Assessing temporary speed restrictions and associated unavailability costs in railway infrastructure

António Ramos Andrade[1] and Paulo Fonseca Teixeira[2]

[1]Invited Assistant Professor, PhD, E-mail: araes@iscte.pt, Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL), Lisboa, Portugal. Address: Av. das Forças Armadas, Ala Autónoma, 1649-026 Lisboa (Portugal). Telephone: (+351) 210 464 014. (corresponding author)

[2] Assistant Professor, PhD, e-mail: pft@tecnico.ulisboa.pt, CESUR, CEris, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal. Address: Av. Rovisco Pais, 1, 1049-001 Lisboa (Portugal). Telephone: (+351) 218 417 000.

Highlights:

- The occurrence of temporary speed restrictions in railway infrastructure is statistically modelled and several regression models are compared;
- A Negative Binomial regression model provides a better fit than the Poisson and 'over-dispersed' Poisson regression models;
- The main quality indicators for railway track geometry degradation proved to have a statistically significant effect in the occurrence of temporary speed restrictions;
- The maintenance and renewal operations also have a statistically significant effect in the occurrence of temporary speed restrictions;

Abstract:

This paper analyses the occurrence of temporary speed restrictions in railway infrastructure associated with railway track geometry degradation. A negative binomial regression model is put forward to estimate the expected number of temporary speed restrictions, controlling for the main quality indicators of railway track geometry degradation and for the maintenance and renewal actions/decisions. The prediction of temporary speed restrictions provides a quantitative way to support the assessment of unavailability costs to railway users. A case study on the Lisbon-Oporto Portuguese line is explored, comparing three statistical models: the Poisson, the 'over-dispersed' Poisson and the proposed negative binomial regression. Main findings suggest that the main quality indicators for railway track geometry degradation are statistically significant variables, apart from the maintenance and renewal actions. Finally, a discussion on the impacts of the unavailability costs associated with temporary speed restrictions is also provided in a regulated railway context.

Keywords: Temporary speed restrictions; railway maintenance; statistical modelling; Negative Binomial regression; unavailability costs.

Background

In transportation infrastructure systems, maintenance and renewal operations might cause some impacts in the availability of the railway system, besides the associated costs related to these operations. One of these impacts is the occurrence of temporary speed restrictions, which affect the normal operation of trains in the railway infrastructure and cause unavailability costs. A main cause of the occurrence of temporary speed restrictions is the degradation of railway track, namely the railway track geometry degradation. In order to study these impacts, performance indicators of the railway infrastructure have to be measured and monitored [1], namely temporary speed restrictions as an availability indicator. Besides, the analysis of such indicators has to take into account the infrastructure influence on rail punctuality (delays), i.e. infrastructure fault datasets should be linked with operational delay datasets to improve railway infrastructure management. Temporary Speed Restrictions are also part of the performance indicators that regulators use to assess the infrastructure manager performance [3]. Moreover, the imposition of temporary speed restrictions has been found to be an influencing factor on train punctuality [4]. Operating speed has also been identified as a key variable in infrastructure design consistency [5].

Many studies on delays in railway infrastructure have focused on quantifying the delays [6], given the occurrence of a delay (e.g. a temporary speed restriction), exploring the impact of a given train delay in the network. These studies do not discuss what caused the occurrence of a temporary speed restriction, and they just assume that it happens and then they are interested in computing the different train delays imposed to a given network. From our perspective, it seems that there is a missing link/connection/dependence between railway track geometry, the maintenance and renewal actions, and the occurrence of temporary speed restrictions in the network. Moreover, in any decision support system for planning maintenance and renewal actions in transportation infrastructures, the assessment of unplanned impacts like the temporary speed restrictions are crucial for the definition of a maintenance/renewal strategy that not only minimizes maintenance and renewal costs but also minimizes delays.

The past research on railway delay modelling has walked a long path, mainly focused on the quantification of train delays. Several studies aimed to model train delays, considering primary delays and knock-on effects or secondary delays, i.e. the propagation of delays to other trains in the network. A delay estimation methodology is put forward in [7], which defines an exponential relation between travel time delay and train mix for single and double track lines, validated with simulation results from a design of experiments and also with real-world delay values from a sub-network existing in the Los Angeles area. Moreover, they provided an excellent review on previous research, identifying two main approaches dominating the research in this topic: the analytical models and the simulation models. Another approach also relied on simulation software (Rail Traffic Controller) results to fit an exponential dependence with the number of trains per day to estimate average delay times for single and double track due to in-service failures of different length (e.g. 1h, 3h and 5h) and associated costs [8]. They also conducted some analysis on the variability of train delays for different traffic volumes. Further micro-simulation/simulation models that support decisions regarding timetabling and railway operations were also explored in [9]. Another statistical estimation approach to model railroad congestion delay from BNSF railway data for eight districts in the western US, using multiple linear regression with an exponential functional form to explain the total train running time (i.e. the free running time plus the congestion-related delay) using as independent/explaining variables: train-related and track-related variables, primary and secondary-effect variables and capacity utilization effect variables [10].

