Methodical Approach to Selecting the Appropriate Distribution for Reliability Analysis : Automotive Application

NAOUFAL BELLA[1,*], *NOHAILA* SALHI[2,**], and *ISMAIL* LAGRAT[3,***]

[1]PhD Student in Mechatronics, passionate about mechatronics and complex systems. Lab of Advanced Systems Engineering, Ibn Tofail University, ENSA, Kenitra, Morocco.

—

[2]PhD Student in Embedded Systems. Lab of Advanced Systems Engineering, Ibn Tofail University, ENSA, Kenitra, Morocco.

—

[3]Professor, qualified to supervise research in industrial technologies, automotive mechatronic systems, And Robotic at Ibn Tofail University. Lab of Advanced Systems Engineering, Ibn Tofail University, ENSA, Kenitra, Morocco.

Abstract. In this study, we propose a methodical approach to selecting an appropriate statistical distribution for reliability analysis. In this approach, we have defined a methodology for testing reliability distributions based on the Kolmogorov Smirnov K-S test for MTBF Data collected from Self-Diagnostic of a sample of 50 critical components part of a complex automotive system. Finally, we proposed two solutions: the first involves migrating from one distribution to another according to the intervals, and the second allows for the selection of the distribution that is representative over a maximum number of intervals. These strategies were developed from the analysis of results after application of the K-S test on the distributions tested. This approach will contribute. to the reliability analysis of complex systems. As a result, in improving the models used to analyze complex systems behavioral analogies such as Petri nets or Markov chains.

Index Terms : Statistical distribution, Kolmogorov Smirnov K-S test, Self-Diagnosis, reliability of complex automotive systems.

1 Introduction

The design of automotive mechatronic systems [1] relies on increasingly sophisticated components, whose reliability is crucial to ensure safety, performance, and customer satisfaction [2]. The reliability of automotive components [3] is not only essential to prevent the expensive and potentially dangerous failures, but also plays a key role in improving experiences for designers and developers [4]. In this context, reliability analysis that contributes to the operational safety of systems designed or to be developed is not simply an option, but a must for vehicle manufacturers [5].

[*]**Contact:** +212 6 06 11 97 78 **Email:** naoufal.bella@uit.ac.ma **ORCID:** 0009-0008-5452-1510
[**]**Contact:** +212 6 28 88 96 66 **Email:** nohaila.salhi@uit.ac.ma
[***]**Contact:** +212 6 65 63 38 51 **Email:** ismail.lagrat@uit.ac.ma

Previous research and studies have highlighted various methods for assessing the reliability of automotive components [6]. These studies are based on statistical models, and durability tests have been used [7] to predict potential failures and estimate component lifetime [8]. These tests allow us to collect the data for the failures and apply statistical models to analyze the data [9]. And, improving the study of models such as Petri net (PT net) [10] and Markov chain [11], ... proposed to study and simulate the behavior of complex systems [12].

Among the commonly used statistical models, we can find Weibull model, which is particularly useful for analyzing component failure times and is widely used because of its flexibility in representing different forms of failure distribution [13]. The exponential model, often used for components with a constant failure rate, is simple but less flexible than the Weibull model [14]. The log-normal model is applied when failure times are log-normal, which is often the case for electronic components where failure depends on multiple multiplicative factors [15].

One of the main difficulties lies in selecting the appropriate reliability distribution for a component [16]. An inappropriate selection can lead to inaccurate estimates of lifetime and, consequently, to sub-optimal maintenance decisions and erroneous informations contributing to decision support. To select the appropriate distribution, we need to search for a methodical approach. This approach includes failure data analysis, where component failure data is collected and analyzed to understand their behavior in the different operating conditions. statistical hypothesis test, such as the Kolmogorov-Smirnov test [17] or the chi-squared test [18], are used to verify the fit of failure data to candidate distributions. Model comparison involves comparing several distribution models using information criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) to identify the model that offers the best compromise between complexity and fit [19]. Cross-validation involves using cross-validation techniques to assess the predictive performance of selected models [20].

