On the estimation of the reliability probabilistic model of railway wheelset

M A Spiroiu¹ and M Nicolescu²

¹University POLITEHNICA of Bucharest, Department of Railway Rolling Stock, Splaiul Independenţei 313, Bucharest, Romania
²Romanian Railway Investigating Agency, Calea Griviţei 393, Bucharest, Romania

marius.spiroiu@upb.ro

Abstract. In the present paper is made a comparison between various estimation methods for the reliability probabilistic model of the railway wheelset. The analysis is based on real field data, observed in actual operations of freight wagons. The Weibull distribution is used to model the reliability and the applied estimation methods are the Maximum Likelihood Method and the Regression Method.

1. Introduction

A railway wheelset consists of a rigid assembly of two wheels on the axle. Besides the traditional functions performed by axles and wheels on all land transport systems - supporting the vehicle weight, ensuring vehicle rolling, transmitting its traction and braking forces, etc. – the railway wheelset performs also the function of guiding the vehicle on the track, thus has an essential role in terms of traffic safety [1,2]. As shown in [3], the failure of axles and wheels are one of the main categories of rolling stock failures that have produced derailments in recent years both on the European and Romanian railway networks. Taking into account that train derailment may have important social and/or economic implications and - in the worst-case scenario – may lead to important material damage or even to fatalities [4], the wheelset is the most critical subsystem of railway vehicles in terms of the reliability [2].

Being such an important issue - both from technical and economic point of view - the railway wheelset reliability is the subject of many researches. For example, the fatigue crack issue is a subject often addressed in scientific papers, such as [5], where is presented a study on the methodology of fatigue reliability degradation of railroad wheels, considering both fatigue crack initiation and crack propagation life. Also, in [6] is investigated the influence of wheel material structure and properties on the wheel rim cracking caused by the fatigue process. The wheel-rail contact phenomena are another source of wear and damages for wheelsets, an investigation on this subject being carried out in [7].

An analysis of freight car wheelset failure modes based on field data is made in [1] and a reliability model is deduced using the regression method. Also, in [2] are analysed the railway wheel failure modes and, for the major four failure modes, is carried out an additional analysis regarding the evolution in time of the failure rates, as well as the estimation of the failure probabilistic models, based on the Weibull distribution.
The present paper aim is to make a comparison between the results obtained by using different estimation methods for the reliability probabilistic model of the railway wheelset. Reliability mathematical model is based on the Weibull distribution and its parameters are estimated on the basis of real field data.

2. Estimation methods for the reliability model parameters
The Weibull distribution is widely used in the reliability field because of its versatility. Usually it is used the two-parameter version of Weibull distribution, according to which the evolution in time of the reliability \( R(t) \) is given by:

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

where \( t \) is the time, \( \beta \) and \( \eta \) are the shape and the scale parameters, respectively. The versatility mentioned above is given by the structure of the probabilistic law (i.e. by the two parameters), which makes it suitable to model the failure law of practically any industrial product.

To estimate the Weibull reliability model implies thus finding the values of the shape and the scale parameters so that the model is appropriate to the experimental data. Among the different estimation methods, the most commonly used are the Regression (Least Squares) Method and the Maximum Likelihood Method. Both methods are producing good results from the point of view of modeling the real failure laws and they are applicable both for complete data sets (i.e. when the failure time is known for all the units of the observed sample) and censored data sets (i.e. when the failure time is known only for a part of the units of the observed sample), the latter being the case of this paper.

2.1. Regression Method
The Regression Method uses a linearized form of the Weibull law and the Least Squares principle in order to estimate the parameters of the Weibull reliability model. The linear form is obtained by applying twice the natural logarithm in both members of equation (1), with the result:

\[
\ln \ln \frac{1}{R(t)} = \beta \ln t - \beta \ln \eta, \tag{2}
\]

so the Weibull probability law can be written under a linear form:

\[
Y = a_0 + a_1 X \tag{3}
\]

where

\[
X = \ln t; \quad Y = \ln \ln \frac{1}{R(t)}. \tag{4}
\]

Consequently, to estimate the parameters of Weibull distribution, can be used the linear regression and the Least Squares Method, which implies to minimize the sum of the squares of the deviations:

\[
S = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \tag{5}
\]

where

\[
x_i = \ln t_i; \quad y_i = \ln \ln \frac{1}{1 - F(i)} \tag{6}
\]

\( n \) being the sample size, \( t_i \) the failure time of wheelset \( i \) and \( F(i) \) an estimation of the failures cumulative distribution function.
Using the conditions for minimizing the expression in equation (5):
\[
\frac{\partial S}{\partial a_0} = \frac{\partial S}{\partial a_1} = 0
\] (7)
and solving the corresponding system, the expressions for coefficients \(a_1\) and \(a_0\) are obtained:
\[
a_1 = \left( \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i \right) \left( \sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i \right)^{-1}; \quad a_0 = \frac{\bar{y} \sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i}
\] (8)
where
\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i.
\] (9)

Therefore, the estimations of parameters \(\beta\) and \(\eta\) are:
\[
\hat{\beta} = a_1; \quad \hat{\eta} = \exp \left( -\frac{a_0}{a_1} \right).
\] (10)

The relationships above apply to the complete data sets (i.e. all \(n\) failure times are known). In the case of a censored data set (i.e. only \(k\) of the \(n\) wheelsets have failed, so only \(k\) failure times are known) the sums in equations (5) and (8) have the upper limit \(k\) instead of \(n\) and the means in equations (9) are calculated only for the corresponding \(k\) terms.

