MODELLING ANIMAL-VEHICLE COLLISION COUNTS ACROSS LARGE NETWORKS USING A BAYESIAN HIERARCHICAL MODEL WITH TIME-VARYING PARAMETERS

Krishna Murthy Gurumurthy, Ph.D.
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
gkmurthy10@utexas.edu

Prateek Bansal, Ph.D.
Assistant Professor
Department of Civil and Environmental Engineering
National University of Singapore, Singapore
prateekb@nus.edu.sg

Kara M. Kockelman, Ph.D., P.E.
Professor and Dewitt Greer Centennial Professor of Transportation Engineering
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin – 6.9 E. Cockrell Jr. Hall
Austin, TX 78712-1076
kkockelm@mail.utexas.edu

Zili Li, Ph.D.
Post-Doctoral Researcher
University of Queensland
Brisbane, Australia
zili.li@uq.edu.au

11 May 2022

Forthcoming in *Analytic Methods in Accident Research*.

ABSTRACT

Animal-vehicle collisions (AVCs) are common around the world and result in considerable loss of animal and human life, as well as significant property damage and regular insurance claims. Understanding their occurrence in relation to various contributing factors and being able to identify high-risk locations are valuable to AVC prevention, yielding economic, social, and environmental cost savings. However, many challenges exist in the study of AVC datasets. These include seasonality of animal activity, unknown exposure (i.e., the number of animal crossings), very low AVC counts across most sections of extensive roadway networks, and computational burdens that come with discrete response analysis using large datasets. To overcome these challenges, a Bayesian hierarchical model is proposed where the exposure is modeled with nonparametric Dirichlet process, and the number of segment-level AVCs is assumed to follow a Binomial distribution. A Pólya-Gamma augmented Gibbs sampler is derived to estimate the proposed model. By using the AVC data of multiple years across about

85,000 segments of state-controlled highways in Texas, U.S., it is demonstrated that the model is scalable to large datasets, with a preponderance of zeros and clear monthly seasonality in counts, while identifying high-risk locations and key explanatory factors based on segment-specific factors (such as changes in speed limit) can be done within the modelling framework, which provide useful information for policy-making purposes.

Keywords: Animal-vehicle collisions, count modelling, seasonality, Pólya-Gamma augmentation, hierarchical models.

INTRODUCTION

Animal-vehicle collisions (AVCs) are common around the world and result in considerable loss of animal and human life, as well as significant property damage and regular insurance claims (Bruinderink and Hazebroek, 2003; Al-Ghamdi and AlGadhi, 2004; Seiler, 2005; Klöcker, Croft and Ramp, 2006; Mountrakis and Gunson, 2009; Sullivan, 2011; Mrtka and Borkovcová, 2013). For such reasons, there is continued research in AVC prediction and the effectiveness of various prevention measures (Gunson, Mountrakis and Quackenbush, 2011).

Special attention has been paid to AVCs' spatial and temporal attributes, due to clustering at certain times of year and times of day, with different species' movements and breeding seasons (see, e.g., Wilkins et al. 2019). In the spatial dimension, the focus is on identifying the relationship between AVC locations and animal habitats (Hurley, Rapaport and Johnson, 2009; Gkritza, Baird and Hans, 2010; Dettki *et al.*, 2011) and nearby landscapes (MALO, SUÁREZ and DÍEZ, 2004; Grilo, Bissonette and Santos-Reis, 2009; Danks and Porter, 2010; Jensen, Gonser and Joyner, 2014). In the temporal dimension, within-day and seasonal activity patterns of both animals and humans vary, affecting vehicle presence and animal presence - and their collisions - on roadways (Haikonen and Summala, 2001; Rowden, Steinhardt and Sheehan, 2008; Dettki *et al.*, 2011; Diaz-Varela *et al.*, 2011). Across the year, migratory patterns, variations in sunrise and sundown, and climatic conditions also play a role (Garrett and Conway, 1999; Rodríguez-Morales, Díaz-Varela and Marey-Pérez, 2013; Hothorn *et al.*, 2015; Niemi *et al.*, 2017).

In terms of AVC prevention, the effectiveness of warning signs (Ujvari, Baagoe and Madsen, 2007), light-reflecting devices (Brieger *et al.*, 2016), fencing and barriers (LEBLOND *et al.*, 2007; Zuberogoitia *et al.*, 2015), overpasses and underpasses (McCollister and van Manen, 2010; Rodriguez *et al.*, 2010), modification of nearby landscapes (Jaeger *et al.*, 2016), overhead lighting, and other treatments have been investigated. Some studies have emphasized the effects of roadway design details on AVCs, like speed limit choices (Found and Boyce, 2011; Meisingset *et al.*, 2014), road widths (Litvaitis and Tash, 2008), shoulder widths (Lao, Wu, *et al.*, 2011), and the number of lanes used (Lao, Zhang, *et al.*, 2011). Most of these studies have done an aggregate level (zonal or corridor level) analysis to avoid discrete counts, which allows researchers to focus on general trends.

A handful of recent studies have also explored the factors affecting the severity levels of AVCs using discrete choice models (Al-Bdairi, Behnood and Hernandez, 2020; Ahmed,

Cohen and Anastasopoulos, 2021), but the results of these models are difficult to apply in developing preventive measures because accident-level factors generally dominate operational and planning factors in determining the severity of an AVC. Instead, high-risk locations are central to assessing AVC clustering on segments in large networks. Two approaches are normally adopted to deal with thousands of distinct locations and network links. A computationally simple approach used by Kolowski and Nielsen (2008) relies on correlation coefficients to define the similarity between road segments with AVC occurrences, and judges high-risk locations according to correlation strengths. Alternatively, kernel-based smoothing can be applied across all segments at once (Ramp, Wilson and Croft, 2006; Snow, Williams and Porter, 2014; Bíl *et al.*, 2016). As noted by Snow et al. (2014), such methods normally require a large number of subjective inputs (like the spatial weights and kernel band-width used) in the implementation process, and can result in unreliable inference. More importantly, the relative importance of each attribute for identifying high-risk locations is generally unknown, and scenario evaluations based on specific attributes can be unreliable or impossible.

This study improves upon such methods and demonstrates how high-risk locations can be identified using a Bayesian binomial regression model on a large-scale network of around eighty-five thousand segments. A Gibbs sampler is derived to estimate the proposed model. The model facilitates scenario evaluations based on any segment-specific attribute (like speed limit and average daily traffic), which could be valuable in developing AVC prevention practices. There are three main challenges in modeling AVC counts. First, traditional count data models cannot be directly used because exposure (i.e., the number of animal road crossings) is unknown. Second, a high proportion of segments have zero AVCs. Third, along with unobserved heterogeneity in the effect of covariates across segments, seasonality and spatial correlation are required to be modeled to capture heterogeneity in AVC counts. No existing discrete response model can address all these challenges simultaneously due to trade-off between computational tractability and flexibility, but several studies could handle them individually. The literature related to each of these modeling challenges and the adopted approach is discussed below.

Whereas traffic volume and its proxies are used as the exposure in crash count data models, exposure is not required in ordered or multinomial response (i.e., injury severity) models. To the best of our knowledge, entirely unknown exposure has not been modeled in crash count data models. However, Crépet and Tressou (2011) demonstrate how a nonparametric Dirichlet process (DP) mixture can be used to model exposure in food risk analysis. DP mixture has also been used in accident analysis, but to model the semi-parametric heterogeneity in multivariate and multilevel count data models (Heydari *et al.*, 2016, 2017). We illustrate the first application of the DP mixture to model exposure in crash count data models.

High proportion of zeros in discrete responses are generally handled using zero-inflated models (Anastasopoulos, 2016; Fountas and Anastasopoulos, 2018; Liu *et al.*, 2018). However, we do not adjust for preponderance of segments with zero crashes because DP mixture inherently handles such situations. Specifically, DP mixture creates clusters of segments in a data-driven manner and segments in same cluster can share the information about the number of animal crossings.

Extant literature has emerged in the last decade on modeling unobserved heterogeneity in discrete response regression models, such as spatiotemporal correlation in intercept term (Liu and Sharma, 2017, 2018), mixture-of-normal-distributed random parameters to represent cross-sectional unobserved heterogeneity (Xiong and Mannering, 2013; Buddhavarapu, Scott and Prozzi, 2016; Mannering, Shankar and Bhat, 2016; Huang *et al.*, 2019), and heterogeneity in mean and variance of mixing distributions (Yu *et al.*, 2019; Yu, Zheng and Ma, 2020; Fanyu *et al.*, 2021; Li, Song and Fan, 2021; Yan *et al.*, 2021). Hou et al. (2021), Krueger et al. (2020), and Mannering et al. (2020) have argued and illustrated that accounting for unobserved heterogeneity improves the predictive ability of discrete response models. However, Krueger et al. (2020) also show that such gains in predictive accuracy are compensated when a linear link function is replaced by a nonparametric counterpart in spatial count data models. Considering that the predictive ability of the crash count model is crucial and nonparametric link function would make the estimation time prohibitable large, accounting for unobserved heterogeneity through random parameters in linear link function is in the interest of this study. However, with the large-scale dataset at hand, DP mixture would lead to challenges in the mixing and convergence of the Gibbs sampler (Hastie, Liverani and Richardson, 2015). Therefore, instead of making link function parameters random, a segment-time-specific random intercept is included in the model (see details in Eq. 3 of "The Modeling Framework" section). Spatial correlation is ignored because its empirical identification is challenging due to a preponderance of zero-AVC segments.

There is limited literature on modeling time-varying parameters in discrete response models, which are essential to capture the temporal variation in covariate effects due to the unobserved factors (e.g., environmental conditions). To this end, previous studies adopted Markov Switching Models (MSMs) in crash frequency (Malyshkina, Mannering and Tarko, 2009; Malyshkina and Mannering, 2010), ordered injury severity (Xiong, Tobias and Mannering, 2014), and multinomial choice estimation (Bansal, Hörcher and Graham, 2020). After the formal introduction of the term "temporal instability" by Mannering (2018), most recent analytical studies in accident research check for its presence using likelihood ratio test (Behnood and Mannering, 2019; Islam, Alnawmasi and Mannering, 2020; Islam and Mannering, 2020, 2021; Yu, Ma and Shen, 2021). However, to conduct such hypothesis testing, the model is required to be estimated as many times as the number of periods. This approach is not feasible to capture temporal instability caused due to seasonality across months of the year. MSMs also offer a very restrictive specification as they only allow parameters to take as many values as the number of latent states and going beyond two latent states increases model complexity. Therefore, in the proposed model, we capture seasonality through time-varying parameters. To the best of our knowledge, this is the first study that accounts for seasonal effects along with unknown exposure in identification of high-risk locations using large-scale data. To check for temporal instability in parameters across years, we estimate three separate models for each year (2014, 2015, and 2016) while capturing monthly variation in collision probability in each year using month-specific parameters.

In sum, the proposed model incorporates exposure (i.e., segment-specific animal crossing) using a nonparametric DP mixture, and the number of segment-level AVCs is assumed to

follow a binomial distribution. The probability of AVC occurrence is a logistic function of time-varying parameters and segment-specific characteristics. Special attention is paid to AVC seasonality and the preponderance of zero-crash segments while avoiding computational burdens that often accompany discrete response analysis for such a large data set obtained for Texas (roughly 85,000 reasonably homogeneous [in design attributes, like curvature, grade, number of lanes, speed limit, and median presence] segments, as distinguished in the Texas Department of Transportation's state-maintained network). A Pólya-Gamma augmented Gibbs sampler is derived for computationally tractable estimation of the proposed model.

ANIMAL-VEHICLE COLLISIONS IN TEXAS

The dataset used here comes from two sources. First, AVC records are from the Crash Records Information System (CRIS) maintained by the Texas Department of Transportation. Second, segment-specific roadway design factors were obtained from the Texas Department of Transportation website. Figures 1 and 2 show the 43,319 AVCs that were reported over the 2010-2016 seven-year period. Figure 1 shows a small increase in total AVCs in more recent years, perhaps as traffic has risen, with the Texas economy bouncing back from a global recession. More interestingly, the months of October through December demonstrate much higher counts. This seasonal pattern comes largely from the white-tailed deer's rutting or breeding season (Bruinderink and Hazebroek, 2003; Sullivan, 2011; Niemi *et al.*, 2017; TPWD, 2019).

