Development of hazard-specific truck crash modification factors for cold-region rural highways

Yasanthi, Rillagoda G. N.; Mehran, Babak; Patnala, Phani Kumar; Regehr, Jonathan D.; Regoui, Chaouki

Publisher's version / Version de l'éditeur:
https://doi.org/10.1139/cjce-2023-0436
Canadian Journal of Civil Engineering, 2024-06-07

Article title: Development of Hazard-specific Truck Crash Modification Factors for Cold-region Rural Highways

Author details (Authors name followed by author affiliation)

- ***Rillagoda G.N. Yasanthi****
  Department of Civil Engineering, University of Manitoba, 15 Gillson Street,Winnipeg MB R3T 5V6, Canada

- ***Babak Mehran***
  Department of Civil Engineering, University of Manitoba, 15 Gillson Street,Winnipeg MB R3T 5V6, Canada

- ***Phani Kumar Patnala***
  Department of Civil Engineering, University of Manitoba, 15 Gillson Street,Winnipeg MB R3T 5V6, Canada

- ***Jonathan D. Regehr***
  Department of Civil Engineering, University of Manitoba, 15 Gillson Street,Winnipeg MB R3T 5V6, Canada

- ***Chaouki Regoui***
  AI for Logistics Supercluster support Program, Digital Technologies Research Centre, National Research Council Canada K1A 0R6

*Corresponding author

Development of Hazard-specific Truck Crash Modification Factors for Cold-region Rural Highways

Abstract

This study attempts to develop (*i*) truck safety performance functions (SPFs), and (*ii*) hazard-specific crash modification factors (CMFs), for cold-region rural highways. Police-reported truck-involved crashes on rural highway segments of Alberta, Canada, were used to develop truck SPFs for four crash severity levels: total, fatal, personal injury (PI), and property damage only (PDO). Three settings of the Poisson-Tweedie Regression modelling approach representing Poisson, geometric Poisson, negative binomial distributions were used to develop truck SPFs; the negative binomial distribution was deemed as the most appropriate distribution to model truck-involved crashes for all crash severity levels. The CMF for poor visibility (CMF=1.5) suggests that poor visibility increases PI type truck-involved crashes on rural two-lane two-way highway segments by 50% as compared to the number of such crashes attributed to crash causes other than transportation hazards. Road safety researchers may adopt the methodology to effectively rank hazard risks to highway freight transportation systems.

Keywords: Poisson-Tweedie distribution; Truck safety performance functions; Transportation hazards; Hazard ranking

1. Introduction

Cold-region rural highways are often associated with hazardous driving conditions and are frequently subjected to transportation hazards such as adverse precipitation conditions (e.g., snow, rain), atypical road surface conditions (e.g., icy/slushy pavements), poor visibility conditions, and wildlife. It is well-documented that the presence of such transportation hazards (hazards hereinafter) may lead to serious consequences including motor vehicle crashes and/or fatalities. For instance, crash statistics for the United States (US) and Canada revealed that nearly 21% of annual total crashes are attributed to adverse road surface conditions prevalent in winter weather conditions (FHWA, 2021; Transport Canada, 2021). Wildlife crashes — motor vehicle crashes due to vehicles' contact with wildlife — is a serious concern for rural highways in Canada. In fact, according to British Columbia's wildlife crash prevention program, the average hourly wildlife crash rate on Canadian roads ranges from four to six (WCPP, 2022), which accrues an annual highway safety maintenance cost of approximately 800 million Canadian Dollars (Desjardins, 2021). Due to their dynamic vehicle performance capabilities (e.g., braking distance, roll stability), large trucks may not be able to safely maneuver in the presence of transportation hazards such as wildlife and adverse road-weather conditions, making them a potential threat to the safety of other vehicles on the road. According to the Federal Highway Administration (FHWA), approximately 71% of fatalities in crashes involving large trucks were occupants of other vehicles (FHWA, 2021). In addition, several past studies have demonstrated that the presence of trucks in traffic composition has a significant effect on crash frequency (Jacob and Beaumelle, 2010; Huang et al., 2011). However, empirical evidence available on the impact of hazards on truck-involved crashes is limited particularly in the context of cold-region rural highways. Despite the scarcity of such evidence, some Canadian transportation agencies continue to predict crash frequencies based on the crash prediction approach proposed in the Highway Safety Manual (HSM) (AASHTO, 2010), which is developed using crash data reported on specific highways mainly in the US (Davis, 2019).

The HSM provides directives on predicting crashes based on safety performance functions (SPFs) and crash modification factors (CMFs) (AASHTO, 2010). SPFs are empirically fitted regression models which predict crash frequency for (*i*) a specific period of analysis (typically one year), and (*ii*) a specific analysis unit (e.g., an intersection). In SPFs, crash frequency can be modelled as a function of (*i*) roadway geometric variables such as lane width, lane count, etc., and/or (*ii*) traffic variables such as annual average daily traffic (AADT), vehicle miles travelled (VMT), or traffic composition. Yet, it is important to note that SPFs are developed for a particular set of geometric conditions in one of the two analysis units, i.e., (*i*) intersections, or (*ii*) highway segments (homogenous highway sections with uniform geometric and traffic characteristics). For instance, an SPF could be developed using crash data reported on a highway with n number of lanes; such an SPF can only be used to predict crashes on highways with a lane count of n. To mitigate the limited applicability of SPFs, the HSM suggests using SPFs with CMFs — multiplicative factors

applied to SPFs. More specifically, CMFs adjust the SPF-predicted crash frequencies to different exposure conditions (e.g., modification in lane width, addition of a new travel lane).

Despite the extensive use of SPFs and CMFs in road safety research, recent studies have suggested several directions of improvement to enhance the applicability of the HSM-based crash prediction approach (Miaou, 2013; Noland and Adediji, 2018). Some of them include: (i) introducing heterogeneity in geometric conditions of highway segments (Park et al., 2014), (ii) adopting advanced modeling approaches to handle over-dispersed, heterogeneous, and zero-inflated crash data (Saha et al., 2020), (iii) mitigating omitted-variable bias (e.g., not including hazards as independent variables in SPFs) and statistical bias (predicted versus observed values) in crash frequency modeling (Ahmed, 2022), (iv) developing vehicle type-specific (e.g., truck-specific) SPFs (Davis, 2019; Park et al., 2014), and (v) development of hazard-specific CMFs for cold-region rural highways (Davis, 2019). In addition to such improvements proposed in HSM-based literature, the US-based SPFs and CMFs suggested in the HSM must be recalibrated to fit the driving conditions of the jurisdiction planning to use the crash models. Such recalibration efforts will help avoid biased crash estimations, which will eventually help transport authorities to accurately predict crash frequencies and thus improve safety of their highways.

The present study focuses on hazard-specific truck involved crashes in cold regions and intends to address the above mentioned limitations of the HSM-based crash prediction approach by developing (i) truck SPFs: SPFs developed using historical truck-involved crash data to specifically predict truck crash frequencies, and (ii) hazard-specific CMFs: multiplicative factors (applied to truck SPFs) reflecting the change in truck crash frequencies due to transportation hazards, for cold region rural highways. The truck SPFs and hazard-specific CMFs presented in this study were developed using (i) truck-involved crash (hereinafter truck crash) data, and (ii) the major contributing hazard for each crash, reported over a period of three years from 2015 to 2017 in the provincial rural highway network of Alberta, Canada. It is important to note that, the term "trucks" in this study refers to any truck with a gross vehicle weight greater than 4,500 kg including tractor-trailer combinations. A series of Poisson-Tweedie regression (PTR) models were developed to generate truck SPFs and hazard-specific CMFs to predict crash frequencies by crash severity type in rural highway (two-lane two-way and multilane) segments. The Tweedie distribution of a PTR model can transform several distributions into a flexible unified mean-variance relation: $variance = mean + dispersion \times (mean)^P$, where P ($P \in \mathbb{R}$) is typically referred to as the "power parameter". In this study, each PTR model was characterized by a unique setting of P: (i) $P = 1$ (Poisson), (ii) $P = 1.5$ (Geometric Poisson), (iii) $P = 2$ (Negative Binomial (NB)), and (iv) P not fixed. The PTR models can also handle over-dispersion, zero-inflation, and heterogeneity in crash data (Kokonendji et al., 2004; Saha et al., 2020; Gaweesh et al., 2022). Contributions of this study are threefold. First, road safety researchers/practitioners may adopt the study methodology to predict and mitigate wildlife truck crashes in cold-region rural highways. Second, the hazard-specific CMFs could be effectively used for developing a highway hazard ranking system to warrant prioritized safety measures. Third, this study contributes to the body of road safety literature by exploring the feasibility of developing truck SPFs based on the class of PTDs — a rather novel statistical distribution in the SPFs/CMFs paradigm.

