Chapter

Accident Prediction Modeling Approaches for European Railway Level Crossing Safety

Ci Liang and Mohamed Ghazel

Abstract

Safety is a core concern in the railway operation. Particularly, in Europe, level crossing (LX) safety is one of the most critical issues for railways. LX accidents often lead to fatalities and weighted injuries and seriously hamper railway safety reputation. Moreover, according to statistics, collisions between trains and motorized vehicles contribute most to LX accidents. With this in mind, we will elaborate on accident prediction modeling for train-vehicle collisions at LXs in this chapter. The methods and findings discussed in this chapter will offer an in-depth insight for interpreting significant aspects underlying collision occurrence and facilitate identifying technical countermeasures to improve LX safety.

Keywords: level crossing safety, train-vehicle collisions, accident prediction modeling, nonlinear least-squares method, negative binomial regression method, Poisson regression method, zero-inflated Poisson regression method, zero-inflated negative binomial regression method, model performance evaluation

1. Introduction

The level crossing (LX) is railway property upon which road users are given permission to cross [1]. Accidents at LXs give rise to serious material and human damage, and the majority of accidents are caused by vehicle driver violations. As demonstrated by accident statistics, LX safety is one of the most critical issues that railway stakeholders need to deal with [2, 3]. In 2012, there were more than 118,000 LXs in the 28 countries of the European Union (E.U.) [4]. In some E.U. countries, LX accidents account for up to 50% of railway accidents [5]. In the UK, LXs account for 11.8 fatalities and weighted injuries on average per year, comprising 8.4% of the total system risk for the railway network [6]. There were 49 collisions between road vehicles and trains at LXs in Australia in 2011 [7]. In France, the railway network incorporates more than 18,000 LXs for 30,000 km of railway lines and around 13,000 LXs show heavy road and railway traffic [8]. In 2016, 111 train-vehicle collisions at French LXs led to 31 deaths [9]. This number was half the total number of collisions per year at LXs a decade ago, but still too large [10]. Due to nondeterministic causes, complex operation background, and the lack of thorough statistical analysis based on detailed accident/incident data, the risk assessment of LXs remains a challenging task. Therefore, there is a pressing need for a series of thorough analyses to understand the potential reasons for these accidents and to identify practical countermeasures to prevent accidents at LXs, thus significantly reducing the LX accidents.

In recent years, the Poisson regression model, negative binomial (NB) regression model, and other variants of the Poisson regression model [11, 12] have gained popularity to deal with risk/accident statistics. Ref. [13] adopted the expressions of the estimated expectation value $\hat{\lambda}$ as shown in Eq. (1) corresponding to the Poisson regression and NB regression models, respectively. Ref. [14] employed the variants of Poisson regression model, namely, the zero-inflated Poisson (ZIP) model and the hurdle Poisson model, to deal with LX accident prediction involving the data in North Dakota. Ref. [15] compared the zero-inflated negative binomial (ZINB) model with the USDOT model [16] by using the LX accident data from Illinois, in terms of accident prediction accuracy. The results of this study show that the ZINB model has higher accuracy of prediction. It is worth noticing that the expressions of estimated $\hat{\lambda}$ as shown in Eq. (1) are not appropriate in our current study, since they are limited to handling zero observations and some impacting variables should not be in the exponential form. Ref. [17] developed another model of $\hat{\lambda}$ as shown in Eq. (2). In this model, the product of the average daily road traffic V and the average daily railway traffic T (known as the conventional traffic moment) is adopted. However, using the conventional traffic moment hinders improving the accuracy of the prediction model:

$$\begin{aligned} \hat{\lambda}_{Poi} &= \exp\left(\sum_{j=1}^{m} \beta_0 + \beta_j x_j\right), \\ \hat{\lambda}_{NB} &= \exp\left(\sum_{j=1}^{m} \beta_0 + \beta_j x_j + \varepsilon\right), \end{aligned} \tag{1}$$

where β is the estimated regression coefficient, x is the impacting variable, and ε is the gamma-distributed error in NB regression model:

$$\hat{\lambda} = (V \times T)^{\beta_1} \exp\left(\sum_{j=1}^{m} \beta_j x_j + \sigma\right), \tag{2}$$

where $\sigma = \beta_0$ in Poisson regression model or $\sigma = \beta_0 + \varepsilon$ in NB regression model.

Based on these investigations, it is clear that there is a pressing need for an appropriate accident prediction model that should comprehensively consider contributing factors toward LX safety. Moreover, such a model should have high predictive accuracy. Therefore, in the present study, a new accident prediction model is developed to predict the accident frequency at LXs. Specifically, we focus on the SAL2 type of LX (i.e., an automated LX system with two half barriers and flashing lights), which is the most widely used type of LX in France and contributed most to the total number of accidents at French LXs from 1974 to 2014.

2. Method

In this section, an advanced accident prediction model is developed, which enables to rank risky LXs accurately and identify the significant impacting parameters efficiently. The model considers the average daily road traffic, the average daily railway traffic, the annual road accidents, the vertical road profile, the horizontal road alignment, the road width, the crossing length, the railway speed limit, and the geographic region. The nonlinear least-squares (NLS) method, Poisson regression method, NB regression method, ZIP regression method, and ZINB regression method are employed to estimate the respective coefficients of parameters in the prediction model.

2.1 Data sources and coding

The dataset used in our study, which cover SAL2 LXs in 21 administrative regions in mainland France from 2004 to 2013, has been provided by SNCF Réseau (the French national railway infrastructure manager). Moreover, the dataset includes 10 years of information about annual LX accident frequency, annual roadway accident statistics and railway, roadway, and LX characteristics. In total, there are 8332 public SAL2 LXs involved in our investigation. The impacting parameters relevant to LX accidents considered in our investigation can fulfill the following characteristics: (1) important in determining accident frequency, (2) more permanent in nature (e.g., sight obstruction noted as a problematic factor due to involved alterable construction topography, vegetation, and other environmental elements), and (3) not accident-dependent [18]. The statistical characterization of parameters considered in this investigation are shown in **Table 1**. It is worth noticing that the road accident factor is reflected by the ratio of the annual number of road accidents in a given year to the average number of road accidents per year over the period of 10 years considered, while the region risk factor is reflected by the general accident frequency per SAL2 in the corresponding region. Overall, the data coding is shown in **Table 2**.

