Statistical Description of Wheel Roughness

G. Squicciarini[1], M.G. Toward[1], D.J. Thompson[1] and C.J.C Jones[2]

[1] Institute of Sound and Vibration Research, University of Southampton

University Road, Southampton SO17 1BJ, UK

E-mail: g.squicciarini@soton.ac.uk

[2] 78 Cliffe Road, Glossop, Derbyshire SK13 8NT, UK

Summary

Wheel roughness measurements available from several different campaigns are presented in terms of average levels and dispersion. The dependence on factors such as brake type and whether the wheel is powered or trailing is also addressed. A method to decide how many wheels from a train are to be measured is then presented. Finally, the main outcomes are described from a round robin test aimed at assessing the effect on wheel roughness measurements of adopting different equipment, used independently by different teams,.

1   Introduction

The role played by the combined wheel and rail roughness in generating rolling noise is of fundamental importance [1]. Along with the mobilities of the rail, of the wheel and of the contact, the roughness can be used in estimating the wheel-rail contact force. The motion of a wheel rolling along a track generates a force with an amplitude that is proportional to the roughness amplitude and at a frequency given by the ratio between the speed of the wheel centre and the roughness wavelength. A method to measure and analyse rail roughness is described in the standard EN15610:2009 [2]; however no equivalent standard exists for the measurement of wheel roughness.

The EU project ACOUTRAIN is exploring the possibility of introducing some elements of virtual testing in the acoustic certification process of new rolling stock. In doing so, it is important to estimate the effect of uncertainty and variability. For example, in rolling noise prediction models, variability in measured wheel roughness will result in variability in the predictions of rolling noise. A measurement procedure has been proposed aimed at reducing this variability [3]. Construction of a database of measured wheel roughness is also desirable. Such a database can be used in the first instance to estimate the statistical distribution of roughness. Moreover, it can be used to identify the important factors influencing wheel roughness, e.g. braking system and powered or unpowered wheels. Finally, since it is not always possible to measure all the

wheels of a train, the database could be used to estimate the number of wheels required to be measured to obtain a representative sample for a whole train.

The paper will first describe wheel roughness data measured in the last two decades among the ACOUTRAIN project partners in terms of statistical properties. The question of how many wheels should be measured from a particular train is addressed in Section 3. Finally, Section 4 is devoted to describing the results of a round robin test performed within the ACOUTRAIN project to assess the effect on wheel roughness measurements of adopting different equipment used independently by different teams.

2   Wheel Roughness Description

2.1   Description of the Database

Wheel roughness data are currently available from seven campaigns between 1999 and 2012 (see Table 1) and are continuously being updated. In total, data from 310 wheels have been input into the database. The database is implemented with MATLAB objects, and the main wheel properties are described in Table 2.

Table 1. Summary of available measurement campaigns with corresponding number of wheels

Campaign	Vehicle	Disc brakes	Cast iron brakes	Composite brakes
A	Regional	64	-	-
B	Regional	26	-	-
C	Regional	16	-	-
D	Freight	32	28	-
E	Freight	-	8	8
F	Freight	-	16	48
G	Regional	-	32	-

Table 2. Summary of database wheel properties

Property	Database stored data
Type of braking system	Disc, cast iron, composite or metal sinter
Tag for traction identification	Powered, trailer
Measurement campaign information	Campaign name, train type, etc.
Mileage since last reprofiling	Not always available
Position of the measured line(s)	e.g. 60-70-80 mm from the flange back
Acoustic roughness level	One-third octave band data
Out of roundness	Harmonics 1-10 included where available
One-third octave band wavelength	Wavelengths

These campaigns took place over a time of 15 years and they were performed by different people with different equipment and with different post-processing software, thus adding a certain amount of additional variability, as will be discussed in Section 4. As a consequence, the wavelength range is not always

consistent among the campaigns; some of them in fact do not report the very long and/or the very short wavelengths.