2. Research Background

Understanding the factors affecting truck crash frequency is vital for establishing reliable highway freight transportation systems (Gaweesh et al., 2022). Although the HSM-based crash prediction approach provides a systematic approach to evaluate the impacts of different factors on truck crash frequency, the SPFs and CMFs presented in the current version of the HSM may not be transferable to some highway segments due to several reasons (Brimley et al., 2012). For instance, the SPFs and CMFs presented in the current version of the HSM do not differentiate crash prediction based on vehicle type (AASHTO, 2010). Thus, transferability of the current SPFs and CMFs to roadway segments with significant truck traffic is questionable (Brimley et al., 2012). To address this issue, past studies have suggested developing SPFs to predict truck crash frequency based on (i) highway geometry (Lee et al., 2015; Caliendo et al., 2007; Cafiso et al., 2021), (ii) traffic exposure (Hadi et al., 1995; Caliendo et al., 2007), and (iii) transportation hazards (Gaweesh et al., 2022; Cafiso et al., 2021). Consistent with the HSM (AASHTO, 2010), most studies suggest that roadway geometric features such as lane width (Das et al., 2021; Lee and Mannering, 2002), horizontal curvature (Das et al., 2021), and segment length (Caliendo et al., 2007) have a significant effect on truck crash frequency. Some studies showed a positive relationship between crash frequency and truck traffic exposure,

when expressed in terms of (*i*) truck AADT (Das et al., 2021), (*ii*) truck miles travelled (Gaweesh et al., 2022), and (*iii*) truck percentage (Wen et al., 2022; Gaweesh et al., 2022). Although some studies focused on the variation of truck crash frequency due to different hazards, the impact of adverse road-weather conditions and wildlife on crash frequency is less explored in truck safety research. More importantly, previous research revealed that adverse road-weather conditions could intensify truck crash vulnerability (Gaweesh et al., 2022; Ahmed et al., 2018). For instance, according to Gaweesh et al. (2022), wind and snow significantly increase HAZMAT truck crash frequency. Moreover, recent studies highlighted that the impact of road-weather conditions on truck crashes also vary according to highway geometric characteristics (e.g., crash frequency in mountainous terrain versus negligible grade) of highway segments (Ahmed et al., 2018).

Identifying the most appropriate statistical distribution to fit crash frequency data is a critical step in developing SPFs (Caliendo et al., 2007). Crashes are discrete, random events represented by non-negative integers. Accordingly, crash data are sometimes assumed to follow a Poisson distribution (Saha et al., 2020) which assumes equi-dispersion in crash data, i.e., the mean is equal to the variance (Hadi et al., 1995). In practice, however, crash data are often found to be over-dispersed (Saha et al., 2020), i.e., the mean is significantly different from the variance (Hadi et al., 1995). Neglecting overdispersion in crash data frequency modelling may cause significant errors such as (*i*) biased model coefficients, and (*ii*) type I error: false rejection of the null hypothesis (i.e., all model coefficients are equal to zero), (Mannering et al., 2016). As a remedy, most past studies developed crash frequency models based on the Poisson-Gamma distribution (i.e., the NB distribution) which incorporates a Gamma distributed error term with a quadratic variance function into the Poisson process (Saha et al., 2020). While the HSM assumes the NB distribution to represent crash data, many studies demonstrated that a Gamma distributed error term may not always best fit crash data (Greene, 2008). Some studies employed different Poisson-mixture distributions including (*i*) Poisson-Weibull (Cheng et al., 2013), (*ii*) Poisson-lognormal (Miranda-Moreno et al., 2005), and (*iii*) Geometric Poisson (Özel and Inal, 2010), to fit crash data. However, evaluating the suitability of a wide variety of distributions is time consuming and computationally challenging. Therefore, recent studies have introduced Poisson-Tweedie regression (PTR) models — a class of Poisson-mixture distributions assuming a Tweedie distributed random error term (Bonat et al., 2018) — into crash prediction (Saha et al., 2020). In fact, Saha et al. (2020) confirmed the suitability of PTR in crash frequency modeling by developing SPFs for intersections in Florida highways.

In summary, recent studies question the applicability of the current SPFs/CMFs to predict truck crash frequencies because (*i*) they are not developed to specifically predict truck crash frequencies, and (*ii*) they are developed by assuming the NB distribution to explain crash data which might not best represent truck crash frequencies. In addition, the literature focusing on incorporating crash trigger factors such as hazards into SPFs is limited, particularly in the context of truck SPFs. Accordingly, this study intends to provide an efficient, robust approach to predict truck crash frequencies in extremely cold region rural highway segments by (*i*) modelling truck crash frequencies using different settings of the PTD, and (*ii*) developing hazard-specific CMFs for truck SPFs. Table 1 summarizes the reviewed literature, highlights research gaps, and presents the proposed research gap fills.

Table 1: Summary of SPF literature, research gaps, and the proposed research gap fills

	Literature review summary	**Research gap**	**Proposed research needs**
Characteristics of data used to develop SPFs	Presence of hazards intensify truck vulnerability (Gaweesh et al., 2022) The HSM (AASHTO, 2010) does not differentiate crash prediction by vehicle type; thus, the HSM's SPFs may not accurately predict truck crash frequencies (Brimley et al., 2012)	Limited studies used truck crash data to develop SPFs/CMFs	SPFs and CMFs are developed using truck crash data
		Data aggregation practices (e.g., using an average AADT to represent traffic exposure in SPFs) is questionable	Annual AADT data used over the study period
		Crash data often collected from a specific region	Tuck crash data collected from the entire provincial rural highway network of Alberta
		Weather hazard data often collected from nearby weather stations which may not accurately represent the microclimate at the crash location	Hazard data including weather hazards (e.g., snow, poor visibility) contributing to each crash are retrieved from crash data records

	Literature review summary	Research gap	Proposed research needs
Methods used in SPF	The HSM models crashes using the NB distribution (AASHTO, 2010) The best-fit distribution type to model crash data may not always be the NB distribution (Greene, 2008) The PTD is a unified framework used to efficiently model over dispersed and/or zero-inflated crash data (Saha et al., 2020)	Given the wide variety of distributions proposed to model crash data, a flexible and convenient alternative to model crash data is required Studies evaluating the suitability of the PTD to model crash data is limited	The PTD — a unified distribution to model crash data — is used to develop a set of truck SPFs. SPFs are developed based on several settings of *P*
Independent variables (IVs) used in SPFs	Most frequently used IVs in SPFs reflect traffic exposure and highway geometric conditions (AASHTO, 2010; Saha et al., 2020) Hazards are often omitted IVs in SPF development (Cafiso et al., 2021) Omitting important covariates may cause omitted-variable bias in SPFs (Davis, 2019)	Past literature on the impact of hazards on crash frequency is less consistent No studies focused on developing hazard-specific CMFs for truck SPFs	Hazard-specific CMFs developed for truck SPFs

3. Methodology

This study focuses on developing truck SPFs and hazard-specific (weather and wildlife) CMFs for rural highway segments of Alberta, Canada. In this study, separate truck SPFs were developed for rural two-lane, two-way highways (R-TL-TW) and rural multilane highways (RM) for four crash severity levels: total, fatal, personal injury (PI), and property damage only (PDO). While the HSM considers two analysis units (intersections and highway segments) for SPFs, this study only focuses on developing truck SPFs for rural highway segments. The highway segments used in this study are homogenous highway segments: road segments possessing uniform highway geometric conditions represented by a segment's number of lanes, pavement surface type (paved/unpaved), horizontal alignment (straight/curve), and vertical alignment (level, hillcrest, sag (i.e., bottom of hill), grade). Such homogenous highway segments lead to smaller prediction errors as compared to using fixed-length segments in SPF development (Akbari et al., 2020). The study methodology consists of three steps: (*i*) data collection and preparation, (*ii*) development of truck SPFs, and (*iii*) extracting hazard-specific CMFs. The following subsections provide a brief description of each step.

