WIT CONFERENCES
This paper is part of the Proceedings of the 15th International Conference on Railway Engineering Design and Operation (CR 2016)
www.witconferences.com

Spatial characteristics of probability distribution of longitudinal irregularity in ballasted track

J. F. Shen[1], Y. D. Xu[1], H. F. Li[1,2], F. Y. Li[1], J. X. Qiu[1] & W. Y. Liu[1]
[1]Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, China
[2]Department of Transportation Engineering, Tongji Zhejiang College, China

Abstract

Because of the discreteness of ballasted track structure, the randomness of train loads and the periodicity of maintenance, the geometric irregularity tends to be stochastic. The foundation of track management is to select a reasonable probability distribution function and use its characteristics to establish a geometric irregularity prediction model. Nowadays, the improvement in maintenance technologies considerably reduces the overrun of irregularity; however, large values of irregularity that does not overrun lead to a poor fitting effect of two-parameter probability distribution. This study used five kinds of three-parameter probability distribution functions to fit samples from six segments, calculated and compared their P values, selected the best probabilistic distribution function to describe the spatial characteristics of longitudinal irregularity in each segment, and analysed the differences in track quality between segments. The results showed that the three-parameter probability distribution functions fitted effectively to show the characteristics of unilateral thick tail of longitudinal irregularity in ballasted track. The distribution of the longitudinal irregularity in the whole line could not be fitted by a single function.
Keywords: longitudinal irregularity, spatial characteristics, unilateral thick tail, three-parameter probability distribution.

www.witpress.com, ISSN 1743-3509 (on-line)
doi:10.2495/CR160291

1 Introduction

High-speed railways are in construction in China on a large-scale. With higher speeds and heavier loads, the state of the track's quality deteriorates more quickly, which makes the maintenance very difficult. The existing prediction models are not stochastic and cannot scientifically guide railway maintenance. Therefore, it would be advantageous to investigate the following in depth: obtaining the probability distribution function of longitudinal irregularity by the mathematical statistics method and studying the stochastic development law of longitudinal irregularity by using its distribution function characteristics. Most research has provided statistical analyses of the probability distribution functions, which different track irregularity characteristic values of different lengths of segments obeyed [1, 2]. Some studies have established deterministic models to predict the deterioration of longitudinal irregularity [3–5], and, since 2010, stochastic restoration models, based on the distribution function characteristics of track irregularity characteristic values, have begun to be established [6–12]. However, most existing literature has used the same distribution function to describe the characteristics of probability distribution of irregularity and have ignored the difference in probability distribution fitting, which is caused by different line types. Zhang *et al*. [1] and Andrade and Teixeira [6] still used the same distribution function after the statistic classification of different line types, which could simplify the prediction model, but some large values of standard deviation of longitudinal irregularity still remained, which did not overrun the limit after maintenance, and the probability distribution function commonly used could not effectively fit the area including these large values.

This paper has carried out the following tasks: divided the Shanghai–Kunming railway line into six kinds of segments and selected three-parameter probability distribution functions to fit the standard deviation of longitudinal irregularity based on the fitting method of probability distribution; selected a distribution function which fitted best to describe the deviation of longitudinal irregularity in one segment by statistical analysis; explored the spatial characteristics of probability distribution of longitudinal irregularity by using the distribution function; and compared the difference in track quality between different segments.

2 Probability distribution fitting method

2.1 Original data processing

Original data processing includes data pre-processing of the track inspection car and segments' divisions of different types of lines.

2.1.1 Data pre-processing of track inspection car

This study chose data measured at a speed of more than 50km/h due to the error generated by the track inspection car at low speeds. The faster the train goes, the faster the longitudinal irregularity deteriorates, which greatly affects the tamping works; therefore, this study chose the data of standard deviation of longitudinal irregularity for statistical analysis.

2.1.2 Segments' divisions of different types of lines

The Shanghai–Kunming railway line is ballasted with a design speed of 200km/h. The operational speed of most segments is more than 160km/h, but the speed in some segments is limited to 160km/h, as the curve radius is too short or the length of the transition curves is insufficient. The values of longitudinal irregularity in these segments are too large to be consistent with the others, so they will be counted separately. The line is divided into 131 switches, 252 bridges and tunnels, 203 transition curves, 309 curves and 674 tangents for the segments where the speed is more than 160km/h.