Moreover, delays incurred by the passengers have been analysed and an overall generalized waiting cost was put forward, comprising: the cost of extra stopping in the stations, the cost of extended transfer times and the cost of deviating from the ideal running time supplements [11]. This approach detailed all passenger flows in train connections, namely: the transfer passengers, the through passengers, the departing passengers and the arriving passengers; and assigning distinct costs to each type of delay. In the same research direction, i.e. focusing in delays suffered by the passengers, passenger delay models were explored, instead of the typical train delay models, in which passengers are adaptive agents that may choose a different route than their planned route (assuming in the most optimistic scenario that they have a complete knowledge of present and future delays in the system) [12]. These two contributions [11, 12] represent the most important steps towards the quantification of railway delays suffered by the passengers (or freight).

Regarding the delays caused by the infrastructure manager, i.e. the infrastructure delays, there is a lack of published references. To the best of our knowledge, the only reference discussing infrastructure delays is [13]. Delay risks associated with train schedules were modelled, detailing three types of delays: track related delays, train dependent delays and terminal/schedule stop delays.

Nevertheless, there has been little research on the impact of maintenance and renewal decisions in railway delays and unavailability costs. For simplicity, let us put forward a classification for different delays, following the idea of quantifying delays depending on the agent responsible for causing it, in order to frame this research work in a larger research framework. The term 'agent' is used considering the vertical separation between the Infrastructure Manager (IM) and the Train Operating Companies (TOC), which means that we may have as agents: the IM, the different TOC and also the passengers or freights (i.e. the final users). Having said that, delays can be classified into three groups:

i) infrastructure delays, i.e. the delays whose responsibility is assigned to the infrastructure manager;

ii) train operating companies delays, i.e. the delays whose responsibility is assigned to the train operating companies;

iii) the passengers or freight delays, i.e. the delays whose responsibility is assigned to the passengers or the final users.

This classification is particularly useful as it emphasizes the need of more research on the link between degradation processes, maintenance and renewal actions of the IM's responsibility and the above-mentioned infrastructure delays. Note that other railway agents that could also be integrated in this conceptual framework for a vertically separated sector would be the regulator or the regulatory entity and the maintenance contractors.

To a certain extent, there is a parallel between this proposed delay classification and the one put forward before in [13], especially in the first two groups, i.e. the infrastructure delays (or track-related delays) and the train operating companies' delays (train dependent delays), respectively. However, the terminal/schedule stop delays from [13] are not necessarily equal to the passenger delays as a passenger can catch a delayed train without incurring into any delay impact in his/her trip.

Let us now focus on the infrastructure delays as they are the most relevant for IM decision-making process regarding maintenance and renewal actions. There are two main types of infrastructure delays: i) the infrastructure delays due to medium-/long-term changes in the maximum permissible speeds (i.e. the planned infrastructure delays) and ii) the infrastructure delays due to temporary speed restrictions (i.e. the unplanned infrastructure delays).

Some of these delays are not even totally perceived by the passengers, by the operators and even by the regulator. These delays were above defined as planned infrastructure delays because they are associated with medium-/long-term downgrades of speed performance due to reductions of the maximum permissible speed. As these changes immediately affect the train schedule production, they are not perceived by the other railway agents and in fact, they may hide a poor performance of the IM in terms of asset management regarding maintenance and renewal actions. However, the aim of this paper is to discuss solely the unplanned infrastructure delays due to temporary speed restrictions and these planned infrastructure delays are left for further research, though some first steps have been taken in [14] within a bi-objective optimization model for maintenance and renewal decisions.

The outline of this paper is as follows: this first section introduces the need to assess the occurrence of temporary speed restrictions in railway infrastructure and reviewed the past research on railway delays, focusing on the delays related with maintenance and renewal actions (or the 'infrastructure delays'). Afterwards, a review is provided on the statistical methodology followed within the Generalized Linear Model (namely the negative binomial regression, the Poisson and the 'over-dispersed' Poisson regressions), in which the different regression models are estimated for our case study and compared using the Akaike Information Criterion (AIC). A context discussion on the impacts of the assessment of temporary speed restrictions and associated costs in the railway regulatory framework is put forward. Finally, the last section highlights the main conclusions and suggests further research in this topic.