2.2   Analysis Procedure

A single analysis is performed on a subset of N wheels and each subset is identified by certain factors. Normally, each wheel was measured along three different running lines. However, only the line corresponding to the nominal contact point (70 mm from the flange back) is considered here.

The quantity observed is the wheel roughness level expressed in decibels as:

$$L_r = 10\log_{10}\left(r_{rms}/r_{ref}\right)^2; \quad r_{ref} = 1\mu\text{m} \tag{2.1}$$

where r_{rms} is the root mean square value of the measured roughness in a particular one-third octave band. The statistical analysis is based entirely on processed data in one-third octaves after application of suitable pre-processing such as removal of pits and spikes [2, 3].

T-Tests [4] are used to compare different sets of data to see if the factors characterising the set are sufficient to indicate a significant difference in the population. The brake type is known to have a significant effect on wheel roughness; statistical tests have confirmed this but results are not reported here. The t-test has also been performed to compare trailer wheels with powered wheels or to compare different campaigns. Once a set is defined, for each wavelength, a Lilliefors normality test [5] at 5 % significance level is used to understand the likelihood of the data being normally distributed.

To summarize the variability of measured wheel roughness data, the measurements of each set will be presented in terms of the arithmetic mean of the levels, the energy average (as the noise level is expected to depend on the energy average contribution of each wheel) and of width of the distribution defined by the 15.8[th] and the 84.2[th] percentile, corresponding, for a normal distribution, to one positive and one negative standard deviation relative to the mean.

2.3   Powered versus Trailer Wheels

The wheels of campaign A of Table 2 are taken from a single type of EMU, with equal numbers of powered and trailer wheels. A t-test has been used to test the null hypothesis that two sub-sets of data (powered versus trailer wheels) are independent random samples from normal distributions with equal means and equal but unknown variances. This shows that for all the wavelengths, apart from one, the null hypothesis of the data belonging to the same population cannot be rejected. This implies that there is no significant difference between trailer and powered wheels. Other analyses on different subsets of data have confirmed this.

The results for the two subsets are summarised in Fig. 1 showing the energy average, the mean value and the standard deviation corridor computed in terms of percentiles as explained above. A standard limit curve for rail roughness [6] is shown for reference. Clearly the average levels and variability of the powered and trailer wheels are similar.
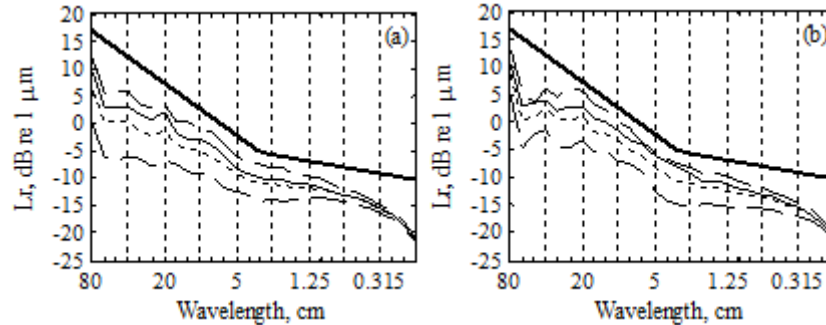
Fig. 1. Roughness levels for disc brake wheels from campaign A. (a): powered; (b): trailer. ———: [6]; ———: energy mean; ·······: sample mean; – – –:± one standard deviation corridor
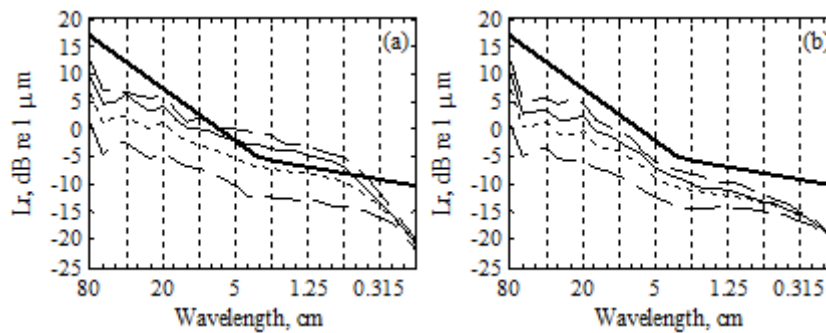