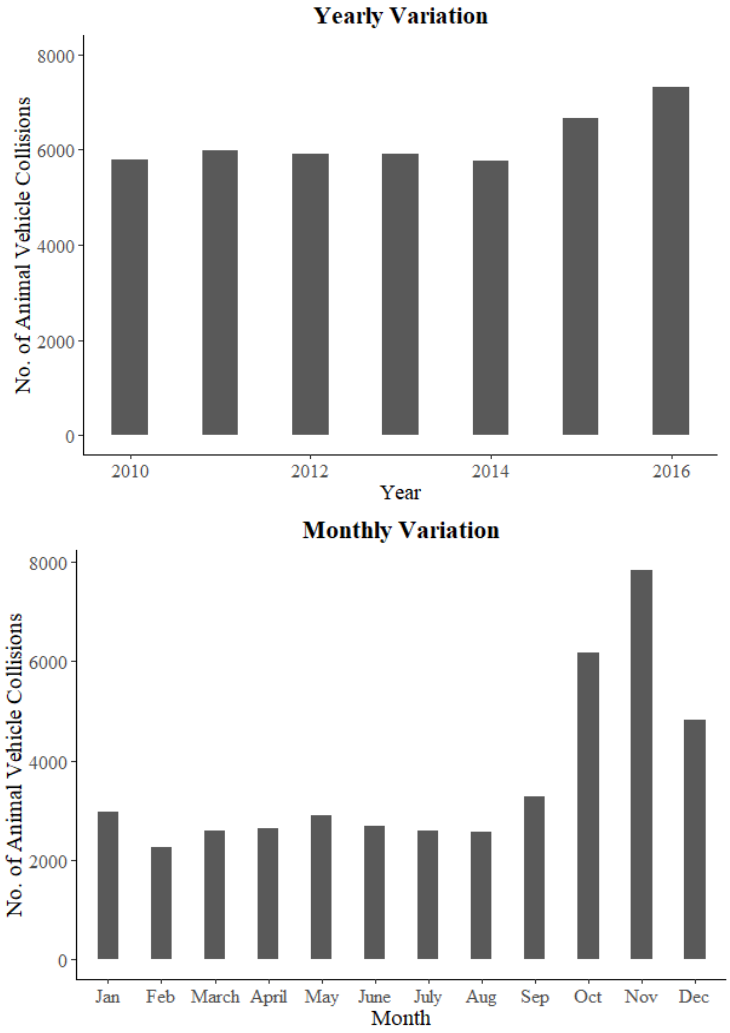


Figure 1: AVC counts by year (top) and month (bottom), as reported between 2010 and 2016 (on Texas' state-maintained highways)

Figure 2 shows the locations of AVCs associated with the 120,726 segments of state-maintained roadway network*. It is clear from the figure that a large portion of the AVCs is located on the east side of the state around several urbanized areas. After filtering for segments shorter than 0.1 mi and segments with posted speed limit below 20 miles per hour, a total of 85,953 remained for further analysis.

AVCs rarely occur on most segments after disaggregating AVCs over all distinctive Texas highway segments. Further temporal disaggregation to the level of monthly data shows that reported AVC counts are very low along all Texas segments. Just 0.4% of the monthly segment-level AVC counts are non-zero. Among the non-zero monthly segment-level AVCs, only 4.8% have more than one AVC (with a maximum monthly count of 6 AVCs). Accordingly, two important challenges can emerge for segment level and monthly AVC counts: the computation involved increases dramatically due to the number of

* A small percentage of AVC displayed didn't occur on the network system. The total number of off-system AVCs is 5930, which account for 12 percent of the total AVCs recorded.

observations (85,953 segments × 12 months = 1,031,436), counts are very sparse (typically zero) and highly variable.

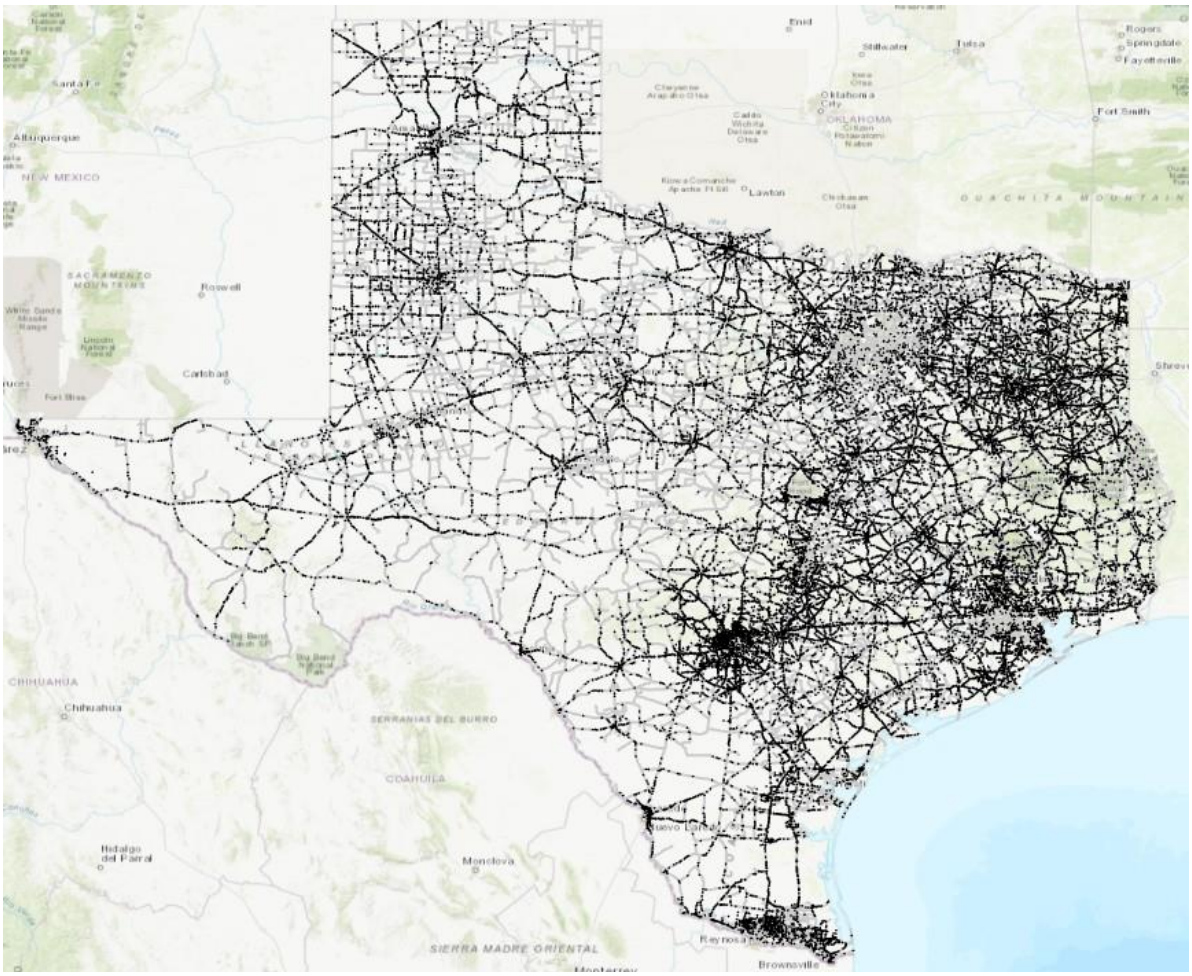


Figure 2: Texas' state-maintained roadway network with reported AVCs (from 2010 through 2016) shown as black dots

Figure 3 shows AVCs recorded on segments in a small section of Texas. In this figure, the AVCs are denoted as circles (one for each occurrence), whereas segments are shown as solid black lines with their endpoints indicated by crossbars. As evident in Figure 3, most segments have zero AVCs recorded, suggesting that spatial autocorrelation will show as near-zero at this highly disaggregated level although AVC clustering is evident at regional and state levels.

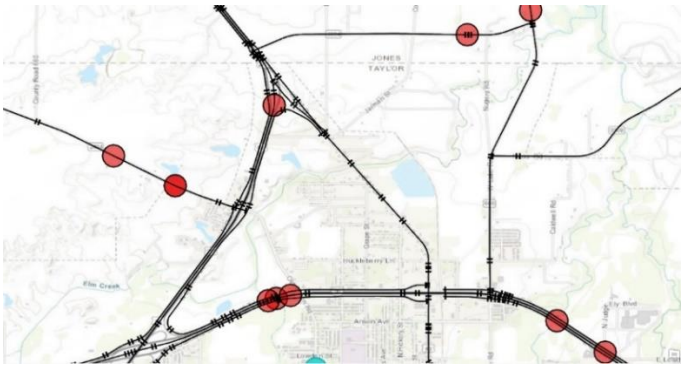


Figure 3: Reported AVCs from 2010 through 2016 in a small area in Texas

Moreover, Figure 3's zero-count segments may indicate heterogeneity across segments. It is possible that many segments located in Figure 3's bottom left (or its top left) are elevated bypasses or have lots of fencing or special underpasses to avoid animals crossing at grade. For this reason, the effect of a specific design factor may only impact AVC counts on the selected subgroup of segments in a large network. In this case, the need to identify the heterogeneity among a large number of segments further complicates the computation.

In summary, the data analysis reveals several important aspects influencing AVC modelling and inference. These include AVC seasonality, the sparse and highly variable nature of AVC count data at the monthly and segment levels, and potential for observed and unobserved heterogeneity among segments. The next section discusses how all these aspects are incorporated in the proposed count data model to find high AVC locations.

THE MODELING FRAMEWORK

To account for several important aspects of AVC modelling in a Bayesian hierarchical framework, the model specification begins with the use of a binomial distribution for the number of AVCs recorded at each segment s in month t, so that the probability of having k reported AVCs for a segment-time pair (s, t) is

$$P(k_{s,t} \mid n_{s,t}, p_{s,t}) = \binom{n_{s,t}}{k_{s,t}} p_{s,t}^{k_{s,t}} (1 - p_{s,t})^{n_{s,t} - k_{s,t}}, \quad (1)$$

where $n_{s,t}$ is the number of animal road crossings depending on animal habitats and seasonality, and $p_{s,t}$ is the probability of an AVC occurrence, both of which vary by location (s) and month (t). Using the binomial distribution with two parameters $n_{s,t}$ and $p_{s,t}$, the number of AVCs, $k_{s,t}$, can be interpreted as the result of repeated $n_{s,t}$ Bernoulli trials. Each trial represents an animal road crossing with probability of $p_{s,t}$ causing an AVC. For this reason, the collision probability, $p_{s,t}$ can be regarded as a quantity that is determined by segment-specific characteristics (like land-use and segment-design factors), and time-varying natural factors (like rainfall) and is assumed to have a logistic functional form:

$$p_{s,t} = \frac{1}{1 + \exp(-\psi_{s,t})}, \tag{2}$$

where $\psi_{s,t}$ determines the probability of causing an AVC when an animal road crossing is made at segment s and month t and can be represented as a linear function of characteristics, namely segment-specific design factors like land use, and time-varying natural factors, such as rainfall (see Equation 3). In addition, there may be heterogeneity in collision probability due to the omission of important segment-specific factors. In order to accommodate this possibility, a two-component clustering is incorporated for the intercept:

$$\psi_{s,t} = \alpha_{0,t} I_{s,t} + \boldsymbol{\beta}' \boldsymbol{x_s} + \boldsymbol{\gamma_t'} \boldsymbol{y_{s,t}}, \tag{3}$$

where $\boldsymbol{x_s}$ is a column vector containing the time-invariant design factors of segment s, $\boldsymbol{\beta}$ is the corresponding conformable parameter vector, $\boldsymbol{y_{s,t}}$ is a column vector containing time-varying parameters, and $\boldsymbol{\gamma_t}$ is the corresponding conformable month-specific parameter vector.[†] More importantly, $I_{s,t}$ is an indicator for the non-zero constant effect $\alpha_{0,t}$ at segment s and month t. In other words, the constant effect of each segment-time pair arising from this specification is either zero or $\alpha_{0,t}$. Nonzero $\alpha_{0,t}$ suggests that some network segments have their collision probabilities affected by some important but unknown factors.

While finding appropriate specification of $n_{s,t}$, it is worth noting that the total number of animal crossings for each segment differs across segments and depends on the seasonality of average animal activity levels as seen in Figure 1. To achieve this flexibility, while simultaneously reflecting very low AVC counts on most segments, a nonparametric Dirichlet process (DP) prior is used for the number of animal road crossing at all segments (s) and at all months (t) of a year: $n_{s,t}$. In summary, the proposed modelling framework for AVCs is presented below:

$$\begin{array}{cc}
P \sim \mathbf{DP}(\vartheta P_0), & q_t \sim \mathbf{Beta}(a_0, b_0), \\
(\mu_{s,t}, \sigma_{s,t}) \sim P, & I_{s,t} \sim \mathbf{Bernoulli}(q_t), \\
n^*_{s,t} \sim \mathbf{Normal}_{(-0.5,\infty)}(\mu_{s,t}, \sigma^2_{s,t}), & \alpha_{0,t}, \boldsymbol{\beta}, \boldsymbol{\gamma_t} \sim \mathbf{MVN}(\mathbf{0}, \boldsymbol{\Sigma_{0,t}}), \\
n_{s,t} = \lfloor n^*_{s,t} \rceil, & p_{s,t} = 1/(1 + exp(-\psi_{s,t})), \\
 & \psi_{s,t} = \alpha_{0,t} I_{s,t} + \boldsymbol{\beta}' \boldsymbol{x_s} + \boldsymbol{\gamma_t'} \boldsymbol{y_{s,t}},
\end{array} \tag{4}$$

$$k_{s,t} \sim \mathbf{Binomial}(n_{s,t}, p_{s,t}).$$

The number of animal road crossings, $n_{s,t}$ and the collision probability, $p_{s,t}$, on segment s in month t are obtained from the left and the right blocks of Equation 4, respectively. The number of observed AVCs on segment s in month t, $k_{s,t}$, is a realization of the binomial distribution with the parameters $n_{s,t}$ and $p_{s,t}$, as shown in the last part of Equation 4.