Statistical modelling of temporary speed restrictions

This section explores and discusses the statistical methodology followed in this paper to predict the occurrence of temporary speed restrictions in railway infrastructure within the Generalized Linear Model framework, namely using the negative binomial regression model, the Poisson and the 'over-dispersed' Poisson regression models.

To assess the temporary speed restrictions related with rail track geometry, a database from the Portuguese IM (REFER), called 'e-LVs', was analysed. This application/database compiles a series of information regarding temporary speed restrictions, namely: the identification details as the line, the direction and the location; the delay details as the theoretical/computed delay, the restriction speed, the maximum permissible speed, the initial and final times, the motive; and other information not relevant for the following discussion.

Of course, many temporary speed restrictions have other motives than the ones related with the rail track subsystem or related with the railway track geometry. Take for instance the example of the temporary speed restrictions due to maintenance actions associated with the catenary subsystem. Those speed restrictions were not included in the following assessment because only the speed restrictions related with rail track geometry condition, maintenance or renewal actions were included in this analysis. In fact, IM is responsible for 20% up to 30% of the total delays in the railway system, and the track system and their faults are responsible for around 3% of the total delays in the railway system [1, 2].

Let Y be the dependent counting variable that counts the number of temporary speed restrictions that a 200-m long track section suffers in a trimester period. Here, a single temporary speed restriction is any reduction of speed for that 200-m long track section, i.e. when the restricted speed is lower than the maximum permissible speed for that 200-m long track section and for any time interval. Then, all temporary speed restrictions are summed up for a given trimester, leading to the observed Y for that 200-m long track section and that trimester. Within the Generalized Linear Model (GLM), two competing models are natural candidates to model the dependent variable Y: i) the Poisson regression model and ii) the Negative Binomial regression model. The logarithmic function is the only link function explored in both cases. A similar statistical modelling approach was also used for the assessment of unplanned/corrective maintenance needs [15].

One of the hypotheses (or motivating ideas) to explore in the following statistical approach was related with whether or not the quality indicators for rail track geometry lead to significant change in Y, i.e. the variability of the expected number of temporary speed restrictions that a 200-m long track section suffers in a trimester period. Therefore, the two usual indicators (for planned/preventive maintenance) were used as explaining variables: the standard deviation of longitudinal level defects (SD_{LL}) and the standard deviation of horizontal alignment defects (SD_{HA}). Additionally, information on all track geometry defects were also provided by the binary variables IAL (Immediate Action Limit), IL (Intervention Limit) and AL (Alert Limit), so that the model could use some information on unplanned/corrective maintenance needs. Finally, two additional binary covariates regarding planned maintenance and renewal operations were included so that one could quantify the impact of a planned maintenance (T_p - tamping operations) and of renewal operations (R_w) in the unplanned infrastructure delays (i.e. temporary speed restrictions). Regarding the spatial and time spread of the sample, it refers to 14 trimesters from July 2005 up to September 2009, with a total of 1358 track sections in each direction, summing up a total of 30,862 observations. All this statistical information was collected within a research project for the Portuguese IM from several databases: i) inspection records from EM-120 inspection vehicle, ii) infrastructure component database; iii) track component database; iv) e-LVs (database dedicated to temporary speed restrictions) and v) train schedule database.

Table 1 details the variables considered in the model and their associated description:

(Table 1)

The Immediate Action Limits (IAL), the Intervention Limits (IL) and the Alert Limits (AL) are set by the European Standard EN 13848-5 [16] for all rail track geometry defects. For further information on these rail track geometry defects and their indicators, the reader is referred to [17-20], while for further details on the railway track system, irregularities and variability of some physical parameters, the reader is referred to [21, 22].

The Poisson distribution is usually parameterized through the parameter λ and has the following probability function:

$$f(y|\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!} \tag{2}$$

For $y = 0,1,2,\ldots$ and $\lambda > 0$. A Poisson distributed random variable Y has mean equal to the variance, i.e. $E[Y] = \lambda = \mu$ and $VAR[Y] = \lambda = \mu$.

Within the Generalized Linear Model (GLM), in a Poisson regression model, the Poisson parameter is estimated as a linear combination of covariates for each observation i, typically using a log link function, i.e. $\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_j X_{ij}$. As the Poisson model has mean equal to the variance, this model is not appropriate if there is over-dispersion, i.e. when the variance is larger than the mean.

