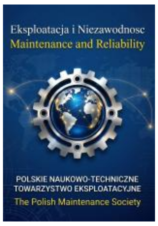


Eksploatacja i Niezawodnosc - Maintenance and Reliability

Volume 29 (2027), Issue 2

https://ein.org.pl

The Polish Maintenance Society

A reliability-oriented method for constructing baseline operating conditions of electric drive systems based on minimum cycle identification

Jingfei Fu[1], Dongdong Zhang[1], Zhen Wang[2], Chunhui Gong[3], Qianbin Liu[4], Shibo Bin[5], Shanchao Wang[5], Li-Hui Zhao[1,*]

[1] School of Mechanical Engineering, University of Shanghai for Science and Technology, China
[2] Base for Postdoctoral Innovation and Practice of Yuxin Electronic Technology Group Co., Ltd., China
[3] Central Research Institute for Product R&D, JianglingMotorsCo.,Ltd, China
[4] School of Mechanical and Automotive Engineering, South China University of Technology, China
[5] Dongfeng Liuzhou MOTOR Co.,Ltd, China
* Corresponding Author: pheigoe@126.com

Received: 25 April 2026
Revised: 31 July 2026
Accepted: 14 August 2026
Online: 10 September 2026

This is an open access article under the CC BY 4.0 license

Abstract

Electric drive systems in new energy vehicles operate under complex and time-varying conditions that accelerate component degradation and affect system reliability. Constructing representative operating conditions is therefore essential for reliability evaluation. This study proposes a method for constructing baseline operating conditions based on minimum cycle identification and category-based sampling. Multiscale load data are segmented using a variable sliding-window method, and indicators from time, frequency and damage domains are used to determine the minimum cycle period through a relative recurrence index. Operating-condition segments are then clustered according to damage-rate characteristics and sampled within categories to maintain statistical consistency with the overall dataset. A case study based on operational data from 30 vehicles shows that 90% of users have minimum cycle periods shorter than 11,587 km. The difference in damage-rate means is below 1%, while the joint probability distribution error is lower than 0.001.

Keywords: *electric drive system, reliability evaluation, baseline operating conditions, minimum cycle period, big data of users*

Article citation:
Fu J, Zhang D, Wang Z, Gong C, Liu Q, Bin S, Wang S, Zhao L-H, A reliability-oriented method for constructing baseline operating conditions of electric drive systems based on minimum cycle identification, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2027: 29(2) http://doi.org/10.17531/ein/231130

Highlights

- Variable windows segment electric-drive loads.
- A 95% threshold identifies the minimum cycle.
- Damage-rate features enable six-category KFCM clustering.
- Category sampling preserves load and damage patterns.
- The sampled data support reliability-oriented baseline testing.

1. Introduction

As the core of new energy vehicle powertrains, electric drive systems experience complex, time-varying operating conditions and multidimensional loads, which accelerate the degradation of critical components and compromise overall system reliability [1,2]. Establishing high-fidelity baseline operating conditions is essential not only for system-level fatigue damage evaluation and accelerated life testing, but also for load-spectrum optimisation and reliability-oriented product design in forward engineering processes [3,4]. However, substantial variability in user driving behaviours and randomness in operating environments lead to pronounced individual differences in users' baseline condition datasets. Furthermore, traditional approaches typically require collecting road-load data over the entire service life, resulting in high acquisition costs and limited user coverage, thereby constraining the reliability assessment of electric drive systems.

More recently, advanced methods such as Integrated-Dispersion Manifold Distance (IMD) [5], distribution-based health indicators [6,7], and reliability modeling approaches under small-sample conditions [8] have been developed to improve sensitivity to degradation evolution and state discrimination.Over long-term service, user operating conditions exhibit measurable repeatability. As the service duration increases, the load distribution characteristics of the electric drive system gradually converge toward those of the full life cycle [9]. In this context, a key question arises: how long should the operational observation window be so that the resulting sample can adequately represent the full operating population? To address this issue, this paper introduces the concept of minimum cycle period, which is defined as the shortest observation horizon at which the discrepancy between the sample operating distribution and the full population operating distribution falls within an acceptable representativeness criterion. It should be noted that the full population distribution is constructed using all available user data over the entire observation period, and the representativeness criterion is defined by measuring the discrepancy between the sample-based distribution and this population-level reference. If the selected cycle is excessively long, the deviation becomes negligible but at the expense of unnecessarily high test duration and data costs. Conversely, an overly short cycle leads to significant discrepancies between sample and population distributions, weakening its representativeness of full-lifecycle loads. Therefore, analysing trend patterns in load characteristics throughout users' service cycles and determining the minimum cycle period are crucial for constructing a reliable baseline operating condition database for electric drive systems [10].

