TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

ACTA TECHNICA NAPOCENSIS

Series: Applied Mathematics, Mechanics, and Engineering
Vol. XX, Issue xx, Month, 20xx

ALTERNATIVE USAGE OF MAXIMUM LIKELIHOOD METHOD TO ESTIMATE THE RAILWAY WHEELSET RELIABILITY PROBABILISTIC MODEL

Marius-Adrian SPIROIU, Sorin ARSENE

Abstract: *The paper proposes an alternative usage of the Maximum Likelihood Method to estimate the parameters of the probabilistic reliability model using a reduced amount of experimental data. The proposed method involves dividing the operating period into time intervals and using the number of failures in each interval and interval limits instead of considering individual failure times. The classical Maximum Likelihood Method is adapted according to these hypotheses. The resulted alternative method is applied to estimate the wheelset reliability probabilistic model, based on operational data of freight cars. The parameters of the model are estimated under several hypotheses regarding the time intervals. The results are evaluated and the influence on the estimates of the assumptions adopted within the method is identified.*
Key words: *reliability, estimation, Weibull parameters, Maximum Likelihood Method, railway wheelset*

1. INTRODUCTION

In railway terminology, a wheelset designates the rigid assembly between the wheels and the axle. According to the vehicle's type, some other elements such as brake discs, gear wheels, etc., may be also mounted on the axle. In railway vehicles, the role of the wheelset is of particular importance compared to that of the axles and wheels of other land transportation systems. This is because besides the traditional functions, common to all land transportation systems, related to supporting vehicle loads, its running, traction and braking, the wheelset additionally ensures the guidance of the vehicle in the track. Taking into account the specificity of the service provided by railway vehicles, the failure to perform any of these functions may result in accidents with important social and/or economic or even catastrophic implications.

To ensure safety against derailment, the geometry of the wheel rolling profile is extremely important. In this regard, there are several dimensions of the wheel profile that must be met. Consequently, a wheelset is considered defective in the case of a deviation from the prescribed limit of any of the regulated dimensions of the wheel [1].

Considering all the above, the wheelset is the most critical component of railway vehicles in terms of reliability. Given the importance from the point of view of traffic safety but also the complexity of the specific phenomena of the wheelset-track interaction, there is a large scientific interest regarding the reliability of the wheels and wheelsets and their failure mechanisms. One of the issues often studied is the fatigue cracking. Some scientific articles in this field evaluate the impact on railway wheels reliability of crack initiation and propagation [2], or address the cracking of the wheel rim, considering the structure and properties of the material [3]. There are also studies, such as [4], regarding the wheel-rail contact related phenomena, as the main cause of wheels wear and damages.

There are researches based on railway cars operational data, which results in statistics of wheelset failures and in the estimation of a reliability model based on Weibull distribution

[1], or in more detailed analysis of the wheel critical failure modes, including failure rates monotony analysis and estimation of individual Weibull reliability models [5]. A comparison between estimation methods of the wheelset reliability mathematical model, based on Weibull distribution is made in [6], the considered methods being the Maximum Likelihood Method and three variants of the Regression Method.

The aim of this paper is to estimate the reliability mathematical model of the freight car wheelset, based on real operational data, by using a modified version of the Maximum Likelihood Method. This original alternative approach to the classical estimation method allows a simplification of the observation or testing procedure.

2. ESTIMATION OF RELIABILITY PROBABILISTIC MODEL

2.1 Theoretical probability distribution

Estimating a reliability mathematical model involves using a theoretical probability distribution and finding its parameters so that the model fits the experimental data. Regarding the theoretical probability distribution, in the field of reliability it is widely used the Weibull distribution. Experience has shown that it is very suitable for modelling the failure law for industrial products. The reason for this is the intrinsic versatility of the Weibull probability law. The reliability equation according to the two-parameter Weibull distribution law is given by:

$$R(t) = \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \tag{1}$$

where $R(t)$ is the reliability, t represents the time, β is the shape parameter, and η the scale parameter.

2.2 Classical methods of estimation for the reliability mathematical model parameters

There are three classical methods that are widely used for the estimation of the reliability mathematical model parameters: the Moments Method, the Regression Method and the Maximum Likelihood Method.