2.2 Advanced accident prediction model

Here, we define that the formula of the conventional traffic moment is given as: Traffic moment = Road traffic frequency × Railway traffic frequency [19]. However, based on some previous analyses [20], we adopt a variant called "corrected moment," or CM for short. $CM = V^a \times T^b$, where $a + b = 1$ and the optimal value of a in terms of fitting is calculated to be $a = 0.354$ according to the statistical analysis performed by SNCF Réseau [21]. Therefore, we consider $\left(V^{0.354} \times T^{0.646}\right)$ as an integrated parameter that reflects the combined exposure frequency of both railway and road traffic.

The developed advanced model takes into account various variables as interpreted in **Table 2**. The general form of the model is shown as follows:

$$\lambda_{10Y} = K \times F_{RAcc} \times \left(V^a \times T^b\right) \times \exp\Big(C_{Profile} \times I_{Profile} + C_{Align} \times I_{Align} + C_{Wid} \times Wid + C_{Leng} \times Leng + C_{RSL} \times RSL + C_{Reg} \times F_{Reg}\Big), \quad (3)$$

Parameter	Description	Mean	Std. dev.
Railway traffic characteristics			
Average daily railway traffic	The average number of trains crossing the LX daily;	26.1	30.2
Railway speed limit	The maximum permission speed of train within the LX section;	92.5	42.4
Roadway traffic characteristics			
Average daily road traffic	The average number of road vehicles crossing the LX daily;	826.8	1.8e+03
Annual road accidents	The number of road accidents in a given year;	7.1e+04	9.7e+03
LX characteristics			
Alignment	Horizontal road alignment shape: "straight", "curve," or "S";	N/A	N/A
Profile	Vertical road profile shape: "normal", "hump," or cavity";	N/A	N/A
Length	The entering road width;	9.7	3.9
Width	The distance that road vehicles need to cross through the LX;	5.5	1.4
Region	The region of the LX considered;	N/A	N/A

Table 1.
Statistical characterization of parameters considered.

where λ_{10Y} represents the annual accident frequency at a given SAL2 for a period of 10 years; F_{RAcc} is the road accident factor, which is a time-dependent variable and reflects the variation of annual road accidents as time advances; K is the coefficient of F_{RAcc}; V denotes the average daily road traffic; T denotes the average daily railway traffic; $I_{Profile}$ is the profile indicator and $C_{Profile}$ is the coefficient of $I_{Profile}$; I_{Align} is the alignment indicator and C_{Align} is the coefficient of I_{Align}; Wid is the LX width and C_{Wid} is the coefficient of Wid; $Leng$ is the crossing length and C_{Leng} is the coefficient of $Leng$; RSL is the railway speed limit and C_{RSL} is the coefficient of RSL; F_{Reg} is the region factor and C_{Reg} is the coefficient of F_{Reg}. Note that this model does not only rank risky LXs accurately but also allow for identifying significant parameters efficiently.

2.2.1 Regression approaches

In this section, several regression approaches are adopted to estimate the coefficients associated with the parameters of our model. The nonlinear least-squares (NLS) technique and Gauss-Newton algorithm [22] are firstly considered to estimate the variable coefficients in our model. Considering a fitting model function $y = f(x, \boldsymbol{\beta})$, where variable x depends on a vector of l parameters: $\boldsymbol{\beta} = (\beta_1, \ \beta_2, \ \ldots, \ \beta_l)$. The goal is to find the vector $\boldsymbol{\beta}$ which can let the model function fit best the actual observed data in the least-squares sense. In other words, minimize the sum of residual squares S expressed as follows:

Parameter	Data coding
Railway traffic characteristics	
Average daily railway traffic	Numerical, used directly;
Railway speed limit	Numerical, used directly;
Roadway traffic characteristics	
Average daily road traffic	Numerical, used directly;
Annual road accidents	Road accident factor: *Annual road accidents in a given year/Average road accidents per year over the period observed*;
LX characteristics	
Alignment	Alignment indicator: 0, 1, and 2 represent "straight", "curve," and "S," respectively;
Profile	Profile indicator: 0 and 1 represent "normal" and "hump or cavity," respectively;
LX width	Numerical, used directly;
Crossing length	Numerical, used directly;
Region	Region risk factor, highlighting the general LX-accident-prone region: *The number of SAL2 accidents over the observation period in the region considered/ The number of SAL2 LXs in the region considered*;

Table 2.
Parameters considered and data coding.

$$S = \sum_{i=1}^{m} r_i^2, \quad m \geq l, \tag{4}$$

where r_i is the residual between the fitting model estimation and the actual observation, $r_i = y_i - f(x_i, \boldsymbol{\beta})$.

The minimum value of S is obtained by solving the gradient function $\partial S/\partial \beta_j = 0$, i.e.,

$$\begin{aligned} &\partial S/\partial \beta_j = 2\sum_i r_i \partial r_i/\partial \beta_j = 0, \\ &\beta_j \approx \beta_j^{k+1} = \beta_j^k + \Delta\beta_j, \end{aligned} \tag{5}$$

where k is the iteration number and $\Delta\beta_j$ is the shift parameter.

At each iteration step, the model is linearized by approximation to the first-order Taylor series expansion about $\boldsymbol{\beta}^k$:

$$f(x_i, \boldsymbol{\beta}) \approx f\left(x_i, \boldsymbol{\beta}^k\right) + \sum_{j=1}^{l} \left(\beta_j - \beta_j^k\right) \partial f\left(x_i, \boldsymbol{\beta}^k\right)/\partial \beta_j \approx f\left(x_i, \boldsymbol{\beta}^k\right) + \sum_{j=1}^{l} J_{ij} \Delta\beta_j, \tag{6}$$

where J_{ij} is the element of Jacobian matrix $\mathbf{J}$ and $\partial r_i/\partial \beta_j = -J_{ij}$.

