RISK ASSESSMENT OF THE SPANISH NATIONAL RAILWAY SYSTEM

A. MATEOS, A. JIMENEZ, H.J. PICHARDO

Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid (UPM), 28660 Boadilla del Monte, Madrid, Spain

The principal risks in the railway industry are mainly associated with collisions, derailments and level crossing accidents. An understanding of the nature of previous accidents on the railway network is required to identify potential causes and develop safety systems and deploy safety procedures. Risk assessment is a process for determining the risk magnitude to assist with decision-making. We propose a three-step methodology to predict the mean number of fatalities in railway accidents. The first is to predict the mean number of accidents by analyzing generalized linear models and selecting the one that best fits to the available historical data on the basis of goodness-of-fit statistics. The second is to compute the mean number of fatalities per accident and the third is to estimate the mean number of fatalities. The methodology is illustrated on the Spanish railway system. Statistical models accounting for annual and grouped data for the 1992-2009 time period have been analyzed. After identifying the models for broad and narrow gauges, we predicted mean number of accidents and the number of fatalities for the 2010-18 time period.

1. Introduction

As in all types of transportation there are risks we have to face in the railway networks. The occurrence of an accident in railways infrastructures has a very low probability, but when an accident occurs it gets a lot of attention[3]. Train accidents sometimes lead to a number of casualties and injuries and entail substantial financial costs, such as damage to equipment, increased insurance premiums, legal costs, fines, compensations, and loss of company reputation.

Developing safety systems and deploying safety procedures requires an understanding of the nature of previous accidents on the railway network to identify potential causes.

Many of the railway risk assessment techniques currently used are comparatively mature tools. The results of using these tools are heavily reliant on the availability and accuracy of the risk data[1]. However, data are usually incomplete or there is a high level of uncertainty involved in such data[1].

Most lines in Europe and North America were built to the so-called *standard* or *international gauge* (1435 mm), although there are some exceptions where broader gauges were chosen, like Spain and Portugal (1668 mm). Narrow gauge lines have also been built all around the world, mainly in mountainous areas or for branch lines. The total length of the broad and narrow gauge networks in Spain in 2009 was 13,354 km and 1,2691 km, respectively.

According to EU Safety Directives, EU member states have established independent national accident investigation bodies. In Spain the task was assigned to the *Railway Accident Investigation Committee* (CIAF). CIAF publishes both the *common safety indicators* and the accident reports on its website and in annual reports.

We propose a methodology that helps to improve the safety of railway systems by reducing the risk of fatal accidents and providing for the possibility of applying new strategies.

2. Methodology

2.1. *Phase 1. Prediction of the Mean Number of Accidents*

The number of accidents is predicted by fitting statistical models to historical data. We have analyzed generalized linear models where the dependent variable is linearly related to the independent variables via a specified link function. These models allow for the dependent variable to have a non-normal distribution such as a Poisson, negative binomial, gamma[5] distribution and COM-Poisson[4], while the most commonly used link functions are log and power.

Parameter estimates are obtained using the principle of maximum likelihood. Selecting the model that best fits to data depends on the following goodness-of-fit statistics: *deviance*, *Pearson* χ^2, *log-likelihood*, *Akaike's information criterion* (AIC), *finite sample corrected AIC* (AICC), *Bayesian information criterion* (BIC) and *consistent AIC* (CAIC).

Original annual data and three-year grouped data accounting for broad and narrow gauges were originally considered. Generalized linear models accounting for *log* and *power* functions were analyzed. We found that most goodness-of-fit statistics were better for both broad and narrow gauge when data were grouped, see Table 1.

Thus, the statistical models that best fitted three-year grouped data were used to predict the mean number of accidents. Specifically, Poisson Power -1 model was selected for broad gauge, see Table 2, while the Gamma Power -1 was the best one for narrow gauge.

Table 1. Three-year period grouped data.