One simple alternative to model over-dispersed Poisson data is considering a dispersion parameter (or scale parameter) ϕ in the variance expression, so that the new expression would be $VAR[Y] = \phi.\mu$. This model, as an extension of the Poisson regression model, is usually called the 'over-dispersed' Poisson model as it adds an additional parameter – the scale parameter ϕ, whereas the simple Poisson regression model can be perceived as an 'over-dispersed' Poisson model, in which the scale parameter is held equal to one ($\phi = 1$).

Another popular alternative to model over-dispersed counting data is the Negative Binomial model. The Negative Binomial distribution is parameterized through the parameters p and r and has the following probability function:

$$f(y|p,r) = \frac{\Gamma(y+r)}{y!\,\Gamma(r)}\, p^r (1-p)^y \tag{3}$$

For $y = 0,1,2,\ldots$ and $r > 0;\ p \in (0,1)$. A Negative Binomial random variable Y has mean $E[Y] = r\frac{1-p}{p} = \mu$ and variance $VAR[Y] = r\frac{1-p}{p^2} = \mu + \frac{1}{r}\mu^2$.

Sometimes, the Negative Binomial regression model is parameterized in an alternative way with parameters μ and r, instead of parameters p and r. The additional term $1/r$ is usually referred as the dispersion parameter or ancillary parameter. The log function is also a common choice for the link function so that $\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_j X_{ij}$.

The main difference between the over-dispersed Poisson regression model and the Negative Binomial regression model is that the variance of the former is a linear function of the mean, while the variance of the latter is a quadratic function of the mean [23]. The negative binomial regression model has been used in several studies related to infrastructure modelling [24], from estimating transition probabilities in highway infrastructure degradation [25] to hurricane-related outages in the electric power systems [26], or even in railway safety [27] and road safety [28].

Table 2 presents the estimated simple Poisson regression model, in which the scale parameter is held equal to 1 in the estimation process of the model.

(Table 2)

First of all from the analysis of table 2, all covariates proved to be statistically significant with p-values lower than 1% in the Wald tests. From a 'ceteris paribus' perspective and interpreting the Incidence Rate Ratios (IRR), computed as the exponential of the coefficients, i.e. e^{β_j} for each covariate j, conducting tamping maintenance operations (T_p = 1) increases the expected number of speed restrictions by a multiplicative factor of 1.083, whereas conducting renewal operations (R_w = 1) increases the expected number of speed restrictions by a multiplicative factor of 6.009. Regarding covariates related with rail track geometry degradation, every unit of increase in the SD_{HA} corresponds to an increase in the expected number of speed restrictions by a factor of 1.417 (i.e. 41.7% of increase); whereas every unit of increase in the SD_{LL} corresponds to a decrease in the expected number of speed restrictions by a factor of 0.840 (i.e. 16% of decrease). Nevertheless, note that the negative sign of the coefficient relative to the SD_{LL} is not intuitive and it should be interpreted as the estimated coefficient resulting from other covariates' effects.

Having in mind the meaning of the covariates IAL, IL and AL from table 1, one may interpret the associated coefficients as the impact of all track geometry defects (besides the SD_{LL} and the SD_{HA}) in the expected number of speed restrictions, but one should have in mind that IAL=1 implies IL=AL=1, whereas IL=1 implies AL=1. Therefore, four cases should be interpreted: i) a track section which does not exceed any Alert Limit (IAL=0, IL=0 and AL=0) that serves as the reference case for comparison purposes; ii) a track section which exceeds only the Alert Limits (IAL=0, IL=0 and AL=1), corresponding to an increase of the expected number of speed restrictions by a multiplicative factor of 1.395 (i.e. $\exp(\beta_{\mathrm{AL}}) = \exp(0.333)$) comparing to the reference case; iii) a track section which exceeds the Intervention Limits (IAL=0, IL=1 and AL=1), corresponding to an increase of the expected number of speed restrictions by a multiplicative factor of 1.579 (i.e. $\exp(\beta_{\mathrm{AL}} + \beta_{\mathrm{IL}}) = \exp(0.333 + 0.124)$) comparing to the reference case; and finally, iv) a track section which exceeds the Immediate Action Limits (IAL=1, IL=1 and AL=1), corresponding to a decrease of the expected number of speed restrictions by a factor of 0.860 (i.e. $\exp(\beta_{\mathrm{AL}} + \beta_{\mathrm{IL}} + \beta_{\mathrm{IAL}}) = \exp(0.333 + 0.124 - 0.607)$) comparing to the reference case. Although the first two comparisons seem intuitive, the last comparison between track sections not exceeding any limit (IAL=IL=AL=0) and track sections exceeding all limits (IAL=IL=AL=1) does not seem at all intuitive. Nevertheless, the reader should again have in mind that its effect must be interpreted as an estimated coefficient resulting from the combination effect of other covariates rather than solely from the effect of that covariate.