A very important issue within the regression method is related to the estimation of the failures cumulative distribution function in equation (6). In most cases this estimation is made using either the mean ranks or the median ranks. In the case of Mean Ranks Regression, the cumulative distribution function is given by [4]:
\[
F(i) = \frac{i}{n+1}
\] (11)
or
\[
F(i) = \frac{i - 0.5}{n},
\] (12)
while in the case of Median Ranks Regression is used the Benard's approximation, according to which the cumulative distribution function is estimated by the following expression [8]
\[
F(i) = \frac{i - 0.3}{n + 0.4}.
\] (13)

2.2. Maximum Likelihood Method.
The principle of the method consists in determining the most likely estimators (i.e. that maximize the likelihood of occurrence of the empirical data set) for the unknown parameters of a distribution. Considering a complete set of experimental data \(x_1, x_2, \ldots, x_n\), the likelihood function is defined as [4,9]:
\[
L(x_1,x_2,\ldots,x_n;\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdot \ldots \cdot f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta),
\] (14)
where \(\theta\) is the parameter to be estimated (there may be more than one) and \(f\) is the probability density function of the assumed distribution.
The function above expresses the probability of observing the data set \(x_1, x_2, \ldots, x_n\) (“how likely” it is to obtain these data values) for a given \(\theta\). The estimator that is most likely to have generated the observed sample is the one that maximizes the likelihood function, so it can be found by solving the equation [4,9]:

\[
\frac{dL(x_1, x_2, \ldots, x_n; \theta)}{d\theta} = 0.
\]

(15)

For practical reasons, usually is used the natural logarithm of the likelihood function instead of the function itself, so the equation to be solved is [4,9]:

\[
\frac{d\ln L(x_1, x_2, \ldots, x_n; \theta)}{d\theta} = 0.
\]

(16)

In the situation of a censored data set, when for the \(n\) observed units only the first \(k\) failure times are known \((k<n)\), the likelihood function is defined as [9]:

\[
L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{k} f(x_i; \theta) \prod_{j=1}^{n-k} R(x_j; \theta),
\]

(17)

where \(R\) is the reliability function.

Considering the case of Weibull distribution, the probability density function is given by:

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

(18)

thus the likelihood function for a censored data set of \(k\) failure times given by equation (17) becomes:

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

(19)

Applying the natural logarithm:

\[
\ln L(t_1, t_2, \ldots, t_k; \beta, \eta) = k \ln \beta - k \beta \ln \eta + (\beta - 1) \sum_{i=1}^{k} \ln t_i - \sum_{i=1}^{k} \left(\frac{t_i}{\eta}\right)^\beta - (n-k) \left(\frac{t_k}{\eta}\right)^\beta,
\]

(20)

where \(t_k\) – the last observed failure time – is the censoring time, thus the time assigned to the \((n-k)\) units which have survived the time \(t_k\).

The maximum conditions are

\[
\frac{\partial \ln L(t_1, t_2, \ldots, t_k; \beta, \eta)}{\partial \beta} = 0 \quad \frac{\partial \ln L(t_1, t_2, \ldots, t_k; \beta, \eta)}{\partial \eta} = 0.
\]

(21)

The condition above are leading, after some calculations, to the following expressions of the estimators for \(\beta\) and \(\eta\):

\[
\hat{\beta} = \left(\frac{\sum_{i=1}^{k} t_i^\beta + (n-k) t_k^\beta}{k}\right)^{1/\beta}
\]

(22)
\[
\frac{1}{\beta} + \frac{1}{k} \sum_{i=1}^{k} \ln t_i - \frac{\sum_{i=1}^{k} t_i^\beta \ln t_i + (n-k)t_k^\beta \ln t_k}{\sum_{i=1}^{k} t_i^\beta + (n-k)t_k^\beta} = 0 .
\] (23)

3. Results and discussions

In this section is performed a comparative analysis of the results obtained regarding the estimation of the reliability probabilistic model of railway wheelset, using the previously presented methods. The analysis is based on real field empirical data, observed in actual operations of freight wagons in the period of their useful (normal) life. The sample size was \( n = 1802 \) units, of which there have been observed failures in the case of \( k = 190 \) units over the observation period of one year.

The Weibull parameters of the reliability model are estimated using the Maximum Likelihood Method (ML) and three variants of the Regression Method, the difference between them consisting in how is estimated the failures cumulative distribution function: by Mean Ranks Regression - given by equations (11) (MnR \(_1\)) or (12) (MnR \(_2\)) or by Median Ranks Regression (MdR).