By applying this methodical approach, it is possible to select the appropriate reliability distribution(s) for each component and improving the accuracy of lifetime estimates and the effectiveness of maintenance strategies. To overcome these challenges, our research proposes a methodical approach to validating the suitable reliability distribution for a specific component in the automotive industry. We will focus on an electronic component X (an intelligent sensor, for example), which is critical to vehicle performance. Our method relies on the use of robust statistical test and simulation techniques to identify the distribution that describes better the mean time between failure (MTBF) data for this component.

2 Assumptions Made for the choice of the most suitable distribution

To select the appropriate reliability distribution for repairable critical components in complex automotive mechatronic systems, we based our approach on a set of key assumptions. These assumptions allow us to structure our analysis and ensure that the results obtained will be accurate and relevant. The main assumptions used in our study are listed in the table below **"Tab. 1"**:

Table 1. HYPOTHESES USED TO IDENTIFY THE MOST APPROPRIATE DISTRIBUTION

Assumption	**Description**	**Justification**
Failure time distribution	We have assumed that the failure times of repairable components can be modeled by statistical distributions commonly used in reliability analysis, such as exponential distributions, log-normal, and gamma, ...	This assumption is commonly adopted to simplify the initial analysis and allow the application of various statistical tests to validate distribution models.
Self-Diagnosis Data	We have assumed that component MTBF data, taken from self-diagnostic systems, is accurate and complete.	Modern Self-Diagnostic systems are designed to provide detailed and reliable information on component performance, which is crucial for rigorous reliability analysis.
Failure Independence	We have assumed that component failures are independent of each other.	This assumption simplifies the model and allows the use of conventional statistical tests. However, it must be verified for each type of component, according to actual conditions of use.
Constant Use Conditions	We have assumed that component operating conditions are constant and homogeneous over the observation period X.	Although operating conditions may vary, this assumption is necessary to establish the base for comparison between the different reliability distributions.
Repairability model	We have assumed that the components studied are repairable, and that the MTBF data correctly reflect the failure and repair cycles.	Taking reparability in consideration is essential for a realistic reliability analysis, because it influences the intervals between successive failures.

The hypotheses of our study are now established. We are now able to describe the methodology used to select the appropriate reliability distribution. The methodology detailed below will guide us through the various stages of our analysis. This will guarantee a rigorous and systematic approach, ensuring the reliability and validity of our conclusions.

3 Methodology

This section describes the methodology adopted to develop a suitable reliability distribution for repairable critical electronic components used in automobiles. The process followed comprises several key steps, from collecting data from the database relating to each component self-diagnosis to reliability analysis using an adjusted law after validating the distribution. The details of each step are described below **"Fig. 1"**.

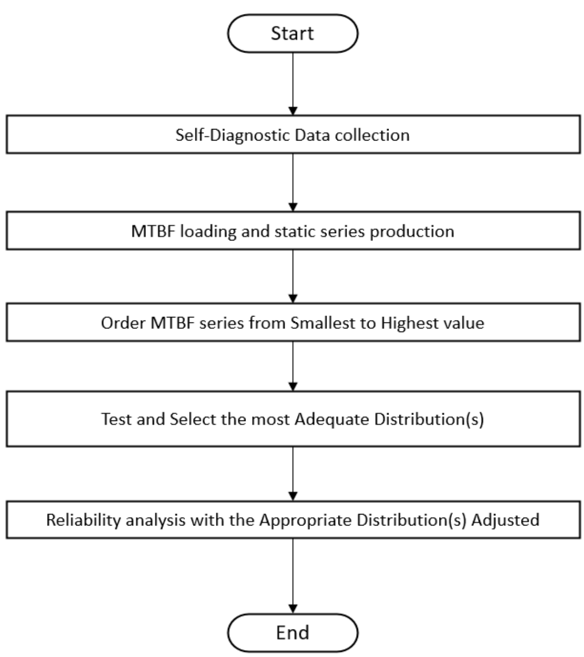
The study starts with the collection of Self-Diagnostic Data for each electronic component. These components have an embedded Self Diagnosis which record faults and allow the calculation of the number of faults with respect to time, then the MTBF to be estimated on a continuous way. The data collected enables us to measure component reliability over time. In fact, we have taken a sample of 50 electronic components that are assumed to be repairable, operating in the same environments (automatic gearbox, for example).