More specifically, in the top-left block of the equations, a discrete distribution P is drawn

[†] Time-varying parameters on time-invariant attributes can be easily incorporated in the proposed model, but are not specified here to avoid explosion of the parameter space. We could afford time-varying coefficients on time-varying attributes since there is just one attribute (rainfall) with monthly variation in this data set.

from DP with scalar precision parameter ϑ and base distribution P_0. Then the cluster locations $\mu_{s,t}$ and scales $\sigma_{s,t}$ are generated from the discrete distribution P for each segment s and month t. Conditional on the cluster locations and scales, a real-valued latent quantity $n^*_{s,t}$ is drawn. Then, the total number of animal crossings, $n_{s,t}$, at site s and month t is set equal to the nearest integer, $\lfloor n^*_{s,t} \rceil$. The truncated normal distribution on $n^*_{s,t}$ ensures the non-negativity of $n_{s,t}$.

For the collision probability, $p_{s,t}$, specification in the top-right block of Equation 4, the indicator probability, q_t, is first drawn from a Beta distribution with prior parameters a_0 and b_0 for each month t. Then the indicator variable, $I_{s,t}$, for all segments and months is generated from a Bernoulli distribution using this indicator probability, q_t. The non-zero constant effect, the effect of segment-specific design factors, and time-varying natural factors, $[\alpha_{0,t}, \boldsymbol{\beta}', \boldsymbol{\gamma_t}']$ are drawn from an uninformative multivariate normal (MVN) distribution with prior mean zero ($\mathbf{0}$) and diagonal covariance, $\boldsymbol{\Sigma}_{0,t}$. Then $\psi_{s,t}$ is determined by the dot product of attributes $[I_{s,t}, \boldsymbol{x_s'}, \boldsymbol{y_{s,t}}']$ and parameters $[\alpha_{0,t}, \boldsymbol{\beta}', \boldsymbol{\gamma_t}']$, which is further transformed to the collision probability, $p_{s,t}$, after passing through a logistic function.

The proposed hierarchical model was estimated using a Markov Chain Monte Carlo simulation. Algorithm 1 shows the step-by-step sampling from the conditional posterior distributions. Key features to note are the Pólya-Gamma data augmentation step to address the non-conjugacy of the logistic probability function (Polson et al., (2013) and the use of a stick-breaking construction to obtain the DP prior (Canale and Dunson, (2011). The complete derivation of the Gibbs sampler is provided in the Appendix.

Initialize parameters – clusters $\{1, \cdots, C\}$, latent variables, and hyper-parameters

Step 1: Draw $n_{s,t}$ using a Metropolis-Hastings (MH) step

→ Step 1a: Assign cluster ID to each segment-month (s, t) pair.

→ Step 1b: Update segment-specific parameters for each time period $(\mu_{s,t},\ \sigma_{s,t}^2)$ from multinomial distribution using cluster parameters μ_l^*, σ_l^{*2} as
$p(\mu_{s,t} = \mu_c^*\ and\ \sigma_{s,t}^2 = \sigma_c^{*2} | \cdot) = \frac{w_c p(n_{s,t}|\mu_c^*, \sigma_c^{*2})}{\sum_{l=1}^{C} w_l p(n_{s,t}|\mu_l^*, \sigma_l^{*2})}$,
where $p(n_{s,t}|\mu_l^*, \sigma_l^*) = \frac{\Phi(n_{s,t}+1/2|\mu_l^*, \sigma_l^{*2}) - \Phi(n_{s,t}-1/2|\mu_l^*, \sigma_l^{*2})}{1-\Phi(-1/2|\mu_l^*, \sigma_l^{*2})}$, and $\Phi(.)$ is a normal cumulative distribution function.

→ Step 1c: Update cluster weights w_l using a stick-breaking construction with Beta-distributed V_l where $V_l|. \sim \mathbf{Beta}(1 + n_l, \vartheta + \sum_{i=l+1}^{C} n_i)$, $w_1 = V_1, w_l = V_l \prod_{i<l}(1 - V_i)$ for $l = 2, \ldots, C$, and n_l is the number of $\mu_{s,t}$ that is equal to μ_l^*. (See Appendix for more details on V_l)

→ Step 1d: Set $n_{s,t}^* = \Phi^{-1}(u_{s,t} \mid \mu_{s,t},\ \sigma_{s,t}^2)$,
where $u_{s,t} \sim \mathbf{Uniform}\left(\Phi\left(n_{s,t} - \frac{1}{2}|\mu_{s,t}, \sigma_{s,t}^2\right), \Phi\left(n_{s,t} + \frac{1}{2}|\mu_{s,t}, \sigma_{s,t}^2\right)\right)$

→ Step 1e: Update cluster (μ_l^*, σ_l^{*2}) from the Normal-Gamma distribution as
$(\sigma_l^*)^{-2}|. \sim \mathbf{Gamma}\left(a_0 + \frac{n_l}{2}, b_0 + \frac{1}{2}\sum_{\{(s,t):\mu_{s,t}=\mu_l^*\}}\left(\left(n_{s,t}^* - \eta\right) + \frac{n_l}{1+n_l}\eta^2\right)\right)$, and
$\mu_l^*|. \sim I_{\left[-\frac{1}{2},\infty\right)}\mathbf{Normal}\left(\frac{\sum_{\{(s,t):\mu_{s,t}=\mu_l^*\}} n_{s,t}^*}{1+n_l}, \frac{\sigma_l^{*2}}{1+n_l}\right)$.

→ Step 1f: Metropolis-Hastings step with
$\mathrm{P}(n_{s,t}|.) \propto \left[\sum_{l=1}^{C} w_l \frac{\Phi\left(n_{s,t} + \frac{1}{2}|\mu_l^*, \sigma_l^{*2}\right) - \Phi\left(n_{s,t} - \frac{1}{2}|\mu_l^*, \sigma_l^{*2}\right)}{1-\Phi\left(-\frac{1}{2}|\mu_l^*, \sigma_l^{*2}\right)}\right] \times \mathbf{Binomial}(k_{s,t}|\, n_{s,t},\ p_{s,t})$.

Step 2: Draw $p_{s,t}$

→ Step 2a: Draw auxiliary variable $\omega_{s,t}|. \sim \mathbf{PólyaGamma}(n_{s,t},\ \alpha_{0,t}I_{s,t} + \boldsymbol{\beta}'\boldsymbol{x_s} + \boldsymbol{\gamma_t'}\boldsymbol{y_{s,t}})$.

→ Step 2b: Draw $\boldsymbol{\beta}|. \sim \mathbf{MVN}(\boldsymbol{m_\beta}, \boldsymbol{V_\beta})$,
where, $\boldsymbol{V_\beta} = \left(\sum_t \sum_s (\omega_{s,t}\boldsymbol{x_s}\boldsymbol{x_s'}) + \boldsymbol{B}_0^{-1}\right)^{-1}$ and
$\boldsymbol{m_\beta} = \boldsymbol{V_\beta}\left(\sum_t \sum_s \boldsymbol{x_s}(\kappa_{s,t} - \omega_{s,t}\boldsymbol{\gamma_t'}\boldsymbol{y_{s,t}} - \omega_{s,t}\alpha_{0,t}I_{s,t})\right)$.

→ Step 2c: Draw $\alpha_{0,t}$, $\boldsymbol{\gamma_t}|. \sim \mathbf{MVN}(\boldsymbol{m_t}.\boldsymbol{V_t})$,
where, $\boldsymbol{V}_t = \left(\sum_s (\omega_{s,t}\boldsymbol{z_{s,t}}\boldsymbol{z_{s,t}'}) + \boldsymbol{D}_0^{-1}\right)^{-1}$ and $\boldsymbol{m}_t = \boldsymbol{V}_t\left(\sum_s \boldsymbol{z_{st}}(\kappa_{s,t} - \omega_{s,t}\boldsymbol{\beta}'\boldsymbol{x_s})\right)$.

→ Step 2d: Draw $I_{s,t}$ with
$P(I_{s,t} = 1|\, \cdot) = \frac{P_{s,t}^1 q_t}{P_{s,t}^0(1-q_t) + P_{s,t}^1 q_t}$, where
$P_s^0 = \exp\left(\kappa_{s,t}\psi_{s,t} - \frac{1}{2}\omega_{s,t}\psi_{s,t}^2\right)\ |\ \psi_{s,t} = \boldsymbol{\beta}'\boldsymbol{x}_s + \boldsymbol{\gamma'}_t\boldsymbol{y_{s,t}}$ and
$P_s^1 = \exp\left(\kappa_{s,t}\psi_{s,t} - \frac{1}{2}\omega_{s,t}\psi_{s,t}^2\right)\ |\ \psi_{s,t} = \alpha_{0,t}I_{s,t} + \boldsymbol{\beta}'\boldsymbol{x}_s + \boldsymbol{\gamma_t'}\boldsymbol{y_{s,t}}$.

→ Step 2e: Draw $q_t|. \sim \mathbf{Beta}\left(1 + \sum_s I_{s,t},\ 1 + \sum_s (1 - I_{s,t})\right)$.

Algorithm 1: Blocked Gibbs Sampler for Detecting Animal-Vehicle Crossings and Collisions
(Note: All draws are for each segment, s, and month, t, as applicable.)

RESULTS

The model described above was estimated using the Texas AVC dataset of the year 2016. Table 1 provides a summary of the segment-level design and environmental features that were used as explanatory variables.

Table 1: Summary statistics for segment-specific factors considered in the model

Variable Name	**Min.**	**Mean**	**Max.**
Segment Length (miles)	0.1	0.93	30.17
Average Daily Traffic, ADT (1000 vehicles per day)	.01	10.3	341.3
Median Width (ft)	0	7.05	710
Inside Shoulder Width (ft)	0	5.10	60
Outside Shoulder Width (ft)	0	6.35	53
Surface Width (ft)	10	32.4	236
% K Factor (for traffic peaking)	4.2	10.8	19.9
Controlled Access?	0	0.09	1
Posted Speed Limit (miles per hour)	20	57.4	85
Urban Area?	0	0.26	1
Barrier Median Present?	0	0.05	1
Terrain Composition			
Water (%)	0	4.5	98.4
Trees (%)	0	10.8	96.3
Open Land (%)	0	52.8	100
Population Density (100 persons/ square mile)	0	0.91	58.4

Algorithm 1 was used to draw 20,000 MCMC samples from the conditional posterior distributions of the model parameters, of which first 10,000 burn-in samples were discarded. The estimation took about 22 hours on a high-performance computer with six Intel Xeon cores operating at 3.4 GHz and having 128 GB of RAM. MCMC chains converged with an average Gelman & Rubin R-hat diagnostic of 1.011. We estimate the model for years 2014, 2015, and 2016 to test the temporal stability of parameters across years. We also estimate a fully specified zero-inflated negative binomial (ZINB) model. This means that both the binary logit specification and the count model specification of ZNIB utilize all explanatory variables that we use in identifying the collision probability of the proposed model. Its root mean square error is compared with the proposed model in Table 2. The results indicate that the proposed model with a non-parametric method to model the exposure performs better than the traditional zero-inflated negative binomial model.

Table 2: Root mean square error (RMSE) comparison

	2014	**2015**	**2016**
Zero-inflated negative binomial	0.077	0.083	0.088
Proposed model	0.027	0.029	0.030

Figure 4 shows the posterior expected AVCs compared against observed AVCs by month

for the year 2016. The relative magnitude of the aggregated quantities resembles the seasonal pattern in the observed data. Such resemblance illustrates good predictive performance of the proposed model.[3]

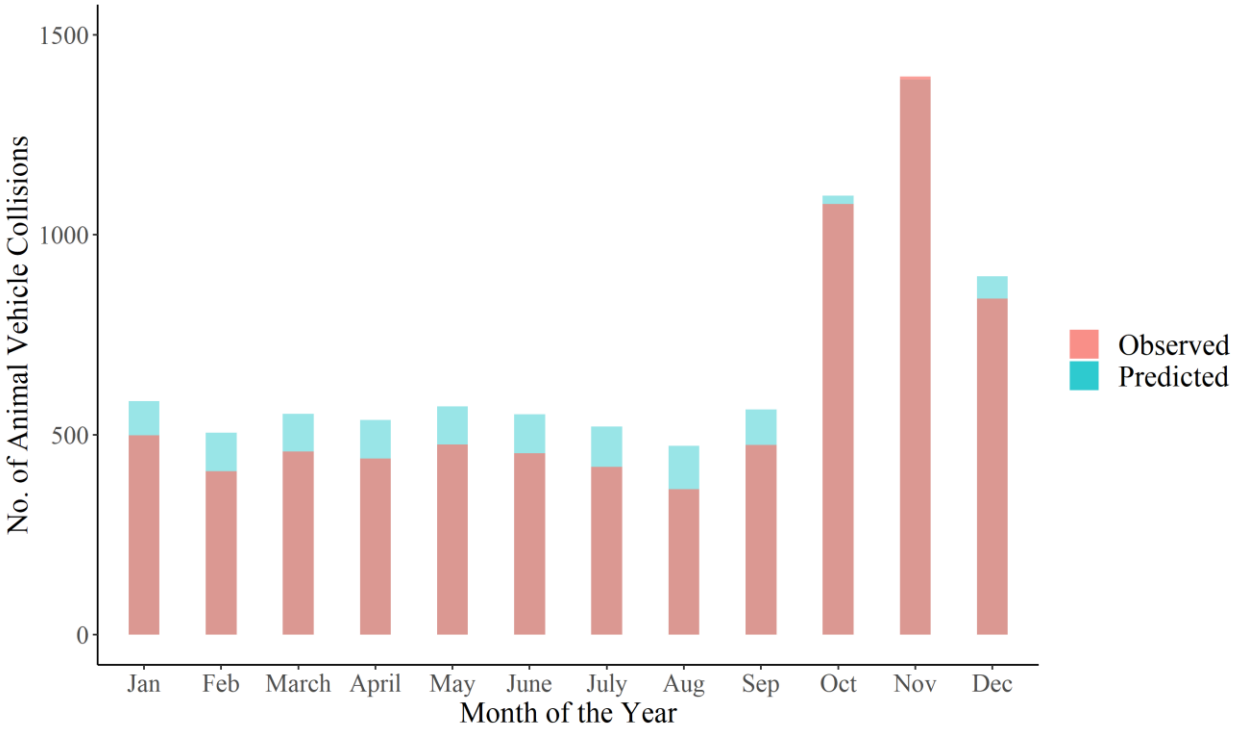


Figure 4: Posterior Expected Versus Observed Number of Animal-Vehicle Collisions by Month.