Therefore, r_i can be rewritten as:

$$r_i = \Delta y_i - \sum_{s=1}^{l} J_{is} \Delta \beta_s, \\ \Delta y_i = y_i - f\left(x_i, \boldsymbol{\beta}^k\right). \tag{7}$$

By substituting the above expressions into the gradient equation in Eq. (5), we obtain the normal equation and its matrix notation:

$$\sum_{i=1}^{m} \sum_{s=1}^{l} J_{ij} J_{is} \Delta \beta_s = \sum_{i=1}^{m} J_{ij} \Delta y_i, \\ \left(\boldsymbol{J}^T \boldsymbol{J}\right) \Delta \boldsymbol{\beta} = \boldsymbol{J}^T \Delta \boldsymbol{y}. \tag{8}$$

For an NLS model, S should be modified as follows:

$$S = \sum_{i=1}^{m} W_{ii} r_i^2, \quad m \geq l. \tag{9}$$

Therefore, the matrix notation of normal equation for an NLS model is expressed as follows:

$$\left(\boldsymbol{J}^T \boldsymbol{W} \boldsymbol{J}\right) \Delta \boldsymbol{\beta} = \boldsymbol{J}^T \boldsymbol{W} \Delta \boldsymbol{y}. \tag{10}$$

These aforementioned equations form the basis of the Gauss-Newton algorithm for solving an NLS problem.

In fact, the Poisson regression model shown as Eq. (11) is a natural choice for modeling accident occurrence:

$$Poi(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0,1,2, \dots, \tag{11}$$

where $Poi(X = k)$ is the probability of k accidents occurring, $k \in \mathbb{N}$, and λ is the expectation value of the number of accidents.

However, [23] indicates that accident frequency is likely to be over-dispersed (see Eq. (12)) and suggests using the negative binomial (NB) regression model as an alternative to the Poisson model:

$$VAR(X) \begin{cases} = E(X) \\ > E(X), & \text{for over-dispersed} \\ < E(X), & \text{for under-dispersed} \end{cases}. \tag{12}$$

The NB model as a special case of Poisson-Gamma mixture model is a variant of the Poisson model designed to deal with over-dispersed data [11, 24, 25]. The over-dispersion could come from several possible sources, e.g., omitted variables, uncertainty in exposure data, covariates, or nonhomogeneous LX environment [26]. The NB model considered in this study has the following expression:

$$P_{NB}(X = k) = \frac{\Gamma\left(k + \frac{1}{\alpha}\right)}{\Gamma(k+1)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{1}{1 + \alpha\lambda}\right)^{1/\alpha} \left(\frac{\alpha\lambda}{1 + \alpha\lambda}\right)^k, \quad k = 0,1,2, \dots, \tag{13}$$

where $P_{NB}(X)$ is the probability of k accidents occurring, $k \in \mathbb{N}$, α is the dispersion parameter, and λ is the expectation of the number of accidents.

The relationship between the mean value and the variance in the NB model is given as follows:

$$VAR(X) = \alpha E(X)^2 + E(X), \tag{14}$$

if $\alpha < 0$, there is an under-dispersion; if $\alpha > 0$, there is an over-dispersion; in the case where $\alpha = 0$, the NB model reduces to the Poisson model.

In practice, the count data may contain extra zeros relative to the Poisson or NB distribution. In this case, the ZIP or ZINB regression model is useful for analyzing such data [27]. The ZIP model is expressed as follows:

$$P_{ZIP}(X = k) = \begin{cases} \omega + (1-\omega)\exp(-\lambda), & \text{for } k = 0 \\ (1-\omega)\exp(-\lambda)\lambda^k/k!, & \text{for } k > 0 \end{cases}, \tag{15}$$

where $P_{ZIP}(X = k)$ is the probability of k accidents occurring, $k \in \mathbb{N}$, λ is the expectation value of the number of accidents, and $log\left(\frac{\omega}{1-\omega}\right) = z'\gamma$ is the ZI link function that z' is the ZI covariate and γ is the corresponding ZI coefficient. The mean value and variance of ZIP model are $E(X) = (1-\omega)\lambda$ and $VAR(X) = (1-\omega)\lambda(1+\omega\lambda)$.

The ZINB model is expressed as follows:

$$P_{ZINB}(X = k) = \begin{cases} \omega + (1-\omega)(1+\alpha\lambda)^{-1/\alpha}, & \text{for } k = 0 \\ (1-\omega)\dfrac{\Gamma\left(k+\frac{1}{\alpha}\right)}{\Gamma(k+1)\Gamma\left(\frac{1}{\alpha}\right)}\left(\dfrac{1}{1+\alpha\lambda}\right)^{1/\alpha}\left(\dfrac{\alpha\lambda}{1+\alpha\lambda}\right)^k, & \text{for } k > 0 \end{cases}, \tag{16}$$

where $P_{ZINB}(X = k)$ is the probability of k accidents occurring, $k \in \mathbb{N}$ and λ is the expectation value of the number of accidents. The mean value and variance of ZINB model are $E(X) = (1-\omega)\lambda$ and $VAR(X) = (1-\omega)\lambda(1+\omega\lambda+\alpha\lambda)$. The ZINB reduces to the ZIP in the limit $\alpha \to 0$.

However, the NB and ZINB models are limited to handling under-dispersed data ($\alpha < 0$) [11]. That is why [13] proposed the Gamma model to handle under-dispersed samples. The Gamma model is given as follows:

$$P_G(X = k) = Gamma(\beta k, \lambda) - Gamma(\beta(k+1), \lambda), \tag{17}$$

where $P_G(X)$ is the probability of k accidents occurring, $k \in \mathbb{N}$, λ is the expectation of the number of accidents, and β is the dispersion parameter. If $\beta > 1$, there is an under-dispersion; while $\beta < 1$, there is an over-dispersion and if $\beta = 1$, the Gamma model reduces to the Poisson model. However, the Gamma model shown in Eq. (18) is limited to the time-dependent observation assumption and zero observations, since general $\Gamma(x)$ restricts discrete responses to positive values:

$$Gamma(\beta k, \lambda) = \begin{cases} 1, & \text{for } k = 0 \\ \dfrac{1}{\Gamma(\beta k)\displaystyle\int_0^{\lambda} u^{\beta k - 1} e^{-u} \mathrm{d}u}, & \text{for } k > 0 \end{cases}. \tag{18}$$

According to the above discussion, the restriction between mean value and variance can be used to identify an appropriate regression model. Therefore, we firstly make preliminary variance analysis by means of group classification. Namely, the annual accidents at a given SAL2 during the 10 years were divided into 100 groups with the same number of samples in each group. Then, the variance and mean value of accidents in each group were calculated, respectively, to analyze the relationship between the group variance and the group mean value. The variance analysis shows that the variance and mean value are very close to each other. Hence, we performed meticulous analyses to assess the NLS regression, the Poisson regression, the ZIP regression, the NB regression, and the ZINB regression methods with regard to SAL2 LXs in our accident dataset so as to identify which model is more effective.