2.2 Selecting theoretical distribution function

It is more difficult for ballasted track than ballastless track to maintain its track geometry status. There are many large values of longitudinal irregularity that, within the limit, cause the unilateral thick tail in frequency distribution histograms, namely, the data are non-negative, and large values cause the left-right asymmetry and thick tail. Therefore, it is difficult for a two-parameter distribution function to exactly fit the phenomenon of the 'peak and tail' of ballasted track, which is commonly used as the theoretical distribution function [1, 2]. As shown in Fig. 1, data of longitudinal irregularity from May 18th 2015 focuses on the average (0.73) and forms the leptokurtic. The data, whose distance from the average value is less than one standard deviation, accounts for 78.3%, which is 68% higher than the value of normal distribution. The numbers of large values of longitudinal irregularity form the right leptokurtosis and fatter tail. Three-parameter lognormal distribution is much better than normal distribution in dealing with the longitudinal irregularity of ballasted track.

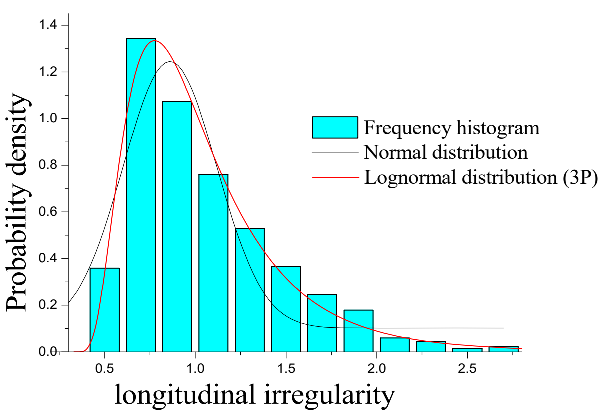

Figure 1: Longitudinal irregularity fitting by two distribution functions.

This paper used three-parameter lognormal distribution, Weibull distribution, loglogistic distribution, Dagum distribution and Burr distribution as theoretical distribution functions to conduct a goodness of fit test of longitudinal irregularity of the ballasted track with a significance level of 0.05. The probability density functions of five theoretical distributions are shown in Table 1.

2.3 Test of goodness of fit

The Kolmogorov-Smirnov test is a kind of nonparametric test based on empirical distribution functions. The number of samples of random variable X is n. X is sorted in ascending order, $x_1 < x_2 < \ldots < x_n$, and the empirical distribution function $F_n(x)$ is shown in Eq. (1). D_n is built as the test statistics, as shown in Eq. (2), $F(x,\theta)$ is the theoretic distribution function.

$$F_n(x) = \begin{cases} 0 & x < x_1 \\ i/n & x_i \ll x < x_{i+1} \\ 1 & x \gg x_n \end{cases} \quad (1)$$

$$D_n = \underset{-\infty \le x \le +\infty}{Sup} |F_n(x) - F(x,\theta)| \quad (2)$$

The P value is usually used to judge the result of hypothesis testing. When the P value is more than 0.05, the sample is considered to obey the theoretic distribution, and the bigger the P value is, the better is the fitting result. Any theoretical distribution may become the optimal fitting distribution of the detection sample when fitting multiple samples of the same segment. On the basis of counting the number of times the optimal fitting distribution is achieved among the five theoretical distributions, this study compared the average and standard deviation of the difference of the P value of the theoretic distribution function and selected the optimal distribution function to describe the characteristics of one segment.

Table 1: Probability density functions of five theoretical distributions.

Distribution function	Probability density function
Lognormal distribution (3P)	$f(x) = \exp(-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2)\Big/(x-\gamma)\sigma\sqrt{2\pi}$
Weibull distribution (3P)	$f(x) = \frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} exp\left(-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)$
Loglogistic distribution (3P)	$f(x) = \frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1}\left(1+\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{-2}$
Dagum distribution (3P)	$f(x) = \alpha \cdot k\left(\frac{x}{\beta}\right)^{\alpha \cdot k-1}\Big/\beta\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{k+1}$
Burr distribution (3P)	$f(x) = \alpha \cdot k\left(\frac{x}{\beta}\right)^{\alpha-1}\Big/\beta\left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{k+1}$

Note: k, α, β, γ, μ are parameters.

3 Calculation and analysis

3.1 Probability distribution fitting of different segments

For statistical analysis in this paper, 21 times of inspection of detection data in the Shanghai–Kunming upline (K103+000~K502+800) during 2015 are used.