Existing studies both domestic and international, have proposed various methods to determine minimum sample sizes based on relationships between samples and population [11]. When the population mean and variance are known, statistical techniques such as Z-tests, t-tests and chi-square tests can be used to estimate the minimum sample size under specified confidence levels and error limits [12]. Gope [13] derived the minimum required sample size based on target survival probability and confidence level for fatigue life estimation, while Xiong [14] provided a theoretical derivation framework for sample-size determination. In the field of vehicle load acquisition, researchers [15] used logarithmic pseudo-damage density under measured load spectra as a statistical indicator, modelling relative errors via the t-distribution. Another study [16] examined cumulative power spectral density and damage-distribution characteristics and developed a Weibull-based sampling model for determining minimum load-spectrum length. These works provide useful methodological references for determining the minimum cycle period in the present study.

As operational data accumulate, quantifying the trend-based relationship between operational samples and the population becomes increasingly important. Existing studies have largely relied on constructing test statistics combined with Markov chain Monte Carlo simulations to analyse trend convergence and thus determine minimum cycle duration [17]. In multi-condition analysis of wheel loader [18], a composite indicator incorporating extreme-load mean, standard deviation and fatigue-life metrics was developed using a hybrid subjective-objective weighting method to estimate the minimum sample size. For determining road-load spectrum acquisition length in vehicles, regression models describing statistical trend characteristics were constructed, and spectrum length was defined based on relative errors between population parameters and regression results [19]. However, for electric drive systems featuring multiple failure modes and heterogeneous component degradation rates, single-feature-based sample-size determination is insufficient. Consequently, constructing trend indicators across multiple domains—such as the time domain, frequency domain and damage domain—becomes necessary.

The minimum cycle period defines the temporal scale for sample selection. To construct a reliable baseline operating condition database, representative operating-condition segments must be extracted from the overall dataset. Classical sampling methods—such as simple random sampling [20], stratified sampling [21] and systematic sampling [22]—are effective for low-dimensional datasets but are limited for complex multi-feature operating conditions. Latin hypercube sampling [23] offers better uniformity across multidimensional features, balancing sampling coverage and efficiency. However, because this method selects optimal samples solely based on load-distribution similarity, without considering component-level failure physics, the resulting operating-condition sets may deviate from true system damage-distribution characteristics.

To address these challenges, this paper proposes a reliability-oriented baseline operating condition library construction method based on minimum cycle period identification and category-based random sampling. Multiscale segmentation of load data is achieved using a variable sliding-window approach. A multidimensional evaluation metric system is developed to analyse trend characteristics in time-domain, frequency-domain and damage-domain features. The concept of relative recurrence is introduced to quantify users' minimum cycle periods. Classification-based random sampling is then applied to extract representative operating-condition segments, ensuring consistency between the sampled and population-level load distributions and damage-rate characteristics. Based on these procedures, a high-fidelity baseline operating condition library is constructed for reliability assessment of electric drive systems.

This study therefore develops a systematic methodology for constructing baseline operating conditions under massive and complex user-load datasets. The proposed framework includes:

(1) multiscale user data segmentation and multidimensional metric construction,

(2) determination of minimum cycle periods using relative recurrence,

(3) category-based random sampling incorporating damage-rate similarity, and

(4) verification of distributional consistency in both load and damage characteristics.

The overall technical workflow is illustrated in Fig. 1.

