Safety Science 101 (2018) 48–59

Contents lists available at ScienceDirect

Safety Science

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

# Developing accident prediction model for railway level crossings

Ci Liang[a,b,c,*], Mohamed Ghazel[b,a,c], Olivier Cazier[d,a], El-Miloudi El-Koursi[b,a,c]

[a] *FCS Railenium, Valenciennes, France*
[b] *IFSTTAR-COSYS/ESTAS, Lille-Villeneuve d'Ascq, France*
[c] *University Lille 1, Lille-Villeneuve d'Ascq, France*
[d] *SNCF Réseau, Paris, France*

ARTICLE INFO

*Keywords:*
Level crossing safety
Train-car collision
Accident prediction modeling
Statistical analysis

ABSTRACT

Railway level crossing (LX) safety continues to be one of the most critical issues for railways, despite an ever-increasing focus on improving design and application practices. Accidents at European LXs account for about one-third of the entire railway accidents and result in more than 300 deaths every year in Europe. Due to the non-deterministic causes, the complex operation background and the lack of thorough statistical analysis based on accident/incident data, the risk assessment of LXs remains a challenging task. In the present paper, some LX accident prediction models are developed. Such models allow for highlighting the influence of the main impacting parameters, i.e., 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 Ordinary Least-Squares (OLS) and Nonlinear Least-Squares (NLS) methods are employed to estimate the respective coefficients of variables in the prediction models, based on the LX accident/incident data. The validation and comparison process is performed through statistical means to examine how well the estimation of the models fits the reality. The outcomes of validation and comparison attest that the improved accident prediction model has statistic-based approbatory quality. Moreover, the improved accident prediction model combined with the NB distribution shows relatively high predictive accuracy of the probability of accident occurrence.

## 1. Context and related works

Accidents at railway level crossings (LXs) often give rise to serious material and human damage and hamper railway safety reputation, although the majority of accidents are caused by vehicle driver violations. LX safety is one of the most critical issues for railways which needs to be tackled urgently (Ghazel, 2009; Mekki et al., 2012; Liu et al., 2016). In 2012, there were more than 118,000 LXs in the 28 countries of the European Union (E.U.) which correspond to an average of 5 LXs per 10 line-km (ERA, 2014). Accidents at European LXs account for about one-third of the entire railway accidents. They result in more than 300 deaths every year in Europe (Liu et al., 2016). In some European countries, accidents at LXs account for up to 50% of railway accidents (Ghazel and El-Koursi, 2014; Evans, 2011b). In the entire E.U. zone, the overall number of deaths per fatal accident in railways from 1990 to 2009 is 4.10, with no apparent long-term change over time (Evans, 2011a). In France, the railway network shows more than 18,000 LXs for 30,000 km of railway lines, which are crossed daily by 16 million vehicles on average, and around 13,000 LXs show heavy road and railway traffic (SNCF Réseau, 2011). Despite numerous measures already taken to improve the LX safety, SNCF Réseau (the French national railway infrastructure manager) counted 100 collisions at LXs which led to 25 deaths in 2014. This number was half the total number of collisions per year at LXs a decade ago, but still too large (SNCF Réseau, 2015). In order to significantly reduce the accidents and their related consequences at LXs, it is crucial to establish a high quality accident prediction model and carry out a thorough analysis to understand the potential reasons for accidents occurring at LXs. Indeed, this paves the way for making appropriate safety diagnoses at LXs.