3.1.1 Segments where speed<160km/h

Five theoretical distributions are used to fit the 21 times of inspection samples; the calculating results of P values are shown in Fig. 2. Weibull distribution (3P) becomes the optimal fitting distribution for 14 times of inspection. Although it does not fit best for the other seven times of inspection, the P values are all larger than 0.05, and the acceptance rate of Weibull distribution (3P) is 100%.

The average and standard deviations between the P value of Weibull distribution (3P) and the P value of the optimal fitting distribution are 0.07 and 0.16, respectively. In contrast, the average and standard deviations between the P value of lognormal distribution (3P), which becomes the optimal fitting distribution for the other seven times of inspection, and the P value of the optimal fitting distribution are 0.2 and 0.2, respectively. Also, it is feasible to use Weibull distribution (3P) to describe the stochastic characteristics of longitudinal irregularity in segments where the speed < 160km/h.

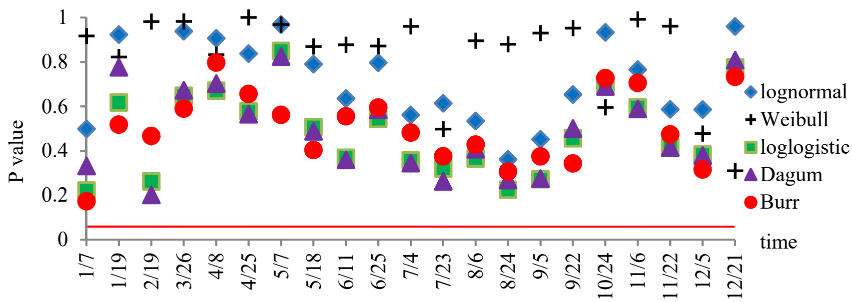

Figure 2: P value of segments where speed < 160km/h.

3.1.2 Segments in tangents

The results from fitting the 21 times of inspection samples and calculating the P values are shown in Fig. 3. Lognormal distribution (3P) becomes the optimal fitting distribution for 18 times of inspection; the P values are all larger than 0.05 in the other three times of inspection, and the acceptance rate of lognormal distribution (3P) is 100%.

The average and standard deviations between the P value of lognormal distribution (3P) and the P value of the optimal fitting distribution are 0.01 and 0.03, respectively, while the average and standard deviations between the P values of the other four theoretical distributions and the P value of the optimal fitting distribution are between among 0.25–0.65 and 0.13–0.22, respectively. Weibull distribution (3P), Dagum distribution (3P) and Burr distribution (3P) are not obeyed by some samples. It is feasible to use lognormal distribution (3P) to describe the stochastic characteristics of longitudinal irregularity in tangents.

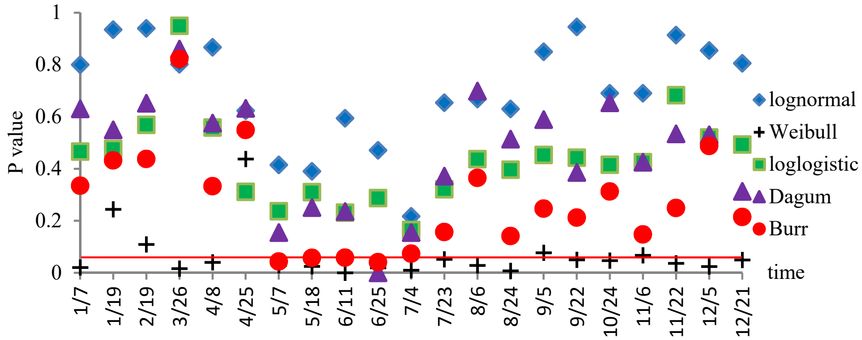

Figure 3: P value of segments in tangents.

3.1.3 Segments in transition curves

P values of samples in transition curves are shown in Fig. 4. The P values of all theoretical distributions are larger than 0.05, and all the 21 times of inspection samples obey them.

Dagum distribution (3P) becomes the optimal fitting distribution for 10 times of inspection, while the standard deviation between the P value of Dagum distribution (3P) and the P value of the optimal fitting distribution is 0.21, as shown in Table 2, which is the largest of all theoretical distributions. The P values of samples on June 25th and July 4th are 0.09 and 0.24, respectively, for Dagum distribution (3P); they lead to a large difference in standard deviation between the P value of Dagum distribution (3P) and the P value of the optimal fitting distribution. The standard deviation between the P value of Dagum distribution (3P) and the P value of the optimal fitting distribution will be 0.04 if the two P values on June 25th and July 4th are included, and Dagum distribution (3P) can be used to describe the stochastic characteristics of longitudinal irregularity in transition curves.