2.2.2 Regression modeling results

NLS regression:

When applying the NLS regression, the form of λ_{10Y} is given by Eq. (3). The estimated coefficients computed by NLS regression are provided in **Table 3**. $|t - statistic| > 1.96$ is introduced to identify the significant parameters corresponding to a 95% confidence level. As a result, the railway speed limit, the average daily railway traffic, the average daily road traffic, the annual road accidents, the LX-accident-prone region, the road alignment, the LX width, and the crossing length have been shown to have significant and positive influence on SAL2 accident frequency. However, the test shows that the road profile is not a significant factor ($|t - statistic| = 0.635 < < 1.96$); thus, the impact of road profile could be neglected. Moreover, the coefficients of the considered variables with the exponential form can reflect the sensitive degrees of the SAL2 accident frequency to these variables, respectively. According to these sensitive degrees (rank indicated in brackets), the LX-accident-prone region factor is the most sensitive contributor among these variables.

In order to assess the predictive accuracy of accident occurrence estimated by the NLS regression model λ_{10Y} combined with the NB and ZINB distributions (see Section 3.1), we adopt the maximum likelihood estimation (MLE) method to estimate the dispersion parameter α of the dataset [28]. As expressed by Eq. (19) and Eq. (20), the values of α in NB and ZINB distributions are estimated, respectively, using R language to solve $\partial l / \partial \alpha = 0$:

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K	2.703e-05	5.078e-06	5.322	×
$I_{Profile}$	$C_{Profile}$	3.626e-02	5.706e-02	0.635	
I_{Align}	C_{Align}	3.427e-01 (2)	2.942e-02	11.648	×
Wid	C_{Wid}	9.847e-02 (3)	1.494e-02	6.589	×
$Leng$	C_{Leng}	2.084e-02 (4)	4.284e-03	4.865	×
RSL	C_{RSL}	3.089e-03 (5)	7.586e-04	4.072	×
F_{Reg}	C_{Reg}	4.962e-01 (1)	1.722e-01	2.882	×

Table 3.
Results of the λ_{10Y} NLS regression model.

$$l(\alpha)_{NB} = ln\left(\prod_{i}^{n} P_{NB}(X_i = y_i)\right) = \sum\left(y_i\, ln\,(\lambda_i) - (y_i + \alpha^{-1})\, ln\,(1+\alpha\lambda_i) + \sum_{v=0}^{y_i-1} ln\,(1+\alpha v)\right), \quad (19)$$

$$l(\alpha)_{ZINB} = ln\left(\prod_{i}^{n} P_{ZINB}(X_i = y_i)\right) = \begin{cases} \sum ln\,(\omega_i) + (1-\omega_i)\left(\frac{1}{1+\alpha\lambda_i}\right)^{1/\alpha}, \text{if} y_i = 0 \\ \sum ln\,(\omega_i) + ln\,\Gamma\left(\frac{1}{\alpha} + y_i\right) - ln\,\Gamma(1+y_i) - ln\,\Gamma\left(\frac{1}{\alpha}\right). \\ +\frac{1}{\alpha}\, ln\left(\frac{1}{1+\alpha\lambda_i}\right) + y_i\, ln\left(1 - \frac{1}{1+\alpha\lambda_i}\right), \text{if} y_i > 0 \end{cases} \quad (20)$$

Poisson regression:

When applying the Poisson regression, the general form of λ_{10Poi} is given by $e^{\sum_{j=1}^{m}\beta_0+\beta_j x_j}$. Therefore, we need to transform Eq. (3) into the following expression:

$$\lambda_{10Poi} = \begin{cases} 0, \text{if } F_{RAcc} = 0, \; V = 0 \text{ or } T = 0 \\ \exp(K_1 + C_F \times F_{RAcc} + C_{CM} \times CM + C_{Profile} \times I_{Profile} + C_{Align} \times I_{Align} + \\ C_{Wid} \times Wid + C_{Leng} \times Leng + C_{RSL} \times RSL + C_{Reg} \times F_{Reg}), \text{if} F_{RAcc} \neq 0, \\ V \neq 0, \text{and } T \neq 0 \end{cases} \quad (21)$$

The results estimated through the Poisson regression approach are shown in **Table 4**. According to these results, being similar to the NLS case, one can notice that the road profile is not significant ($|t - statistic| = 0.621 < < 1.96$). On the other hand, with an exponential form, the impact of road accident factor F_{RAcc} is weakened, namely the impact of F_{RAcc} with an exponential form is not significant when using Poisson regression approach ($|t - statistic| = 1.913 < 1.96$). Furthermore, according to

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K_1	−9.562	0.440	−21.714	×
F_{RAcc}	C_F	0.636	0.332	1.913	
CM	C_{CM}	0.005 (6)	2.949e-04	17.144	×
$I_{Profile}$	$C_{Profile}$	−0.076	0.122	−0.621	
I_{Align}	C_{Align}	0.326 (2)	0.069	4.756	×
Wid	C_{Wid}	0.206 (3)	0.026	8.051	×
$Leng$	C_{Leng}	0.030 (4)	0.009	3.232	×
RSL	C_{RSL}	0.011 (5)	0.001	7.895	×
F_{Reg}	C_{Reg}	1.725 (1)	0.334	5.165	×

Table 4.
Regression results of λ_{10Poi}.

the sensitive degrees of these parameters with the exponential form (rank indicated in brackets), once again the LX-accident-prone region factor is the most sensitive contributor among these parameters.