Once the Self-Diagnostic Data has been collected, it is loaded for further analysis. The values in the MTBF series are then ordered from smallest to highest. This step is crucial to facilitate statistical analysis and modeling of reliability data. Ranking the data also helps to identify trends and variations in component performance.

The next step is to test the MTBF Data set using a statistical test. Various statistical distributions, such as the exponential, Weibull, or log-normal, will be tested by comparing their cumulative distribution functions with the empirical data. The distributions that pass the test will be selected, while those that fail will be rejected. Once the suitable distributions are identified, we can proceed with the reliability analysis based on the fitted distribution. This will allow us to estimate key reliability parameters like the failure rate and mean time between failures, which are crucial for enhancing the design and maintenance of electronic components in complex automotive systems.

The steps described above are illustrated in the figure below. This figure shows the complete methodological process, from self-diagnostic data collection to reliability analysis using a fitted law. Each step is represented in clear way the workflow and the progression of the analysis, enabling understanding of the approach adopted to develop the reliability distribution for the 50 electronic components studied.

Figure 1. METHODOLOGY USED TO ELABORATE THE RELIABILITY DISTRIBUTION



After describing in detail, the methodology adopted to develop the reliability distribution, we will explain now how this methodology was applied practically in our study. We will explain the results obtained, the analyses carried out and the conclusions extract from the application of our methodological approach to the electronic components tested.

4 Elaboration of the Reliability distribution: Application of the methodology

4.1 Implementation Of Reliability Distributions For Collected Data

At the beginning, we detail the application of the methodology developed to elaborate a reliability distribution adapted to the electronic components studied. We began by collecting Self-Diagnostic Data for each electronic component. The selected sample comprised 50 similar components studied during an observation period X, operating in the same environments such as automatic gearboxes. The MTBFs were extracted and ordered to begin statistical analysis for the evaluation of distributions. The ordered series DATA of MTBF used in this study is as follows:

[232, 345, 410, 581, 603, 699, 740, 894, 923, 1012, 1123, 1189, 1269, 1381, 1411, 1631, 1697, 1811, 1923, 2056, 2196, 2023, 2241, 2322, 2491, 2601, 2711, 2834, 2894, 2976, 3087, 3115, 3135, 3167, 3216, 3246, 3266, 3268, 3302, 3413, 3426, 3465, 3486, 3511, 3523, 3583, 3678, 3721, 3743, 3811];

This ordered data forms the basis of our study for statistical analysis and evaluation of reliability distributions. To model adequately the reliability of this base, we will present the statistical distributions used in reliability analysis and their parameters to be estimated. The following table **"Tab. 2"** recapitulates the different statistical distributions used in reliability analysis: Weibull, Log-Normal, Exponential, Normal, Gamma, etc., as well as their own parameters to be estimated, which will help us frame our analysis.

Table 2. DIFFERENT STATISTICAL DISTRIBUTIONS USED IN RELIABILITY ANALYSIS

Distribution	Parameters	Probability Density Function (PDF)	Cumulative Distribution Function (CDF)
Weibull	β (shape), η (scale)	$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}}$	$F(t) = 1 - e^{-\left(\frac{t}{\eta}\right)^{\beta}}$
Log Normal	μ (Mean), σ (standard deviation)	$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t-\mu)^2}{2\sigma^2}}$	$F(t) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\ln t-\mu}{\sigma\sqrt{2}}\right)$
Exponential	λ (failure rate)	$f(t) = \lambda e^{-\lambda t}$	$F(t) = 1 - e^{-\lambda t}$
Normal	μ (Mean), σ (standard deviation)	$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$	$F(t) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{t-\mu}{\sigma\sqrt{2}}\right)\right]$
Gamma	k (shape), θ (scale)	$f(t) = \frac{t^{k-1}e^{-t/\theta}}{\theta^k\Gamma(k)}$	$F(t) = \frac{1}{\Gamma(k)}\gamma\left(k, \frac{t}{\theta}\right)$

After presenting the different distributions used in reliability analysis and their elements to be considered, we now need to assess the suitability of these distributions to the observed data. For this step, we'll use the Kolmogorov-Smirnov Test (K-S), a non-parametric statistical tool used to compare cumulative distributions with theoretical distributions.