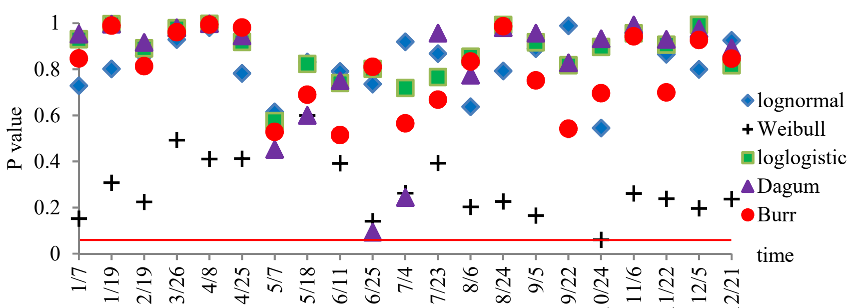

Figure 4: P value of segments in transition curves.

Table 2: Average and standard deviations of differences in P values of theoretic distribution functions in transition curves.

Distribution / Statistical item	Lognormal distribution (3P)	Weibull distribution (3P)	Loglogistic distribution (3P)	Dagum distribution (3P)	Burr distribution (3P)
Times of inspection to be optimal	6	0	3	10	2
Standard deviation	0.10	0.61	0.05	0.10	0.13
Mean	0.11	0.20	0.06	0.21	0.13

3.1.4 Segments in curves

P values of samples in curves are shown in Fig 5. Loglogistic distribution (3P) becomes the optimal fitting distribution for 12 times of inspection, and all the 21 times of inspection samples obey it.

The average and standard deviations between the P value of loglogistic distribution (3P) and the P value of the optimal fitting distribution are 0.03 and 0.05, respectively, while the average and standard deviations between the P value of the other four theoretical distributions and the P value of the optimal fitting distribution are between 0.1–0.8 and 0.09–0.24, respectively. It is feasible to use loglogistic distribution (3P) to describe the stochastic characteristics of longitudinal irregularity in curves.

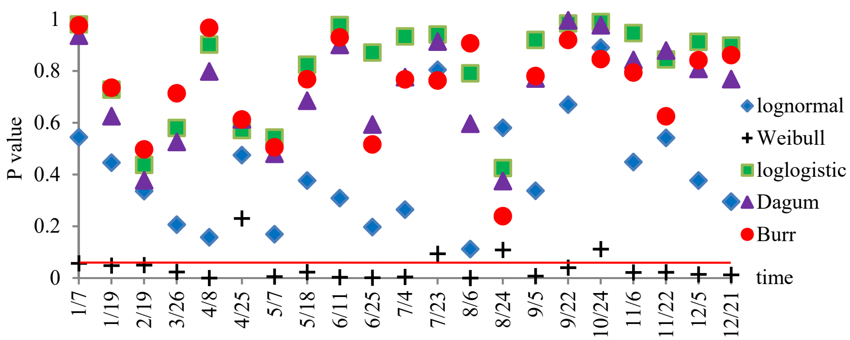

Figure 5: P value of segments in curves.

3.1.5 Segments in switches

Burr distribution (3P) becomes the optimal fitting distribution for seven times of inspection, and all of the samples obey it; the P values of the samples in switches are shown in Fig. 6.

The average and standard deviations between the P value of Burr distribution (3P) and the P value of the optimal fitting distribution are 0.07 and 0.09, respectively, while the average and standard deviations between the P value of the other four theoretical distributions and the P value of the optimal fitting distribution are between 0.07–0.28 and 0.08–0.28, respectively. The average and standard deviations among the theoretical distributions are not large. In particular, the standard deviation between the P value of Dagum distribution (3P) and the P value of the optimal fitting distribution is 0.08. It is smaller than Burr distribution (3P), while it becomes the optimal fitting distribution for only four times of inspection. It is feasible to use Burr distribution (3P) to describe the stochastic characteristics of longitudinal irregularity in switches for the reason that it is the optimal fitting distribution for most times of inspection.