NB regression:

When applying the NB regression, the general form of λ_{10NB} is given by $e^{\sum_{j=1}^{m}\beta_0+\beta_j x_j+\varepsilon}$, and it still requires to be expressed by Eq. (21). The dispersion parameter α is estimated at 3.2394 in our study through the iterative estimation algorithm automatically. The estimated results of the NB regression are shown in **Table 5**. According to the results associated with the NB regression approach, it is worth noticing that the road profile is still not significant ($|t-statistic| = 0.850 < < 1.96$). One can also notice that the impact of F_{RAcc} with an exponential form is not significant as well, when using the NB regression approach ($|t-statistic| = 1.793 < 1.96$). Moreover, according to the sensitive degrees of these parameters with the exponential form (rank indicated in brackets), the LX-accident-prone region factor is still the most sensitive contributor among these parameters.

ZIP regression:

When applying the ZIP regression, the general form of λ_{10ZIP} is given by $e^{\sum_{j=1}^{m}\beta_0+\beta_j x_j}$, and it still requires to be expressed by Eq. (21). The estimated results of the ZIP regression are shown in **Table 6** and (for nonzero observations) and **Table 7** (for zero-inflation observations).

According to the results associated with the ZIP regression approach, it is worth noticing that, as for the nonzero related model, F_{RAcc}, $I_{Profile}$, I_{Align}, and $Leng$ are not significant (<1.96). Moreover, according to the sensitive degrees of other significant parameters with the exponential form (rank indicated in brackets), the LX-accident-prone region factor is still the most sensitive contributor among these parameters. While as for the zero-inflation model, only the Wid, RSL, and F_{Reg} are significant (>1.96).

ZINB regression:

When applying the ZINB regression, the general form of λ_{10ZINB} is given by $e^{\sum_{j=1}^{m}\beta_0+\beta_j x_j+\varepsilon}$, and it still requires to be expressed by Eq. (21). The values of dispersion parameter α for nonzero observations and zero-inflation observations are estimated at

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K_1	−9.424	0.457	−20.615	×
F_{RAcc}	C_F	0.616	0.343	1.793	
CM	C_{CM}	0.006 (6)	3.762e-04	16.493	×
$I_{Profile}$	$C_{Profile}$	−0.107	0.126	−0.850	
I_{Align}	C_{Align}	0.298 (2)	0.072	4.159	×
Wid	C_{Wid}	0.199 (3)	0.028	7.173	×
$Leng$	C_{Leng}	0.031 (4)	0.010	3.201	×
RSL	C_{RSL}	0.010 (5)	0.001	7.034	×
F_{Reg}	C_{Reg}	1.508 (1)	0.351	4.294	×

Table 5.
Regression results of λ_{10NB}.

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K_1	−1.128e+01	7.586e-01	−14.867	×
F_{RAcc}	C_F	3.717e-01	4.202e-01	0.885	
CM	C_{CM}	6.221e-03 (4)	4.336e-04	14.347	×
$I_{Profile}$	$C_{Profile}$	−1.855e-01	1.513e-01	−1.226	
I_{Align}	C_{Align}	1.483e-01	8.786e-02	1.688	
Wid	C_{Wid}	4.397e-01 (2)	6.625e-02	6.636	×
$Leng$	C_{Leng}	3.971e-02	1.725e-02	1.904	
RSL	C_{RSL}	1.432e-02 (3)	2.069e-03	6.921	×
F_{Reg}	C_{Reg}	2.319 (1)	6.655e-01	3.484	×

Table 6.
Count model regression results of λ_{10ZIP}.

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K_1	−1.574e+01	4.276	−3.680	×
F_{RAcc}	C_F	−1.104	1.646	−0.671	
CM	C_{CM}	1.584e-03	1.450e-03	1.093	
$I_{Profile}$	$C_{Profile}$	−4.355e-01	6.531e-01	0.505	
I_{Align}	C_{Align}	−1.185	6.141e-01	−1.931	
Wid	C_{Wid}	1.024 (2)	2.241e-01	4.571	×
$Leng$	C_{Leng}	8.231e-02	4.190e-02	1.964	
RSL	C_{RSL}	4.117e-02 (3)	1.449e-02	2.840	×
F_{Reg}	C_{Reg}	5.861 (1)	1.748	3.353	×

Table 7.
Zero-inflation model regression results of λ_{10ZIP}.

3.8102 and 1.4069, respectively, in our study through the iterative estimation algorithm automatically. The estimated results of the ZINB regression are shown in **Table 8** (for nonzero observations) and **Table 9** (for zero-inflation observations). According to the results associated with the ZINB regression approach, it is worth noticing that, as for the nonzero related model, CM, I_{Align}, and Wid are significant (>1.96). One can also notice that according to the sensitive degrees of the three parameters (rank indicated in brackets), the LX width is the most sensitive contributor among them. While as for the zero-inflation model, only the F_{RAcc} and CM are significant (>1.96).

3. Model performance evaluation and discussion

In this section, we will assess the performance of our prediction models while determining an appropriate statistical distribution to be combined with the models, in such a way as to ensure the most accurate estimation of the probability of accidents

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K_1	−7.128	0.734	−9.709	×
F_{RAcc}	C_F	0.671	0.413	1.624	
CM	C_{CM}	4.486e-03 (3)	4.991e-04	8.990	×
$I_{Profile}$	$C_{Profile}$	−5.886e-02	0.144	−0.406	
I_{Align}	C_{Align}	0.371 (1)	8.274e-02	4.495	×
Wid	C_{Wid}	0.145 (2)	4.558e-02	3.175	×
$Leng$	C_{Leng}	3.219e-03	1.203e-02	0.268	
RSL	C_{RSL}	2.558e-03	1.954e-03	1.309	
F_{Reg}	C_{Reg}	0.795	0.446	1.783	

Table 8.
Count model regression results of λ_{10ZINB}.