In the case of the Regression Method it is sufficient to simply apply the procedure described above, but in the case of the Maximum Likelihood Method the two estimates are determined using iterative procedures, by considering an initial value for the form parameter in equation (23) and to adjust this value until the expression becomes null. As initial value is used one of the values previously determined through Regression Method. Once the form parameter estimation found, it is inserted in equation (22) to find the estimation of the scale parameter.

In table 1 are presented the estimations of the parameters of the Weibull reliability model for the four considered cases.

| Method | \( \hat{\beta} \) | \( \hat{\eta} \) (days) |
|--------|-----------------|---------------------|
| MnR \(_1\) | 1.03073 | 2633.93 |
| MnR \(_2\) | 1.06424 | 2425.38 |
| MdR | 1.04940 | 2514.61 |
| ML | 0.94632 | 3535.66 |

The Weibull shape parameter \( \beta \), has a special significance, its value setting the monotony of the failure rate. For \( \beta = 1 \) the failure rate is constant, while for values higher and lower than 1, the failure rate is increasing and decreasing, respectively. It can be seen in Table 1 that all \( \beta \) values are close to 1, so the wheelset failure rate has not important variations over time (i.e. is almost constant), these results confirming the general theory regarding the behaviour of systems during their normal life.

It can be seen that there are no major differences between the results obtained for both parameters by using the three versions of the Regression Method. Regarding the results obtained through the Maximum Likelihood Method, there is a notable difference concerning the scale parameter, which is significantly higher. On the other hand, the scale parameter is about 10% lower than those obtained by regression. It is worth noting that these results indicate that the use of the Maximum Likelihood Method leads to a more optimistic model of the wheelset reliability.

The question that arises is which of the estimated models is most appropriate for the experimental data. To answer that, in this paper is used the root mean square error (RMSE) to overall assess the deviation of the estimated reliability models in relation to actual values.
\[ RMSE = \sqrt{\frac{\sum_{t=1}^{k} [R(t) - \hat{R}(t)]^2}{k}}. \]  

In table 2 are presented the RMSE of the four estimated models of wheelset reliability. It can be seen that there are not very important differences between the four models in terms of model accuracy. It is, however, to be noted that the model obtained by the Maximum Likelihood Method proves the best fit to the empirical data, thus it seems to be the most appropriate.

**Table 2.** RMSE of estimated reliability models.

|         | RMSE     |
|---------|----------|
| MnR1    | 2.806 \cdot 10^{-2} |
| MnR2    | 2.885 \cdot 10^{-2} |
| MdR     | 2.849 \cdot 10^{-2} |
| ML      | 2.282 \cdot 10^{-2} |

4. Conclusion
The aim of the present paper was to estimate the reliability probabilistic model of railway wheelset, acknowledged as the most critical subsystem of railway vehicles from reliability point of view, and to make a comparison between various estimation methods or variants thereof. The analysis was based on real field empirical data, observed for one year in actual operations of freight wagons, in the period of their useful life.

The estimation of the Weibull reliability models was made using the Maximum Likelihood Method and three variants – given by the manner of estimation of failures cumulative distribution function - of the Regression Method. The shape parameter estimations were close to 1 in all four cases, confirming an almost constant failure rate of railway wheelset – a usual (expected) behaviour during the normal life. The scale parameter estimation through the Maximum Likelihood Method was significantly higher than those obtained through the Regression Method.

The results of the estimation indicated that the Maximum Likelihood Method leads to the most optimistic model of the wheelset reliability, however the RMSE analysis indicated this model as the most suitable for the field data. The models obtained by using Regression Method proved also good accuracy so, although the Maximum Likelihood Method seems to be the most suitable for the analysed case, it is difficult to claim that it is the appropriate method in all situations.

5. References
[1] Spiroiu MA 2015 *Appl. Mech. Mater.* **809-810** 1097-1102
[2] Spiroiu MA and Nicolescu M 2018 *MATEC Web of Conferences* **178** 06005
[3] Nicolescu M and Spiroiu MA 2019 *U.P.B. Sci. Bull. Series D* **81**(3) 61-68
[4] Spiroiu M 2006 *Reliability and maintenance of railway vehicles (in Romanian)* (Bucharest: MatrixRom)
[5] Liu Y, Liu L, Stratman B and Mahadevan S 2008 *Reliab. Eng. Syst. Safe.* **93** 456-67
[6] Liu ZX and Gu HC 2000 *J. Mater. Eng. Perform.* **9** 580-84
[7] Mazzu A, Solazzi L, Lancini M, Petrogalli C, Ghidini A and Faccoli M 2015 *Wear* **342-343** 22-32
[8] McCool JI 2012 *Using the Weibull Distribution: Reliability, Modeling and Inference* (Hoboken: John Wiley & Sons)
[9] Rausand M and Høyland A 2004 *System Reliability Theory Models, Statistical Methods and Applications* (Hoboken: John Wiley & Sons)