Table 3 shows the posterior mean estimates and 95% credible intervals for homogeneous probability parameter ($\boldsymbol{\beta}$) for years 2014, 2015, and 2016. We observe that 95% credible intervals of all parameters across three years have substantial overlap. Thus, we cannot reject null hypotheses at 5% significance level that the parameter estimates across years are the same. To substantiate our claims, we have used the posterior standard deviation (analogous to standard errors), and posterior mean of parameter estimates from each of the three years and conducted pairwise t-tests. The results of t-tests are aligned with our visual inspection of credible intervals as p-values for all comparisons are way above 0.05 (in fact, most p-values are above 0.50).

To avoid redundancy, we only discuss results for the year 2016, but detailed results for 2014 and 2015 are also available upon request. The estimated effects for several design factors are also insightful. Namely, speed limit is positively associated with higher probability of causing an AVC by an animal road crossing. Segments in urban areas tend to have lower probabilities of observing an AVC by animal-road crossings. A median barrier tends to decrease the probability of an AVC, perhaps because animals are unable to see the other side of the segment and may not cross at such locations. Large median widths correspond to a higher likelihood of an AVC, which can correspond to a correlated

[3] Figure 4's absolute magnitudes are generally smaller than those shown in Figure 1 since Figure 44 estimates the sums over all segments for a given year, whereas Figure 1 aggregates AVCs across segments and over seven years.

increase in roadbed width. Outside shoulder widths also show a similar trend, but the effect of shoulder width on AVC probability is not statistically significant. Busy segments or those with a continuous peak traffic flow have a lower likelihood of AVCs, and an increase in average daily traffic corresponds to an increase in AVCs when controlling for several opposing trends (like peak factor and urban areas, for example). Controlled access highways like freeways have a significantly lower likelihood of an AVC, as expected. Land use characteristics also have a significant effect on AVC probability. Segments located near open areas, adjacent to water bodies, or surrounded by trees are more likely to observe an AVC compared to segments near buildings. The association of population density with the likelihood of an AVC is not statistically significant.

Table 3: Posterior summary of the homogeneous probability parameters ($\boldsymbol{\beta}$)

Variables	2014		2015		2016		2014 vs 2015	2014 vs 2016	2015 vs 2016
	Posterior mean	95% credible intervals	Posterior mean	95% credible intervals	Posterior mean	95% credible intervals	p-value from paired t-tests of posterior means		
Segment Length (mi)	0.479**	(0.426, 0.533)	0.468**	(0.418, 0.517)	0.464**	(0.417, 0.512)	0.75	0.67	0.91
Average Daily Traffic (1000 vehicles per day)	0.003	(-0.011, 0.017)	0.003	(-0.01, 0.015)	0.006	(-0.006, 0.019)	0.97	0.76	0.72
Median Width (ft)	0.009**	(0.005, 0.014)	0.009**	(0.005, 0.013)	0.007**	(0.004, 0.012)	0.99	0.66	0.65
Inside Shoulder Width (ft)	0.004	(-0.016, 0.024)	0.00007	(-0.018, 0.019)	-0.010	(-0.028, 0.007)	0.79	0.32	0.44
Outside Shoulder Width (ft)	0.062**	(0.042, 0.082)	0.066**	(0.047, 0.084)	0.074**	(0.056, 0.092)	0.79	0.37	0.51
Surface Width (ft)	0.022**	(0.017, 0.028)	0.020**	(0.015, 0.025)	0.018**	(0.013, 0.022)	0.51	0.25	0.6
Peak Period (%)	-0.080**	(-0.107, -0.056)	-0.070**	(-0.092, -0.048)	-0.075**	(-0.096, -0.054)	0.56	0.77	0.74
Controlled Access?	-1.411**	(-1.769, -1.089)	-1.356**	(-1.643, -1.066)	-1.317**	(-1.604, -1.031)	0.81	0.68	0.85
Posted Speed Limit (mph)	0.026**	(0.02, 0.034)	0.023**	(0.017, 0.029)	0.025**	(0.019, 0.031)	0.45	0.77	0.6
Urban Area?	-0.487**	(-0.652, -0.329)	-0.493**	(-0.64, -0.345)	-0.551**	(-0.692, -0.411)	0.96	0.56	0.58
Barrier Median Present?	-0.592**	(-0.892, -0.276)	-0.403**	(-0.679, -0.12)	-0.391**	(-0.652, -0.123)	0.38	0.34	0.95
Terrain Composition									
% Water	0.008**	(0.002, 0.014)	0.011**	(0.006, 0.017)	0.012**	(0.006, 0.017)	0.73	0.46	0.68
% Trees	0.020**	(0.015, 0.024)	0.018**	(0.014, 0.023)	0.017**	(0.013, 0.021)	0.87	0.67	0.78
% Open Land	0.008**	(0.005, 0.012)	0.009**	(0.005, 0.012)	0.009**	(0.006, 0.012)	0.86	0.73	0.86
Population Density (100 persons/ square mile)	0.021	(-0.021, 0.065)	0.026	(-0.014, 0.067)	0.031	(-0.006, 0.069)	0.75	0.67	0.91

**95% posterior credible interval does not contain zero, i.e., the covariate's effect is statistically different than zero at a 0.05 significance level.

Table 4 shows the estimates and statistics associated with parameters ($\alpha_{0,t}, I_{s,t}$ and $\boldsymbol{\gamma_t}$), which form the time-varying component of the collision probability link function. Three main insights are drawn from these estimates. First, around 53% of the segments have non-zero $I_{s,t}$, i.e., they carry an inherent non-zero constant effect (unexplained by the observed covariates) on the AVC probability. The posterior means of this non-zero constant effect ($\alpha_{0,t}$) is negative for most months, except from October to December. Second, Table 4 also shows that around 1.04% to 2.11% of segments have non-zero animal crossings ($n_{s,t}$) in each month, which is consistent with small proportion of segments with non-zero AVCs in the data. This result implies the DP process could properly cluster segments with no exposure. Third, rainfall effect on collision probability ($\boldsymbol{\gamma_t}$) is generally positive and is statistically significant in January-April and September-November at a significance level of 0.1 or lower.

Table 4: Time-Varying Quantities and Parameter Estimates for year 2016

Month	**Percent of Non-Zero** $n_{s,t}$	**Percent of Non-Zero** $I_{s,t}$	**Posterior mean of constant effect** ($\alpha_{0,t}$)	**Posterior mean of Rainfall (inches) effect** ($\boldsymbol{\gamma_t}$)
January	1.19%	53%	-0.28**	0.12*
February	1.08%	53%	-0.44**	0.12*
March	1.14%	52%	-0.36**	0.14*
April	1.13%	53%	-0.37**	0.24**
May	1.17%	53%	-0.37**	0.12
June	1.13%	52%	-0.39**	0.03
July	1.10%	53%	-0.42**	-0.02
August	1.04%	52%	-0.53**	0.05
September	1.16%	53%	-0.32**	0.27**
October	1.79%	54%	0.31**	0.32**
November	2.11%	54%	0.46**	0.14**
December	1.54%	53%	0.1	0.03

*90% posterior credible interval does not contain zero, i.e., the covariate's effect is statistically different than zero at a 0.1 significance level. **95% posterior credible interval does not contain zero, i.e., the covariate's effect is statistically different than zero at a 0.05 significance level.

Using the posterior draws from conditional distribution of probability parameters, the probability of observing an AVC from an animal road crossing on each segment can be evaluated. Figure 5 shows the probability distribution for January and October to observe the difference in crash probabilities by month. The variation in collision probability for both months across segments can be attributed to the differences in road design factors. The likelihood of observing a crash is markedly higher in October than in January, further confirming the proposed model could capture seasonality.

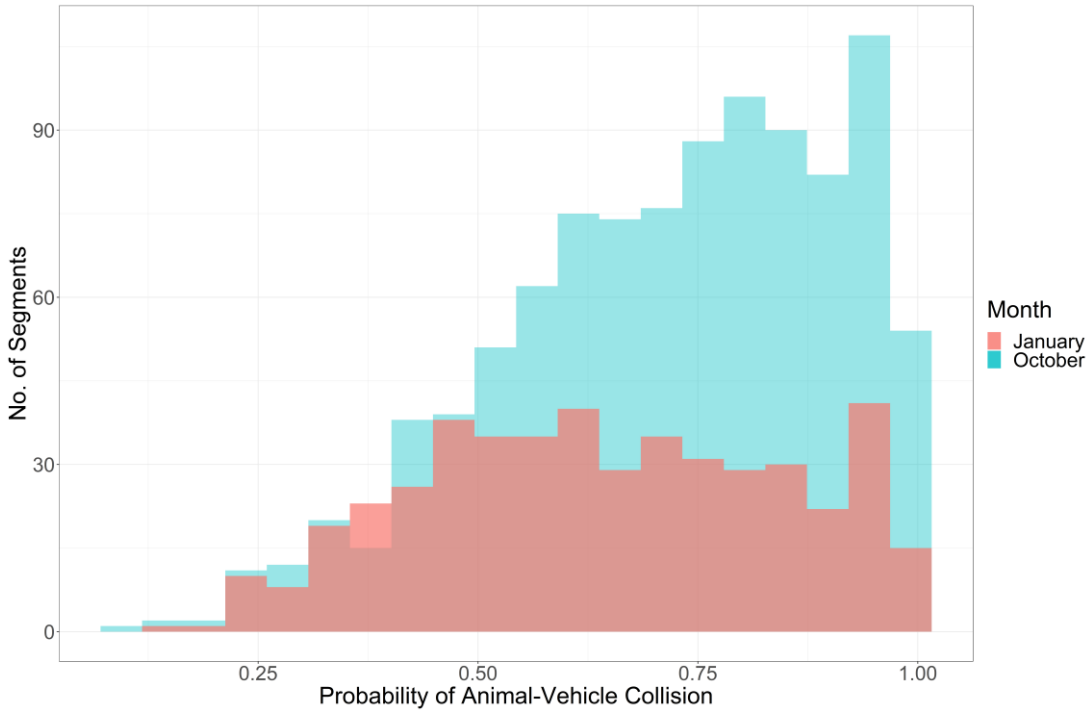


Figure 5: Empirical Posterior Distribution of Collision Probability, $p_{s,t}$ for January and October

The spatial distribution of posterior collision probabilities ($p_{s,t}$) in November is shown in Figure 6 across the Texas highway network. Figure 6 shows several clusters of light-colored segments, which correspond to the network around major urban areas. In contrast, darker segments are typically major highways that span the entire state. This pattern is a manifestation of the positive effect of speed limit, among others. This segment-level spatial distribution is useful for appreciating AVC contributions of road design decisions.

The number of animal road crossings on any segment is also key in determining expected AVCs. Segments with high crash probability but having zero or very few animal road crossings are relatively less of a concern as compared to segments having a high crash probability and many animal crossings. The segments of the latter type can be regarded as high-risk locations, meriting AVC prevention considerations. Identifying such high-risk locations can guide investments and intervention decisions. For the same purpose, the posterior means of the expected AVCs ($n_{s,t} \times p_{s,t}$) are shown in Figure 7 for two different months of a year. By using the number of crossing ($n_{s,t}$) in the calculation for the expected AVCs, Figure 7 shows the prominent effect of seasonality. More specifically, the expected AVC values are higher (darker colored) for more segments in October than that in January. This result shows that the model could capture seasonality as the model-identified high-risk locations vary across months or seasons. Another apparent feature evident in Figure 7 is that segments with higher expected AVCs are only a small share of all segments and scattered across the Texas network. This small share is because over 98% of $n_{s,t}$ values are zero in any month (as noted in Table 4).