Many existing works dealing with LX safety are devoted to developing qualitative approaches, in order to understand the potential reasons causing accidents at LXs, such as surveys (Wigglesworth, 2001), interviews (Read et al., 2016), focus group methods (Stefanova et al., 2015) or driving simulators (Larue et al., 2015), rather than collecting real field data. In recent years, a systems analysis framework (Leveson, 2011; Read et al., 2016; Wilson, 2014) and a psychological schema theory (Salmon et al., 2013; Stanton and Walker, 2011) have been used to analyze the contributory factors underlying the accidents occurring

* Corresponding author at: IFSTTAR, Lille-Villeneuve d'Ascq, 20 Rue Élisée Reclus, BP 70317, F-59666 Villeneuve d'Ascq Cedex, France.
*E-mail address:* ci.liang@railenium.eu (C. Liang).

http://dx.doi.org/10.1016/j.ssci.2017.08.013
Received 15 December 2016; Received in revised form 24 June 2017; Accepted 14 August 2017
0925-7535/ 

at LXs. A study presented by Salmon et al. (2013) described a collision between a loaded semi-trailer truck and a train, which occurred in North Victoria, Australia, when the truck crossed the LX while the LX is occupied by railways without lights flashing. According to the investigation of the Office of the Chief Investigator (OCI), the truck driver in this study was not aware of the train and the activated state of the level crossing until it was too late to stop the truck. A study conducted by Davey et al. (2008) discussed the intentional violation of vehicle drivers crossing LXs, particularly focusing on vehicle driver's complacency due to the high level of familiarity. Tey et al. (2011) conducted an experiment to measure vehicle drivers' responses to LXs equipped with stop signs (passive), flashing lights and half barriers with flashing lights (active) respectively. In this study, the vehicle drivers' responses result from both the field survey and a driving simulator. Although these available qualitative approaches are beneficial to understand factors causing LX accidents, they do not allow for predicting the number or the probability of accident occurrence, or quantifying the contribution degree of the various impacting factors. Thereby, quantitative safety analysis approaches are crucial to thoroughly understand the impacting factors and enable the identification of practical design and improvement recommendations to prevent accidents at LXs.

One can notice that a number of quantitative studies on statistical models to predict LX accident frequency open a significant vista on understanding the risk related to LX accidents. In 1941, L. E. Peabody and T. B. Dimmick of the U.S. Bureau of Public Roads developed one of the earliest railway-highway crossing accident prediction models to estimate the number of accidents at railway-highway crossings in 5 years, named Peabody-Dimmick Formula (Ogden, 2007). This formula was developed based on the accident data of rural railway-highway crossings in 29 states in the U.S. and was utilized through the 1950s. As shown in Eq. (1), the parameters considered in this formula are the average daily road traffic $V$, the average daily railway traffic $T$, and the protection coefficient indicative of warning devices adopted $P$. $K$ is an additional parameter.

$$A_5 = \frac{1.28 \times (V^{0.170} \times T^{0.151})}{P^{0.171}} + K \tag{1}$$

However, advances in both warning device technologies and LX design features quickly led to an unavailability of the predefined formula form and coefficients that reflected the conditions pertaining to LX accidents in 1941.

The next evolutionary step in LX accident prediction was the New Hampshire Index (Oh et al., 2006) which is given as follows:

$$HI = V \times T \times P_f \tag{2}$$

where $HI$ represents the hazard index; $V$ is the average daily road traffic; $T$ is the average daily railway traffic and $P_f$ is the protection factor indicative of the warning devices adopted.

The New Hampshire model is a relative formula which can be used to rank the importance of crossing upgrades. Due to its simplicity, it has been widely used across the U.S. However, it is limited in that it does not predict the expected number of collisions, but only gives some indications about the priorities in terms of LX safety.

The accident prediction formula developed by the U.S. Department of Transportation (USDOT) in the early 1980s sought to overcome the limitations of earlier models (Chadwick et al., 2014). This comprehensive formula comprises three primary equations:

$$a = K \times EI \times MT \times DT \times HP \times MS \times HT \times HL \tag{3}$$

$$B = \frac{T_0}{T_0 + T} \times a + \frac{T}{T_0 + T} \times \left(\frac{N}{T}\right),\ T_0 = \frac{1}{0.05 + a} \tag{4}$$

$$A = \begin{cases} 0.7159 \times B, & \text{for passive devices;} \\ 0.5292 \times B, & \text{for flashing lights;} \\ 0.4921 \times B, & \text{for gates;} \end{cases} \tag{5}$$