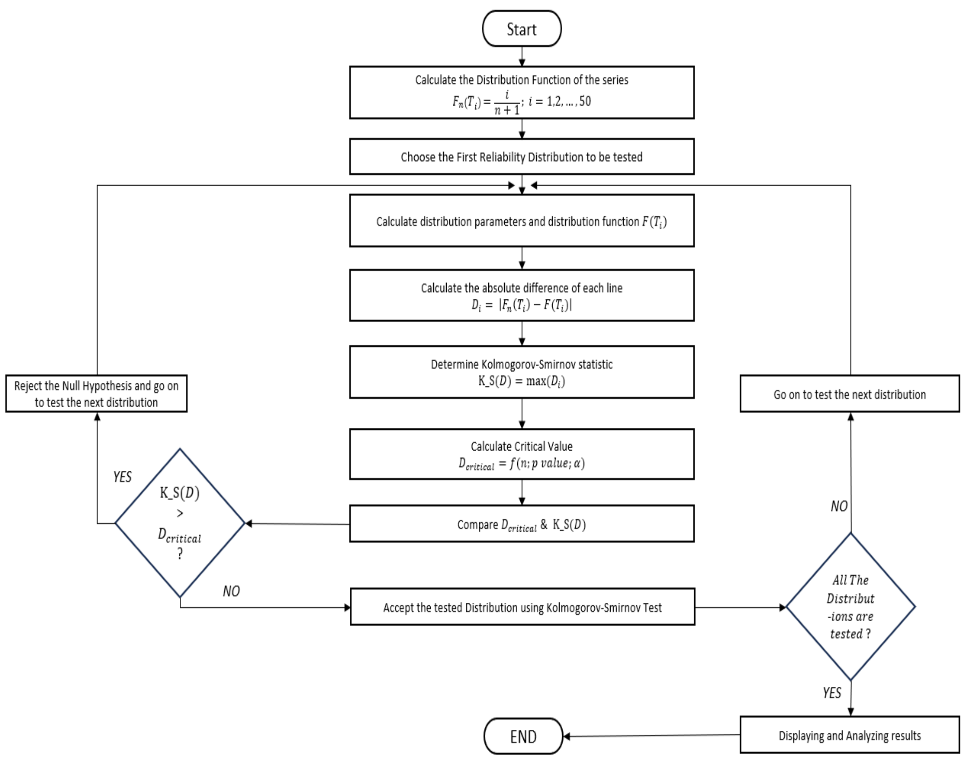
4.2 Kolmogorov Smirnov Test (K-S) For Evaluating Reliability Distributions

The Kolmogorov-Smirnov test is a non-parametric test that is particularly useful for assessing the suitability of different reliability distributions [17]. The test calculates the statistic $K_S(D)$, which represents the greatest absolute difference between the cumulative distribution functions $F_n(x_i)$ relative to the data series and the theoretical $F(T_i)$ of each distribution tested. This test is generally useful for the following reasons:

- It is a non-parametric test that initially makes no specific assumptions about the data distribution.
- It compares distributions against cumulative and theoretical divergences.
- It is a simple test to apply with statistical tools.

The following figure **"Fig. 2"** illustrates our methodology for using the test to evaluate reliability distributions:

Figure 2. METHODOLOGY USED TO EVALUATE THE RELIABILITY DISTRIBUTION.



4.3 Evaluation Of The Theoretical Distributions Of The Data Series Studied By The K-S Test

In our work, we began by estimating the parameters of various distributions (Weibull, Exponential, Normal, Log-Normal, and Gamma) to calculate the absolute differences between each distribution and the cumulative distribution function (CDF) of the data. To do this, we developed a MATLAB program that estimates the parameters of the distributions and plots two curves: the first (Figure 3) represents the absolute differences from the critical value, while the second (Figure 4) illustrates the comparison between the theoretical distributions and the cumulative distribution of the data. The results obtained for estimating the parameters of the various distributions are as follows:

- For the Weibull distribution, the estimated parameters are $\eta = 2594.3$ and $\beta = 2.1717$.
- For the Exponential distribution, the estimated parameter is $\lambda = 4.3335 \times 10^{-4}$.
- For the Log-Normal distribution, the mean and standard deviation are $\mu = 7.559$ and $\sigma = 0.7065$, respectively.
- For the Normal distribution, the mean and standard deviation are $\mu = 2307.62$ and $\sigma = 1121.0984$, respectively.
- Finally, for the Gamma distribution, the parameters are $k = 2.8582$ and $\theta = 807.3665$.