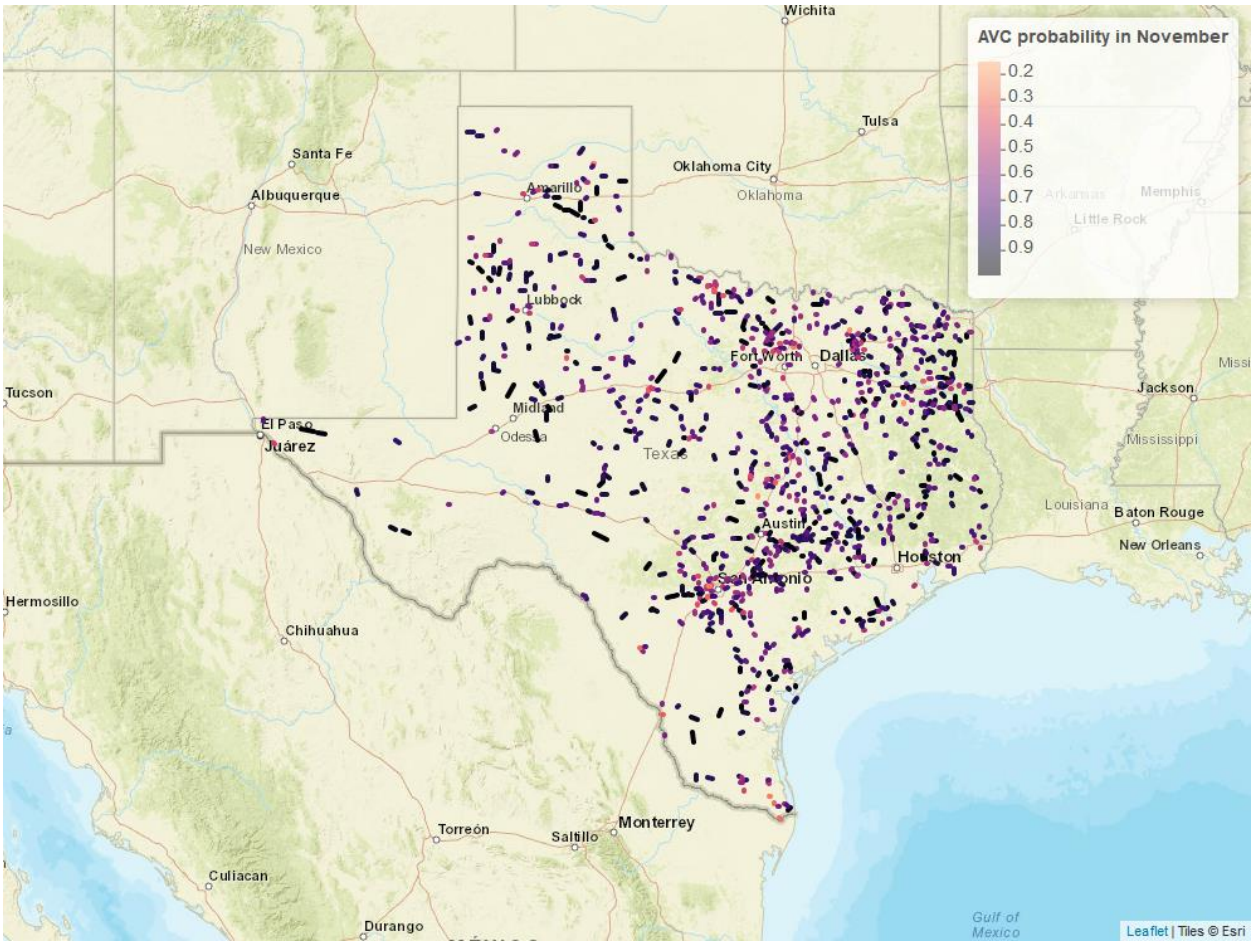
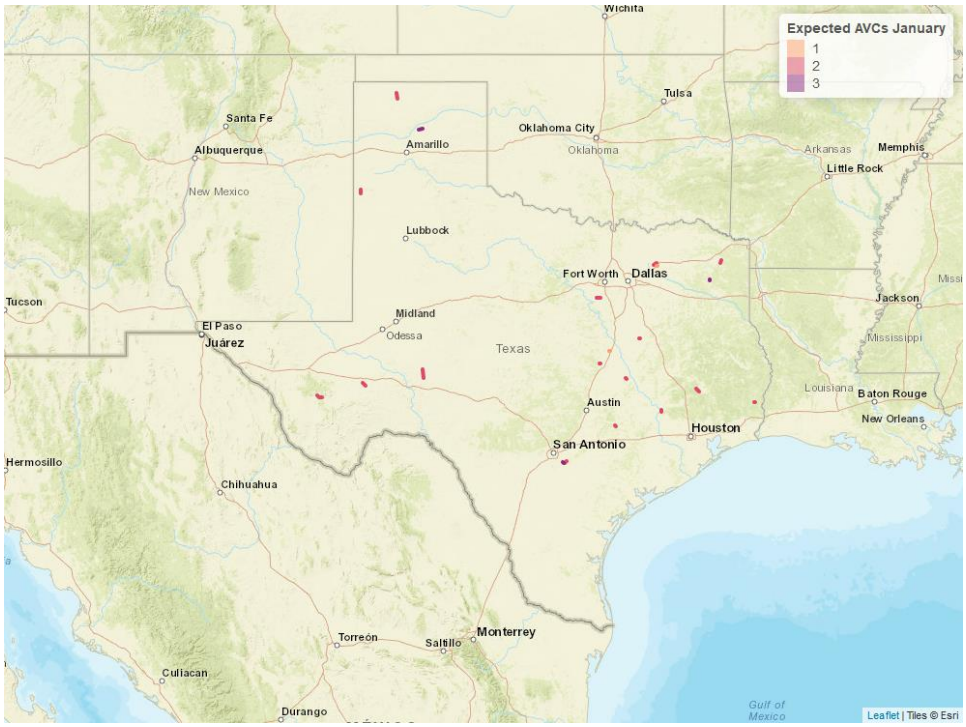
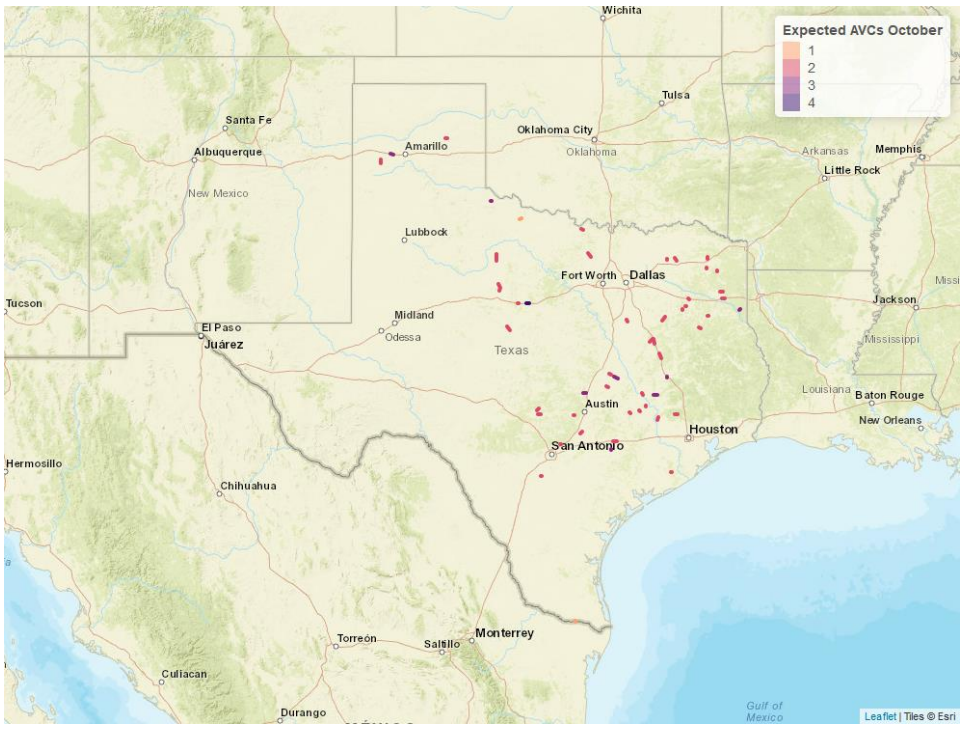


Figure 6: Posterior Mean of Collision Probability $p_{s,t}$ for All Segments in the Texas Roadway Network for November

Since segments with a high number of expected AVCs are scattered across the network, a smooth change in expected AVCs from one segment to other nearby segments may be preferred, through spatial autocorrelation. However, in doing so the advantage of having segments with zero animal crossings play no role in the parameter estimation is lost, causing parameter estimates to be biased low from averaging effects over nearby segments. Moreover, this contradicts the goal of inferring segment-specific design factor effects and shifts the focus onto higher levels of spatial aggregation.



(a) January



(b) October

Figure 7: Posterior Expected AVCs for All Segments in the State-Controlled Texas Roadway Network

Given the segment-level focus of this analysis, another interesting aspect that is easy to identify from this model is the effect of a change in any of the design factors on collision probability. To compute the marginal effect of a variable, the posterior means of collision probabilities are recalculated for each segment by changing that variable with a specific amount for all segments. Among various design factors, speed limit reductions are relatively easier to implement than infrastructure-related measures to lower AVCs. Figure 8 shows the histograms of changes in collision probability (i.e., marginal effects) resulting from the decrease in speed limit. There is a stark difference in the effect of speed limits across months, and the effect almost linearly increases with the increase in the magnitude of the speed limit change. Thus, marginal effect plots are valuable in highlighting the countermeasures' seasonal effectiveness and benefits of implementing them at higher intensity. To further illustrate this point, we also calculate the marginal effect of ADT in Figure 9, showing seasonality effect and diminishing reduction in collision probability with the increase in ADT magnitude. Similar marginal effect plots can be generated for other variables.

Figure 10 shows the spatial nature of the decrease in collision probability due to a decrease in posted speed limit by 10 miles per hour, where marginally larger reductions are observed in urban areas. Conditional on the availability of data on the segment-level fencing, warning signs, overpasses/underpasses, light-reflecting devices, and overhead lighting, spatiotemporal effects of these countermeasures can be computed, which could further assist the State of Texas in saving human and animal lives while avoiding injuries and property loss.

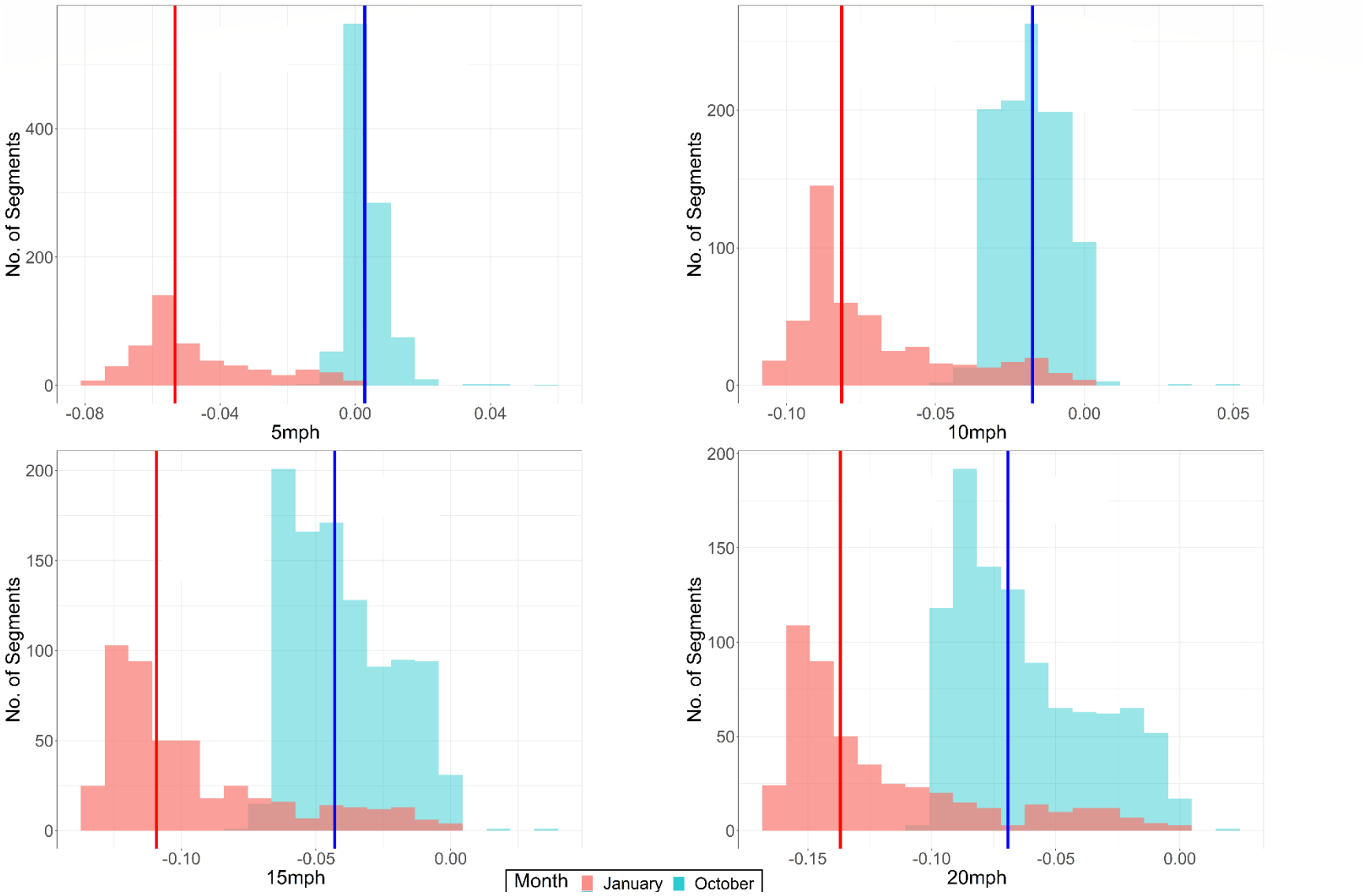


Figure 8: Marginal effect histograms of the speed limit for January and October (the decrease in speed limit is indicated at the bottom of the subplot and median of the distribution is indicated by the solid vertical lines).

