Reliability computer modeling and evaluation method of cryogenic valve using operation and maintenance data

Yi Lu¹, JianMing Zheng¹,*, Chen Zhao¹, Gang Ren²

¹School of Mechanical and Precision Instrument Engineering, Xi’an University of Technology, Xi’an 710048, China
²The Boiler & Pressure Vessel Safety Inspection Institute of Henan Province, Zhengzhou 450016

*Corresponding author: mingzheng@xaut.edu.cn

Abstract. In order to accurately obtain the fault distribution law and operation reliability model of low-temperature globe valves, this paper obtained operation fault data of 1114 groups of low-temperature globe valves running in low-temperature tank trucks through investigation, and used Kaplan-Meier method and average rank method to estimate the reliability model parameters of fault data. In view that the Kaplan–Meier method fails to analyze the fault data in the same period of time effectively and the average rank method fails to consider the censored data, this paper proposes a modified average rank method based on the fact that it can analyze the fault data in the same period of time and consider the influence of censored data, in which a reliability model by considering the fault data in the same period sorted by rank and the censured data. The reliability obtained by this method was compared with that obtained by Kaplan–Meier method and average rank method. The two-parameter Weibull model selected was verified by K-S (Kolmogorov-Smirnov) test method, and the maximum test value D value before and after the improvement of average rank method was compared. The results showed that the modified mean rank method not only overcomes the shortcomings of Kaplan–Meier method that cannot effectively analyze the fault data occurred in the same period of time, but also supplements the shortcomings of the mean rank method that does not consider the deleted data. Through the K-S test method, the maximum test amount D value of the modified mean rank method is less than that of the mean rank method, which proves the effectiveness and accuracy of the modified method.

Keywords: Kaplan–Meier method; average rank method; the modified mean rank method; K–S test method.

1. Introduction

In 2020, the consumption of LNG (Liquefied natural gas) in China was approximately 230 million tons. Among the supporting storage and transportation facilities, the number of LNG tankers has reached 13,196, there are more than 40 large LNG receiving stations, and the storage capacity of LNG filling stations is approximately 3900 [1-2], and the number is still surging. Low temperature stop valve is one of the key components. Its main function is to cut off or adjust and throttle the use of cryogenic liquid.
or gas, and its reliability directly affects the safety of these cryogenic equipment and tank cars, causing the leakage of cryogenic liquid and gas, personnel damage, serious cause explosion, vehicle destruction and death. Therefore, the intensive research on the reliability analysis and evaluation of cryogenic valves has a great impact on reducing accidents and improving equipment safety performance.

At present, the research on the reliability modeling of cryogenic valves at home and abroad is limited, and the research on reliability modeling methods is concentrated on the use of the average rank, least square, and Monte Carlo methods and other methods. To improve the reliability of aviation equipment and seek the best solution to its parameter estimation problem, Shen Anwei [3] et al. used the average rank method and the expectation maximization EM (Expectation Maximization Algorithm) algorithm to estimate the parameters under random censored data, respectively. In the case of different sample sizes, the mean square error of the results calculated by the two estimation methods was compared and analyzed, but the deficiency of the mean rank method in the case of censored data was not pointed out.