3.1. Data Collection and Preparation

- Data collection

The study data include three years (2015 to 2017) of: (*i*) truck-involved crash data (*ii*) traffic exposure data (e.g., AADT, VMT), and (*iii*) highway geometry data for R-TL-TW and RM highway segments of Alberta, Canada. The crash data were collected from Alberta Transportation and the dataset contains detailed information of truck-involved crashes such as (*i*) description of the crash, (*ii*) crash time and date, (*iii*) crash severity (fatal, PI, PDO), (iv) geographic coordinates of the crash location, and (v) highway type (R-TL-TW, RM). The traffic exposure dataset includes yearly AADT and truck percentage values on each highway segment; these data were also collected by Alberta Transportation. The national road network series of Statistics Canada (2020) includes a version of Alberta's provincial highway network which is divided into homogenous highway segments. Thus, Alberta's highway segments provided in the national road network files (Statistics Canada, 2020) are used to represent the analysis unit (i.e., homogenous highway segments) for the truck SPFs developed in this study. Geometric characteristics of each homogeneous segment including: (*i*) segment length, (*ii*) number of lanes, (*iii*) pavement surface condition (paved/unpaved), (*iv*) highway horizontal alignment (straight/curve), and (*v*) highway vertical alignment (level, hillcrest, sag (i.e., bottom of hill), grade) were collected from the national road network files of Statistics Canada (2020).

- Data preparation

The collected data were prepared for analyses in three steps. First, all crashes due to human errors (e.g., speeding, impaired driving) and/or vehicle errors (e.g., worn tires) were labelled as "No hazards" as such crashes are not

attributed to any transportation hazards. Therefore, the main hazard associated with each crash was identified by text mining the crash description entries which represent the major contributing factor for each crash. The text mining process extracted 12 types of hazards using specific word stems: either whole or parts of words. For instance, if the crash causation for a crash is "vehicle swerve due to contact with a deer", the hazard contributing to the crash is labelled as wildlife. Sample word stems and their respective hazards identified during the text mining process are tabulated in Table 2. Study data were further cleaned by (*i*) removing data entries with missing and/or erroneous information (e.g., "NA" entries, unrealistic AADT values), and (*ii*) removing segments smaller than 0.1 miles (~161 meters) to minimize calculation errors (AASHTO, 2010).

Table 2. Word stems and their respective hazards identified during the text mining process.

Hazard/major contributing factor	**Sample word stem (s)**
Wildlife	Deer, Horse, Elk, Cow, Moose, Calf, Sheep, Animal, Bison, Eagle, Antelope, Bear, Bird, Prairie chicken, Donkey, Wolf, Wildlife
Fog	Fog
Snow	Snow, hail, blizzard
Windstorm	Wind
Poor visibility	Reduced visibility, low visibility
Icy/Slushy pavement	Icey, Icy, Icy and Slush, Icey and Slush, Slush
Snow and Icy/Slushy pavement	Snow and Icy, Snow and Icey, Snow and Slushy
Snow and wet pavement	Snow and wet, Snow and hydroplane, Snow and hydroplaning, Snow and water, Snow and moisture
Construction zone	Construction, CNSTRCTN
Debris from preceding vehicle	Debris, bale
Wet pavement	Wet, hydroplane, hydroplaning, moisture, water
Weather	Weather, climate

Second, all three datasets (i.e., crash data, traffic exposure data, and highway geometric data) were spatially aggregated using ArcGIS® software, such that each highway segment includes information about its geometric conditions, yearly traffic exposure conditions, and crash details for all crashes reported in the respective segment in each year of the study period. Thereafter, the aggregated dataset was divided into two separate datasets based on the highway type of each segment, i.e., R-TL-TW or RM. Third, annual truck crash frequencies for each major contributing factor were extracted for every year in the study period (2015 to 2017). Table 3 presents details of the study data in terms of the (*i*) descriptive statistics of traffic exposure conditions and highway geometric conditions, and (*ii*) details of hazards present, for the two highway types considered in this study. As presented in Table 3, the number of R-TL-TW highway segments (i.e., 11,119) is significantly larger than the segment count of RM highways (i.e., 630). This observation is intuitive as two-lane, two-way highways comprise the majority of Alberta's rural highway network (Statistics Canada, 2020). Accordingly, the number of hazards observed for R-TL-TW highways (i.e., 12 hazards) is greater than that observed for RM highways (i.e., nine hazards).

Table 3: Descriptive statistics and details of covariates

Road type	**Variable type**	**Covariate**		**Nomenclature**	**Minimum**	**Maximum**	**Mean**	**Standard deviation**	**Sample size**
Rural two-lane, two-way highways (R-TL-TW)	Continuous	Segment length (meters)		L	162.8	26,546.5	2,123.3	1,988.9	11,119
		AADT (vehicles per day)		$AADT$	50	99,560	7,264	11,065	11,119
		Truck percentage (%)		TrP	0.3	50.1	13.56	6.9	11,119
	Categorical[1]	Pavement status	Paved[2]	PS_P	Not applicable				10,179
			Unpaved	PS_{UP}	Not applicable				1,020
		Highway horizontal alignment	Straight[2]	HA_{St}	Not applicable				9,699
			Curve	HA_C	Not applicable				1,500
		Highway vertical alignment	Level[2]	VA_L	Not applicable				9,081
			Grade	VA_G	Not applicable				1,542
			Hillcrest	VA_{HC}	Not applicable				417

Road type	Variable type	Covariate		Nomenclature	Minimum	Maximum	Mean	Standard deviation	Sample size
Rural two-lane, two-way highways (R-TL-TW)	Categorical[1]	Highway vertical alignment	Sag	VA_{Sg}	Not applicable				159
		Hazards present		H	No hazards[2,3], Wildlife, Construction zone, Fog, Snow, Poor visibility, Wet pavements, Icy/slushy pavements, Snow and icy/slushy pavements, Snow and Wet pavements, Debris from preceding vehicle, Weather, Windstorm				
Rural multilane highways (RM)	Continuous	Segment length (meters)		L	167.5	6395.7	1,798.7	1467.2	630
		AADT (vehicles per day)		$AADT$	530	99,560	29,156	29112.2	630
		Truck percentage (%)		TrP	0.7	28.9	10.9	5.8	630
		Number of lanes		N	3	4	3	0.2	630
	Categorical[1]	Highway horizontal alignment	Straight[2]	HA_{St}	Not applicable				519
			Curve	HA_C	Not applicable				111
	Categorical[1]	Highway vertical alignment	Level[2]	VA_L	Not applicable				405
			Grade	VA_G	Not applicable				159
			Hillcrest	VA_{HC}	Not applicable				54
			Sag	VA_{Sg}	Not applicable				15
		Hazards present		H	No hazards[2,3], Wildlife, Construction zone, Fog, Snow, Poor visibility, Icy/slushy pavements, Snow and icy/slushy pavements, Debris from preceding vehicle, Windstorm				

Notes:
[1]Categorical variable is treated as a dummy variable in SPF development
[2]Considered as base conditions (i.e., included in the intercept of SPFs)
[3]Includes segments with zero crashes or segments with crashes due to causes other than transportation hazards considered in this study. No hazards condition include all crash causation factors (e.g., impaired driving, driver fatigue, brake failure, etc.) other than the hazards listed in Table 3.

In total, 4,986 truck-involved crashes were observed on Alberta's rural highways during the study period (2015 to 2017) with 4,694 crashes reported in R-TL-TW highways and 290 crashes reported in RM highways. Fig. 1 presents the number of crashes reported under each combination road type and crash severity level considered. Most crashes reported in both road types are PDO type crashes followed by PI type crashes and fatal crashes (in case of R-TL-TW). It should be noted that no fatal truck crashes observed on RM segments were attributed to hazards; thus, SPFs were not developed for fatal crashes on RM segments in this study. The number of truck crashes on R-TL-TW highways are higher than that on RM highways for all crash severity levels (Fig. 1), which is intuitive given the comparatively high number of R-TL-TW segments as compared to the RM segment count (Table 3).