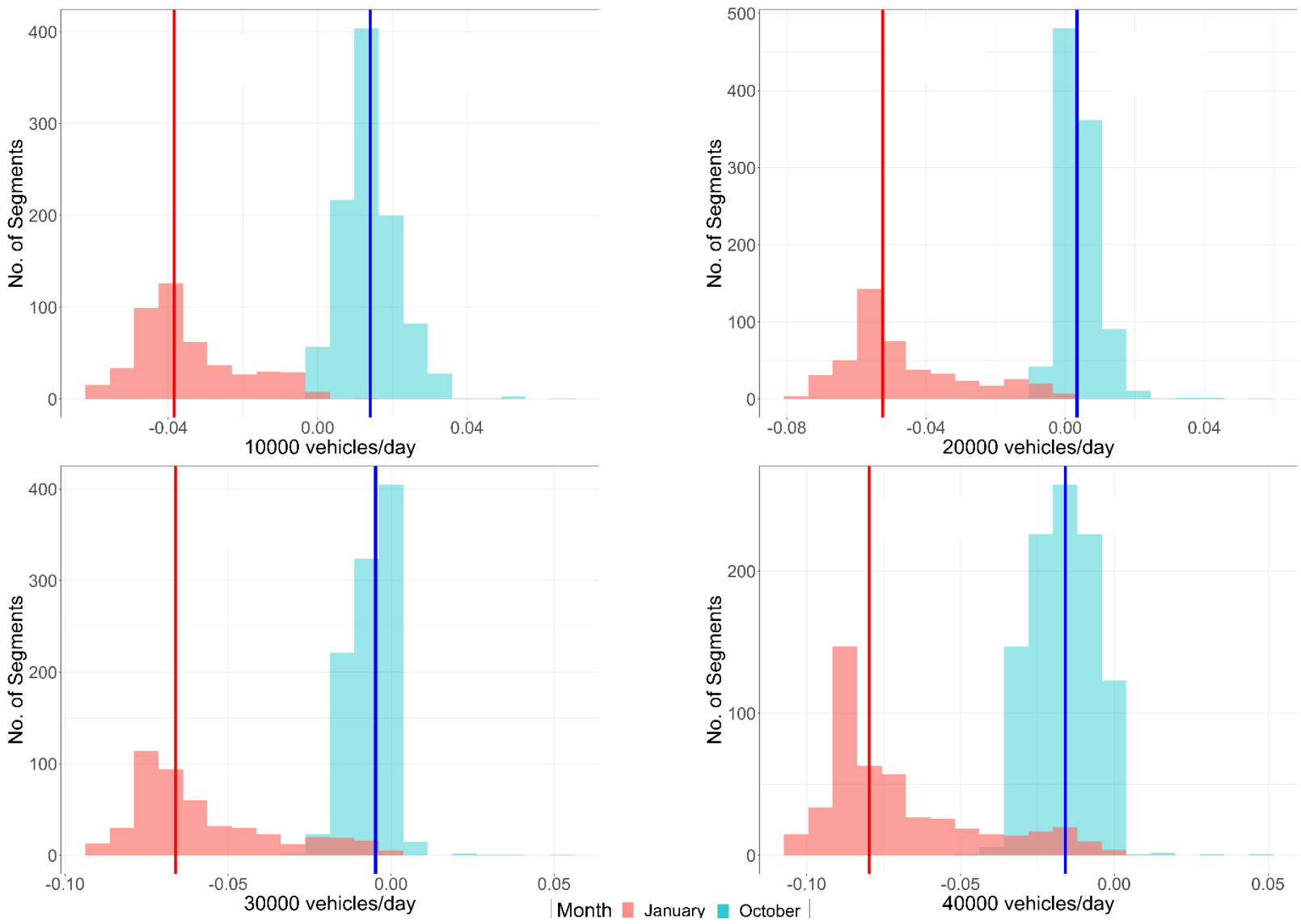


Figure 9: Marginal effect histograms of the average daily traffic (ADT) for January and October months (the increase in average daily traffic is indicated at the bottom of the subplots and median of the distribution is indicated by the solid vertical lines).

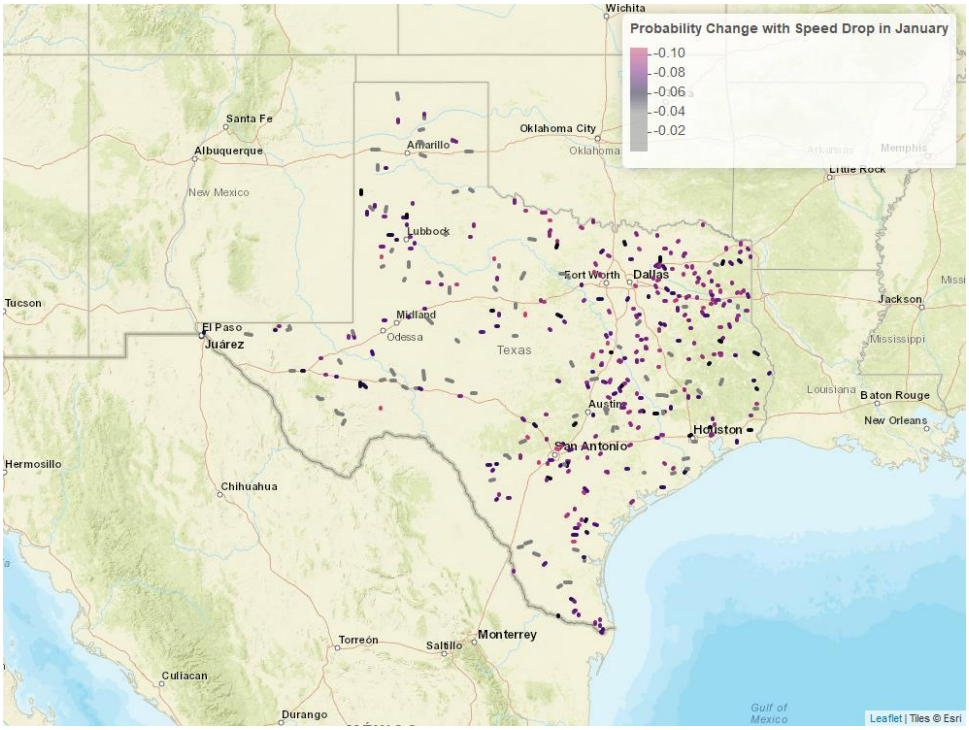


Figure 10: Posterior Change in Collision Probability following a 10 mph Speed Limit Reduction on Texas Highways

CONCLUSIONS

This study develops and demonstrates a new method for analysis of low count values across many points in time and space, using a large network of roadway segments. The model is validated using Texas' animal-vehicle collision (AVC) dataset. The proposed model uses binomial distribution to specify the AVC counts and allows the number of animal crossings to be governed by a Dirichlet process (DP). The collision probability is represented using a logistic function that depends upon segment-specific factors, monthly rainfall, and segment-month random effects. DP enables the modelling of segments with zero AVCs because it creates clusters of segments nonparametrically that can share information. Time-varying probability specification helps in capturing seasonal effects. To address the non-conjugacy of posterior updates of the parameters associated with the logistic probability function, Pólya-Gamma data augmentation is adopted.

Several advantages of the proposed modeling framework become clear in the case study. First, this new specification enables the identification of high-risk locations over time points (for example, months or seasons), not just space. Second, the impacts of various segment-specific attributes are inferred directly across all locations. The proposed modeling framework thus allows policymakers to dive deep into factors that impact AVCs. The impact of purposeful modifications in any segment-specific factor (like speed limit or lane width) on AVC counts can be estimated in a relatively straightforward fashion.

In summary, AVCs are challenging to predict due to the interactions of complex vegetative, climatic, traffic, and human factors. The inclusion of more variables like driver sight distances, availability of underground tunnels for animal crossings, and clear zone dimensions alongside highways may be helpful. Extending the model to include the time-of-day variability can also help improve estimates since a large proportion of AVCs occurs at night[§]. Moreover, accounting for unobserved heterogeneity in the effect of covariates on crash probability is also likely to improve prediction accuracy. However, increased model complexity will require advanced techniques to speed up model estimation and convergence, such as the use of Variational Bayes (Bansal, Krueger and Graham, 2021). Finally, the loglikelihood ratio test (Mannering, 2018; Pang *et al.*, 2022) to check the temporal stability in parameters does not apply to the Bayesian non-parametric models as the number of parameters is growing with observations, and therefore, developing a new statistical test to check temporal stability in such models is also an important avenue for research.

[§] FHWA study (https://www.fhwa.dot.gov/publications/research/safety/humanfac/94156.cfm) suggests a large proportion of AVCs occur early in the day between 4 and 6 am and at night between 6 and 11 pm.

REFERENCES

Ahmed, S.S., Cohen, J. and Anastasopoulos, P.Ch. (2021) 'A correlated random parameters with heterogeneity in means approach of deer-vehicle collisions and resulting injury-severities', *Analytic Methods in Accident Research*, 30, p. 100160.

Al-Bdairi, N.S.S., Behnood, A. and Hernandez, S. (2020) 'Temporal stability of driver injury severities in animal-vehicle collisions: A random parameters with heterogeneity in means (and variances) approach', *Analytic Methods in Accident Research*, 26, p. 100120.

Al-Ghamdi, A.S. and AlGadhi, S.A. (2004) 'Warning signs as countermeasures to camel-vehicle collisions in Saudi Arabia', *Accident Analysis and Prevention*, 36(5), pp. 749–760.

Anastasopoulos, P.Ch. (2016) 'Random parameters multivariate tobit and zero-inflated count data models: Addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis', *Analytic Methods in Accident Research*, 11, pp. 17–32.

Bansal, P., Hörcher, D. and Graham, D.J. (2020) 'A Dynamic Choice Model with Heterogeneous Decision Rules: Application in Estimating the User Cost of Rail Crowding', *arXiv:2007.03682 [econ, stat]* [Preprint]. Available at: http://arxiv.org/abs/2007.03682 (Accessed: 13 June 2021).

Bansal, P., Krueger, R. and Graham, D.J. (2021) 'Fast Bayesian estimation of spatial count data models', *Computational Statistics and Data Analysis*, 157, p. 107152.

Behnood, A. and Mannering, F. (2019) 'Time-of-day variations and temporal instability of factors affecting injury severities in large-truck crashes', *Analytic Methods in Accident Research*, 23, p. 100102.

Bíl, M. *et al.* (2016) 'The KDE+ software: a tool for effective identification and ranking of animal-vehicle collision hotspots along networks', *Landscape Ecology*, 31(2), pp. 231–237.

Brieger, F. *et al.* (2016) 'Effectiveness of light-reflecting devices: A systematic reanalysis of animal-vehicle collision data', *Accident Analysis and Prevention*, 97, pp. 242–260.

Bruinderink, G.W.T.A.G. and Hazebroek, E. (2003) 'Ungulate Traffic Collisions in Europe', *Conservation Biology*, 10(4), pp. 1059–1067.

Buddhavarapu, P., Scott, J.G. and Prozzi, J.A. (2016) 'Modeling unobserved heterogeneity using finite mixture random parameters for spatially correlated discrete count data', *Transportation Research Part B: Methodological*, 91, pp. 492–510.

Canale, A. and Dunson, D.B. (2011) 'Bayesian Kernel Mixtures for Counts', *Journal of the American Statistical Association*, 106(496), pp. 1528–1539.

Crépet, A. and Tressou, J. (2011) 'Bayesian nonparametric model with clustering individual co-exposure to pesticides found in the French diet', *Bayesian Analysis*, 6(1), pp. 127–144.

Danks, Z.D. and Porter, W.F. (2010) 'Temporal, Spatial, and Landscape Habitat Characteristics of Moose–Vehicle Collisions in Western Maine', *Journal of Wildlife Management*, 74(6), pp.

1229–1241.

Dettki, H. *et al.* (2011) ‘Difference in spatiotemporal patterns of wildlife road-crossings and wildlife-vehicle collisions’, *Biological Conservation*, 145(1), pp. 70–78.

Diaz-Varela, E.R. *et al.* (2011) ‘Assessing methods of mitigating wildlife-vehicle collisions by accident characterization and spatial analysis’, *Transportation Research Part D: Transport and Environment*, 16(4), pp. 281–287.