These parameters enable us to model the data in question as closely as possible and to make comparisons with the empirical cumulative distribution (CDF).

After estimating the parameters of the various distributions, we now need to apply the Kolmogorov-Smirnov (K-S) test to calculate the absolute differences between the theoretical Weibull distribution functions $F(T_i)$ and the cumulative distribution function $F_n(x_i)$. Validation of distribution tests according to Kolmogorov-Smirnov requires that the maximum absolute difference between all distributions does not exceed the critical value, determined as follows:

$$K_S(D) < D_{\text{critical}}$$

With:

$$K_S(D) = \max |F_n(T_i) - F(T_i)|,$$
$$D_{\text{critical}} = \frac{1.63}{\sqrt{n}} \approx \frac{1.63}{\sqrt{50}} \approx 0.2305 \text{ for a significance level } \alpha = 0.01.$$

The graph below (Figure 3) has been plotted in MATLAB to show the differences between the cumulative distribution and the various distributions, considering the calculated parameters.

4.4 Results Analysis

Analysis of the results presented in Figures 3 and 4 shows that all the distributions tested are validated by the Kolmogorov-Smirnov test. Indeed, the points in "Fig. 3" representing maximum absolute differences are all below the critical value. However, validating all distributions alone does not guarantee the accuracy required to select the best distribution. This is why we propose a complementary approach to refine the choice of the most appropriate distribution and ensure its reliability. This approach consists of migrating to the distribution closest to the empirical distribution function of the data.

By analyzing Figure 4 in the light of the results obtained in Figure 3, which show the maximum absolute differences between the empirical distribution function and the theoretical distributions tested, it is possible to determine with greater precision the most appropriate distribution for different ranges of MTBF values.

For the interval [232; 670], the Normal Distribution appears to be the most appropriate. Then, for the interval [670; 1108], the Log-Normal Distribution shows a better fit. For values between [1108; 1546], the Gamma Distribution is the most appropriate, while for the interval [1546; 1911], the Weibull Distribution offers the best fit. From [1911; 3007], the Normal

Figure 3. ABSOLUTE DIFFERENCES BETWEEN EMPIRICAL CDF AND VARIOUS DISTRIBUTIONS

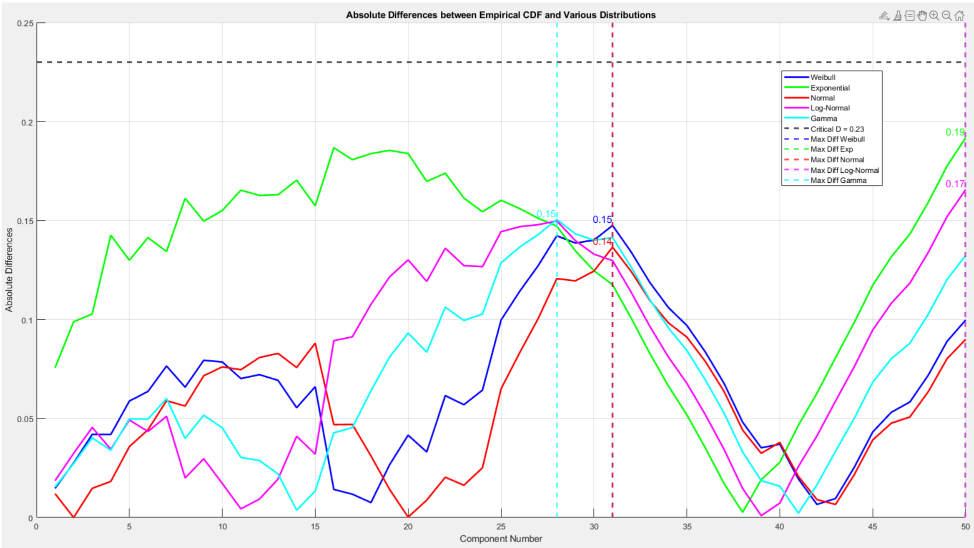
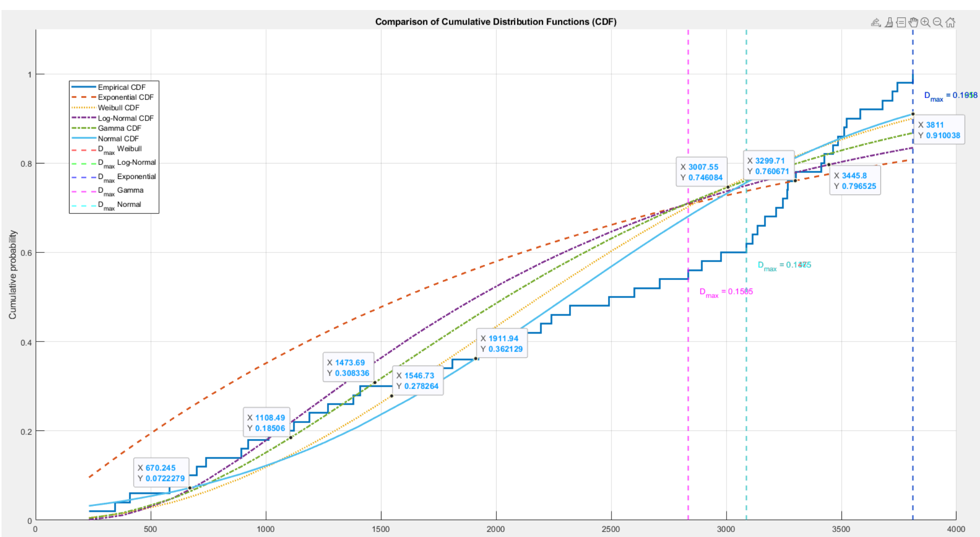