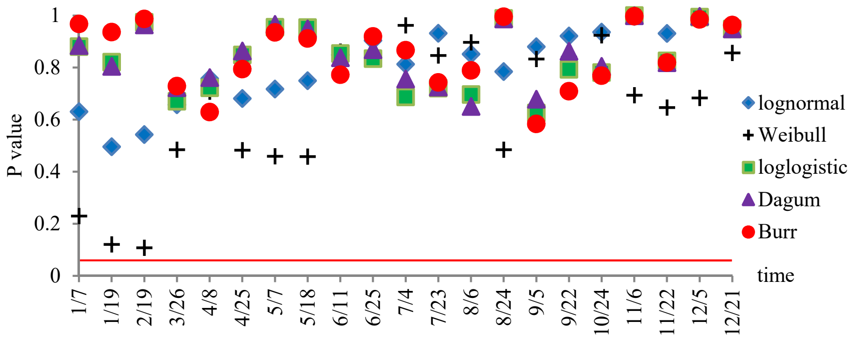

Figure 6: P value of segments in switches.

3.1.6 Segments in bridges and tunnels

The P values of the samples in bridges and tunnels are shown in Fig. 7. Lognormal distribution (3P) becomes the optimal fitting distribution for six times of inspection. The standard deviation between the P value of lognormal distribution (3P) and the P value of the optimal fitting distribution is 0.11, as shown in Table 3. It is the smallest of all theoretical distributions but it is not obeyed by the three samples from March 26th, December 5th and December 21st.

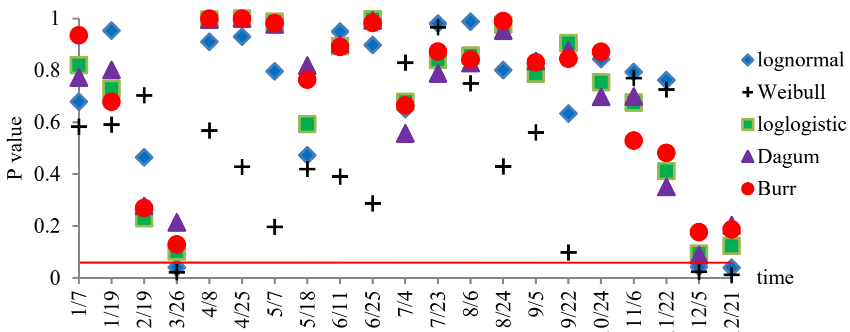

Figure 7: P value of segments in bridges and tunnels.

The average and standard deviations between the P value of Burr distribution and the P value of the optimal fitting distribution are the smallest among the other four theoretical distributions. While it becomes the optimal fitting distribution for only five times of inspection, Burr distribution (3P) can be used to describe the stochastic characteristics of longitudinal irregularity in bridges and tunnels after comprehensive comparison.

Table 3: Average and standard deviations of differences in P values of theoretic distribution functions in bridges and tunnels.

Distribution / Statistical item	Lognormal distribution (3P)	Weibull distribution (3P)	Loglogistic distribution (3P)	Dagum distribution (3P)	Burr distribution (3P)
Times of inspection to be optimal	6	2	4	4	5
Standard deviation	0.12	0.32	0.12	0.11	0.09
Mean	0.11	0.26	0.12	0.13	0.12

3.2 Comparison of track quality among different segments

The spatial characteristics of longitudinal irregularity can be analysed based on the optimal probability distribution function; thus, the track quality of different segments can be compared.

The cumulative probability distribution curves of six different segments in the Shanghai–Kunming upline on July 23rd, 2015, are shown in Fig. 8. The larger the standard deviation of longitudinal irregularity is, the worse the track quality is. As shown in Fig 8, if cumulative probability distribution curves can quickly reach 1.0, the track quality of the corresponding segment will be better. The track quality of the segments in bridges and tunnels is the best, attributed to the great structural integrity and large lateral resistance. What is more, over 70% of the segments of bridges and tunnels have been maintained. The track quality of segments where speed < 160km/h is the worst, with many large values, proving that it should not be calculated with the other segments. The track quality of the segments in curves,

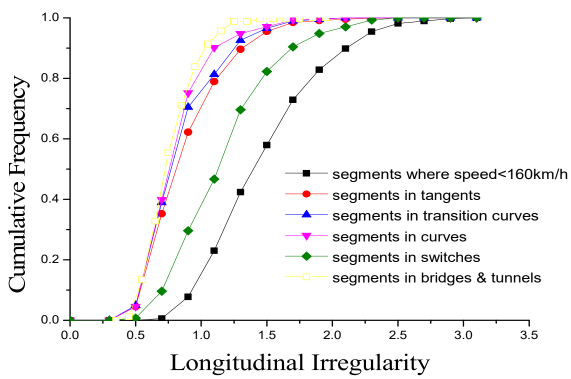

Figure 8: Cumulative probability distribution curves of six different segments in the Shanghai–Kunming upline on July 23rd, 2015.

transition curves and tangents is between the two segments mentioned above, and the level of quality decreased gradually, although the difference is not large; the track quality of the segments' switches is the second worst because machines have difficulty entering the area because of its special structure, so it is difficult to maintain.