Fanyu, M. *et al.* (2021) ‘Temporal instability of truck volume composition on non-truck-involved crash severity using uncorrelated and correlated grouped random parameters binary logit models with space-time variations’, *Analytic Methods in Accident Research*, 31, p. 100168.

Found, R. and Boyce, M.S. (2011) ‘Predicting deer–vehicle collisions in an urban area’, *Journal of Environmental Management*, 92(10), pp. 2486–2493.

Fountas, G. and Anastasopoulos, P.Ch. (2018) ‘Analysis of accident injury-severity outcomes: The zero-inflated hierarchical ordered probit model with correlated disturbances’, *Analytic Methods in Accident Research*, 20, pp. 30–45.

Garrett, L.C. and Conway, G.A. (1999) ‘Characteristics of Moose-vehicle Collisions in Anchorage, Alaska, 1991-1995’, *Journal of Safety Research*, 30(4), pp. 219–223.

Gkritza, K., Baird, M. and Hans, Z.N. (2010) ‘Deer-vehicle collisions, deer density, and land use in Iowa’s urban deer herd management zones’, *Accident Analysis and Prevention*, 42(6), pp. 1916–1925.

Grilo, C., Bissonette, J.A. and Santos-Reis, M. (2009) ‘Spatial-temporal patterns in Mediterranean carnivore road casualties: Consequences for mitigation’, *Biological Conservation*, 142(2), pp. 301–313.

Gunson, K.E., Mountrakis, G. and Quackenbush, L.J. (2011) ‘Spatial wildlife-vehicle collision models: A review of current work and its application to transportation mitigation projects’, *Journal of Environmental Management*, 92(4), pp. 1074–1082.

Haikonen, H. and Summala, H. (2001) ‘Deer-vehicle crashes’, *American Journal of Preventive Medicine*, 21(3), pp. 209–213.

Hastie, D.I., Liverani, S. and Richardson, S. (2015) ‘Sampling from Dirichlet process mixture models with unknown concentration parameter: mixing issues in large data implementations’, *Statistics and Computing*, 25(5), pp. 1023–1037.

Heydari, S. *et al.* (2016) ‘Multilevel Dirichlet process mixture analysis of railway grade crossing crash data’, *Analytic Methods in Accident Research*, 9, pp. 27–43.

Heydari, S. *et al.* (2017) ‘Using a flexible multivariate latent class approach to model correlated outcomes: A joint analysis of pedestrian and cyclist injuries’, *Analytic Methods in Accident Research*, 13, pp. 16–27.

Hothorn, T. *et al.* (2015) 'Temporal patterns of deer–vehicle collisions consistent with deer activity pattern and density increase but not general accident risk', *Accident Analysis and Prevention*, 81, pp. 143–152.

Hou, Q. *et al.* (2021) 'Comparative analysis of alternative random parameters count data models in highway safety', *Analytic Methods in Accident Research*, 30, p. 100158.

Huang, H. *et al.* (2019) 'Modeling unobserved heterogeneity for zonal crash frequencies: A Bayesian multivariate random-parameters model with mixture components for spatially correlated data', *Analytic Methods in Accident Research*, 24, p. 100105.

Hurley, M. V., Rapaport, E.K. and Johnson, C.J. (2009) 'Utility of Expert-Based Knowledge for Predicting Wildlife–Vehicle Collisions', *Journal of Wildlife Management*, 73(2), pp. 278–286.

Ishwaran, H. and James, L.F. (2001a) 'Gibbs Sampling Methods for Stick-Breaking Priors', *Journal of the American Statistical Association*, 96(453), pp. 161–173.

Ishwaran, H. and James, L.F. (2001b) 'Gibbs Sampling Methods for Stick-Breaking Priors', *Journal of the American Statistical Association*, 96(453), pp. 161–173.

Islam, M., Alnawmasi, N. and Mannering, F. (2020) 'Unobserved heterogeneity and temporal instability in the analysis of work-zone crash-injury severities', *Analytic Methods in Accident Research*, 28, p. 100130.

Islam, M. and Mannering, F. (2020) 'A temporal analysis of driver-injury severities in crashes involving aggressive and non-aggressive driving', *Analytic Methods in Accident Research*, 27, p. 100128.

Islam, M. and Mannering, F. (2021) 'The role of gender and temporal instability in driver-injury severities in crashes caused by speeds too fast for conditions', *Accident Analysis and Prevention*, 153, p. 106039.

Jaeger, J.A.G. *et al.* (2016) 'Reducing Moose Vehicle Collisions through Salt Pool Removal and Displacement: an Agent-Based Modeling Approach', *Ecology and Society*, 14(2), p. art17.

Jensen, R.R., Gonser, R.A. and Joyner, C. (2014) 'Landscape factors that contribute to animal-vehicle collisions in two northern Utah canyons', *Applied Geography*, 50, pp. 74–79.

Klöcker, U., Croft, D.B. and Ramp, D. (2006) 'Frequency and causes of kangaroo - vehicle collisions on an Australian outback highway', *Wildlife Research*, 33(1), p. 5.

Kolowski, J.M. and Nielsen, C.K. (2008) 'Using Penrose distance to identify potential risk of wildlife–vehicle collisions', *Biological Conservation*, 141(4), pp. 1119–1128.

Krueger, R. *et al.* (2020) 'Variational Bayesian Inference for Mixed Logit Models with Unobserved Inter- and Intra-Individual Heterogeneity', *arXiv:1905.00419 [econ, stat]* [Preprint]. Available at: http://arxiv.org/abs/1905.00419 (Accessed: 13 June 2021).

Krueger, R., Rashidi, T.H. and Vij, A. (2020) 'A Dirichlet process mixture model of discrete choice: Comparisons and a case study on preferences for shared automated vehicles', *Journal of Choice Modelling*, p. 100229.

Lao, Y., Zhang, G., *et al.* (2011) 'Modeling animal–vehicle collisions considering animal–vehicle interactions', *Accident Analysis and Prevention*, 43(6), pp. 1991–1998.

Lao, Y., Wu, Y.J., *et al.* (2011) 'Modeling animal-vehicle collisions using diagonal inflated bivariate Poisson regression', *Accident Analysis and Prevention*, 43(1), pp. 220–227.

LEBLOND, M. *et al.* (2007) 'Electric Fencing as a Measure to Reduce Moose–Vehicle Collisions', *Journal of Wildlife Management*, 71(5), pp. 1695–1703.

Li, Y., Song, L. and Fan, W. (David) (2021) 'Day-of-the-week variations and temporal instability of factors influencing pedestrian injury severity in pedestrian-vehicle crashes: A random parameters logit approach with heterogeneity in means and variances', *Analytic Methods in Accident Research*, 29, p. 100152.

Litvaitis, J.A. and Tash, J.P. (2008) 'An Approach Toward Understanding Wildlife-Vehicle Collisions', *Environmental Management*, 42(4), pp. 688–697.

Liu, C. *et al.* (2018) 'Multivariate random parameters zero-inflated negative binomial regression for analyzing urban midblock crashes', *Analytic Methods in Accident Research*, 17, pp. 32–46.

Liu, C. and Sharma, A. (2017) 'Exploring spatio-temporal effects in traffic crash trend analysis', *Analytic Methods in Accident Research*, 16, pp. 104–116.

Liu, C. and Sharma, A. (2018) 'Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity', *Analytic Methods in Accident Research*, 17, pp. 14–31.

MALO, J.E., SUÁREZ, F. and DÍEZ, A. (2004) 'Can we mitigate animal-vehicle accidents using predictive models?', *Journal of Applied Ecology*, 41(4), pp. 701–710.

Malyshkina, N.V. and Mannering, F.L. (2010) 'Zero-state Markov switching count-data models: An empirical assessment', *Accident Analysis and Prevention*, 42(1), pp. 122–130.

Malyshkina, N.V., Mannering, F.L. and Tarko, A.P. (2009) 'Markov switching negative binomial models: An application to vehicle accident frequencies', *Accident Analysis and Prevention*, 41(2), pp. 217–226.

Mannering, F. (2018) 'Temporal instability and the analysis of highway accident data', *Analytic Methods in Accident Research*, 17, pp. 1–13.

Mannering, F. *et al.* (2020) 'Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis', *Analytic Methods in Accident Research*, 25, p. 100113.

Mannering, F.L., Shankar, V. and Bhat, C.R. (2016) 'Unobserved heterogeneity and the statistical analysis of highway accident data', *Analytic Methods in Accident Research*, 11, pp. 1–

16.

McCollister, M.F. and van Manen, F.T. (2010) ‘Effectiveness of Wildlife Underpasses and Fencing to Reduce Wildlife–Vehicle Collisions’, *Journal of Wildlife Management*, 74(8), pp. 1722–1731.

Meisingset, E.L. *et al.* (2014) ‘Targeting mitigation efforts: The role of speed limit and road edge clearance for deer-vehicle collisions’, *The Journal of Wildlife Management*, 78(4), pp. 679–688.

Mountrakis, G. and Gunson, K. (2009) ‘Multi-scale spatiotemporal analyses of moose-vehicle collisions: A case study in northern Vermont’, *International Journal of Geographical Information Science*, 23(11), pp. 1389–1412.

Mrtka, J. and Borkovcová, M. (2013) ‘Estimated mortality of mammals and the costs associated with animal-vehicle collisions on the roads in the Czech Republic’, *Transportation Research Part D*, 18(1), pp. 51–54.

Niemi, M. *et al.* (2017) ‘Temporal patterns of moose-vehicle collisions with and without personal injuries’, *Accident Analysis and Prevention*, 98, pp. 167–173.

Pang, J. *et al.* (2022) ‘A temporal instability analysis of environmental factors affecting accident occurrences during snow events: The random parameters hazard-based duration model with means and variances heterogeneity’, *Analytic Methods in Accident Research*, 34, p. 100215.

Polson, N.G., Scott, J.G. and Windle, J. (2013) ‘Bayesian inference for logistic models using Pólya-Gamma latent variables’, *Journal of the American Statistical Association*, 108(504), pp. 1339–1349.

Ramp, D., Wilson, V.K. and Croft, D.B. (2006) ‘Assessing the impacts of roads in peri-urban reserves: Road-based fatalities and road usage by wildlife in the Royal National Park, New South Wales, Australia’, *Biological Conservation*, 129(3), pp. 348–359.

Rodriguez, M.I. *et al.* (2010) ‘Cost–benefit analysis of state- and hospital-funded postpartum intrauterine contraception at a university hospital for recent immigrants to the United States’, *Contraception*, 81(4), pp. 304–308.

Rodríguez-Morales, B., Díaz-Varela, E.R. and Marey-Pérez, M.F. (2013) ‘Spatiotemporal analysis of vehicle collisions involving wild boar and roe deer in NW Spain’, *Accident Analysis and Prevention*, 60, pp. 121–133.

Rowden, P., Steinhardt, D. and Sheehan, M. (2008) ‘Road crashes involving animals in Australia’, *Accident Analysis and Prevention*, 40(6), pp. 1865–1871.

Seiler, A. (2005) ‘Predicting locations of moose-vehicle collisions in Sweden’, *Journal of Applied Ecology*, 42(2), pp. 371–382.

Shirazi, M. *et al.* (2016) ‘A semiparametric negative binomial generalized linear model for modeling over-dispersed count data with a heavy tail: Characteristics and applications to crash

data', *Accident Analysis and Prevention*, 91, pp. 10–18.

Snow, N.P., Williams, D.M. and Porter, W.F. (2014) 'A landscape-based approach for delineating hotspots of wildlife-vehicle collisions', *Landscape Ecology*, 29(5), pp. 817–829.

Sullivan, J.M. (2011) 'Trends and characteristics of animal-vehicle collisions in the United States', *Journal of Safety Research*, 42(1), pp. 9–16.

TPWD (2019) *The Rut in White-tailed Deer., Texas Parks and Wildlife Division, in Austin, Texas.* Available at: https://tpwd.texas.gov/huntwild/hunt/planning/rut_whitetailed_deer/#map.

Ujvari, M., Baagoe, H.J. and Madsen, A.B. (2007) 'Effectiveness of Wildlife Warning Reflectors in Reducing Deer-Vehicle Collisions: A Behavioral Study', *The Journal of Wildlife Management*, 62(3), p. 1094.

Xiong, Y. and Mannering, F.L. (2013) 'The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach', *Transportation Research Part B*, 49, pp. 39–54.

Xiong, Y., Tobias, J.L. and Mannering, F.L. (2014) 'The analysis of vehicle crash injury-severity data: A Markov switching approach with road-segment heterogeneity', *Transportation Research Part B*, 67, pp. 109–128.

Yan, X. *et al.* (2021) 'Temporal analysis of crash severities involving male and female drivers: A random parameters approach with heterogeneity in means and variances', *Analytic Methods in Accident Research*, 30, p. 100161.

Yu, H. *et al.* (2019) 'A latent class approach for driver injury severity analysis in highway single vehicle crash considering unobserved heterogeneity and temporal influence', *Analytic Methods in Accident Research*, 24, p. 100110.

Yu, M., Ma, C. and Shen, J. (2021) 'Temporal stability of driver injury severity in single-vehicle roadway departure crashes: A random thresholds random parameters hierarchical ordered probit approach', *Analytic Methods in Accident Research*, 29, p. 100144.