Figure 4. COMPARISON OF CUMULATIVE DISTRIBUTION FUNCTIONS (CDF)



Distribution again becomes the most relevant, followed by the Exponential Distribution for the range [3007; 3292]. Between [3292; 3445], the Log-Normal Distribution again provides the best fit, and finally, for the interval [3445; 4000], the Normal Distribution emerges as the most appropriate.

In summary, we can propose two solutions. The first solution is based on the migration from one distribution to another according to the intervals: let X be the random variable associated with the MTBF values, then:

- For $X \in [232, 670]$: $X \sim \mathcal{N}(\mu_1, \sigma_1)$

 https://doi.org/10.1051/e3sconf/202560100075

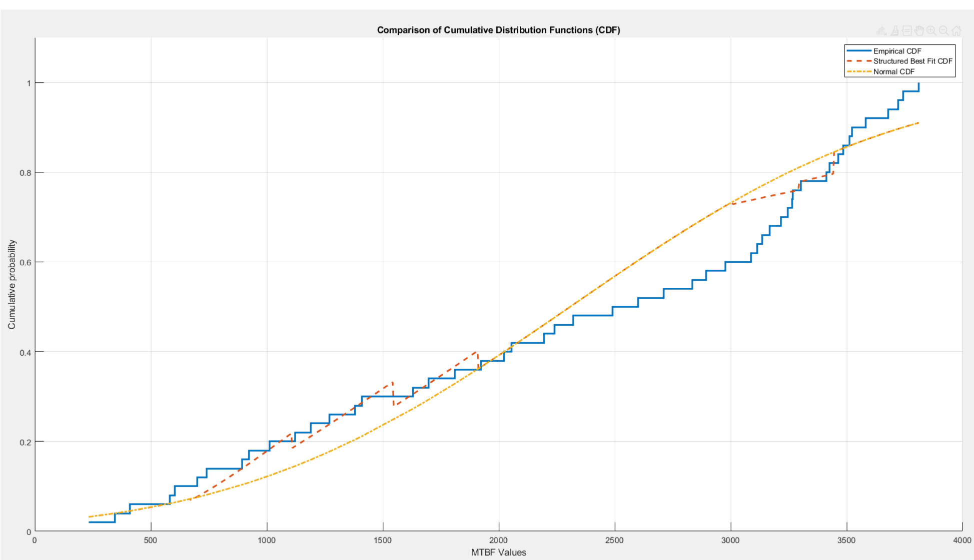
- If $X \in [670, 1108]$: $X \sim \text{Log-Normal}(\mu_2, \sigma_2)$
- If $X \in [1108, 1546]$: $X \sim \Gamma(k, \theta)$
- If $X \in [1546, 1911]$: $X \sim \text{Weibull}(\beta, \eta)$
- If $X \in [1911, 3007]$: $X \sim \mathcal{N}(\mu_1, \sigma_1)$
- If $X \in [3007, 3292]$: $X \sim \text{Exp}(\lambda)$
- If $X \in [3292, 3445]$: $X \sim \text{Log-Normal}(\mu_2, \sigma_2)$
- If $X \in [3445, 4000]$: $X \sim \mathcal{N}(\mu_1, \sigma_1)$