4 Conclusions

The main work of this research is outlined below:

1. There are many large values of longitudinal irregularity owing to the structural characteristics of the ballasted track, causing the phenomenon of the right leptokurtosis and fatter tail. The three-parameter probability distribution function can effectively solve the problem of deviation of the actual value in large values area. It has a better fitting effect on the right leptokurtosis and fatter tail than two-parameter probability distribution.
2. Different kinds of segments have different probability distributions of standard deviation of longitudinal irregularity and need different optimal probability distribution functions to describe the stochastic characteristics of longitudinal irregularity. In the Shanghai–Kunming upline, longitudinal irregularity in segments where speed<160km/h obeyed three-parameter Weibull distribution; where speed≥160km/h, it obeyed three-parameter lognormal distribution in tangents, three-parameter loglogistic distribution in curves, Dagum distribution in transition curves, and Burr distribution in switches, bridges and tunnels, respectively.
3. The quality of longitudinal irregularity in different segments can be evaluated by analysing the spatial characteristics of the probability distribution that they obeyed. The order of longitudinal irregularity in the Shanghai–Kunming upline was bridges and tunnels, curves, spirals, tangents, switches and segments where speed<160km/h.

References

[1] Zhang, Y., M. El-Sibaie and S. Lee, FRA track quality indices and distribution characteristics, in AREMA 2004 Annual Conference. 2004: Nashville, TN.
[2] El-Sibaie, M. and Y. J. Zhang, Objective track quality indices. Proceedings of 83rd TRB Annual Meeting, January 11–15, 2004, Washington, D.C.
[3] Oyama, T. and M. Miwa, Mathematical modeling analyses for obtaining an optimal railway track maintenance schedule. Japan Journal of Industrial and Applied Mathematics, 2006. 23(2): pp. 207–224.
[4] Sadeghi, J., Development of railway track geometry indexes based on statistical distribution of geometry data. Journal of Transportation Engineering – ASCE, 2010. 136(8): pp. 693–700.
[5] Jianxi, Wang, Li, Haifeng and Xu, Yude, Evaluation and prediction method for railway track geometric state based on probability distribution change [J]. China Railway Science, 2008. 29(5): pp. 31–34.

[6] Andrade, A.R. and P.F. Teixeira, Uncertainty in rail-track geometry degradation: Lisbon–Oporto line case study. Journal of Transportation Engineering – ASCE, 2011. 137(3): pp. 193–200.

[7] Quiroga, L.M. and E. Schnieder, Monte Carlo simulation of railway track geometry deterioration and restoration. Proceedings of the Institution of Mechanical Engineers Part O – Journal of Risk and Reliability, 2012. 226(O3): pp. 274–282.

[8] Quiroga, L. and E. Schnieder, Modelling of high speed railroad geometry ageing as a discrete-continuous process, in Stochastic Modeling Techniques and Data Analysis International Conference. 2010: Chania, Crete, Greece.

[9] Quiroga, L. and E. Schnieder, Heuristic Forecasting of Geometry Deterioration of High Speed Railway Tracks. 2012, Springer-Verlag: Berlin. pp. 609–616.

[10] Vale, C. and S.M. Lurdes, Stochastic model for the geometrical rail track degradation process in the Portuguese Railway Northern Line. Reliability Engineering & System Safety, 2013. 116: pp. 91–98.

[11] Andrade, A.R. and P.F. Teixeira, Hierarchical Bayesian modelling of rail track geometry degradation. Proceedings of the Institution of Mechanical Engineers Part F – Journal of Rail and Rapid Transit, 2013. 227(4SI): pp. 364–375.

[12] Audley, M. and J.D. Andrews, The effects of tamping on railway track geometry degradation. Proceedings of the Institution of Mechanical Engineers Part F – Journal of Rail and Rapid Transit, 2013. 227(4SI): pp. 376–391.