td>0.22</td>
<td>0.4</td>
</tr>
<tr>
<td>z8</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z9</td>
<td>0.095</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z10</td>
<td>10</td>
<td>20</td>
<td>12.5</td>
<td>1.88</td>
</tr>
<tr>
<td>z11</td>
<td>0</td>
<td>18</td>
<td>2.44</td>
<td>2.04</td>
</tr>
<tr>
<td>z12</td>
<td>0</td>
<td>20</td>
<td>4.03</td>
<td>3.52</td>
</tr>
<tr>
<td>z13</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z14</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z15</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z16</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z17</td>
<td>0.328</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z18</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z19</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Specific to carcass removal data only; *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 elk.
- Roadway 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 \( \delta \) 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 \( \rho^2 \) (Ben-Akiva and Lerman, 1985), and Akaike’s Information Criterion (AIC) (Akaike, 1974). Adjusted \( \rho^2 \) (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:

\[
\rho^2 = 1 - \frac{\ln(L(\hat{\theta})) - K}{\ln(L(0))} \tag{18}
\]

where \( L(\hat{\theta}) \) 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:

\[
\text{AIC} = 2K - 2\ln(L) \tag{19}
\]

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, \( P_{fr} \), 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 \( P_{fr} \), the probability of the animal’s failure to escape from being hit, \( P_{fr} \), is explicitly formulated. One variable, the sex of animal, is identified significant by \( P_{fr} \). 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):

\[
E_{ik}^{x_{ik}} = \frac{\partial x_{ik}}{\partial x_{ik}} \times \frac{x_{ik}}{\lambda_i} = \beta_k x_{ik} \tag{20}
\]

where \( \lambda_i \) 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 \( \beta_k \) 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):

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

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...
that the GOF of these two models are almost the same: both the adjusted $\rho^2$ 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 $\delta$ value is 1.66 in both MP and VAIP models and their p value is 0.00, which verifies that $\delta$ 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 $P_{af}$, 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 2

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Coeff</th>
<th>st. err</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables affecting the probability of a hazardous crossing of an animal ($P_o$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>$-16.359$</td>
<td>$0.268$</td>
<td>$-60.945$</td>
</tr>
<tr>
<td>Median width (&gt;6 feet: 1; others: 0)</td>
<td>$1.016$</td>
<td>$0.137$</td>
<td>$7.444$</td>
</tr>
<tr>
<td>Total number of lanes</td>
<td>$-0.290$</td>
<td>$0.057$</td>
<td>$-5.119$</td>
</tr>
<tr>
<td>Terrain type (Rolling: 1; Otherwise: 0)</td>
<td>$0.248$</td>
<td>$0.070$</td>
<td>$3.525$</td>
</tr>
<tr>
<td>Rural area (Rural: 1; Urban: 0)</td>
<td>$1.890$</td>
<td>$0.133$</td>
<td>$14.197$</td>
</tr>
<tr>
<td>White-tailed deer habitat (Yes: 1; No: 0)</td>
<td>$1.516$</td>
<td>$0.056$</td>
<td>$26.963$</td>
</tr>
<tr>
<td>Animal sex (Male: 1; Female: 0)</td>
<td>$-0.720$</td>
<td>$0.056$</td>
<td>$-12.876$</td>
</tr>
<tr>
<td>Variables affecting the probability of ineffective response of the driver ($P_{af}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed level (&gt;50 mph: 1; otherwise: 0)</td>
<td>$1.954$</td>
<td>$0.277$</td>
<td>$7.042$</td>
</tr>
<tr>
<td>Truck percentage level (&gt;5%: 1; otherwise: 0)</td>
<td>$-1.219$</td>
<td>$0.183$</td>
<td>$-6.646$</td>
</tr>
<tr>
<td>Model evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC at base model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC at convergence with Poisson regression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC at convergence with standard NB regression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC at convergence with NB regression ($\delta = 1.66$)</td>
<td>$19,777$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| $\rho^2$ | | | 0.36

Table 3

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

*aCoefficients in the model; bstandard error; caverage elasticity value; $\delta$ is referred to as the overdispersion parameter; $\rho^2$ was calculated by comparing the log-likelihood with the base model; base model$^d$; $\delta$ approaches zero and $\beta = 0$; 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 $P_a$. 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, $P_a$ (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 $P_a$ 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 $P_a$ (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 $P_{af}$

Among the factors affecting the probability of the driver's ineffective response, $P_{af}$, 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 $P_{af}$. 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):

$$X^2 = -2[LL(\beta_1) - LL(\beta_0) - LL(\beta_b)]$$

where $LL(\beta_1)$ is the log likelihood at convergence of the model using the data from both regions (or time periods), $LL(\beta_0)$ 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 $\chi^2$ distribution with degrees of freedom equal to the
Table 4
Spatial and temporal transferability test results for AI model 3.

<table>
<thead>
<tr>
<th></th>
<th># of segments</th>
<th># of accidents</th>
<th>log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial transferability test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First five routes</td>
<td>10,415</td>
<td>1,290</td>
<td>−3.369</td>
</tr>
<tr>
<td>*Second five routes</td>
<td>9,993</td>
<td>2,607</td>
<td>−5.132</td>
</tr>
<tr>
<td>Overall data</td>
<td>20,408</td>
<td>3,897</td>
<td>−8.585</td>
</tr>
<tr>
<td>$X^2$</td>
<td>$−2[LL(\beta_1) − LL(\beta_2)] = −2[−8585 + 5132 + 3369] = 168$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 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         |
| $X^2$                     | $−2[LL(\beta_1) − LL(\beta_2)] = −2[−8585 + 4572 + 3992] = 42$ |


The model was not significantly transferable in region b, but the number of coefficients in region a and b minus 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 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 is attributed to the unsatisfactory 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.

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 McAlister 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