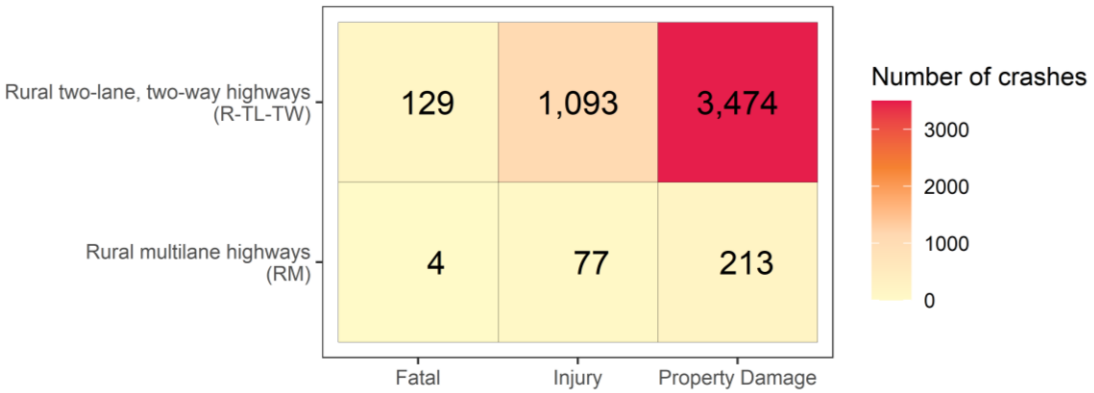

Fig. 1: Crash count for each road type and crash severity level combination in Alberta (2015-2017, inclusive)

3.2. Development of Truck SPFs

Truck SPFs developed in this study attempt to predict y_i: annual truck crash frequency in segment i, based on a set of covariates representing potential hazards to trucks, highway geometry, and traffic exposure conditions for segment i. In this study, truck SPFs are developed based on a rather neoteric class of discrete generalized linear models named Poisson-Tweedie models. Despite its novelty in road safety research, modelling crash data using the Poisson-Tweedie class of models offers a wide variety of benefits. First, fundamentally, the PTD is a Poisson-mixture distribution which incorporates a Tweedie distributed error term into the Poisson process (Bonat et al., 2018). Therefore, the PTD can be used to model crash data manifesting overdispersion and/or zero- inflation (Saha et al., 2020; Bonat et al., 2018) — common phenomena observed in crash data. In fact, some special cases of PTDs include frequently used distributions in crash data modelling (e.g., Poisson distribution, NB distribution). Such unification allows road safety practitioners to adopt the PTD as a solitary distribution function to effectively model crash data irrespective of the nature of crash data. In fact, adopting such a holistic framework to model crash data prevents the need to specify a modelling distribution used to predict crashes (e.g., Poisson distribution), thus mitigating misjudgement of the best-fit distribution for crash data (Bonat et al., 2018). Yet, the application of PTD in road safety analysis and modelling is still in its infancy with limited attention in truck safety research.

In the context of this study, a distribution representing the family of PTDs can be specified using three parameters (see Jørgensen and Kokonendji (2016) for a complete definition): (i) $\mu_i > 0$: mean annual truck crash frequency in segment i, (ii) $\tau_i > 0$: dispersion parameter for segment i, and (iii) $P \geq 1$: the Tweedie power parameter. Developing truck SPFs based on the PTD involves modelling μ_i as a function of X_i: a vector of covariates representing crash trigger factors such as hazards, AADT and segment length (Washington et al., 2020). The covariates included in X_i for R-TL-TW highways and RM highways are presented in Table 3. The Poisson-Tweedie class of models belong to the exponential family (Kaas, 2005). Therefore, a generalized linear regression framework can be conveniently adopted to model crash data based on the PTD (Dunn and Smyth, 2008). Accordingly, the relationship between μ_i and X_i can be formulated using an inverse log link function — the mathematical equation relating μ_i to the covariates (Hadi et al., 1995) — as

$$\mu_i = \exp\,(X_i^T \beta) \tag{1}$$

where β is a vector of estimable parameters (i.e., regression coefficients). The expectation and variance properties of y_i, when modelled using the PTD, are expressed in Equations 2 and 3 respectively.

$$E(y_i) = \mu_i \tag{2}$$

$$Var\,(y_i) = \mu_i + \tau_i \mu_i^P \tag{3}$$

- Truck SPF types

It is well-documented that PTDs with $1 \leq P \leq 2$ are appropriate to model count data with exact zeros, i.e., zero truck crash frequencies (Dunn and Smyth, 2008). Accordingly, a set of regression models (i.e., truck SPFs) was developed for four different settings of P: (i) $M1$ in which $P = 1$, i.e., Poisson distribution a.k.a. NB1 model (Greene, 2008), (ii) $M2$ in which $P = 1.5$, i.e., Geometric Poisson distribution (Özel and Inal, 2010), (iii) $M3$ in which $P = 2$, i.e., Poisson-Gamma distribution a.k.a. NB2 model, and (iv) $M4$ in which P is not pre-defined. In $M4$, P is estimated using maximum likelihood estimation approach (Dunn, 2013) which is a rigorous approach used to estimate P (Dunn and Smyth, 2008).

- Model format and variable selection for truck SPFs

It is important to note that, for each setting of P, different model formats and variable combinations were examined. For example, one set of models was developed with one continuous independent variable (e.g., AADT) treated as an explanatory variable as opposed to treating it as offset variable; another set of models was developed by treating all continuous independent variables (i.e., AADT and TrP) were treated as offset variables. Further, different variable combinations were also tested. For example, one set of models was developed by including the product of continuous variables (e.g., product of AADT and TrP which corresponds to the truck AADT) as an independent variable in the truck SPFs; a separate set of models was developed by including each continuous variable as a separate

independent variable. Among all such model formats and variable combinations, the following model format and explanatory variable combination was selected to model truck crash frequencies in this study:

$$\mu_i = \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j X_j + \sum_{m=1}^{q} \beta_m^{r=1\ to\ s} X_m^{r=1\ to\ s}\right) \tag{4}$$

where,

β_0 = $\mu_i\ for$ base conditions (see Table 3 for base conditions)

k = $\begin{cases} 3\ for \text{ rural two} - \text{lane, two} - \text{way highways} \\ 4\ for \text{ rural multilane highways} \end{cases}$

j = $\begin{cases} 1\ for \text{ segment length} \\ 2\ for \text{ AADT} \\ 3\ for \text{ Truck Percentage} \\ 4\ for \text{ number of lanes (in case of rural multilane highways)} \end{cases}$

q = $\begin{cases} 4\ for \text{ rural two} - \text{lane, two} - \text{way highways} \\ 3\ for \text{ rural multilane highways} \end{cases}$

m = $\begin{cases} 1\ for \text{ highway horizontal alignmnet} \\ 2\ for \text{ highway vertical alignmnet} \\ 3\ for \text{ the hazard contributing to the crash} \\ 4\ for \text{ pavement surface condition (in case of rural multilane highways)} \end{cases}$

s = $\begin{cases} 1\ for\ \mathrm{m} = 1 \\ 3\ for\ \mathrm{m} = 2 \\ 12\ for\ \mathrm{k} = 1 \text{ and } \mathrm{m} = 3 \\ 9\ for\ \mathrm{k} = 2 \text{ and } \mathrm{m} = 3 \\ 1\ for\ \mathrm{m} = 4 \text{ (in case of rural multilane highways)} \end{cases}$

X_j = Continuous independent variable (see Table 3)

$X_m^{r=1\ to\ s}$ = Categorical variable level (e.g., HA_C) (see Table 3)

β_j = Regression coefficient for continuous independent variable X_j

$\beta_m^{r=1\ to\ s}$ = Regression coefficient for categorical variable level $X_m^{r=1\ to\ s}$

In this study, all truck SPFs were developed as generalized linear regression models (Eq. 4) where all available parameters (see Table 3) were included as covariates in the regression models. Prior to developing the truck SPFs, the impact of potential multicollinearity of independent variables on model coefficients was estimated using the variance inflation factor (VIF): a statistical tool used to quantify inflation in the variance of an independent variable's coefficient (Daoud, 2017). As compared to the conventional correlation indices (e.g., Pearson correlation coefficient) which measure linear correlation between two variables, the VIF indicates the impact of multicollinearity on regression model coefficients. Thus, VIF is considered as a robust indicator for testing correlation in regression models (Daoud, 2017). The VIFs for all independent variables presented in Table 3 were estimated according to Equation 5.

$$VIF_i = \frac{1}{1 - R_i^2} \tag{5}$$

where,

VIF_i : VIF for the ith independent variable

R_i^2 : Coefficient of determination of a regression model where the ith covariate is the dependent variable, and all other covariates are treated as independent variables (see Craney and Surles, 2002)

VIF for each independent variable presented in Table 3 was found to be less than 1.5 which corresponds to minimal impact of potential multicollinearity on the truck SPFs developed (Daoud, 2017). Therefore, all independent variables (Table 3) were included in the truck SPFs. While developing the truck SPFs, all categorical covariates are treated as dummy variables with two or more levels; base conditions are included in the intercept (see Table 3).