Fig. 2. Roughness levels disc brake wheels. (a): all campaigns; (b): campaign A. ———: [6]; ———: energy mean; ·······: sample mean; – – –:± one standard deviation corridor

2.4   Disc Braked Wheels

For sake of simplicity one may include all 69 wheelsets with disc brakes in one single set and analyse it as a whole. However, there might arguably be other important factors to consider and the campaigns may be a subset of different populations defined, for example, by the mileage. Unfortunately the information of mileage since last re-profiling is only available for campaign B. However, a t-test between the sets defined by two campaigns at a time can be performed. By comparing Campaign A with B the null hypothesis that the two subsets are from the same population must be rejected for 16 wavelength bands out of 30, and the two campaigns are likely to be part of two different populations. The same test performed between other pairs of campaigns confirms this.

Despite this outcome all the disc-braked wheels are presented as one single set in Fig. 2a, and then the same analysis has later been repeated for each single campaign (but it is not reported here). In fact, for the purpose of producing input data for a Monte Carlo analysis, it is more convenient to have one single mean value and one single standard deviation describing all the disc-braked wheels. The presence of more campaigns will then result in an increased estimated variance of

the considered population, but this might not be normally distributed. A normality test on the entire set of available disc braked wheels rejects the null hypothesis that the data is normally distributed for 15 wavelengths bands out of 30. Fig. 2a summarises average levels and standard deviation distributions for these wheels.

As an example Fig. 2b shows the same type of analysis performed on wheels with disc brakes from campaign A only. By comparing Fig. 2a with Fig. 2b it can be observed how the variability increases if several campaigns are analysed together and how the energy mean can assume significantly different values (up to 7-8 dB in this example).

2.5   Wheels with Cast-Iron Block Tread Brakes

Fig. 3a presents average wheel roughness levels and variability for all 84 wheels with cast-iron brake blocks. In this case not all the datasets are available over the same wavelength range; therefore the variability appears not to be consistent, being quite low at short wavelengths. Again, introducing more data in the analysis results in a wider range of variability. The null hypothesis of the data being normally distributed is to be rejected for seventeen wavelength bands out of thirty. It can be observed that the shape of the roughness curve as a function of wavelength shows a peak at around 5 cm. In three cases a similar behaviour can be found by considering the campaigns one at time. However, both the normality test (null hypothesis is often to be rejected) and the t-test between two sets defined by the campaigns suggest that it is not straightforward to consider all the available wheels with cast-iron brake blocks as members of one single population. Again, other factors, such as mileage or running condition, might be too important to be discarded.

Fig. 3b presents average roughness levels for wheels with cast-iron brake blocks of campaign G. These measurements were on a single vehicle type and consequently show a much smaller variability.
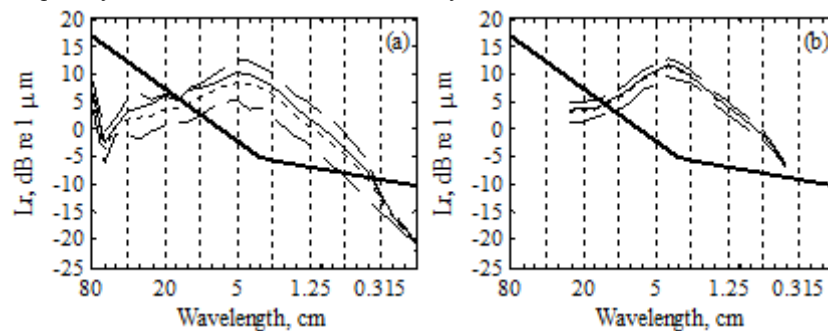