This approach allows us to segment the data according to the optimal distributions for each interval, thus ensuring better modeling of the data.

The second solution is to conclude that the Normal distribution is the most appropriate overall. In fact, it significantly approximates the cumulative distribution function over a large proportion of the intervals, making it the most suitable for modeling the data.

The following figure **"Fig. 5"** illustrates the results of the two proposed solutions. The red curve represents the migration from one distribution to another, aimed at getting as close as possible to the cumulative distribution function (CDF). The yellow curve corresponds to the theoretical normal distribution.

Figure 5. REPRESENTATION OF PROPOSED SOLUTIONS FOR FITTING TO THE CUMULATIVE DISTRIBUTION FUNCTION (CDF).



5 Conclusion and prospects

Validating the most appropriate distribution for the data under study offers several advantages for reliability analysis and modeling of repairable components by choosing the right distribution. Key elements to consider when validating the chosen distribution include improving the accuracy of reliability laws, optimizing maintenance parameters, facilitating predictions, and enhancing risk management.

The use of the Kolmogorov-Smirnov (K-S) test allows us to represent the observed data more accurately, leading to better prediction of future failures. This precision helps in calculating failure rates and optimizing maintenance plans, ensuring more effective preventive actions. Additionally, by simulating future scenarios, we can assess component performance in different environments, making predictions easier and more reliable. Lastly, the validated distribution allows for the identification of high-risk periods, contributing to a more effective risk management strategy that minimizes potential failures and associated risks.

To enrich the conclusion, several perspectives can improve the accuracy of the choice of statistical distributions for reliability analysis. In addition to the Kolmogorov-Smirnov test, the use of complementary tests such as Anderson-Darling or Chi-square offers more robust validation. Artificial intelligence and machine learning can refine distribution fitting by analyzing large quantities of data to identify complex patterns. Bayesian approaches enable dynamic modeling by integrating prior knowledge and updating estimates with new data. The integration of Monte Carlo simulations improves the understanding of the impact of distributions on component performance, optimizing maintenance strategies. Finally, interdisciplinary collaboration, involving experts in engineering, statistics, and computer science, can offer innovative perspectives for the selection and validation of distributions. These complementary approaches can lead to significant improvements in the accuracy and efficiency of reliability models, ensuring more proactive risk management and optimization of maintenance resources.

References

[1] Werner Dieterle, "Mechatronic systems: Automotive applications and modern design methodologies", *Annual Reviews in Control*, Volume 29, Issue 2, 2005, Pages 273-277.

[2] Amna Pir Muhammad, Eric Knauss, Jonas Bärgman, "Human factors in developing automated vehicles: A requirements engineering perspective", *Journal of Systems and Software*, Volume 205, November 2023, 111810.

[3] Rodrigue Sohoin, Abdelkhalak El Hami, Fabrice Guerin, Hassen Riahi, Djelali Attaf, "A novel approach based on meta-modeling technique and time transformation function for reliability analysis of upgraded automotive components", *Reliability Engineering & System Safety*, Volume 207, March 2021, 107357.

[4] Yang Yu, Shibo Wu, Yiqin Fu, Xiaowei Liu, Qingze Zeng, Hongyu Ding, Yu Pan, Yuke Wu, Hao Guo, Yuheng Yang, "Human reliability analysis of offshore high integrity pressure protection system based on improved CREAM and HCR integration method", *Ocean Engineering*, Volume 307, 1 September 2024, 118153.