- Model selection

To compare the performance of the four truck SPF types (i.e., $M1$ through $M4$), this study uses two widespread model performance measures used in crash modelling: Akaike Information Criterion (AIC) (Akaike, 1998) and Mean Absolute Deviation (MAD) (Neath and Cavanaugh, 2012). AIC and MAD are estimated according to Eq. 6 and Eq. 7 respectively.

$$AIC = 2V - 2LL \tag{6}$$

$$MAD = \sum_{i=1}^{n} \frac{|\mu_i - y_i|}{n} \tag{7}$$

where V: number of model parameters, LL: log likelihood at convergence, and n: number of observations (i.e., number of segments). The truck SPFs with the minimum AIC and/or MAD are selected as the final crash prediction models.

3.3. Development of Hazard-specific CMFs

A CMF reflects the difference in predicted crash frequency due to a change in conditions from the conditions used in developing an SPF (AASHTO, 2010). Accordingly, the hazard-specific CMFs developed in this study represent the increase or decrease in truck crash frequencies due to different transportation hazards as compared to the "no hazards" condition (see Table 3). In regression modelling, such comparisons (e.g., comparing the effect of wildlife on truck crash frequency with respect to the no hazards condition) can be quantified by including base conditions (e.g., no hazards, straight highways) in a model's intercept. In this study, the "no hazards" condition is included in the intercept of the crash prediction models (i.e., β_0) developed in Step 2 (Eq. 4). Therefore, CMF_{H_t}: the hazard-specific CMF for the transportation hazard H_t, can be estimated using the exponent of β_{H_t}: the regression model coefficient associated with H_t, as

$$CMF_{H_t} = \exp(\beta_{H_t}) \tag{8}$$

4. Modeling Results

This study focuses on developing truck SPFs for R-TL-TW and RM highway segments for four different crash severity levels (total, fatal, PI, and PDO). Notably, no hazard-specific truck-involved fatal crashes were reported on RM highway segments during the study period. Hence, the modelling results are presented for seven models combining road type and crash severity levels. First, the best fit PTR setting (P) for each road type and crash severity level combination is presented. Second, the model coefficients of the truck SPFs corresponding to the best fit PTRs are presented.

4.1. The best-fit Poisson-Tweedie distribution settings to model truck crash data

The best-fit P to model truck crash data, for each combination of road type and crash severity, was identified by adopting different settings of P (i.e., $M1$ through $M4$). Accordingly, 28 PTR models (7×4) were developed. Fig. 2 presents the goodness-of-fit measures (AIC and MAD) of the models developed. Looking at AIC and MAD values, $P = 2$ (i.e., M3 type models) was deemed to be best fit for all combinations of road type and crash severity levels which affirms the HSM's approach of modelling SPFs using the NB distribution (i.e., $P = 2$). For fatal crashes on R-TL-TW highways, all P settings were found to provide similar AIC and MAD values. This implies that the fatal crash data do not demonstrate overdispersion. Such equality in the mean-variance relationship allows fatal truck crashes to be modelled using any discrete statistical distribution including the Poisson distribution and/or NB2 distribution. In contrast, the occurrence of PI and PDO type crashes is versatile. In fact, both PI and PDO type crashes exhibited overdispersion. Similarly, total crashes may also demonstrate overdispersion because most crashes in the total crashes category are PDO type crashes (Fig. 1). Therefore, unlike for fatal truck crashes, the Poisson distribution is not appropriate to model PI, PDO, or total crashes. This phenomenon is well-reflected in the study results. For instance,

the value of the information criterion (AIC) for M1 models is substantially high for less severe crashes (e.g., $AIC = 14{,}858$ for PDO crashes in R-TL-TW), implying that there is a need to account for overdispersion when modelling such crashes.

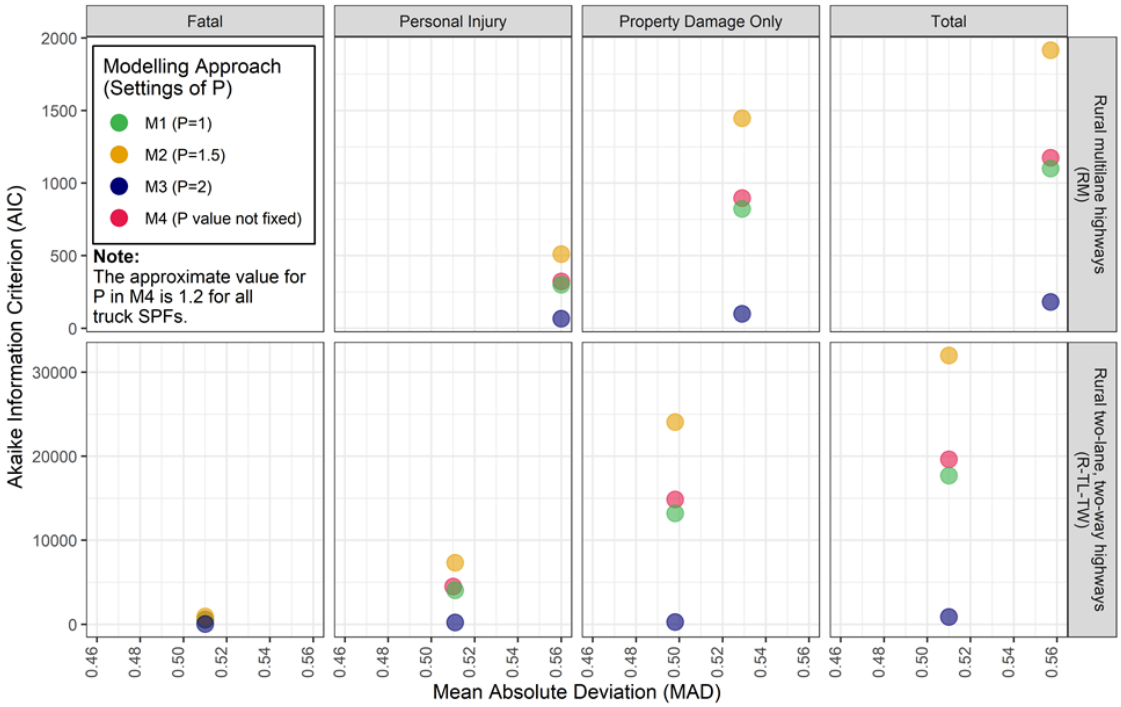

Fig. 2: AIC and MAD values for truck SPFs

Although the best modelling approach to model truck crashes is deemed M3 (i.e., $P = 2$), the maximum likelihood estimate (MLE) for P was estimated as 1.2 for all combinations of road type and crash severity levels considered. In fact, MLEs for P between 1 and 2 are anticipated in truck crash modelling because the candidate PTDs for discrete count data with exact zeros (e.g., zero crash counts for a set of highway segments) are often characterized by P values between 1 and 2 (Dunn and Smyth, 2008). From a modelling perspective, the NB2 distribution is often suggested as the most appropriate candidate for crash data characterized by PTDs with $1 < P < 2$ (Dunn and Smyth, 2008), which is consistent with the study results. In fact, past literature (Saha et al., 2020) also suggests that, in some cases, NB2 ($P = 2$) models may outperform the PTDs in which P is not constrained to a certain value. In contrast, by developing a set of PTR models, Gaweesh et al. (2022) recommended developing truck SPFs without pre-defining P when developing SPFs to predict crash frequencies for trucks transporting hazardous materials. Therefore, it is important to highlight that truck SPFs developed based on special cases of the PTD (e.g., $P = 2$) may not always outperform the statistical performance of truck SPFs developed based on the MLE for P, although this study's results suggest developing truck SPFs based on the Poisson-Gamma distribution.