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K_1	−4.036	2.190	−6.709	×
F_{RAcc}	C_F	0.260 (1)	1.456	2.179	×
CM	C_{CM}	6.685e-02 (2)	1.838e-02	3.636	×
$I_{Profile}$	$C_{Profile}$	0.705	0.544	1.296	
I_{Align}	C_{Align}	0.535	0.328	1.632	
Wid	C_{Wid}	8.873e-02	0.180	0.491	
$Leng$	C_{Leng}	0.114	6.639e-02	1.725	
RSL	C_{RSL}	5456e-03	6.629e-03	0.823	
F_{Reg}	C_{Reg}	1.632	1.679	0.972	

Table 9.
Zero-inflation model regression results of λ_{10ZINB}.

occurring at a given SAL2 in a given year. The Bayesian information criterion (BIC) [29], Akaike's information criterion (AIC) [30], the Pearson chi-square statistic (PCS) test [31], and the degree of freedom (DF) are used to evaluate the goodness of fit (GOF) of the model. They can be respectively expressed as follows:

$$\mathrm{BIC} = n + n \times \ln(2\pi) + n \times \ln(\mathrm{RSS}/n) + (l+1)\ln(n), \tag{22}$$

$$\mathrm{AIC} = n + n \times \ln(2\pi) + n \times \ln(\mathrm{RSS}/n) + 2(l+1), \tag{23}$$

$$\mathrm{PCS} = \sum_{i=1}^{n} \frac{(O_i - \lambda_i)^2}{\lambda_i}, \tag{24}$$

$$\mathrm{DF} = n - (l+1), \tag{25}$$

where RSS is the sum of the squares of residuals between the annual accident frequencies observed and the annual accident frequencies estimated, n is the sample size, l is the number of independent exponential parameters, λ_i is the annual accident frequency expected, and O_i is the annual accident frequency observed.

The BIC and AIC are used to test the relative quality of models for a given dataset. Smaller BIC and AIC values indicate a better model fitting. The PCS test is used to determine if there is a significant difference between the values expected and the values observed. The PCS is roughly equal to DF if the model fits the data perfectly without any dispersion. Namely, the closer the PCS is to the DF, the better the model fits the data [14].

The log-likelihood statistic test (LL) is adopted to assess the GOF of the accident frequency prediction model combined with a statistical distribution. The larger the LL, the more preferred the model [14]. The mathematical expression of the LL is given as follows:

$$\text{LL} = \sum_{i=1}^{n} ln\left(\hat{P}_i\right), \tag{26}$$

where n is the sample size and $\hat{P}_i$ is the estimated probability of accident frequency observed. $\hat{P}_i$ is computed respectively according to the accident frequency prediction model combined with the Poisson or the NB distribution.

3.1 Model performance comparison among variants of λ_{10Y}

The results of AIC, BIC, and PCS statistical tests are shown in **Table 10** with the goodness ranked in brackets. The following findings are obtained: 1) considering AIC and BIC, the λ_{10Y} model gives better results, since the AIC and BIC values corresponding to the λ_{10Y} model are much smaller than those for the λ_{10Poi}, λ_{10NB}, λ_{10ZIP}, and λ_{10ZINB} models; 2) in terms of PCS test, the λ_{10Y} model is also the most effective one, since the PCS of λ_{10Y} model is closer to DF (DFs of λ_{10Y}, λ_{10Poi}, λ_{10NB}, λ_{10ZIP}, and λ_{10ZINB} are considerably approximative).

LL test results are shown in **Table 10**. One can notice that, for the λ_{10Y} model combined with either the Poisson or NB distribution, its GOFs are significantly better than λ_{10Poi} and λ_{10NB} models' GOFs according to the LL test. Furthermore, the GOF of λ_{10Y} combined with the NB distribution (NB-λ_{10Y}) is better than when combined with the Poisson distribution (POI-λ_{10Y}).

Test	**POI-λ_{10Y}**	**NB-λ_{10Y}**	$\lambda_{10\text{Poi}}$	$\lambda_{10\text{NB}}$	$\lambda_{10\text{ZIP}}$	$\lambda_{10\text{ZINB}}$
AIC	−190,744 (1)	−190,744 (1)	−187,804 (5)	−189,942 (2)	−188,312 (4)	−189,826 (3)
BIC	−190,670 (1)	−190,670 (1)	−187,720 (5)	−189,858 (3)	−188,176 (4)	−189,935 (2)
PCS	65,796 (1)	65,796 (1)	125,495 (5)	123,715 (4)	118,185 (3)	110,496 (2)
DF	83,313	83,313	83,311	83,311	83,311	83,311
LL	−2599 (2)	−2596 (1)	−2732 (6)	−2711 (5)	−2701 (4)	−2631 (3)
Goodness score						
(the lower, the better)	5	4	21	14	15	10

Table 10.
Model GOF comparison among variants of λ_{10Y}.

3.2 A comparison between λ_{10Y} and two existing reference models

In this section, we compare the present model λ_{10Y} with other two models which are widely used in existing related works. As mentioned in Section 1, the first widely used model is given in Eq. (1) [13, 14, 18]. In our study, this model can be specified as follows:

$$\lambda_{TV} = \begin{array}{l} \exp\Big(K_2 + C_V \times V + C_T \times T + C_F \times F_{RAcc} + C_{Profile} \times I_{Profile} + C_{Align} \\ \times I_{Align} + C_{Wid} \times Wid + C_{Leng} \times Leng + C_{RSL} \times RSL + C_{Reg} \times F_{Reg}\Big), \end{array} \tag{27}$$

where the average daily road traffic V and the average daily railway traffic T are applied separately in exponential form.

The second model as shown in Eq. (2) (e.g., [17, 32]) is specified as Eq. (28) in our study:

$$\lambda_{Mon} = \begin{array}{l} \exp\Big(K_3 + C_M \times ln\,(V \times T) + C_F \times F_{RAcc} + C_{Profile} \times I_{Profile} + C_{Align} \\ \times I_{Align} + C_{Wid} \times Wid + C_{Leng} \times Leng + C_{RSL} \times RSL + C_{Reg} \times F_{Reg}\Big), \end{array} \tag{28}$$

where the conventional traffic moment $V \times T$ is applied.