Table 3 presents and compares two estimated models: the 'over-dispersed' Poisson and the negative binomial regression models. Note that the coefficients have similar estimates under both the Poisson and negative binomial models. Regarding the goodness-of-fit, AIC values indicate that the negative binomial model is considerably better than the Poisson model, i.e. 62,266.37 is considerably lower than 97,601.89, indicating better relative fit of the Negative Binomial regression model against the 'over-dispersed' Poisson regression model. However, a word of caution must be included at this time: the AIC does not provide any indication of absolute fit. On the opposite, it provides a relative indication of goodness-of-fit for competing models. Another aspect that favors the Negative Binomial model against the Poisson is the considerable over-dispersion. For instance, both the Deviance/df and the Pearson/df are larger than one, which clearly indicate over-dispersion in the data.

(Table 3)

From the analysis of table 3, all the covariates explored before in these models present again statistically significant coefficients with p-values lower than 1% in the Wald tests, except for the Tamping (T_p) covariate ($p = 0.055$) and the IL covariate ($p = 0.012$) in the 'over-dispersed' Poisson regression model, and except for the Tamping covariate ($p = 0.018$) in the negative binomial regression model. Moreover, note that the estimated coefficients for the initial Poisson regression model (in table 2) exhibit the same value in the over-dispersed Poisson regression model (in table 3). Regarding the coefficients for the negative binomial regression model, they exhibit the same sign for each covariate (i.e. positive or negative) as in the Poisson regression, but with slightly different values. For instance, for the Tamping covariate the coefficient exhibits different values: 0.140 instead of 0.080, which would mean that conducting tamping maintenance operations (Tp = 1) corresponds to an increase in the expected number of speed restrictions by a multiplicative factor of 1.150 instead of 1.083; whereas conducting renewal operations (i.e. Rw = 1) increases the expected number of speed restrictions by a factor of 5.748 instead of 6.009.

Regarding the covariates related to the rail track geometry degradation, every unit increase in the SD_{HA} would correspond to an increase in the expected number of speed restrictions by a multiplicative factor of 1.822 (i.e. 82.2% of increase); whereas every unit of increase in the SD_{LL} would correspond to a decrease in the expected number of speed restrictions by a multiplicative factor of 0.785 (i.e. 21.5% of decrease). Again, this negative sign does not seem intuitive and this coefficient should be interpreted as the estimated coefficient regarding all the other covariates' effects.

Moreover, and following the previous strategy to interpret the coefficients from the covariates IAL, IL and AL (associated with all track geometry defects or with unplanned maintenance needs), four cases are distinguished: i) a track section that does not exceed any Alert Limit, i.e. AL=0 (which implies IAL=IL=0), which is the reference case; ii) a track section which, at its maximum, exceeds the Alert Limits, i.e. AL=1 but IAL=IL=0, which corresponds to an increase of the expected number of speed restrictions by a multiplicative factor of 1.275 (i.e. $\exp(\beta_{\mathrm{AL}}) = \exp(0.243)$) comparing to the reference case; iii) a track section which exceeds the Intervention Limits, i.e. IL=AL=1 but IAL=0, which corresponds to an increase of the expected number of speed restrictions by a multiplicative factor of 1.570 (i.e. $\exp(\beta_{\mathrm{AL}} + \beta_{\mathrm{IL}}) = \exp(0.243 + 0.208)$) comparing to the reference case; and finally iv) a track section which exceeds the Immediate Action Limits, i.e. IAL=IL=AL=1, corresponding to a decrease of the expected number of speed restrictions by a factor of 0.861 (i.e. $\exp(\beta_{\mathrm{AL}} + \beta_{\mathrm{IL}} + \beta_{\mathrm{IAL}}) = \exp(0.243 + 0.208 - 0.602)$) comparing to the reference case. Again, although the first two comparisons seem intuitive, the last comparison between track sections not exceeding any limit (IAL=IL=AL=0) and track sections exceeding all limits (IAL=IL=AL=1) does not seem intuitive.