The Moments Method is based on the statistical moments of the empirical and the theoretical distributions. Within the Regression Method the estimation of the Weibull distribution parameters is done by using linear regression and the Least Squares method.

The principle of the Maximum Likelihood Method, to which the present paper refers, is to find the most likely estimators (i.e., which are maximizing the probability of occurrence of the empirical data) for the unknown parameters of a distribution.

The Maximum Likelihood Method can be used in the case of complete data sets - when the failure time is known for every unit of the observed sample, but also for incomplete (censored) data sets - when the failure time is known only for some of the units.

In the situation of a complete data set the likelihood function is given by [7]:

$$\begin{aligned} &L(x_1, x_2, ..., x_n; \theta) = \\ &= f(x_1;\theta)\cdot f(x_2;\theta)\cdot ...\cdot f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta) \end{aligned} \tag{2}$$

θ being the parameter to be estimated (there may be several), $x_1, x_2, \ldots x_n$ the experimental data and f the probability density function of the considered distribution.

The function in equation (2) is the product of the probability densities of all experimental data of the sample, thus gives the probability of observing the data set $x_1, x_2, \ldots x_n$ for a given θ. The estimator with the highest probability of having generated the given data set is the one for which the likelihood function is maximum, so it is given by the solution of the equation [7]:

$$\frac{dL(x_1, x_2, ..., x_n; \theta)}{d\theta} = 0 \tag{3}$$

To ease the calculations, it is common practice to use the natural logarithm of the likelihood function instead of the function itself. The following equation is therefore used instead of equation (3) [7]:

$$\frac{d \ln L(x_1, x_2, ..., x_n; \theta)}{d\theta} = 0 \tag{4}$$

In the hypothesis of a censored data set, when only k out of the n units have failed ($k<n$), thus only the first k failure times are known, the likelihood function is given by [7]:

$$L\left(x_1, x_2, ..., x_n; \theta\right) = \prod_{i=1}^{k} f\left(x_i; \theta\right) \prod_{j=1}^{n-k} R\left(x_j; \theta\right) \quad (5)$$

R being the reliability function.

Applying the Maximum Likelihood Method to estimate the Weibull distribution parameters requires the use of its corresponding probability density function, given by:

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \quad (6)$$

thus, the likelihood function for a censored data set of k failure times ($t_1, t_2, \dots t_k$) given by equation (5) becomes:

$$L\left(t_1, t_2, ..., t_k; \beta, \eta\right) = \\ = \prod_{i=1}^{k} \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t_i}{\eta}\right)^{\beta}\right] \prod_{j=1}^{n-k} \exp\left[-\left(\frac{t_j}{\eta}\right)^{\beta}\right] \quad (7)$$

Applying the natural logarithm to (7):

$$\ln L\left(t_1, t_2, ..., t_k; \beta, \eta\right) = k \ln\beta - k\beta \ln\eta + \\ +(\beta-1)\sum_{i=1}^{k} \ln t_i - \sum_{i=1}^{k}\left(\frac{t_i}{\eta}\right)^{\beta} - \left(n-k\right)\left(\frac{t_k}{\eta}\right)^{\beta} \quad (8)$$

where t_k – the last observed failure time – is the censoring time, thus the time assigned to the *(n-k)* surviving units at time t_k.

The maximum conditions are:

$$\frac{\partial \ln L\left(t_1, t_2, ..., t_k; \beta, \eta\right)}{\partial \beta} = 0 \\ \frac{\partial \ln L\left(t_1, t_2, ..., t_k; \beta, \eta\right)}{\partial \eta} = 0 \quad (9)$$

The conditions in equations (9) lead, after some calculations, to the expressions of the estimators for Weibull parameters β and η [7]:

$$\hat{\eta} = \left(\frac{\sum_{i=1}^{k} t_i^{\hat{\beta}} + \left(n-k\right) t_k^{\hat{\beta}}}{k}\right)^{1/\hat{\beta}} \quad (10)$$

$$\frac{1}{\hat{\beta}} + \frac{1}{k}\sum_{i=1}^{k} \ln t_i - \frac{\sum_{i=1}^{k} t_i^{\hat{\beta}} \ln t_i + (n-k) t_k^{\hat{\beta}} \ln t_k}{\sum_{i=1}^{k} t_i^{\hat{\beta}} + (n-k) t_k^{\hat{\beta}}} = 0 \quad (11)$$

To determine the two estimates in equations (10) and (11) an iterative procedure must be performed, by considering an initial value for the shape parameter in equation (11) and by adjusting it until the expression becomes null. Once the shape parameter is determined in this manner, the scale parameter is obtained from equation (10).