where $a$ is the initial collision prediction (collisions per year at a given LX); $K$ is the formula constant; $EI$ is the exposure index (a variant of traffic moment) based on the product of highway and railway traffic; $MT$ is the index for the number of main tracks; $DT$ is the index for daily through trains during daylight; $HP$ is the index for highway paved; $MS$ is the index for maximum train speed; $HT$ is the index for highway type; $HL$ is the index for highway lanes. $B$ is the adjusted accident frequency; $T_0$ is the weighting factor and $N$ is the number of accidents observed in $T$ years at a given LX. Finally, $A$ is the normalized accident frequency.

The USDOT formula is the most commonly used model in the U.S. today. A specified table of USDOT provides each of the indexes for LXs equipped with passive controls, flashing lights and gates (Austin and Carson, 2002). Although the formula is comprehensive, its current definition makes it difficult to identify or prioritize design or improvement activities that will most effectively address LX safety-related problems, since it does not provide the magnitude of the characteristics' contribution to the LX safety.

The Australian Level Crossing Assessment Model (ALCAM) is a location specific and parameterized risk model which provides a method for assessing risks to LX users, train passengers and train staff (Woods et al., 2008). The ALCAM model is given as follows:

$$\begin{aligned} \text{ALCAM Risk Score} = {} & \text{Infrastructure Factor} \times \text{Exposure Factor} \\ & \times \text{Consequence Factor} \end{aligned} \tag{6}$$

where the Infrastructure Factor is the output of a complex scoring algorithm that assesses how the physical properties at each LX site will affect human behavior; the Exposure Factor is a function of the LX control type, vehicle (or pedestrian) volumes and train volumes (i.e., the Peabody-Dimmick Formula is used as the Exposure Factor function) to address the combined exposure of trains and road vehicles (or pedestrians) pertaining to various LX control types; the Consequence Factor is the expected consequence of a collision which includes deaths and injuries involving both railway and roadway. The Infrastructure Factor adjusts the accident probability per year to reflect the actual LX site conditions. Multiplying the Infrastructure Factor by the Exposure Factor will give the actual annual likelihood of an accident occurring at a particular LX (National ALCAM Committee, 2012). The Consequence Factor is expressed in terms of an expected number of equivalent fatalities per year. An equivalent fatality is a combination of all types of harm using the ratio: 1 fatality = 10 serious injuries = 200 minor injuries. The ALCAM has been applied across all Australian states and in New Zealand since 2003, and overseen by a committee of representatives from the various jurisdictions of these countries to ensure its consistency in terms of development and application. However, the ALCAM does not cover all kinds of LX accidents, since its main focus is deliberate and accidental collisions involving user errors but excluding vandalism and suicide. It should be noticed that some LX physical properties considered in ALCAM show a high correlation between each other, which implies the existence of a kind of redundancy between the model inputs, and consequently a bias in terms of the outputs.

In recent studies, authors tended to adopt the Poisson regression model, the NB regression model or variants of the Poisson regression model (e.g., zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB)) combined with the estimated $\hat{\lambda} = e^{\sum_{j=1}^{m} \beta_j x_j + \sigma}$ ($x_j$ is the independent variable considered and $\beta_j$ is the estimated coefficient of $x_j$) (Cameron and Trivedi, 1986, 1990; Lawless, 1987; Miaou, 1994; Austin and Carson, 2002; Chang, 2005; Lu and Tolliver, 2016) to deal with accident statistics. However, this form of estimated $\hat{\lambda}$ is not appropriate in our case. According to the constraints between the LX accident frequency and impacting variables, presented in Section 3.2, some variables (e.g., the average daily railway traffic, the average daily road traffic and the road traffic accidents) should not be used in an exponential form, due to the logical assumption that the case where these variables are equal to 0, would directly lead to 0 accident occurrence. Therefore, these aforementioned approaches combined with