Moreover, for the 'over-dispersed' Poisson model, an additional goodness-of-fit statistic is presented: the adjusted log-likelihood, which is based on the estimated scale parameter (computed based on the deviance). Note that the Log-likelihood for the first Poisson regression model in table 2 and for the 'over-dispersed' Poisson regression model in table 3 are the same. Finally, when the overall models are tested against the intercept model, all the three investigated models are statistically significant against the intercept-only model ($p < 0.001$).

Although other models have not been explored in this section, the reader should consider that other distributions frequently used for over-dispersed count variables are the generalized Poisson distribution and some zero-inflated models. The exploration of these approaches and comparison between them and the ones here proposed is left for further research.

According to the negative binomial regression model presented in table 3, an estimation of the expected number of temporary speed restrictions $(\hat{\mu})$ for a 200-m track section in a given trimester can then be provided using the following expression, depending on certain track geometry degradation indicators, i.e. the SD_{HA} (σ_{HA}) and the SD_{LL} (σ_{LL}) and IAL (IAL), IL (IL) and AL (AL) defects, and also on Tamping (T_p) and Renewal (R_w) decisions:

$$\hat{\mu} = \exp\left(\hat{\beta}_0 + \hat{\beta}_{\sigma_{HA}}\sigma_{HA} + \hat{\beta}_{\sigma_{LL}}\sigma_{LL} + \hat{\beta}_{IAL}IAL + \hat{\beta}_{IL}IL + \hat{\beta}_{AL}AL + \hat{\beta}_{T_p}T_p + \hat{\beta}_{R_w}R_w\right) \qquad (4)$$

Values for the estimated parameters from expression 4 were presented in table 3. Figure 1 provides the expected value of temporary speed restrictions $(\hat{\mu})$ for a 200-m track section in a given trimester, and contrasts it with the recommended range from the European Standard EN 13848-5 for the planned maintenance criteria, i.e. the SD_{LL} (1.4-2.4 mm) and the SD_{HA} (1.0-1.3 mm) defects represented by a rectangle, for the speed (S) group 120<S<160 km/h. Figure 1 assumes that the track geometry indicators for unplanned maintenance needs (IAL, IL and AL) are zero $(IAL = IL = AL = 0)$, i.e. all indicators for track geometry defects are below the Alert Limits, and no tamping nor renewal actions are conducted in that trimester, i.e. $T_p = R_w = 0$.

(Figure 1)

Therefore, considering the lower left corner of the rectangle, an Infrastructure Manager which sets as planned criteria 1.4 and 1.0 for the SD_{LL} and SD_{HA} limits, the expected number of speed restrictions that a given track section will suffer in a trimester period using expression 4: $\hat{\mu} = \exp(-0.720 + 0.600 \cdot 1.0 - 0.243 \cdot 1.4) = 0.631$; whereas a strategy corresponding to the right upper corner of the rectangle (i.e. (2.4, 1.3)) leads to an expected number of speed restrictions of 0.592; and a strategy corresponding to the left upper corner of the rectangle (i.e. (1.4, 1.3)) leads to an expected number of speed restrictions of 0.756 and finally, a strategy corresponding to the right lower corner of the rectangle (i.e. (2.4, 1.0)) leads to the expected number of speed restrictions of 0.495. Similar calculations can be conducted for the limits of the recommended range for each speed (S) group.

Towards a more complete assessment of unavailability costs in a regulated railway industry

After the introduction of the European Union (EU) directive 91/440 (EC 1991), whose aim was to 'facilitate the adoption of the Community railways to the needs of the Single Market and increase their efficiency', the relation between different railway stakeholders dramatically changed. The EU directive 91/440 mainly required a separation between the provision of transport services and the operation of infrastructure managers by introducing separate accounts between both activities and the state. It should be mentioned that though the separation of accounts was compulsory, organizational and institutional separation was optional. Over the past 20 years, different policy implementation took place in each EU country and, unsurprisingly, each state benefitted from different efficiency effects [29].

Under vertical separation, the Infrastructure Manager (IM) is responsible for the maintenance and renewal actions of railway infrastructure, and for operating the control and safety systems, whereas the operators are responsible for the transport service for goods and/or passengers. The IM then charges a fee for the use of the railway infrastructure to the operators, and the operators would sell tickets for their transport service to passengers or freight (i.e. the end users). Ideally, the regulatory entity or Regulator is then responsible to ensure non-discriminatory access of operators to railway infrastructure, to ensure that safety and economic regulations are complied, and to provide advice to the Government/State on the asset management plans of the Infrastructure Manager, particularly on its maintenance, renewal and investment plans.