2.3 Alternative Maximum Likelihood Method

As mentioned before, the Maximum Likelihood Method requires the values of the probability density function at each failure time, so the moment of failure must be known for every tested or observed unit. This is not very convenient in the case of units observed in real operation, especially when it comes to large samples.

For this reason, the idea of the present paper is to use the Maximum Likelihood Method principle to estimate the Weibull parameters without being necessary to know all the failure times of the observed units. It is proposed instead to divide the testing period into time intervals and to apply the method using as data the limits of these intervals and the number of failures in each interval. This would simplify the procedure of obtaining data during the entire testing or observation period.

In the hypothesis of a censored data set, when for the n observed units only the first k failure times are known ($k<n$), considering that the observation period is divided into m intervals and k_i is the number of failures in interval i, $i=1,2,\dots m$, with

$$\sum_{i=1}^{m} k_i = k \quad (12)$$

the likelihood function in equation (5) can be written in this case:

$$L(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{m} \left[f(x_i; \theta) \right]^{k_i} \prod_{j=1}^{n-k} R(x_j; \theta) \quad (13)$$

Equation (8) becomes

$$\ln L(t_1, t_2, ..., t_m; t_k; \beta, \eta) = k \ln \beta - k\beta \ln \eta +$$
$$+(\beta - 1)\sum_{i=1}^{m} k_i \ln t_i - \sum_{i=1}^{m} k_i \left(\frac{t_i}{\eta}\right)^{\beta} - (n-k)\left(\frac{t_k}{\eta}\right)^{\beta} \quad (14)$$

where $t_1, t_2, \ldots t_m$ are the times assigned to the m intervals and t_k is the censoring time.

The estimators for β and η are given by:

$$\hat{\eta} = \left(\frac{\sum_{i=1}^{m} k_i t_i^{\hat{\beta}} + (n-k) t_k^{\hat{\beta}}}{k} \right)^{1/\hat{\beta}} \quad (15)$$

$$\frac{1}{\hat{\beta}} + \frac{1}{k}\sum_{i=1}^{m} k_i \ln t_i - \frac{\sum_{i=1}^{m} k_i t_i^{\hat{\beta}} \ln t_i + (n-k) t_k^{\hat{\beta}} \ln t_k}{\sum_{i=1}^{m} k_i t_i^{\hat{\beta}} + (n-k) t_k^{\hat{\beta}}} = 0 \quad (16)$$

Within the proposed alternative method, the difference from the classic Maximum Likelihood Method is that the time corresponding to an interval is assigned to all failures occurring in that interval instead of considering individual failure times.

This approach leads to changes in the formulas used in the estimation method – see equations (13) … (16), where the summations and the products are made for the m intervals and each interval time t_i is associated with a multiplication factor k_i corresponding to the number of failures in interval i.

3. METHODOLOGY

This paper proposes the estimation of the railway wheelset reliability model based on an original alternative usage of the Maximum Likelihood Method. To apply the previously described estimation method experimental data are required. The data used in this paper originates from the real operation of freight wagons wheelsets in the period of their useful (normal) lifetime. The sample size was n=1802 units, and the number of failures observed during almost a year (360 days) was k=190.

Since the method involves dividing the observation period into intervals and using the corresponding times, for the application of the method it is necessary to decide:

- the size of the interval;
- the time corresponding to each interval.

The first problem requires deciding how to divide the observation period. Assuming the natural hypothesis of equal intervals, the problem comes down to deciding the duration of an interval. For the latter, two solutions may be considered, namely:

- to use the interval upper limit;
- to use the interval midpoint.