4.2. Hazard-specific crash modification factors

Fig. 3 presents the exponents of model coefficients for each M3 type ($P = 2$) truck SPF developed for a particular combination of road type and crash severity level considered in this study. Most model coefficients in the seven truck SPFs were found to be statistically insignificant (Fig. 3), which is deemed as a common phenomenon in crash data analysis (Abdulhafedh, 2016). Such statistical insignificance implies the absence of a statistically significant linear relationship between covariates with statistically insignificant coefficients and the mean truck crash frequency. In fact, statistically insignificant model coefficients are anticipated for some hazards (e.g., road-weather hazards: hazards representing adverse road-weather conditions) due to several reasons. First, truck throughput is substantially reduced in extreme cold conditions (Hernandez et al., 2017). In fact, large trucks such as longer combination vehicles (i.e.,

rocky mountain doubles, turnpike doubles, and triple trailer combinations) are restricted from travelling in adverse road-weather conditions including rain, snow, fog, sleet, and ice (Woodrooffe, 2001). Such weather- and vehicle-specific travel regulations reduce trucks operations in inclement road-weather conditions. Accordingly, truck crash frequencies for distinct road-weather hazards are often small, and thus lead to limited sampling conditions. Such limited sampling conditions, i.e., less number of truck crashes due to some hazards (e.g., snow, fog), may lead to statistically insignificant coefficients in truck SPFs (Abdulhafedh, 2016). Second, due to the weather-specific travel regulations imposed on trucks, trucking organizations may tend to reroute and/or reschedule trips to prevent the impact of adverse road-weather conditions on truck operations. In fact, information on most hazards corresponding to statistically insignificant coefficients could be attained prior to travelling. For instance, trucking organizations retrieve weather data predictions for their trucking routes prior to deploying trucks; construction zones are often scheduled, and road users are typically notified of road construction activities. Such safety precautions (e.g., travel restrictions, notifications of hazards) might have prevented truck crashes attributed to some hazards (e.g., road-weather hazard, construction zones) leading to statistically insignificant truck SPF coefficients. In contrast, wildlife pose unanticipated threats for which truck drivers are not prepared for. As a result, the model coefficient for wildlife is found to be statistically significant implying that wildlife has a significant impact on annual average truck crash frequency (AATCF). Third, serval studies (Yasanthi and Mehran, 2020; Yasanthi and Mehran, 2022) have suggested that truck drivers drive vigilantly in adverse road-weather conditions. For instance, according to Yasanthi and Mehran (2022), the average speed of empty multi-trailer trucks travelling on ice warning and moderate/heavy snow is approximately 10 km/h less than the mean speed of such trucks traveling on dry pavements and no precipitation. Truck drivers' cautious driving patterns in the presence of adverse road-weather conditions may have also contributed to the statistically insignificant model coefficients for most road-weather hazards. Fourth, it is also possible that the relationship between road-weather hazards and AATCF is not linear. Thus, only the modelling results corresponding to statistically significant hazard-specific CMFs are discussed.

Fig. 3: Exponent of model coefficients for M3 type truck SPFs

The hazard-specific CMFs correspond to the exponents of model coefficients associated with hazards (Eq. 7); thus, each hazard-specific CMF reflects the increase/decrease of annual truck crash frequencies associated with a specific hazard. Of all hazard-specific CMFs, two hazard-specific CMFs were deemed statistically significant at a confidence level of 90%: (*i*) CMF for wildlife for PDO and total crashes, and (*ii*) CMF for poor visibility for PI crashes, in R-TL-TW highway segments. The statistically significant CMF for wildlife (CMF-W) implies that wildlife imposes a substantial threat to trucks. More specifically, the CMF-W (0.9) suggests that, for both total and PDO type crashes in R-TL-TW highway segments, the AATCF of wildlife-truck crashes is 0.9 times that of no hazard crashes: truck crashes attributed to crash causation factors other than transportation hazards. No hazard crashes, in this study, mainly correspond to defective vehicle conditions (e.g., malfunctioning brakes, poor cargo loading, engine failure), or human

errors/conditions (e.g., speeding, driver fatigue, distracted driving). In fact, two non-hazard crash contributing factors: improper driver conditions (e.g., fatigue, distraction) and driver actions (e.g., driving too fast for conditions), are found to be the predominant crash contributing factors for truck crashes in Canada (Mayhew, 2004). Therefore, the CMF-W of 0.9 is alarming because it implies that the solitary presence of wildlife on a level, paved, and straight R-TL-TW highway segment can lead to 90 PDO type wildlife-truck crashes every year, while the no hazards PDO type annual crash frequency in the same segment is 100. In other words, the CMF-W of 0.9 implies that AATCF due to wildlife alone is comparable to AACTF of no hazard crashes for both PDO type and total crashes. Such a comparatively high wildlife-truck crash frequency indicate cost to wildlife as well. For instance, some wildlife-truck crashes may increase animal mortalities which may pose a significant threat to endangered species or animal populations with low population densities. Therefore, road safety countermeasures focusing on mitigating truck crashes attributed to wildlife on R-TL-TW highways warrant considerable attention. In fact, road mitigation structures — wildlife-vehicle crash mitigation measures incorporated to highway systems (Taylor and Goldingay, 2010) — such as wildlife crossing structures (e.g., underpasses, overpasses), fences, and/or jumpouts (i.e., structures that mitigate wildlife from entering fenced highway segments) are found to be highly effective in reducing wildlife-vehicle crashes (Sugiarto, 2022). For instance, the presence of underpasses and jumpouts with fencing substantially reduced the wildlife-vehicle crash frequency along the Trans Canada highway in Bow Valley, Alberta (Edwards et al., 2022).

Unlike the CMF-W, CMF-PV — the CMF for poor visibility for PI type truck crashes in R-TL-TW highway segments — predicted an increase in such crashes (Fig. 3) compared to no hazard crashes. In fact, CMF-PV (1.5) suggests that PI crashes in R-TL-TW segments will increase by 50% in the presence of poor visibility, as compared to the "no hazards" condition. Such a significant increase in PI type crashes is alarming particularly in the context of crashes involving heavy vehicles. In fact, this observation is rather interesting because some trucks are prohibited to travel on Alberta highways under certain extreme road-weather conditions (Alberta Transportation, 2022). For instance, longer combination vehicles are not allowed to travel on Alberta's two-lane highways in low visibility conditions (Alberta Transportation, 2022). Thus, the predicted increase in truck crash frequencies in poor visibility (Fig. 3) emphasizes the need to develop and/or enforce restrictive travel regulations to limit truck travel in adverse road-weather conditions involving poor visibility. Yet, limiting truck travel also limits the economic performance of trucks which eventually reduces truck productivity. Therefore, we recommend developing robust strategies to mitigate truck crashes triggered by adverse road-weather events resulting in poor visibility. For instance, truck-specific weather-responsive variable speed limit systems (see Yasanthi et al., 2022) could be developed to improve truck safety by regulating truck speed in different visibility conditions. Other strategies focusing on mitigating poor visibility related truck crashes include but not limited to using (*i*) connected vehicle technology to inform truck drivers about visibility conditions along their routes (Raddaoui and Ahmed, 2019), (*ii*) active traffic management strategies such as dynamic lane utilization techniques which restrict truck travel to shoulder lane in poor visibility conditions (Bhatia, 2003), (*iii*) infrared detector-based night vision systems to help truck drivers navigate through poor visibility conditions (Bhatia, 2003), and (*iv*) advanced crash detection and warning systems which could detect objects in the surrounding of a truck's position and warn about potential crashes (Bhatia, 2003).

A comparison of the study results with the findings of a similar study (Gaweesh et al., 2022) reveals that some adverse weather conditions may significantly affect truck crash frequencies. More specifically, by developing a set of truck SPFs, Gaweesh et al. (2022) revealed that high wind speeds significantly increase crash risk for trucks transporting hazardous materials (HAZMAT) for all crash severity levels. In contrast, this study did not find a statistically significant relationship between the presence of windstorms and truck crash frequencies for any crash severity level (Figure 3). It should be noted that, unlike this study, the SPFs developed in Gaweesh et al. (2022) only considered crashes involving trucks transporting HAZMAT. Therefore, it is possible that the comparatively poor roll stability of some HAZMAT transport trucks (e.g., tankers transporting larger quantities of HAZMAT) are more susceptible to crashes caused by high wind speeds (Gaweesh et al., 2022). In addition, it should be noted that this study considered truck crashes reported on highways located in Alberta, Canada (mostly comprising of prairie terrain) while Gaweesh et al. (2022) considered HAZMAT truck crashes in Wyoming, United States (comprising of a mix of mountainous and prairie terrains). Accordingly, the difference between the geographic areas (and thus the variations in the prevalent wind patterns) of the two studies may have contributed to the difference in correlation between high wind speeds and truck crash frequencies. Consistent with the findings of this study, Gaweesh et al. (2022) did not find a statistically significant relationship between the presence of snow and truck crash frequency.