Regarding the cost flows and necessary separate accounts, the vertical separation plan is well drawn, though its implementation proved difficult as many railway companies have faced hard financial situations. Regarding the delay flows, there is still a lack of understanding on this dimension, though there is a belief that from a regulatory perspective, the delay dimension is crucial for performance assessment of all railway agents. For instance, the Office of Rail Regulation (ORR) – the regulator for Britain's railways (one of the most active European regulators), monitors the performance of railway agents through several indicators, according to its National Rail Trends (NRT) website, namely: (i) temporary speed restrictions, (ii) cancellations and significant lateness, (iii) possession disruption for passengers and freight trains, (iv) the punctuality performance measures and (v) passengers above capacity. These five dimensions provide different performance assessment for the agents: Infrastructure Manager and Operators. Note that temporary speed restrictions mainly focus in the IM performance, whereas the remaining indicators are more focused in the operators' performance.

In fact, the Portuguese Infrastructure Manager (REFER) has been trying to implement such a system, in which REFER is responsible in the operating control centers to input which agent is responsible for a given delay, as well as to compute the knock-on delays assigned to that delay. In case REFER fails to assign an agent responsible, REFER is considered responsible for it. This specific rule provides a clear incentive to REFER to fill that information. However, as the information flow is almost completely dominated by one single agent – the Infrastructure Manager, the risk that the Regulator is captured by the IM must not be neglected. Mostly because, for the case of infrastructure delays, the impacts on train schedule production are immediately incorporated and are not as visible as other delays, i.e. operators and passenger delays are more visible though less controllable and lead to unplanned consequences and knock-on delays in the train schedules. However, if an operator does not agree that an assigned delay was its responsibility, it can always protest against it to the regulator, and in many situations agreements on the split of the delay consequences can be negotiated.

The regulatory entity as an independent safety and economic regulator monitors the performance of each agent and defends the public interest. In that sense, it is extremely important to monitor all dimensions of delays, especially the changes in the maximum permissible speed. Having said that, and taking for example the case of the British Regulator – ORR, it seems that the changes in the maximum permissible speed have been overlooked, and the regulator should take action to include them in the information flow for asset management and performance monitoring.

Finally, in terms of detailed calculation from an economical perspective, unavailability impacts should ideally distinguish each pair O-D per time period (e.g. rush hour an non-rush hour) and market segment (e.g. urban, intercity and long-distance, freight) in an overall assessment of the willingness to pay for different operators and other agents in the railway market. This assessment is rather complex and simplified approaches to integrate these ideas should be sought in future research, particularly regarding the signals of the regulator entity on the cost quantification of delays. For instance, the Portuguese regulator URF fixed a cost quantification of delays of 4 €/min for urban passenger trains, and 60% of that value for intercity and long-distance trains and 5% of that value for freight trains (URF 2011). Note that these values refer to costs per minute of delay per train, and not per passenger. The rationale behind the setting of this values for the delays is still not transparent enough and an economic study is lacking, at least for the Portuguese case.

Conclusions and further research

The present paper explored statistical regression approaches to model the expected number of temporary speed restrictions in railway infrastructure within the Generalized Linear Framework. A negative binomial regression model was considerably better than the Poisson and the 'overdispersed' Poisson regression models. The quality indicators for planned maintenance, i.e. the standard deviations of longitudinal level defects and of the horizontal alignment defects proved to be statistically significant predictors for the expected number of temporary speed restrictions. Moreover, the variables controlling all other rail track geometry defects, i.e. IAL, IL and AL, also proved to be statistically significant predictors. Finally, the maintenance (T_p) and renewal (R_w) decisions were also statistically significant predictors, and exhibited Incidence Rate Ratios equal to 1.150 and 6.009, respectively. Some discussion on the need to include unavailability costs associated with temporary speed restrictions was also provided. Regarding further research, the final objective of the present model is the integration of the expected number of temporary speed restrictions and associated delays in an objective function in order to optimize the Alert Limits that trigger preventive maintenance actions, as part of a planned maintenance strategy.

References

[1] Stenström, C., Parida, A., & Galar, D. (2014). Performance indicators of railway infrastructure. Int J Railw Technol, 1(3), 1-18.

[2] Veiseth, M., Olsson, N., & Saetermo, I. A. F. (2007). Infrastructure's influence on rail punctuality. WIT Transactions on The Built Environment, 96.

[3] Vickerman, R. (2004). Maintenance incentives under different infrastructure regimes. Utilities Policy, 12(4), 315-322.

[4] Olsson, N. O., & Haugland, H. (2004). Influencing factors on train punctuality—results from some Norwegian studies. Transport policy, 11(4), 387-397.