It should be noted that the ZIP and ZINB models were also investigated for λ_{TV} and λ_{Mon} but resulted in no higher goodness-of-fit values and a quite small number of significant parameters compared with the Poisson and NB models and, hence, were not reported in this section. The Poisson and NB regression results of the λ_{TV} and λ_{Mon} are shown in **Tables 11–14**, respectively. One can notice that the impacts of road profile and road accident are still not significant in the λ_{TV} and λ_{Mon}. The AIC, BIC, PCS, and LL tests and observed/estimated accident frequency comparison are given in **Table 15**. According to the quality test results discussed in Section 3.1, the λ_{10Y} combined with the NB distribution (NB-λ_{10Y}) shows the best prediction performance among the four investigated combinations. Therefore, we will only compare the NB-

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K_2	−9.807	0.413	−22.223	×
V	C_V	1.098e-04 (7)	1.613e-05	6.811	×
T	C_T	8.777e-03 (6)	1.115e-03	7.869	×
F_{RAcc}	C_F	0.636	0.333	1.913	
$I_{Profile}$	$C_{Profile}$	−1.445e-01	1.209e-01	−1.195	
I_{Align}	C_{Align}	3.319e-01 (2)	6.747e-02	4.919	×
Wid	C_{Wid}	2.059e-01 (3)	2.483e-02	8.292	×
$Leng$	C_{Leng}	3.952e-02 (4)	7.868e-03	5.024	×
RSL	C_{RSL}	1.154e-02 (5)	1.487e-03	7.759	×
F_{Reg}	C_{Reg}	1.750 (1)	3.463e-01	5.053	×

Table 11.
Poisson regression results of λ_{TV}.

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K_2	−9.882	4.531e-01	−21.810	×
V	C_V	1.155e-04 (7)	1.683e-05	6.861	×
T	C_T	9.152e-03 (6)	1.234e-03	7.416	×
F_{RAcc}	C_F	0.607	3.402e-01	1.784	
$I_{Profile}$	$C_{Profile}$	−1.532e-01	1.243e-01	−1.232	
I_{Align}	C_{Align}	3.240e-01 (2)	6.988e-02	4.636	×
Wid	C_{Wid}	2.212e-01 (3)	2.579e-02	8.575	×
$Leng$	C_{Leng}	3.895e-02 (4)	8.415e-03	4.629	×
RSL	C_{RSL}	1.160e-02 (5)	1.529e-03	7.589	×
F_{Reg}	C_{Reg}	1.739 (1)	3.575e-01	4.864	×

Table 12.
NB regression results of λ_{TV}.

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K_2	−11.816	4.540e-01	−26.023	×
$ln\,(V \times T)$	C_M	4.036e-01 (2)	2.776e-02	14.538	×
F_{RAcc}	C_F	6.359e-01	3.325e-01	1.913	
$I_{Profile}$	$C_{Profile}$	−6.279e-02	1.205e-01	−0.521	
I_{Align}	C_{Align}	2.875e-01 (3)	6.799e-02	4.228	×
Wid	C_{Wid}	1.185e-01 (4)	3.296e-02	3.596	×
$Leng$	C_{Leng}	2.213e-02 (5)	9.530e-03	2.322	×
RSL	C_{RSL}	8.811e-03 (6)	1.350e-03	6.527	×
F_{Reg}	C_{Reg}	1.446 (1)	3.358e-01	4.307	×

Table 13.
Poisson regression results of λ_{Mon}.

Parameter	Coefficient	Estimated value	Standard error	t-statistic	Significant
	K_2	−11.850	4.628e-01	−26.603	×
$ln\,(V \times T)$	C_M	4.034e-01 (2)	2.822e-02	14.297	×
F_{RAcc}	C_F	6.368e-01	3.382e-01	1.883	
$I_{Profile}$	$C_{Profile}$	−7.103e-02	1.230e-01	−0.578	
I_{Align}	C_{Align}	2.848e-01 (3)	6.960e-02	4.092	×
Wid	C_{Wid}	1.214e-01 (4)	3.361e-02	3.612	×
$Leng$	C_{Leng}	2.204e-02 (5)	9.752e-03	2.260	×
RSL	C_{RSL}	8.892e-03 (6)	1.368e-03	6.500	×
F_{Reg}	C_{Reg}	1.480 (1)	3.428e-01	4.316	×

Table 14.
NB regression results of λ_{Mon}.

Test	NB-λ_{10Y}	POI-λ_{TV}	NB-λ_{TV}	POI-λ_{Mon}	NB-λ_{Mon}
AIC	−190,744 (1)	−177,914 (5)	−179,842 (4)	−183,714 (3)	−186,532 (2)
BIC	−190,670 (1)	−177,610 (5)	−179,738 (4)	−183,587 (3)	−186,191 (2)
PCS	65,796 (1)	121,715 (5)	119,133 (4)	118,511 (3)	115,634 (2)
DF	83,313	83,310	83,310	83,311	83,311
LL	−2596 (1)	−2722 (5)	−2703 (3)	−2705 (4)	−2683 (2)
Goodness score					
(the lower, the better)	4	20	15	13	8

Table 15.
Model GOF comparison among λ_{10Y}, λ_{TV}, and λ_{Mon}.

λ_{10Y} with the λ_{TV} and λ_{Mon} combined with the Poisson and NB distributions, respectively, in the following content.

As shown in **Table 15**, the AIC, BIC, and PCS results related to the λ_{10Y} model are better than those for the λ_{TV} and λ_{Mon} models. Moreover, in terms of the LL test, the NB-λ_{10Y} is still the most preferred one.

4. Conclusions

Based on our study, some remarks need to be highlighted as follows:

1. The corrected traffic moment proposed is more effective in estimating automobile-involved LX accidents frequency compared with the conventional traffic moment, single average daily railway traffic or single average daily road traffic. It is worth mentioning that the average daily railway traffic with a power of 0.646 has a more decisive impact on the LX accident frequency than the average daily road traffic with a power of 0.354. Moreover, the higher the combined exposure of railway and roadway traffic, the higher the likelihood of an accident occurring.

2. According to the analyses above, the form of λ_{10Y} highlights the impact of road accident factor F_{RAcc}, while the impact of F_{RAcc} is neglected in λ_{10Poi}, λ_{10NB}, λ_{TV}, and λ_{Mon} models (see **Tables 4, 5, 11–14**). The impact of road accidents on the risk level was likely to be ignored in the previous studies related to LX safety analysis.

3. We originally introduce the region LX-accident-prone factor (see **Table 2**) in this study to interpret the variation of LX accident statistics with regard to various regions. According to the sensitive degrees of variables ranked in **Table 3**, among the LX characteristics, the risk of LX accidents is most sensitive to the region LX-accident-prone factor. However, in many past studies, the impact of LX local region is neglected. In fact, the regional accident history varies from one region to another, which correspondingly has varying degrees of impact on the LX accident frequency in different regions.