Years	Broad Gauge			Narrow Gauge		
	Train-kms	Accidents	Fatalities	Train-kms	Accidents	Fatalities
1992-1994	496.5	421	93	22.7	56	41
1995-1997	485.6	304	80	24.1	62	26
1998-2000	501.9	203	58	29.3	96	20
2001-2003	519.7	170	84	30.0	46	20
2004-2006	524.8	137	138	29.8	16	17
2007-2009	540.2	135	135	29.6	16	13

Table 2. Goodness-of-fit statistics for grouped data and broad gauge.

Goodness-of-fit	Poisson Log	Poisson Pow-1	Negative Binomial Log	Negative Binomial Pow-1	Gamma Log	Gamma Pow-1	COM-Poisson
Deviance	30.0	22.4	0.6	0.5	0.6	0.5	34.3
Pearson χ^2	30.8	22.1	0.6	0.5	0.6	0.5	35.2
Log Likel.	69.6	65.8	94.5	94.4	70.8	69.3	70.9
AIC	145.2	137.6	195.0	194.9	149.6	146.5	149.7
AICC	146.9	139.3	196.7	196.6	152.6	149.6	152.8
BIC	147.9	140.3	197.6	197.5	153.1	150.1	153.3
CAIC	150.9	143.3	200.6	200.5	157.1	154.1	157.3

Regarding **broad gauge**, based on the Poisson Power – 1 model, predictions for the next 19 years (2010-2018) can be computed as follows: $E(y_{1t}) = (0.00108867013927t + 8.764125618932E\text{-}0.6tkm_t - 0.00306052078)^{-1}$, where t is the time period and broad gauge train-kilometers for the 2010-18 period have to be previously estimated to predict the number of accidents in such period. A linear regression model from 1992-2009 data was used to estimate train-kms in 2010-2018: $tkm_t = 10.111428574285714285t + 476.06$.

For **narrow gauge**, the selected Gamma Power -1 model is used to make predictions for the number of accidents in the coming 9 years: $E(y_{2t}) = (0.0177103395923t - 0.00597228124517tkm_t + 0.132157197445)^{-1}$, where narrow gauge train-kilometers for the 2010-18 period is again previously estimated assuming a linear regression model applied to 1992-2009: $tkm_t = 1.488171428571t + 22.3914$, where t is the time period.

Table 3 shows the corresponding estimation for train-kilometers and the prediction of the mean number of broad and narrow gauge accidents for the

2010-2018 period, respectively. Finally, to predict the total number of accidents in the 2010-18 period we just have to add the predictions for broad and narrow gauges in Table 3.

Table 3. Prediction of the mean number of accidents.

Years	Broad Gauge		Narrow Gauge	
	Train-kms (millions)	Accidents	Train-kms (millions)	Accidents
2010-2012	546.8	106.9	32.8	16.6
2013-2015	557.0	95.0	34.3	14.5
2016-2018	567.1	85.4	35.8	12.8

2.2. *Phase 2. Computation of the Mean Number of Fatalities per Accident*

The mean number of fatalities per accident can vary hugely depending on the characteristics of the system. Therefore, it might be worthwhile finding out how to reduce the number of fatalities, which we do by changing some characteristics. Suppose that we classify the accidents in r groups, according to the characteristics we have determined. Then, c_j is the mean number of fatalities per accident given characteristic j, j =1... r. Within the j group there could be other sub-characteristics. Suppose we have r_j subgroups within group j. Within each group r_j there could be other subgroups, and so on. The mean number of fatalities per accident in period t is defined as $c_t = \sum_{j=1}^{r} p_{jt} c_{jt}$, where p_{jt} represents the proportion of a selected characteristic that belong to a group of characteristics j in period t and $c_{jt} = \sum_{k=1}^{r_j} p_{jkt} c_{jkt}$, where p_{jkt} is the proportion of the selected characteristic that belong to subgroup jk in period t. c_{jkt} follows the same process, and more c can be calculated until there are no more subgroups to consider. If there are not more subgroups to consider we propose two possible methods to calculate c:

1. If all data are considered equally significant and reliable, then c_j (c_{jk}) is the mean number of fatalities per accident given the selected characteristic in group j (jk) in period t.
2. Recent data are more representative and significant than the older information. With this assumption, c_j (c_{jk}) is computed as a weighted mean of the number of fatalities per accident in group j (jk) in period t. We used the *rank-order centroid function*[2] to weight the values.