Fig. 3. Roughness levels wheels with cast-iron brakes block. (a): all campaigns; (b): campaign G. ——: [6]; ——: energy mean; ·······: sample mean; − − −:± one standard deviation corridor

2.6   Wheels with Composite Brake Blocks

Fig. 4a presents average wheel roughness levels and variability for all 56 wheels with composite brake blocks. The null hypothesis of the data being normally

distributed is to be rejected for only three wavelength bands out of thirty. The strange difference in terms of standard deviation above and below 0.2 m is, again, because the two campaigns were not post-processed over the same wavelength range, therefore the wavelengths between 0.2 and 0.8 m are here represented by only eight wheels from campaign E. A similar phenomenon occurs at wavelengths shorter than 0.003 m.

Wheels with composite brake blocks from campaign F are presented in Fig. 4b. Measurements in this campaign were taken from a train composed of different vehicle types, which is the reason for the high variability shown.
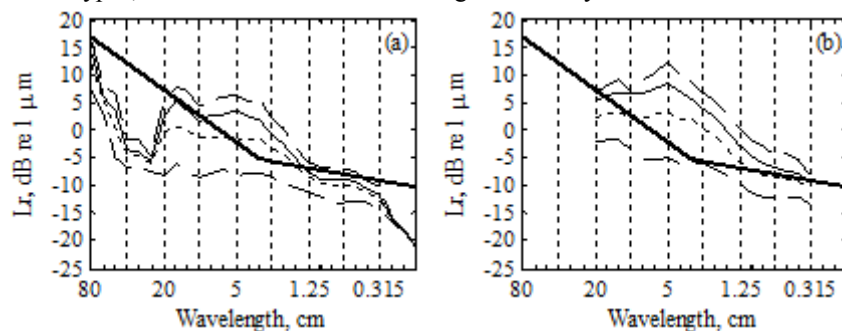

Fig. 4. Roughness level, wheels with composite brakes. (a): all campaigns; (b): campaign F. ▬▬: [6]; ——: energy mean; ·······: sample mean; – – –:± one standard deviation corridor

3   Estimation of Sample Size

When planning wheel roughness measurements an important decision is the number of wheels to be measured necessary to give a representative evaluation of a whole train. Measuring all the wheels will give the exact representation of the specific train under test, but this might not always be possible. In the final report of the NOEMIE project [7] a suggestion was given for the minimum number and positions of wheels to be measured in a campaign. However, this was based on experience rather than on statistical considerations. The aim of this section is to add some statistical considerations which may also be taken into account in deciding the number of wheels that would be representative of a specific set.

Considering the simplest example of a population whose variance, σ^2, is known a priori, the number of samples n to be measured in order for the error between the sample mean and the real mean of the population not to exceed the amount E, with confidence 1-α, is:

$$n = \left(z_{\alpha/2}\sigma/E\right)^2 \tag{3.1}$$

where $z_{\alpha/2}$ is the 'standard normal deviate', or 'standard normal random variable' [4], corresponding to an upper tail of $\alpha/2$ (for example: $z_{0.025}$=1.96).

Where the variance of the population is not known, the Student T distribution should be used instead of the normal one and an estimator for the variance should be adopted. However, as the Student T distribution itself is dependent on the

sample size, n, a closed form cannot be found (see e.g. [4]). Instead, the information acquired from the previous sections can be used to estimate the value of the variance of each population (disc brakes, cast-iron brake blocks etc.) and such an estimator can be used to replace the actual value of σ. This can be either the upper bound of the standard deviation confidence interval or the sample standard deviation itself.

A calculation for n should be done for each wavelength and the greatest value will give the actual sample size. Eq. (3.1) should be applied only at those wavelengths where data appear to be normally distributed.