Taking these into account, the method is applied for several variants, given by the various combinations of the above options. Regarding the interval length, only the 10 and 20 days variants were considered, as being more plausible but, obviously, the method can be applied for any interval size. Thus, considering both hypotheses regarding the time assigned to each interval – upper limit and midpoint and interval lengths of 10 and 20 days, the four possible cases are:

- Case 1: 10 days interval, upper limit;
- Case 2: 10 days interval, midpoint;
- Case 3: 20 days interval, upper limit;
- Case 4: 20 days interval, midpoint.

To evaluate the estimated reliability probabilistic models in the four cases it is used the root mean square deviation (*RMSD*) which is an overall measure of the deviation of the values given by the estimated reliability models in relation to real values of the wheelset reliability *R(t)*:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{k} \left[R(t) - \hat{R}(t) \right]^2}{k}} \quad (17)$$

4. RESULTS AND DISCUSSION

In this section the previously described alternative estimation method is applied to estimate the Weibull distribution parameters on the basis of wheelset empirical failure data and the evaluation of the resulting mathematical models of reliability is made. By applying the iterative procedure described in section 2.2 the estimations of the two parameters of the Weibull reliability model are found. The values for the four above cases are shown in table 1. The estimates obtained by applying the classical Maximum Likelihood Method are also included as reference values.

Table 1

Estimated parameters of Weibull reliability model distribution.

Case	$\hat{\beta}$	$\hat{\eta}$ **(days)**
Case 1	1.000	3134
Case 2	0.9324	3676
Case 3	1.0324	3007
Case 4	0.9126	3976
Classical MLM	0.9463	3539

It is to be noted, first, that there are some notable differences between the results obtained in the four cases. However, it can be said that these results are comparable to (and around) the reference values corresponding to the classical Maximum Likelihood Method.

A second important aspect is related to the shape parameter β, which is the parameter that has the greatest impact on the reliability model, its value setting the monotony of the failure rate. It can be observed in Table 1 that the estimations of β are close to the value of 1 in all four considered cases. This is an expected result, as it indicates an almost constant failure rate, which is a normal feature during the normal lifetime of systems, such as the current case of the railway wheelsets.

Analysing the results, it can be seen that the choice of the interval time (see case 1 versus case 2, case 3 versus case 4) has a greater influence on the parameter estimations than the duration of the interval (see case 1 versus case 3, case 2 versus case 4).

Using the upper limit of the interval results in more pessimistic reliability models (higher β, lower η), compared to the situation when interval midpoint is used. As for the influence on each parameter, it is difficult to draw a clear conclusion. For a larger interval size, the β value is higher when the upper limit is used, but lower when the midpoint is used. The effect on the η value is the opposite: for a larger interval size, the η value is lower when the upper limit is used and higher when the midpoint is used.

Regarding the comparison with the reference values, it can be observed that the use of a smaller interval and of the midpoint (case 2) results in the closest values of both estimated parameters to those corresponding to the classical Maximum Likelihood Method. In the same purpose of evaluating the four estimated reliability models the *RMSD* values, given by equation (17), are calculated (see Table 2).

Table2

RMSD of estimated reliability models.

Case	**RMSD**
Case 1	$4.585 \cdot 10^{-3}$
Case 2	$2.836 \cdot 10^{-3}$
Case 3	$6.715 \cdot 10^{-3}$
Case 4	$3.526 \cdot 10^{-3}$

This analysis confirms that the most appropriate model (i.e., with a better fit to the empirical data), is the one for which the smaller interval and midpoint time were used. This was somehow expected, since a smaller interval allows a more precise allocation of the failure times, and the midpoint value is normally closer (on average) to the actual values of failure times in the interval.

5. CONCLUSION

In the present paper an alternative usage of the classical Maximum Likelihood Method was proposed, in order to reduce the amount of experimental data required for the estimation of the reliability mathematical model. The proposed original method involves dividing the testing period into intervals and to use the number of failures in each interval instead of the individual failure times. This is achieved by assigning the time corresponding to an interval to all failures occurring in that interval.