[5] Liu Jia-Qi, Feng Yun-Wen, Lu Cheng, Pan Wei-Huang, "Decomposed-coordinated framework with intelligent extremum network for operational reliability analysis of complex system", *Reliability Engineering & System Safety*, Volume 242, February 2024, 109752.

[6] E.V. Golikova, "Assessment of Reliability Indicators of the Automotive Equipment to Ensure its Sustainable Operation", *Transportation Research Procedia*, Volume 68, 2023, Pages 249-256.

[7] Hayder M. Issa, "Streamlining aromatic content detection in automotive gasoline for environmental protection: Utilizing a rapid and simplified prediction model based on some physical characteristics and regression analysis", *Results in Engineering*, Volume 21, March 2024, 10177.

[8] Andreas Theissler, Judith Pérez-Velázquez, Marcel Kettelgerdes, Gordon Elger, "Predictive maintenance enabled by machine learning: Use cases and challenges in the au-

tomotive industry", *Reliability Engineering & System Safety*, Volume 215, November 2021, 107864.

[9] Chenghao Liu, Kai Zhang, Zhongwei Deng, Xiaowei Zhao, Xinyu Zhang, Zhenyu Wang, "A failure risk assessment method for lithium-ion batteries based on big data of after-sales vehicles", *Engineering Failure Analysis*, Volume 163, Part B, September 2024, 108559.

[10] Alessandro Giua, Manuel Silva, "Modeling, analysis and control of Discrete Event Systems: a Petri net perspective", *IFAC-PapersOnLine*, Volume 50, Issue 1, July 2017, Pages 1772-1783.

[11] Junyu Jiang, Yuanbin Yu, Haitao Min, Qiming Cao, Weiyi Sun, Zhaopu Zhang, Chunqi Luo, "Trip-level energy consumption prediction model for electric bus combining Markov-based speed profile generation and Gaussian processing regression", *Energy*, Volume 263, Part D, 15 January 2023, 125866.

[12] Mohammad Hassan Bahmani, Mostafa Esmaeili Shayan, Davide Fioriti, "Assessing electric vehicles behavior in power networks: A non-stationary discrete Markov chain approach", *Electric Power Systems Research*, Volume 229, April 2024, 110106.

[13] Laura Attardi, Maurizio Guida, Gianpaolo Pulcini, "A mixed-Weibull regression model for the analysis of automotive warranty data", *Reliability Engineering & System Safety*, Volume 87, Issue 2, February 2005, Pages 265-273.

[14] Artur J. Lemonte, "A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function", *Computational Statistics & Data Analysis*, Volume 62, June 2013, Pages 149-170.

[15] Saralees Nadarajah, Jiahang Lyu, "New bivariate and multivariate log-normal distributions as models for insurance data", *Results in Applied Mathematics*, Volume 14, May 2022, 100246.

[16] Lechang Yang, Xinyao Zhang, Zitong Lu, Yuqiang Fu, David Moens, Michael Beer, "Reliability evaluation of a multi-state system with dependent components and imprecise parameters: A structural reliability treatment", *Reliability Engineering & System Safety*, Volume 250, October 2024, 110240.

[17] Olayinka S. Ohunakin, Emerald U. Henry, Olaniran J. Matthew, Victor U. Ezekiel, Damola S. Adelekan, Ayodele T. Oyeniran, "Conditional monitoring and fault detection of wind turbines based on Kolmogorov–Smirnov non-parametric test", *Energy Reports*, Volume 11, June 2024, Pages 2577-2591.

[18] José María Luna-Romera, María Martínez-Ballesteros, Jorge García-Gutiérrez, José C. Riquelme, "External clustering validity index based on chi-squared statistical test", *Information Sciences*, Volume 487, June 2019, Pages 1-17.

[19] George Almpanidis, Constantine Kotropoulos, "Phonemic segmentation using the generalised Gamma distribution and small sample Bayesian information criterion", *Speech Communication*, Volume 50, Issue 1, January 2008, Pages 38-55.

[20] Aritz Adin, Elias Teixeira Krainski, Amanda Lenzi, Zhedong Liu, Joaquín Martínez-Minaya, Håvard Rue, "Automatic cross-validation in structured models: Is it time to leave out leave-one-out?", *Spatial Statistics*, Volume 62, August 2024, 100843.