In order to verify the effectiveness of the average rank method in estimating the parameters of the complete failure data distribution function, Wang Zhuojian [4] and others designed a Monte Carlo simulation method that clarifies the application scope of the average rank method; Wang Zhenxing [5] proposed the joint least square and average rank methods to estimate the Weibull distribution model parameters. To address the reliability uncertainty of water conservancy equipment during operation, the Weibull distribution was used to establish a water conservancy equipment reliability analysis model, but these two articles do not provide a solution in the case of censored data. Gu Jilian [6] et al. addressed the problem of small sampling and random censorship in the reliability evaluation of aero-engine field data. They compared and analyzed the non-parametric estimation methods in terms of reliability evaluation and applied the Monte Carlo method in the competition. Under the fault model, a simulation method was designed to compare the advantages and disadvantages of reliability evaluation methods, but no specific reliability model was proposed. Shen Guixiang [7] et al. discussed the series connection of several independent subsystems. When the life of the system follows the two-parameter Weibull distribution, the average rank method is used to correct the fault sequence, and model parameter estimation and the fit test are carried out on this basis, which can reduce modeling errors and improve the fitting accuracy. Gu Dongwei et al. [8] based on the field test data of 40 sets of the spindle system of a certain domestic CNC machine tool, determined that the fault process was subject to the two-parameter Weibull distribution through parameter estimation, D test and run-distance test, and used the mean rank method to evaluate the indicators affecting customer satisfaction. Junshan Shen [9] et al. proposed a new EM algorithm for double-checking data constrained by non-parametric moments and established the empirical likelihood confidence area, and their results show that the corresponding empirical likelihood ratio converges with the standard chi-square random variable. Zhang Haibo [10] et al. applied the least square method to the failure interval for the first time based on the one-year failure data of a certain type of China system. The parameters of the time distribution model were estimated, and the hypothesis test was carried out using the D test method, which proved that the time between failures of the China system follows the Weibull distribution. Xu Yan [11] et al. collected on-site operation data related to the life of the relay protection device. These data are analyzed and processed, and the Weibull distribution is used to estimate the reliability parameters. The average rank method is used to estimate the parameter values of the two-parameter Weibull distribution to obtain the reliability function of the relay protection device life. However, the average rank method fails to consider the problem of data deletion. Saeed Ariyafar et al. [12] modeled the reliability of financial risk investment portfolio based on Laplace transform, and obtained the portfolio with the highest reliability. Z. Behboudi [13] et al. used the Monte Carlo method combined with the periodic handover method to evaluate the reliability of the refrigeration standby system. The results showed that the periodic handover method reduced the failure rate of the system, and the Monte Carlo method was used to verify and improve the reliability of the system. Jerrold H. May [14] et al. used the average rank method, which evaluates the model that produces discriminatory behavior and combines the Gaussian polynomial distribution to test the influence of different parameters on the reliability of the model.
In short, the average rank method, least square method, Monte Carlo method and other traditional methods are used for reliability modeling at home and abroad. However, these traditional methods all have their shortcomings. This paper aims at this situation, according to the investigation and collection of the failure data of 1114 cryogenic shut-off valves, the Kaplan–Meier and average rank methods are used to analyze the failure data, and reliability modeling comparison is carried out to determine the shortcomings of the two methods. An improved average rank method not only compensates for the shortcomings of the Kaplan–Meier method, which cannot effectively analyze the failure data that occurs at the same period, but also supplements the average rank method, which does not consider censored data. This improved method is combined with the two-parameter Weibull model to estimate the parameters, verify the accuracy of the two-parameter Weibull model selected by the K-S test method, and compare the effects before and after the improvement of the average rank method, obtaining a more accurate reliability evaluation value.

2. Investigation and collection of cryogenic valve failure data

Low-temperature stop valves are installed in LNG tank cars and supporting facilities of LNG filling stations, and their working process is shown in Figure 1. It is usually located at the tail of the cryogenic tanker, the liquefied gas tank, and the front of the filling station. The connected cryogenic liquid or gas cuts off, regulates, and prevents leakage. Cryogenic shut-off valves are usually used in working conditions where the temperature is lower than -29°C. Due to cavitation, it is easy to cause some problems like valve body damage, valve stem fracture, sealing surface damage and leakage, and packing leakage, thus its reliability directly determines the LNG tanker as well as the safety issues of LNG filling stations. Since the cryogenic shut-off valves are used in tank trucks with uncertain trajectory, it cannot be monitored in real time, the reliability assessment of the valves is carried out using survey data.