In summary, the study results suggest the following observations:

1. The family of Poisson-Tweedie distributions provides a unified framework to model truck crashes.
2. This study recommends modelling truck crashes using a Poisson-Tweedie power parameter (P) of 2 (i.e., the NB distribution). This observation is rather important because it empirically affirms the HSM approach of modelling SPFs using the NB distribution by conducting a comparative analysis on appropriate distribution types to model truck crash data rather than assuming such crash data to follow the NB distribution.
3. The ratio of the annual average truck crash frequency of wildlife-truck crashes to that of no hazard crashes is 0.9 to 1 for both total and PDO type crashes on R-TL-TW highway segments.
4. PI type truck crashes reported on R-TL-TW highway segments are estimated to increase by 50% as compared to no hazard PI type truck crashes.

Implications of the study results are threefold. First, while fatal truck crashes on R-TL-TW highway segments may be modelled using the Poisson distribution, PI and PDO type truck crashes on R-TL-TW highway segments may need to be modelled using a distribution type that addresses overdispersion (e.g., Negative Binomial distribution). Second, the study results emphasize the impact of wildlife on truck crashes and the importance of implementing effective wildlife-vehicle collision mitigation measures particularly on rural highways. Third, the study results imply freight transport trucks' vulnerability in poor visibility conditions, and thus highlight the need for sophisticated weather-responsive traffic management strategies developed to enhance truck safety in rural highways under adverse road-weather conditions.

5. Conclusions and Future Directions

The crash prediction approach proposed in the highway safety manual (AASHTO, 2010) provides guidance on predicting crash frequency on highway segments based on safety performance functions and crash modification factors. While SPFs and CMFs are extensively used in road safety research, the current version of the HSM does not focus on (i) predicting the impact of different transportation hazards (e.g., wildlife, adverse road-weather conditions) on crash frequency, and (ii) differentiating crash prediction by vehicle type (e.g., prediction of truck crashes). Therefore, the applicability of the HSM's SPFs and CMFs to predict truck crashes is questionable, particularly in the context of cold region rural highways which are frequently exposed to several transportation hazards. Further, The HSM assumes that crash data are represented by the Poisson-Gamma distribution (i.e., the NB distribution). Past literature, however, argue that the best-fit statistical distribution to model crash data is dependent on the nature of crash data. This study attempts to provide an effective approach to predict truck crash frequencies in cold region highway segments by developing truck SPFs and hazard-specific CMFs based on the Poisson-Tweedie distribution — a holistic modelling framework unifying several statistical distributions (e.g., Poisson distribution, NB distribution) which are frequently adopted in crash prediction. In fact, using a PTD-based framework to model truck SPFs prevents presuming that truck crash data are represented by the NB distribution and thus mitigates potential misrepresentation of the distribution type used to model crash data. The proposed methodology to develop truck SPFs and hazard-specific CMFs is demonstrated using truck crash data, traffic exposure data (e.g., AADT), and highway geometric data (e.g., segment length, number of lanes), reported on rural two-lane, two-way highways, and rural multilane highways in Alberta, Canada, over a three-year period from 2015 to 2017. Four different crash severity levels (total, fatal, PI, PDO) were considered.

According to the study results, the best-fit Poisson-Tweedie power parameter to model truck crash frequencies reported on cold region rural highway segments is two ($P = 2$) irrespective of the crash severity level considered. Accordingly, the Poisson-Gamma distribution (i.e., NB) is deemed the most appropriate statistical distribution to model truck SPFs for Alberta's rural provincial highway network, which affirms the HSM's assumption of NB-distributed crash data. Yet, it is important to note that the NB distribution may not always be the most appropriate distribution to model truck crashes. Most hazard-specific CMFs were deemed statistically insignificant. Of the statistically significant hazard-specific CMFs, the CMF for poor visibility suggests a 50% increase in truck crashes due to poor visibility conditions, as compared to the no hazards condition (i.e., truck crashes caused by crash contributing factors other than transportation hazards). The CMF for wildlife (0.9) suggests that the ratio of annual wildlife-truck crash frequency to the annual no hazards crash frequency (i.e., the number of truck crashes attributed

to causes other than transportation hazards) is 0.9 to 1. Such alarming impacts of transportation hazards on truck crash frequency warrant implementing crash mitigation measures such as using (*i*) advanced traffic management and/or vehicle technologies to mitigate truck crashes attributed to poor visibility conditions, and (*ii*) wildlife-truck crash mitigation structures such as underpasses, jumpouts and/or fences to reduce wildlife-truck crash frequency, in rural highway segments.

This study contributes to enhancing rural road safety in several aspects. For instance, transport authorities located in cold regions may adopt the study methodology to identify predominant transportation hazards present to trucks, by developing truck SPFs and hazard-specific truck CMFs. In fact, the intensities of such hazard-specific CMFs could be used to develop a solitary ranking system to rank hazards thus identify hazards posing significant safety threats to trucks. Such a hazard ranking system could be effectively used to improve rural road safety in cold region jurisdictions by prioritizing road safety countermeasures designed to mitigate crashes attributed to high-risk hazards with an alarming impact on truck crash frequency. To the extent of our knowledge, no prior study focuses on developing hazard-specific CMFs based on the family of PTDs. Therefore, this study contributes to road safety research by exploring the suitability of modeling truck crashes using the class of Poisson-Tweedie models. The study findings also contribute to road safety research by affirming the SPF's assumption of NB-distributed crash data by conducting a comparative analysis to evaluate the suitability of different distribution functions to model truck crash data. Despite the evident practical applications of this study, the applicability of the proposed crash prediction approach is limited to rural highway segments in cold regions. Therefore, future research is recommended to extend the study methodology to urban highway segments as well as rural/urban intersections. It should be noted that the truck SPFs developed in this study do not consider highway geometric cross-sectional elements (e.g., shoulder and/or lane widths) or the highway segment type (i.e., divided or un-divided for multilane highways). Accordingly, future research may focus on expanding the proposed methodology to include highway geometric cross-sectional elements as well as differentiating SPFs for divided and un-divided RM highways. Further, the truck crash data used in this study were extracted from police-reported crash data, which may have been affected from the different crash reporting thresholds and practices adopted by different police officers. While consistent with similar studies, the statistically insignificant model coefficients in the truck SPFs highlight the need to explore the applicability of other machine learning approaches used in crash modelling (e.g., local sensitivity analysis (LAS), partial dependence plots (PDP)) to develop truck crash frequencies and hazard-specific CMFs.

Acknowledgements

The authors are thankful to Alberta Transportation Department for providing us the study data. This project was supported in part by collaborative research funding from the National Research Council of Canada's Artificial Intelligence for Logistics Program.

Statement of Competing Interests

The authors declare that the research work presented in this paper was not influenced by any known competing financial or personal interests.

Data availability

Data generated or analyzed during this study are not publicly available due to confidentiality agreements with data providers and research collaborators but are available from the corresponding author on reasonable request.

References

Abdulhafedh, A., 2016. Crash frequency analysis. Journal of Transportation Technologies, 6(04), p.169.

Ahmed, I.U., 2022. Advanced Statistical Modeling of the Frequency and Severity of Traffic Crashes on Rural Highways (Doctoral dissertation, University of Wyoming).

Ahmed, M.M., Franke, R., Ksaibati, K. and Shinstine, D.S., 2018. Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models. Accident Analysis & Prevention, 117, pp.106-113.

Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle. In Selected papers of hirotugu aka ike (pp. 199-213). Springer, New York, NY.

Akbari, M., Shafabakhsh, G. and Ahadi, M.R., 2020. The impact of segmentation method on the aggregate goodness-of-fit measurements of non-linear crash prediction models. SN applied sciences, 2(10), pp.1-13.

Alberta Transportation, 2022. Traffic safety act. Attached Conditions for the Operation of Long Combination Vehicles, Section 3Commercial vehicle dimension and weight regulation. pp. 323-24 Available at https://www.alberta.ca/assets/documents/trans-conditions-long-combination-vehicle.pdf

American Association of State Highway and Transportation Officials (AASHTO), Highway Safety Manual (1st Edition) with Supplement, 2010

Bhatia, P., 2003. Vehicle technologies to improve performance and safety.

Bonat, W.H., Jørgensen, B., Kokonendji, C.C., Hinde, J. and Demétrio, C.G., 2018. Extended Poisson–Tweedie: properties and regression models for count data. Statistical Modelling, 18(1), pp.24-49.