To sum up, the develop model λ_{10Y} has trustworthy goodness of fit. Moreover, it shows relatively high prediction accuracy for LX accident frequency prediction when combined with the NB distribution.

Author details

Ci Liang[1,2]* and Mohamed Ghazel[2]

1 Harbin Institute of Technology, Harbin, China

2 Université Gustave Eiffel (ex-IFSTTAR) - COSYS/ESTAS, Villeneuve-d'Ascq, France

*Address all correspondence to: ciliang.lc@gmail.com

References

[1] Read GJM, Salmon PM, Lenné MG, Stanton NA. Walking the line: Understanding pedestrian behaviour and risk at rail level crossings with cognitive work analysis. Applied Ergonomics. 2016;**53**:209-227

[2] Ghazel M. Using stochastic petri nets for level-crossing collision risk assessment. IEEE Transactions on Intelligent Transportation Systems. 2009;**10**(4):668-677

[3] Liu B, Ghazel M, Toguyéni A. Model-based diagnosis of multi-track level crossing plants. IEEE Transactions on Intelligent Transportation Systems. 2016;**17**(2):546-556

[4] ERA. Railway safety performance in the European Union. 9(2) Agency Regulation 881/2004/EC. 2014

[5] Ghazel M, El-Koursi EM. Two-half-barrier level crossings versus four-half-barrier level crossings: A comparative risk analysis study. IEEE Transactions on Intelligent Transportation Systems. 2014;**15**(3):1123-1133

[6] Silmon J, Roberts C. Using functional analysis to determine the requirements for changes to critical systems: Railway level crossing case study. Reliability Engineering and System Safety. 2010; **95**(3):216-225

[7] Australian Transport Safety Bureau. Australian Rail Safety Occurrence Data: 1 July 2002 to 30 June 2012 (ATSB Transport Safety Report RR-2012-010). Canberra, Australia: ATSB; 2012

[8] SNCF Réseau. World Conference of Road Safety at Level Crossings. 2011. Available from: http://www.planetoscope.com/automobile/1271-nombre-de-collisions-aux- passages-a-niveau-en-france.html

[9] Plesse G. Des détecteurs d'obstacles déployés aux passages à niveau. 2017. Available from: http://www.leparisien.fr/info-paris-ile-de-france-oise/transports/des-detecteurs-d-obstacles-deployes-aux-passages-a-niveau-02-06-2017-7011714.php

[10] Liang C, Ghazel M, Cazier O, El-Koursi EM. Analyzing risky behavior of motorists during the closure cycle of railway level crossings. Safety Science. 2018;**110**:115-126

[11] Lord D, Mannering F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice. 2010;**44**(5):291-305

[12] Guikema SD, Quiring SM. Hybrid data mining-regression for infrastructure risk assessment based on zero-inflated data. Reliability Engineering and System Safety. 2012;**99**: 178-182

[13] Oh J, Washington SP, Nam D. Accident prediction model for railway-highway interfaces. Accident; Analysis and Prevention. 2006;**38**(2): 346-356

[14] Lu P, Tolliver D. Accident Analysis & Prevention. Accident prediction model for public highway-rail grade crossings. 2016;**90**:3-81

[15] Medina JC, Benekohal RF. Macroscopic models for accident prediction at railroad grade crossings: Comparisons with US Department of Transportation accident prediction formula. Transportation Research

Record: Journal of the Transportation Research Board. 2015;**2476**:85-93

[16] Chadwick SG, Zhou N, Saat MR. Highway-rail grade crossing safety challenges for shared operations of high-speed passenger and heavy freight rail in the US. Safety Science. 2014;**68**:128-137

[17] Miranda-Moreno L, Fu L, Saccomanno FF, Labbe A. Alternative risk models for ranking locations for safety improvement. Transportation Research Record: Journal of the Transportation Research Board. 2005; **1908**:1-8

[18] Austin RD, Carson JL. An alternative accident prediction model for highway-rail interfaces. Accident; Analysis and Prevention. 2002;**34**(1):31-42

[19] Liang C, Ghazel M, Cazier O, El-Koursi EM. Risk analysis on level crossings using a causal Bayesian network based approach. Transportation Research Procedia. 2017; **25**:2172-2186

[20] Liang C, Ghazel M, Cazier O, El-Koursi EM. Developing accident prediction model for railway level crossings. Safety Science. 2018;**101**:48-59

[21] SNCF Réseau. SNCF. Statistical Analysis of Accidents at LXs. France: SNCF Réseau; 2010

[22] Madsen K, Nielsen HB, Tingleff O. Methods for non-linear least squares problems. In: Informatics and Mathematical Modelling. 2nd ed. Denmark: Technical University of Denmark; 2004

[23] Chang LY. Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. Safety Science. 2005;**43**(8): 541-557

[24] Buddhavarapu P, Scott JG, Prozzi JA. Modeling unobserved heterogeneity using finite mixture random parameters for spatially correlated discrete count data. Transportation Research Part B: Methodological. 2016;**91**:492-510

[25] Utkin LV, Coolen FPA, Gurov SV. Imprecise inference for warranty contract analysis. Reliability Engineering and System Safety. 2015;**138**:31-39

[26] Miaou SP. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident; Analysis and Prevention. 1994;**26**(4):471-482

[27] Ridout M, Hinde J, DeméAtrio C. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. Biometrics. 2001;**57**(1):219-223

[28] Dai H, Bao Y, Bao M. Maximum likelihood estimate for the dispersion parameter of the negative binomial distribution. Statistics & Probability Letters. 2013;**83**(1):21-27

[29] Weakliem DL. A critique of the Bayesian information criterion for model selection. Sociological Methods & Research. 1999;**27**(3):359-397

[30] Bozdogan H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika. 1987;**52**(3): 345-370

[31] Pearson KX. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 1900; **50**(302):157-175

[32] Saccomanno FF, Fu L, Miranda-Moreno L. Risk-based model for identifying highway-rail grade crossing blackspots. Transportation Research Record: Journal of the Transportation Research Board. 2004;**1862**:127-135