The survey data is not entirely failure data but includes censored data caused by incomplete valve failure information due to time constraints and some random unpredictable conditions. The censored data [15] are divided into three categories: left, interval, and right censored data. Left censored data indicate that the sample failed before the start of the experiment; interval censored data indicate that all samples are operating normally at the initial moment, but some samples fail at a later time interval, and the specific time of sample failure is unknown; right censored data indicate that the sample has not expired at the beginning of the experiment. However, when the experiment ends, the sample has not expired or the sample’s failure has not been observed, and the failure data of the sample is lost for some reason. Right censoring data are divided into time and fixed number censoring data. Time censoring mainly focuses on censoring time, and fixed number censoring focuses on the number of sample failures.

When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. Should authors use tables or figures from other Publications, they must ask the corresponding publishers to grant them the right to publish this material in their paper.
This article is based on the actual inspection and maintenance of the low-temperature tank car inspection station of the Henan Boiler and Pressure Vessel Safety Inspection Institute. The low-temperature tank car and its valve must be inspected annually through the low-temperature tank car inspection station, and damages must be recorded. The operation data of 1114 valves from January 1, 2017 to December 30, 2018 were selected for data collection, and the DN50 low temperature stop valve was selected for data collection. This article tracks two years of low-temperature shut-off valve operating data. Thus, the data is timed right censored data. The operating data of these 1,114 valves consist of 56 failure and 1,058 ending samples. Among those, 56 failed samples contained a 53 failed valves, and 3 valves were withdrawn from inspection due to missing data. The time corresponding to the specific valve failure is shown in Table 1:

| Time/day | 1 | 2 | 5 | 8 | 8 | 13 | 18 | 18 | 20 | 20 |
|----------|---|---|---|---|---|----|----|----|----|----|
| 26       | 29 | 36 | 40 | 43 | 44 | 44 | 63 | 63 | 63 |
| 65       | 65 | 67 | 69 | 71 | 75 | 75 | 85 | 86 | 95 |
| 100      | 105 | 108 | 111 | 124 | 131 | 140 | 150 | 175 | 207 |
| 281      | 339 | 372 | 400 | 419 | 447 | 477 | 506 | 533 | 544 |
| 566      | 595 | 631 | 667 | 705 | 727 |    |    |    |    |

The data in Table 1 represents the time when the valve failed. Two 8s represent two valve failures on the 8th day, and one 1 represents one valve failure on the 1st day. Among them, the deleted data recorded on the 207th, 419th and 544 days due to the loss of valve base materials are deleted. For the reliability evaluation of timed right truncated data, the number of normally operating valves before and after each failure time point must be determined. According to the data in Table 1, the number of failed valves increases as time changes, and the number of remaining normally operating valves increases with time. The changes are shown in Figure 2.

Figure 2 shows that in the early observation period, the number of remaining valves decreased rapidly, indicating that the number of early failure valves increased. After 175 days, the number of remaining valves decreased slowly, it conforms to the actual situation that the valve enters the accidental failure period.

3. Research on reliability modeling methods
To evaluate the reliability of cryogenic valves, the Kaplan–Meier and average rank methods are commonly used for the analysis of reliability data with fault and censored data. This article focuses on
these two methods for evaluating the reliability of cryogenic valves. To address the shortcomings, an improved average rank method is proposed and compared with the Kaplan–Meier and average rank methods.

3.1. Kaplan–Meier method

The Kaplan–Meier method evaluates the survival rate corresponding to each actual observation event. When estimating the product life or reliability, the method can analyze the complete failure sample or the sample containing the end data. This method follows a specific process. Given sample size n and m failure data, the failure time of m failure data samples are arranged in increasing order, $t_1 < t_2 < t_3 < t_m$. Then the ratio of the number of cryogenic valves in normal operation after $t_j$ to the number of cryogenic valves in normal operation before $t_j$ can be obtained. Then the product of all these ratios at the time $t_j$ is the obtained reliability, and the formula is as follows:

$$R(t) = \prod_{j=1}^{n} \frac{n_j - d_j}{n_j}$$  \hspace{1cm} (1)