[5] Luque, R. & Castro, M., Highway Geometric Design Consistency: Speed Models and Local or Global Assessment, Int. J. Civ. Eng. (2016). doi:10.1007/s40999-016-0025-2

[6] Hansen, I. A. (2010). Railway Network Timetabling and Dynamic Traffic Management. International Journal of Civil Engineering, 8(1).

[7] Murali, P., Dessouky, M., Ordóñez, F., & Palmer, K. (2010). A delay estimation technique for single and double-track railroads. Transportation Research Part E: Logistics and Transportation Review, 46(4), 483-495.

[8] Schlake, B., Barkan, C., & Edwards, J. (2011). Train delay and economic impact of in-service failures of railroad rolling stock. Transportation Research Record: Journal of the Transportation Research Board, (2261), 124-133.

[9] Marinov, M., Şahin, İ., Ricci, S., & Vasic-Franklin, G. (2013). Railway operations, time-tabling and control. Research in Transportation Economics, 41(1), 59-75.

[10] Gorman, M. F. (2009). Statistical estimation of railroad congestion delay. Transportation Research Part E: Logistics and Transportation Review, 45(3), 446-456.

[11] Vansteenwegen, P., & Van Oudheusden, D. (2007). Decreasing the passenger waiting time for an intercity rail network. Transportation Research Part B: Methodological, 41(4), 478-492.

[12] Nielsen, O. A., Landex, A., & Frederiksen, R. D. (2008). Passenger delay models for rail networks. Schedule-Based Modeling of Transportation Networks: Theory and applications, 46, 27.

[13] Higgins, A., Kozan, E., & Ferreira, L. (1995). Modelling delay risks associated with train schedules. Transportation Planning and Technology, 19(2), 89-108.

[14] Andrade, A. R., & Teixeira, P. F. (2011). Biobjective optimization model for maintenance and renewal decisions related to rail track geometry. Transportation Research Record: Journal of the Transportation Research Board, (2261), 163-170.

[15] Andrade, A. R., & Teixeira, P. F. (2014). Unplanned-maintenance needs related to rail track geometry. In Proceedings of the Institution of Civil Engineers-Transport, Vol. 167, No. 6, pp. 400-410.

[16] EN 13848-5 (2008): Track geometry quality – Geometric quality levels.

[17] Andrade, A. R., & Teixeira, P. F. (2012). A Bayesian model to assess rail track geometry degradation through its life-cycle. Research in transportation economics, 36(1), 1-8.

[18] Andrade, A. R., & Teixeira, P. F. (2013). Hierarchical Bayesian modelling of rail track geometry degradation. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of rail and rapid transit, 227(4), 364-375.

[19] Andrade, A. R., & Teixeira, P. F. (2015). Statistical modelling of railway track geometry degradation using Hierarchical Bayesian models. Reliability Engineering & System Safety, 142, 169-183.

[20] Andrade, A. R., & Teixeira, P. F. (2016). Exploring Different Alert Limit Strategies in the Maintenance of Railway Track Geometry. Journal of Transportation Engineering, 04016037.

[21] Sadeghi, J. (2010). Field investigation on vibration behavior of railway track systems. International Journal of Civil Engineering, 8(3), 232-241.

[22] Dahlberg, T. (2010). Railway Track Stiffness Variations – Consequences and Countermeasures. International Journal of Civil Engineering, 8(1), 1-12.

[23] Ver Hoef, J. M., & Boveng, P. L. (2007). Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data?. Ecology, 88(11), 2766-2772.

[24] Guikema, S. D., & Coffelt, J. P. (2009). Practical considerations in statistical modeling of count data for infrastructure systems. Journal of Infrastructure Systems, 15(3), 172-178.

[25] Madanat, S., & Ibrahim, W. H. W. (1995). Poisson regression models of infrastructure transition probabilities. Journal of Transportation Engineering, 121(3), 267-272.

[26] Liu, H., Davidson, R. A., Rosowsky, D. V., & Stedinger, J. R. (2005). Negative binomial regression of electric power outages in hurricanes. Journal of infrastructure systems, 11(4), 258-267.

[27] Evans, A. W. (2013). The economics of railway safety. Research in transportation economics, 43(1), 137-147.

[28] Noland, R. B. (2013). From theory to practice in road safety policy: Understanding risk versus mobility. Research in transportation economics, 43(1), 71-84.

[29] Friebel, G., Ivaldi, M. and Vibest, C. (2010). Railway (De)Regulation: A European Efficiency Comparison. Economica, Vol. 77, pp. 77-91.