The formulas of the classical Maximum Likelihood Method were adapted according to these hypotheses. The resulting alternative method was applied to estimate the Weibull reliability probabilistic model of the railway wheelset, based on real operational data, using diverse interval durations and assigned time (upper limit or midpoint).

It was found that the interval time has a greater influence on the parameter estimations. Regarding the influence on each parameter, it was difficult to identify a clear trend, the results not being consistent from this point of view.

The resulting reliability models were evaluated by comparing the estimated parameters values to those corresponding to the classical Maximum Likelihood Method, and by using the root mean square deviation. According to both criteria it resulted that the most appropriate model was the one using the smaller interval and midpoint time.

5. REFERENCES

[1] Spiroiu, M.A. *Reliability Analysis of Railway Freight Wagon Wheelset.* IManE 2015, Appl Mech Mater, vols. 809-810, pp. 1097-1102, ISBN 978-3-03835-663-9, Iaşi, Romania, May 2015, https://doi.org/10.4028/www.scientific.net/AMM.809-810.1097

[2] Liu, Y., Liu, L., Stratman, B., Mahadevan, S. *Multiaxial fatigue reliability analysis of railroad wheels*. Reliab Eng Syst Saf, 93, 456-67, 2008, https://doi.org/10.1016/j.ress.2006.12.021

[3] Liu, Z.X., Gu, H.C. *Failure modes and materials performance of railway wheels.* J Mater Eng Perform, 9, 580-84, 2000, https://doi.org/10.1361/105994900770345728

[4] Mazzu, A., Solazzi, L., Lancini, M., Petrogalli, C., Ghidini, A., Faccoli, M. *An experimental procedure for surface damage assessment in railway wheel and rail steels.* Wear, 342-43, 22-32, 2015, https://doi.org/10.1016/j.wear.2015.08.006

[5] Spiroiu, M.A., Nicolescu, M. *Failure modes analysis of railway wheel.* IManE&E 2018, MATEC Web of Conferences, vol. 178, Article number 06005, ISSN: 2261236X Chisinau; Moldova, 2018, https://doi.org/10.1051/matecconf/201817806005

[6] Spiroiu, M.A., Nicolescu, M. *On the estimation of the reliability probabilistic model of railway wheelset.* IOP Conf Ser: Mater Sci Eng, 682, 012001, Bucharest, Romania, 2019, https://doi/org/10.1088/1757-899X/682/1/012001

[7] Rausand, M., Høyland, A.. *System Reliability Theory Models, Statistical Methods and Applications*. ISBN 0-471-47133-X, John Wiley & Sons, Hoboken, 2004

UTILIZARE ALTERNATIVĂ A METODEI VEROSIMILITĂȚII MAXIME PENTRU ESTIMAREA MODELULUI PROBABILISTIC AL FIABILITĂȚII OSIEI MONTATE

În lucrare se propune o utilizare alternativă a metodei verosimilității maxime în scopul estimării parametrilor modelului probabilistic al fiabilității cu utilizarea unui volum redus de date experimentale. Metoda propusă presupune împărțirea în intervale de timp a perioadei de funcționare și utilizarea numărului de defecte din fiecare interval și limitele intervalelor, nemaifiind necesară cunoașterea timpilor de defectare. Metoda clasică este adaptată în conformitate cu aceste ipoteze, iar metoda alternativă rezultată este aplicată pentru estimarea modelului probabilistic al fiabilității osiei montate, pe baza datelor obținute în cadrul exploatării vagoanelor de marfă. Parametrii modelului sunt estimați în mai multe ipoteze privind intervalele de timp iar rezultatele obținute sunt evaluate și se identifică influența asupra estimațiilor a ipotezelor adoptate în cadrul metodei.

Marius-Adrian SPIROIU, Associate Professor, PhD Eng., University POLITEHNICA of Bucharest, Department of Railway Rolling Stock, Splaiul Independenţei 313, Bucharest, Romania, Email: marius.spiroiu@upb.ro, +4021.402.96.95.

Sorin ARSENE, Associate Professor, PhD Eng., University POLITEHNICA of Bucharest, Department of Railway Rolling Stock, Splaiul Independenţei 313, Bucharest, Romania, Email: sorin.arsene@upb.ro, +4021.402.96.95.