In the formula, $d_j$ represents the number of valve failures at $t_j$, $n_j$ represents the number of normal valves at time $t_j$, and after $n_j = n-j+1$ after $t_j$. When using the Kaplan–Meier method to model cryogenic valves, the number of valve failures between two adjacent failure sample points is zero, so the reliability value remains unchanged. Only when failure data exist, the reliability jumps. Change, use Formula 1 to obtain the reliability of the cryogenic valve. As shown in Figure 3:

![Figure 3. Kaplan-Meier method reliability change diagram.](image)

Figure 3 shows that the reliability of the valve continues to decline as time increases. In the first 175 days of the observation period, the reliability of the valve dropped from 1 to 0.965, and the reliability dropped the fastest, indicating that the cryogenic valve had an early failure period. After 175 days, the reliability decline flattened, indicating that the cryogenic valve entered the accidental failure period.

3.2. Average rank method

The average rank method is an effective method for improving the accuracy of the empirical distribution function. By principle, the specific failure time of the truncated sample is unknown, the failure of its real time may appear in any moment during ob-servation period, and this makes the sample of known failure data in order to become very uncertain, so need to estimate all the known failure data observed in samples of final order. The problem of uncertain censored data is solved by solving the average failure order of the failure data. The average rank of the failure data can be obtained by calculating the increment between the two true failure data. The increment is related to the number of censored data
between the two failure data. Finally, the average failure rank of the failure samples is calculated. The specific formula is as follows:

\[ A_i = A_{i-1} + \Delta A_i \]  
(2)

\[ \Delta A_i = \frac{n+1-A_{i-1}}{n-k+2} \]  
(3)

In the formula, \( i \) is the serial number of the faulty device, \( n \) is the total number of samples, \( k \) is the sequence number of all devices, and \( A_i \) is the average rank of the faulty device. The average rank is obtained and substituted into the approximate median rank formula to calculate the cumulative failure distribution function of the cryogenic valve failure data as follows:

\[ F(t_i) = \frac{A_i-0.3}{n+0.4} \]  
(4)

The reliability of cryogenic valves is described. The relationship between the reliability and cumulative failure distribution functions is \( F(t) + R(t) = 1 \), so the reliability function of the cryogenic valve is solved by solving the cumulative failure distribution function as follows:

\[ R(t_i) = 1 - F(t_i) \]  
(5)

The average rank method is used to model the reliability of the cryogenic valve failure data samples, considers the censored data in the samples, estimates the average failure order of each sample, and uses Equation (5) to obtain its reliability, which is similar to the Kaplan–Meier method. The obtained reliability values are compared, and the result is shown in Figure 4:

![Figure 4](image)

**Figure 4.** Comparison of reliability between average rank method and Kaplan-Meier method.

Figure 4 shows that the reliability of the valve in the average rank method decreases with time. In the first 175 days, the reliability dropped from 1 to 0.963, indicating that the cryogenic valve had an early failure period, and the average rank was used. The overall reliability obtained by the second method is lower than that of the Kaplan–Meier method. Because when the Kaplan–Meier method is used to estimate the system reliability, the censored data is treated as the failure data. At the same time, for different failure samples on the same day, the average reliability estimation of adjacent failure sample points is used, and the reliability estimation of each failure sample point is not carried out.

### 3.3. Modified average rank method

The above studies using the Kaplan–Meier and average rank methods only considered the fault or defect data unilaterally and did not make full use of the fault and defect data simultaneously. The average rank method only considers the location of each data’s censored point and does not consider the specific
value of the censored data and the influence of the interval length formed by each two adjacent fault data on the results, which may lead to the same parameter estimation in different situations.

To compensate for the shortcomings of the traditional average rank method, an improved average rank method is proposed to solve the empirical distribution function. The core of the future improvement is the increase of the probability of failure of the protection device censored item to correct the accuracy of the parameter evaluation results, which can produce near-reality evaluation results.