To give an example, for a hypothetical campaign on wheels with disc brakes (see Fig. 2a), by setting $E = 4$ dB and using the upper bound of the standard deviation confidence interval (6.5 dB) the number of wheels to be measured should be 10, in accordance with [7]. Adopting the lower bound of the standard deviation (4.9 dB) would result in 6 wheels to be measured.

4   ACOUTRAIN Round Robin Test

Wheel roughness measurements have been carried out independently by four partners of the ACOUTRAIN project for the purpose of comparing five systems available in the consortium (one of the partners contributed with two systems) and the post-processing methodology used. Two freight vehicles were made available; both were tread braked, one equipped with cast-iron brake blocks, the other with composite brake blocks. The measurements and the post-processing of data were made by following the same procedure [3].
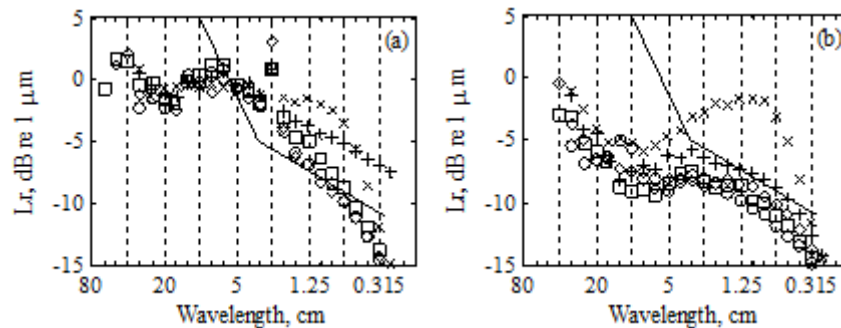

Fig. 5. Wheel roughness from the ACOUTRAIN project test. (a): cast-iron brakes; (b): composite brakes. ——: [6]; ○: P1; □: P2; +: P3; ×: P4; ◊: P5

Measurements made by the five teams (P1 to P5) are shown in Fig. 5a for the vehicle with cast-iron brakes and in Fig. 5b for the vehicle with composite brakes. Lines P1, P2 and P5 agree, apart from two one-third octave bands, within 2.5 dB in the entire range for both the brake types. Measurements labelled as P3 and P4, to differing extents, represent anomalous values, indicating either a malfunction of the equipment and/or incorrect data post-processing.

5   Conclusions

With the aim of estimating the effect of uncertainty and variability in preparing input data for rolling noise models, this paper describes, in terms of the average and standard deviation, wheel roughness data for seven different campaigns collected in a database. By addressing their statistical distribution, considering one campaign at a time, it was shown to be likely that the roughness level is normally distributed. If all the available campaigns are analysed at once, however, the likelihood of data being normally distributed decreases. However, the likelihood is expected to increase again by introducing into the database more measurements, although, as a side effect, the estimated deviation from the average would increase.

A simple method to estimate the minimum number of wheels necessary to be measured to represent, with enough confidence, the whole train has been addressed. The method has been shown to give results in agreement with previous research projects [7].

Finally, a round robin test performed within the ACOUTRAIN project, has shown that, in the particular case considered, the amount of measurement variability due to different equipment used by different teams following the same protocol can be estimated to be within 2.5 dB, provided that no major equipment malfunctions or post-processing errors occur.

Acknowledgement

References

[1] Thompson, D.J.: Railway noise and vibration: Mechanisms, modelling and means of control, 1st Edition. Elsevier Science, 2009.

[2] European Standard EN 15610:2009.

[3] Thompson, D.J.: Proposed analysis method for wheel roughness. Deliverable 2.4 for ACOUTRAIN project 2013.

[4] Montgomery, D.C.: Design and analysis of experiments, 8th Edition. John Wiley & Sons, 2012.

[5] Lilliefors, H.W.: On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association 62 (318), 399-402, 1967.

[6] European Standard EN ISO 3095:2013.

[7] Fodiman, P.: NOEMIE Project Final report (Project 2002/EU/1663), AEIF, July 2005, see Appendix D.