Yu, M., Zheng, C. and Ma, C. (2020) 'Analysis of injury severity of rear-end crashes in work zones: A random parameters approach with heterogeneity in means and variances', *Analytic Methods in Accident Research*, 27, p. 100126.

Zuberogoitia, I. *et al.* (2015) 'Testing pole barriers as feasible mitigation measure to avoid bird vehicle collisions (BVC)', *Ecological Engineering*, 83, pp. 144–151.

APPENDIX: DERIVATION OF THE GIBBS SAMPLER

The sampling from conditional posterior distributions in the MCMC estimation of the proposed hierarchical model can be divided into two blocks. Whereas the first block contains the sampling of the number of animal road crossings, $n_{s,t}$, the second block includes the sampling of collision probability, $p_{s,t}$, and related parameters for all segment-month pairs.

In the first block, a stick-breaking construction is considered for the Dirichlet process that enables the estimation of the probability of each cluster containing segments. These probabilities help estimate cluster parameters, and eventually the continuous $n^*_{s,t}$ that is then truncated and discretized to obtain segment-month-level animal road crossing count $n_{s,t}$. Conditional posterior distributions of all parameters in block 1 are in closed-form, except the final step for which the Metropolis-Hastings algorithm is used. In the second block, since the binomial distribution with logistic probability function does not have a conjugate prior, Pólya-Gamma data augmentation is adopted to transform the model likelihood to the Gaussian likelihood (Polson et al., (2013). For the notational simplicity, $P(A|\cdot)$ is used to denote the probability of A conditioning on the rest of the parameters and data.

A.1 Posterior sampling of $n_{s,t}$ and the related parameters

Conditioning on the starting value of exposure, $n_{s,t}$, a blocked Gibbs sampler for the Dirichlet process is used to sample $n_{s,t}$ (Ishwaran and James, 2001a).[5] Here, the kernel function used for representing clusters is a truncated normal density function with the truncation made at -0.5 from below to ensure the non-negativity of resulting $n_{s,t}$. For simplicity in cluster-specific distribution, the precision parameter ϑ is set to one, with the following base distribution:

$$P_0(\mu, \sigma^2) = \mathbf{Normal}(\mu|\mu_0, \sigma^2)\mathbf{Gamma}(1/\sigma^2\ |d_0, e_0), \tag{5}$$

The prior parameters are chosen to be weakly informative ($\mu_0 = 0$, $d_0 = 2$ and $e_0 = 10$). Due to the discrete nature and the limited number of distinct values of $n_{s,t}$, the maximum number of distinct clusters (C) is set to 3.[6] For a thorough review of the stick-breaking construction, readers can refer to Ishwaran and James (2001b). Using the above prior specification, the Gibbs sampler proceeds via the following sampling steps:

- For each $s = 1, \dots, S$ and $t = 1, \dots, T$, update $\mu_{s,t}$ and $\sigma^2_{s,t}$ by sampling from a multinomial distribution with

$$p\left(\mu_{s,t} = \mu^*_c \text{ and } \sigma^2_{s,t} = \sigma^{*2}_c \middle| \cdot\right) = \frac{w_c p\left(n_{s,t}\middle|\mu^*_c, \sigma^{*2}_c\right)}{\sum_{l=1}^{C} w_l p\left(n_{s,t}\middle|\mu^*_l, \sigma^{*2}_l\right)}, \tag{6}$$

  where w_l is the weight and μ^*_l and σ^*_l are parameters for cluster l. The kernel function for each cluster is as follows:

[5] See Shirazi et al. (2016) for an application in safety research.

[6] We tried to estimate the model with the higher number of clusters, but convergence issues are encountered due to a limited range of $n_{s,t}$.

$$p\left(n_{s,t}\middle|\mu_l^*,\sigma_l^{*2}\right)=\frac{\Phi\left(n_{s,t}+1/2|\mu_l^*,\sigma_l^{*2}\right)-\Phi(n_{s,t}-1/2|\mu_l^*,\sigma_l^{*2})}{1-\Phi(-1/2|\mu_l^*,\sigma_l^{*2})}, \tag{7}$$

where Φ(.) is a normal cumulative distribution function

- A stick-breaking construction of Dirichlet process is used to compute the probabilities or weights for each cluster. Assuming V_l for each cluster l is independent $\mathbf{Beta}(1,\vartheta)$, the weights are $w_1=V_1$ and $w_l=V_l\prod_{i<l}(1-V_i)$ for $l=2,\dots,C$. By setting V_C to 1, the weights across all clusters are guaranteed to sum to 1. The posterior of V_l accounts for the number of segment-time pairs that belong to cluster l. It can be sampled from:

$$V_l\sim\mathbf{Beta}\left(1+n_l,\vartheta+\sum_{i=l+1}^{C}n_i\right),\text{for } l=1,\dots C-1, \tag{8}$$

where n_l is the number of $\mu_{s,t}$ that is equal to μ_l^*.

- For each $s=1,\dots,S$ and $t=1,\dots,T$, draw $n_{s,t}^*$ by first sampling from the following:

$$u_{st}\sim\mathbf{Uniform}\left(\Phi\left(n_{s,t}-\frac{1}{2}\middle|\mu_{s,t},\sigma_{s,t}^2\right),\Phi\left(n_{s,t}+\frac{1}{2}\middle|\mu_{s,t},\sigma_{s,t}^2\right)\right), \tag{9}$$

and then set $n_{s,t}^*=\Phi^{-1}\left(u_{s,t}\middle|\mu_{s,t},\sigma_{s,t}^2\right)$.

- Update the cluster-specific parameters using their conditional distributions for $l=1,\dots C$:

$$1/\sigma_l^{*2}\sim\mathbf{Gamma}\left(a_0+\frac{n_l}{2},b_0+\frac{1}{2}\sum_{\{(s,t):\mu_{s,t}=\mu_l^*\}}\left(\left(n_{s,t}^*-\eta\right)+\frac{n_l}{1+n_l}\eta^2\right)\right), \tag{10}$$

$$\mu_l^*\sim I_{\left[-\frac{1}{2},\infty\right)}\mathbf{Normal}\left(\frac{\sum_{\{(s,t):\mu_{s,t}=\mu_l^*\}}n_{s,t}^*}{1+n_l},\frac{\sigma_l^{*2}}{1+n_l}\right), \tag{11}$$

where $\eta=\sum_{\{(s,t):\mu_{s,t}=\mu_l^*\}}\frac{n_{s,t}^*}{n_l}$.

- The above steps include update of all parameters associated with the distribution of $n_{s,t}$. Conditional on these parameters and the parameters related to collision probability $p_{s,t}$, the conditional marginal probability of $n_{s,t}$ is as follows:

$$P\left(n_{s,t}\middle|\cdot\right)\propto\left[\sum_{l=1,\dots,C}w_l\frac{\Phi\left(n_{s,t}+\frac{1}{2}\middle|\mu_l^*,\sigma_l^{*2}\right)-\Phi\left(n_{s,t}-\frac{1}{2}\middle|\mu_l^*,\sigma_l^{*2}\right)}{1-\Phi\left(-\frac{1}{2}\middle|\mu_l^*,\sigma_l^{*2}\right)}\right]\mathbf{Binomial}\left(k_{s,t}\middle|n_{s,t},p_{s,t}\right) \tag{12}$$

Utilizing the above expression, a Metropolis-Hastings algorithm is used to sample $n_{s,t}$ for all segment-month pairs.

A.2 Posterior sampling of $p_{s,t}$ and the related parameters

The posterior sampling for the parameters related to collision probability, $p_{s,t}$, can be divided into two parts: The first part is concerned with the parameters for defining $p_{s,t}$,

which are the regression parameters $\alpha_{0,t}, \boldsymbol{\beta}$ **and** $\boldsymbol{\gamma_t}$, and the second part is related to the parameters associated with indicators, $I_{s,t}$.

Conditional on both $n_{s,t}$ and $I_{s,t}$, marginal posterior distributions of $\alpha_{0,t}, \boldsymbol{\beta}$ **and** $\boldsymbol{\gamma_t}$ are not of well-known form. To address this non-conjugacy challenge, a Pólya-Gamma-distributed auxiliary variable, $\omega_{s,t}$, is introduced for each segment-month pair (Polson et al., (2013).[7] After conditioning on $\omega_{s,t}$ and other parameters, the resulting marginal posterior distributions of $\alpha_{0,t}, \boldsymbol{\beta}$ and $\boldsymbol{\gamma_t}$ become Gaussian. The detailed sampling steps are as follows:

- Conditional posterior distribution of $\omega_{s,t}$ is:

$$\omega_{s,t}|. \sim \textbf{PólyaGamma}\left(n_{s,t}, \alpha_{0,t} I_{s,t} + \boldsymbol{\beta}' \boldsymbol{x}_s + \boldsymbol{\gamma_t'} \boldsymbol{y_{s,t}}\right), \tag{13}$$

- It is worth nothing that posterior updates for time-invariant, $\boldsymbol{\beta}$, and time-varying parameters, $[\alpha_{0,t}, \boldsymbol{\gamma_t}]$, differ substantially, and therefore, we update them separately as detailed below:

  a. We follow Polson et al. ((2013) to obtain the posterior update for time-invariant parameters, $\boldsymbol{\beta}$**,** which turns out to be Gaussian:

$$\boldsymbol{\beta} \sim \textbf{MVN}\left(\boldsymbol{m}_\beta, \boldsymbol{V}_\beta\right), \tag{14}$$

  where,

$$\begin{aligned} \boldsymbol{V}_\beta &= \left(\sum_t \sum_s \left(\omega_{s,t} \boldsymbol{x_s} \boldsymbol{x_s'}\right) + \boldsymbol{B}_0^{-1}\right)^{-1}, \\ \boldsymbol{m}_\beta &= \boldsymbol{V}_\beta \left(\sum_t \sum_s \boldsymbol{x_s} (\kappa_{s,t} - \omega_{s,t} \boldsymbol{\gamma_t'} \boldsymbol{y_{s,t}} - \omega_{s,t} \alpha_{0,t} I_{s,t})\right), \end{aligned} \tag{15}$$

  $\kappa_{s,t} = k_{s,t} - \frac{n_{s,t}}{2}$ and $\boldsymbol{B}_0$ is the prior uninformative covariance matrix (subset of $\boldsymbol{\Sigma_{0,t}}$, a diagonal matrix with large values) for $\boldsymbol{\beta}$.

  b. Similarly, time-varying parameters, $[\alpha_{0,t}, \boldsymbol{\gamma_t'}]$**,** are drawn from:

$$[\alpha_{0,t}, \boldsymbol{\gamma_t'}] \sim \textbf{MVN}\ (\boldsymbol{m_t}, \boldsymbol{V_t}) \tag{16}$$

  where,

$$\begin{aligned} \boldsymbol{V}_t &= \left(\sum_s \left(\omega_{s,t} \boldsymbol{z_{s,t}} \boldsymbol{z_{s,t}'}\right) + \boldsymbol{D}_0^{-1}\right)^{-1}, \\ \boldsymbol{m}_t &= \boldsymbol{V}_t \left(\sum_s \boldsymbol{z_{st}} \left(\kappa_{s,t} - \omega_{s,t} \boldsymbol{\beta}' \boldsymbol{x_s}\right)\right), \end{aligned} \tag{17}$$

  where $\boldsymbol{z_{st}} = [I_{s,t}\ , \boldsymbol{y_{s,t}'}]'$ and D_0 is the prior uninformative covariance matrix (subset of $\boldsymbol{\Sigma_{0,t}}$, a diagonal matrix with large values) for

[7] See Buddhavarapu et al. (2016) and Buddhavarapu et al. (2021) for applications in safety research.

$[\alpha_{0,t}, \boldsymbol{\gamma_t}]$.

For all segment-month pairs, the indicator $I_{s,t}$ is drawn from its conditional posterior distribution:

$$P(I_{s,t} = 1 | \cdot) = \frac{P_{s,t}^1 q_t}{P_{s,t}^0 (1 - q_t) + P_{s,t}^1 q_t}, \tag{18}$$

where

$$\begin{aligned} P_{s,t}^0 &= \exp\left(\kappa_{s,t}\psi_{s,t} - \frac{1}{2}\omega_{s,t}\psi_{s,t}^2\right) \mid \psi_{s,t} = \boldsymbol{\beta}'\boldsymbol{x}_s + \boldsymbol{\gamma}'_{\boldsymbol{t}}\boldsymbol{y}_{\boldsymbol{s,t}} \\ P_s^1 &= \exp\left(\kappa_{s,t}\psi_{s,t} - \frac{1}{2}\omega_{s,t}\psi_{s,t}^2\right) \mid \psi_{s,t} = \alpha_{0,t} I_{s,t} + \boldsymbol{\beta}'\boldsymbol{x}_s + \boldsymbol{\gamma}'_{\boldsymbol{t}}\boldsymbol{y}_{\boldsymbol{s,t}} \end{aligned} \tag{19}$$

Lastly, assuming prior distribution for q_t to be $\mathbf{Beta}(1,1)$, its conditional posterior distribution is:

$$q_t \sim \mathbf{Beta}\left(1 + \sum_{s=1}^{S} I_{s,t}\,, 1 + \sum_{s=1}^{S}(1 - I_{s,t})\right). \tag{20}$$