The following is a set of specific examples for illustration. Ten failure data of mechanical product operation are divided into two groups, namely, mechanical products 1 and 2. The failure data is recorded according to the running time. Three groups of failure data and two groups of censored data exist. The censored data, where F refers to failure data, and S refers to censored data, vary as shown in Tables 2 and 3:

| Serial number | operation hours(t/h) | situation |
|---------------|----------------------|-----------|
| 1             | 4000                 | F         |
| 2             | 4200                 | S         |
| 3             | 5500                 | S         |
| 4             | 8000                 | F         |
| 5             | 10000                | F         |

| Serial number | operation hours(t/h) | situation |
|---------------|----------------------|-----------|
| 1             | 4000                 | F         |
| 2             | 6000                 | S         |
| 3             | 7500                 | S         |
| 4             | 8000                 | F         |
| 5             | 10000                | F         |

The comparison of the two sets of data in Tables 2 and 3 shows that the censored data of Nos. 2 and 3 are different. However, according to the average rank method, shape parameters $m$ and $\eta$ are the same. The main reason is that the average rank method does not consider the influence of the specific value of the censored data on the evaluation result. It is considered that the probability of future failure of the censored items is equal among the failure items larger than the censored data. According to the probability theory, the probability of future failure of the censored item is different between each two adjacent fault data in the future, which is not only related to the specific value of censored time, but also related to the interval length formed by the adjacent two fault data. If the length of the interval formed by two adjacent fault data is longer, the failure probability of the censored item in the corresponding interval is greater.

Therefore, given the shortcomings of the average rank method, this article increases the probability that the protection device censored item will fail between every two adjacent fault data to correct the fault rank. Then, an estimate is made using the Weibull parameter.

Here, $I_{m(j)}$ is regarded as the increment of the rank of invalid data $t_j$ in interval $[t_{j-1}, t_j]$ of the censored data. If $t_j$ has $m_j$ censored data before, the increment of $m_j$ deleted data to the failure data $t_j$ rank is:

$$I_j = 1 + \sum_{m=1}^{M_{m(j)}} I_{m,j}$$

The purpose is to improve the estimation accuracy of empirical distribution function. Let the deleted data be denoted by $t_{sm}$, where $m = 1, 2, \ldots, M$, which is the deleted number. The invalid data larger than the deleted data is denoted as $t_k, t_{(k+1)}, t_{(k+2)}, \ldots, t_n$. When a product is deleted at $t_{sm}$, it is deleted in $[t_{ij},$
The future failure probability within \( n \) is denoted by \( I_{(m)} \). The conditional probability formula is expressed as follows:

\[
I_{m,j} = \int_{t_{j-1}}^{t_j} f(t) \, dt = \left. \exp\left[-\frac{(t_j-1)}{\alpha}\right] - \exp\left[-\frac{(t_j)}{\alpha}\right] \right|_{t_{j-1}}^{t_j} \exp\left[-\frac{(t_{j-1})}{\alpha}\right]
\]

(7)

Therefore, the average rank of the \( j \)th failure data is:

\[
A_j = A_{j-1} + I_j
\]

(8)

By substituting Equation (8) into Equation (4), the empirical distribution function is obtained as follows:

\[
F(t_j) = \frac{A_{j-1} + I_j - 0.3}{n+0.4}
\]

(9)

The improved reliability calculation formula of the average rank method is:

\[
R(t_j) = 1 - F(t_j)
\]

(10)

The reliability of the cryogenic shut-off valve obtained by using the Kaplan–Meier and average rank methods and the improved average rank method is shown in Figure 5:

Figure 5. Kaplan-Meier, average rank method, and improved average rank method reliability comparison chart.

The figure shows that the reliability obtained by using the improved average rank method is obviously different using censored data and shows a downward trend. This phenomenon occurs because the position of censored points of each data, the specific value of censored data and the influence of the interval length formed by each two adjacent fault data on the evaluation results are taken into account.

4. Conclusions

On the basis of the failure data of 1114 cryogenic shut-off valves, this paper uses the Kaplan–Meier and average rank methods to analyze the failure data and proposes an improved average rank method. The following conclusions as drawn:

1. The comparison of the reliability of the Kaplan–Meier and average rank methods indicates that when the Kaplan–Meier method estimates the reliability of the system and uses the average estimation
of the reliability for the failure samples that appear in the same period, failing to estimate the reliability of each failure sample point.