Now, we estimate the mean number of fatalities per accident in broad and narrow gauge, denoted by c_1 and c_2 respectively. We consider three possibilities:

- c_1 and c_2 are the mean number of fatalities per broad and narrow gauge accident in the 1992-2009 period ($c_1 = 0.42919708$, $c_2 = 0.469178082$).

- c_1 and c_2 are the weighted mean number of fatalities per broad and narrow gauge accident in the 1992-2009 period. The centroid function was used to weight the different years, i.e., recent years are assigned a higher weight than years further in the past ($c_1 = 0.692831966$, $c_2 = 0.52795136$).
- c_1 and c_2 are computed in the 2004-2009 period, in which accident investigations were more exhaustive ($c_1 = 1.003676471$, $c_2 = 0.9375$).

The mean number of fatalities per accident, c_t, is computed as follows: $c_t = p_t\, c_1 + (1 - p_t)\, c_2$, with p_t as the proportion of broad gauge train-kilometers.

2.3. *Phase 3. Prediction of the Mean Number of Fatalities*

The total number of fatalities, f_t, in the 2010-2018 period are estimated using the following model: $f_t = E(y_t)c_t$. The total predicted number of fatalities per 3-year period is shown in Table 4.We have noticed broad gauge is safer than narrow gauge, and narrow gauge train-kilometers are expected to grow. If we reduce their proportion by 1/1000 and take this to be the proportion for broad gauge train-kilometers, the results would be as follows, see Table 4.

Table 4. Prediction of the mean number of fatalities.

	Mean number of fatalities (f_t)			f_t applying changes		
	Mean	Weighted mean	Mean for the last 6 years	Mean	Weighted mean	Mean for the last 6 years
2010-2012	53.3	84.4	123.5	51.3	81.5	119.2
2013-2015	47.2	74.8	109.4	45.2	71.8	104.9
2016-2018	42.4	67.1	98.2	40.4	64.2	93.7

As we can see, the proposed strategy reduces the number of fatalities. With the different methods, i.e., weighted mean, mean and mean for last six years, we reduce fatalities by a total of 9, 6 and 13 respectively. We consider the last method to be the best for the reasons explained above. The final decision has to be made by the experts, who are the ones to evaluate whether or not the cost of such changes are worth it.

3. Conclusions

We have proposed a methodology to predict the mean number of fatalities in railway network accidents, based on three phases. The methodology has been applied on the Spanish railway system on the basis of annual and grouped data for the 1992-2009 time period. In both cases, we made a distinction between broad and narrow gauges.

The proposed methodology could be a starting point for a more complete analysis of the Spanish railway network, accounting for other variables apart from gauge, since CIAF has conducted fuller investigations in Spain from 2004 on wards. Variables such as the gauge traffic level, manual or automatic block options or centralized traffic control can add meaningful analyses in the future.

References

1. M. An, A Review of Design and Maintenance for Railway Safety the Current Status and Future Aspects in the UK Railway Industry, *World Journal of Engineering* 163, 3, 99-119 (2005).
2. Barron, F., Barret, B. Decision Quality Using Ranked Attribute Weights. Management Science 42 (11), 1515-1523 (1996)
3. A.W. Evans, Fatal Train Accidents on Europe's Railways: 1980-2009, *Accident Analysis and Prevention* 43, 10-23 (2011).
4. D. Lord, S. Guikema and S. Geedipally, Application of the Conway-Maxwell-Poisson Generalized Linear Model for Analyzing Motor Vehicle Crashes, *Accident Analysis and Prevention* 40, 1123-1134 (2008)
5. R. Winkelmann and K. Zimmermann. Recent Developments in Count Data Modeling: Theory and Applications, *J. of Economy Surveys* 9, 1-24 (1995).