Brimley, B.K., Saito, M. and Schultz, G.G., 2012. Calibration of Highway Safety Manual safety performance function: development of new models for rural two-lane two-way highways. Transportation research record, 2279(1), pp.82-89.

Cafiso, S., Montella, A., D'Agostino, C., Mauriello, F. and Galante, F., 2021. Crash modification functions for pavement surface condition and geometric design indicators. Accident Analysis & Prevention, 149, p.105887.

Caliendo, C., Guida, M. and Parisi, A., 2007. A crash-prediction model for multilane roads. Accident Analysis & Prevention, 39(4), pp.657-670.

Cheng, L., Geedipally, S.R. and Lord, D., 2013. The Poisson–Weibull generalized linear model for analyzing motor vehicle crash data. Safety science, 54, pp.38-42.

Craney, T.A. and Surles, J.G., 2002. Model-dependent variance inflation factor cutoff values. Quality engineering, 14(3), pp.391-403.

Daoud, J.I., 2017, December. Multicollinearity and regression analysis. In Journal of Physics: Conference Series (Vol. 949, No. 1, p. 012009). IOP Publishing.

Das, S., Geedipally, S.R. and Fitzpatrick, K., 2021. Inclusion of speed and weather measures in safety performance functions for rural roadways. IATSS research, 45(1), pp.60-69.

Davis, G.A., 2019. Explaining crash modification factors: Why it's needed and how it might be done. Accident Analysis & Prevention, 131, pp.225-233.

Debrabant, B., Halekoh, U., Bonat, W.H., Hansen, D.L., Hjelmborg, J. and Lauritsen, J., 2018. Identifying traffic accident black spots with Poisson-Tweedie models. Accident Analysis & Prevention, 111, pp.147-154.

Desjardins, 2021. Wildlife-Vehicle Collisions: What You Need to Know. Available at https://www.desjardinsgeneralinsurance.com/blog/-/wildlife-vehicle-collisions-what-you-need-to-know

Dunn, P. K. (2013). tweedie: Tweedie exponential family models. R package version 2.1.7.

Dunn, P.K. and Smyth, G.K., 2008. Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. Statistics and Computing, 18(1), pp.73-86.

Edwards, H.A., Lebeuf-Taylor, E., Busana, M. and Paczkowski, J., 2022. Road mitigation structures reduce the number of reported wildlife-vehicle collisions in the Bow Valley, Alberta, Canada. *Conservation Science and Practice*, *4*(9), p.e12778.

FHWA. 2021. How do weather events impact roads? US Department of Transportation. Federal Highway Administration. https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm.

Gaweesh, S.M., Ahmed, I.U., Ahmed, M.M. and Wulff, S.S., 2022. Developing Statewide Safety Performance Functions for Commercial Trucks Transporting Hazardous Materials on Interstate Rural Roads in Wyoming. Transportation Research Record, p.03611981221103231.

Greene, W., 2008. Functional forms for the negative binomial model for count data. Economics Letters, 99(3), pp.585-590.

Hadi, M.A., Aruldhas, J., Chow, L.F. and Wattleworth, J.A., 1995. Estimating safety effects of cross-section design for various highway types using negative binomial regression. Transportation Research Record, 1500, p.169.

Hernandez, S., Akter, T. and Diaz, K., 2017. *The Effect of Weather Events on Truck Traffic Patterns Using Fixed and Mobile Traffic Sensors* (No. SPTC 15.1-20-F). Southern Plains Transportation Center.

Huang, H., Siddiqui, C. and Abdel-Aty, M., 2011. Indexing crash worthiness and crash aggressivity by vehicle type. Accident Analysis & Prevention, 43(4), pp.1364-1370

Jacob, B. and Feypell-de La Beaumelle, V., 2010. Improving truck safety: Potential of weigh-in-motion technology. IATSS research, 34(1), pp.9-15.

Jørgensen, B. and Kokonendji, C.C., 2016. Discrete dispersion models and their Tweedie asymptotics. AStA Advances in Statistical Analysis, 100(1), pp.43-78.

Kaas, R., 2005. Compound Poisson distribution and GLM's-Tweedie's distribution. MATHEMATICS DAY, 3.

Kokonendji, C.C., Dossou-Gbété, S. and Demétrio, C.G., 2004. Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. SORT-Statistics and Operations Research Transactions, pp.201-214.

Lee, C., Abdel-Aty, M., Park, J. and Wang, J.H., 2015. Development of crash modification factors for changing lane width on roadway segments using generalized nonlinear models. Accident Analysis & Prevention, 76, pp.83-91.

Lee, J. and Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis & Prevention, 34(2), pp.149-161

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.

Mayhew, D.R., Simpson, H.M. and Beirness, D.J., 2004. Heavy trucks and road crashes. Traffic Injury Research Foundation.

Miaou, S. P. (2013). Some limitations of the models in the highway safety manual to predict run-off-road crashes. Transportation research record, 2377(1), 38-48.

Miranda-Moreno, L.F., Fu, L., Saccomanno, F.F. and Labbe, A., 2005. Alternative risk models for ranking locations for safety improvement. Transportation Research Record, 1908(1), pp.1-8.

Neath, A.A. and Cavanaugh, J.E., 2012. The Bayesian information criterion: background, derivation, and applications. Wiley Interdisciplinary Reviews: Computational Statistics, 4(2), pp.199-203.

Noland, R.B. and Adediji, Y., 2018. Are estimates of crash modification factors mis-specified?. Accident Analysis & Prevention, 118, pp.29-37.

Özel, G. and Inal, C., 2010. The probability function of a geometric Poisson distribution. Journal of Statistical Computation and Simulation, 80(5), pp.479-487.

Park, J., Abdel-Aty, M. and Lee, C., 2014. Exploration and comparison of crash modification factors for multiple treatments on rural multilane roadways. Accident Analysis & Prevention, 70, pp.167-177.

Raddaoui, O. and Ahmed, M., 2019, June. User Experience and Human Machine Interface Design for Connected Heavy Trucks: Lessons Learned in Support of the Wyoming DOT Connected Vehicle Pilot. In *2019 ITS America Annual Meeting*. ITSWC.

Raihan, M.A., Alluri, P., Wu, W. and Gan, A., 2019. Estimation of bicycle crash modification factors (CMFs) on urban facilities using zero inflated negative binomial models. Accident Analysis & Prevention, 123, pp.303-313.

Saha, D., Alluri, P., Dumbaugh, E. and Gan, A., 2020. Application of the Poisson-Tweedie distribution in analyzing crash frequency data. Accident Analysis & Prevention, 137, p.105456.

Srinivasan, R., Carter, D. and Bauer, K.M., 2013. Safety performance function decision guide: SPF calibration vs SPF development (No. FHWA-SA-14-004). United States. Federal Highway Administration. Office of Safety.

Statistics Canada, Road Network File, 2020. Catalogue no. 92-500-X

Sugiarto, W., 2022. The Impact of Wildlife Crossing Structures on Wildlife-Vehicle Collisions. Available at SSRN 4025079.

Transport Canada, 2021. Canadian motor vehicle traffic collision statistics: 2020. https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2020.

Washington, S., Karlaftis, M., Mannering, F. and Anastasopoulos, P., 2020. Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC.

WCPP. 2022. Wildlife vehicle collision facts. Wildlife Collision Prevent Program. https://www.wildlifecollisions.ca/collision/collision-facts.htm.

Wen, X., Xie, Y., Jiang, L., Li, Y. and Ge, T., 2022. On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development. Accident Analysis & Prevention, 168, p.106617.

Woodrooffe, J., 2001. Long combination vehicle (LCV) safety performance in Alberta 1995 to 1998. *Woodrooffe & Associates*.

Yasanthi, R.G. and Mehran, B., 2020. Modeling free-flow speed variations under adverse road-weather conditions: case of cold region highways. *Case studies on transport policy*, *8*(1), pp.22-30.

Yasanthi, R.G. and Mehran, B., 2022. Application of Different Data Analytics for Evaluation of Heavy Vehicle Vulnerability in Cold-Region Rural Highways. *Transportation Research Record*, p.03611981221111353.

Yasanthi, R.G., Mehran, B. and Alhajyaseen, W.K., 2022. A reliability-based weather-responsive variable speed limit system to improve the safety of rural highways. Accident Analysis & Prevention, 177, p.106831.