2. The improvement of the average rank method indicates that the average rank method only considers the location of each data censored point and does not consider the specific value of the censored data and the length of the interval formed by every two adjacent fault data. The impact on the results may lead to the same consequence as the parameter estimation results in different situations.

Acknowledgments
This research has been supported by the National Natural Science Foundation of China (Grant No. 51475367), National Science and Technology Major Project of China (Grant No. 2013ZX04009011), Natural Science Foundation of Department of Education of Shaanxi Province (Grant No. 2018JM5113), and Key Laboratory Project of Shaanxi Education Department (Grant No. 17JS093).

References
[1] S.M. Ultra-low temperature valve for liquefied natural gas. Valve, 2012, 02:24-26. doi: 10.3969/j.issn.1006-9348.2013.09.055
[2] L.Q. Overview of foreign reliability technology development. Automation Instrumentation, 1980, 6:10-12+10. doi: 10.3969/j.issn.1001-506X.2013.05.39
[3] S.Z; Comparison study of parameter estimation methods Computer Simulation, 2013, 30(09):239-243. doi: 10.3969/j.issn.1004-373X.2013.02.034
[4] W.Z; The effectiveness analysis of the average rank method based on the cluster simulation model. System Engineering and Electronic Technology,2013,35(05):1128-1132. doi: 10.13224/j.cnki.jasp.2016.01.007
[5] Wang,Z. Reliability analysis model and application of water conservancy equipment based on Weibull distribution.Modern Electronic Technology,2013,36(02):107-109.doi: 10.13229/j.cnki.jxdt.201401018
[6] Shen ,Z; Guo J; Wang Z; Non-parametric estimation method for reliability evaluation of competitive failure model.Journal of Aeronautical Dynamics,2016,31(01):49-56.doi: CNKI:SUN:ZNGD.0.2013-02-016
[7] Shen ,G; Meng S; Zhang Y; The Application of Average Rank Method in Subsystem Reliability Modeling.Journal of Jilin University(Engineering and Technology Edition),2014,44(01):101-105.
[8] Gu,D; Shen,G; Zhang,Y; Wang,Z; Chen B; Zheng. S; Reliability Evaluation of CNC Machine Tool Spindle System.Journal of Central South University (Natural Science Edition),2013,44(02):540-545.
[9] Shen ,J , Yuen ,K C , Liu C ; Empirical likelihood confidence regions for one- or two- samples with doubly censored data. Computational Statistics & Data Analysis, 2016, 93(C):285-293.doi:10.1016/j.csda.2015.01.010
[10] Zhang H; Jia Y; Zhou G. Research on Time Distribution Model of NC System Failure Interval. Journal of Harbin Institute of Technology, 2005, 37(002):198-200.doi: 10.3321/j.issn:0367-6234.2005.02.016
[11] Xu, Y; Bai, J; Dai Z.; A new method for reliability analysis of relay protection devices based on Weibull distribution. Journal of North China Electric Power University (Natural Science Edition), 2012(4): 15-19.doi: 10.3969/j.issn.1007-2691.2012.04.004
[12] Ariyafar ,S , Tata M , Rezapour M , et al. Comparison of aggregation, minimum and maximum of two risky portfolios with dependent claims. Journal of Multivariate Analysis, 2020:104620.doi: 10.1016/j.jmva.2020.104620
[13] Asadi M. Reliability Modeling of Two-unit Cold Standby Systems: A Periodic Switching Approach. Applied Mathematical Modelling, 2020.doi: 10.1016/j.apm.2020.11.001
[14] May J H, Vargas L G . An exact test of the accuracy of binary classification models based on the probability distribution of the average rank. Mathematical & Computer Modelling, 2009, 50(7-
8):975-983. doi: 10.1016/j.mcm.2009.04.002