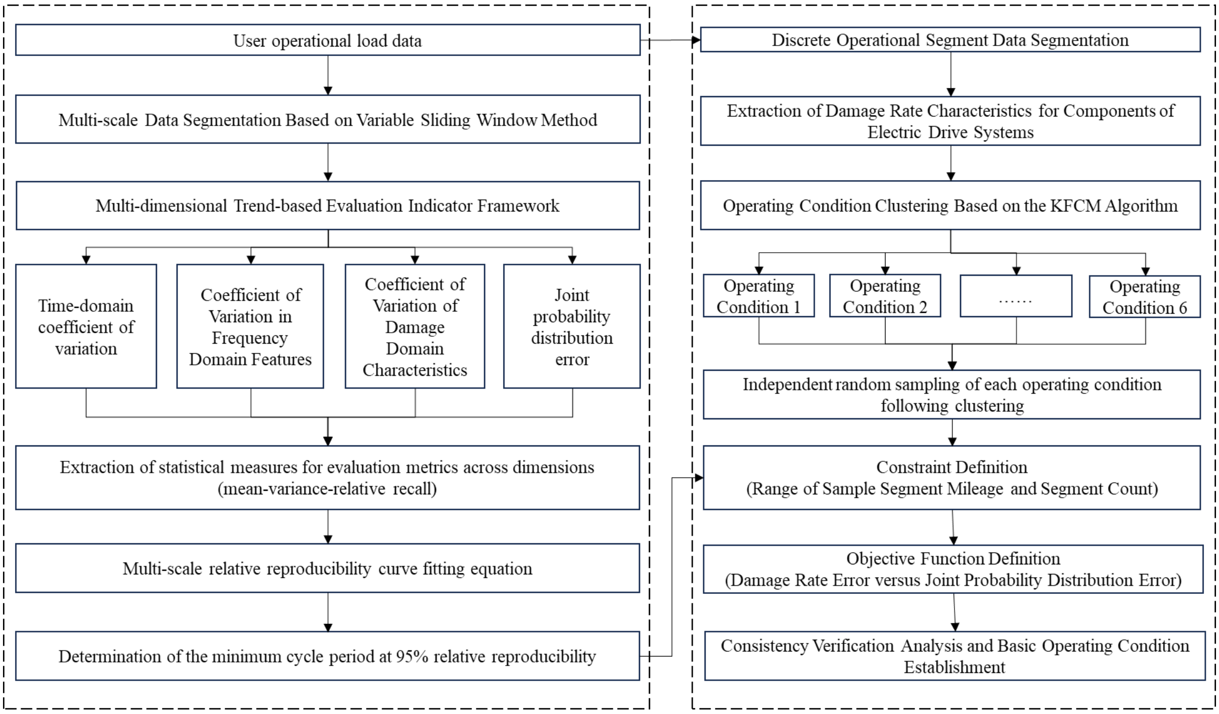


Figure 1. Process for establishing baseline operating conditions for electric drive system reliability.

2. Method for determining the minimum cycle period

2.1. Fragment segmentation based on the variable sliding window method

Here, the minimum cycle period is defined as the shortest observation horizon at which the operational characteristics extracted from a sample window can adequately represent those of the full population. Let $P(x)$ denote the joint distribution of operating variables estimated from the full dataset, and let $PW(x)$ denote the corresponding joint distribution estimated from an observation window of length W. The discrepancy between these two distributions is quantified by an error functional $E(W)$. Accordingly, the minimum cycle period is defined as the smallest W satisfying a prescribed representativeness criterion. In terms of the error threshold, it can be written as:

$$W_{\min} = \inf\{W : E(W) \leq \varepsilon\} \tag{1}$$

When expressed in terms of relative recurrence $R(W)$, the criterion becomes:

$$W_{\min} = \inf\{W : R(W) \geq \eta\} \tag{2}$$

where η is a predefined recurrence threshold.

The variable sliding-window method is a segmentation technique designed for multidimensional time-series data. By dynamically adjusting the window size and stride, the method enables adaptive segmentation of load data across different temporal scales. Let denote the multidimensional time-series load data of an electric drive system, where represents the total sequence length and is the load vector at time point including variables such as rotational speed, torque, current, voltage, and temperature. Given an initial window size and a sliding stride, the total number of sample fragments obtained from the original load sequence is computed as:

$$N = \left\lfloor \frac{n-w}{s} \right\rfloor + 1 \tag{3}$$

where$\lfloor \cdot \rfloor$ denotes the floor operator.

Beginning from the start of the raw data, a sliding window of size w moves forward along the timeline in increments of s. For the *k-th* fragment, the corresponding load segment is expressed as $\boldsymbol{X}_k = \{\boldsymbol{x}_{(k-1)s+1}, \boldsymbol{x}_{(k-1)s+2}, \cdots, \boldsymbol{x}_{(k-1)s+w}\}$ .After traversing the entire sequence, the complete set of segmented sample fragments is given by:

$$\boldsymbol{Y}_w^s = \Omega_{k=1}^N \{\boldsymbol{X}_k\} \tag{4}$$

By adjusting parameters such as the window size w and stride s, sample sets corresponding to different temporal scales can be obtained. The segmentation procedure is illustrated in Fig. 2.

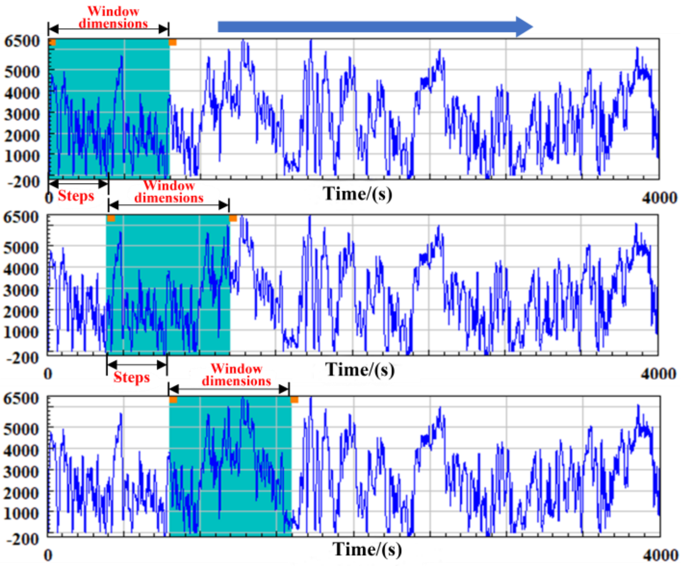


Figure 2. Schematic diagram of dataset partitioning using the variable sliding window method.

2.2. Construction of a multi-dimensional evaluation indicator system

To overcome the limitations of defining the minimum cycle period based solely on a unidimensional characteristic, this study develops a multidimensional evaluation metric system incorporating time-domain, frequency-domain, damage-domain and load-distribution perspectives. As the sample scale increases, the trend evolution of each metric is analysed. However, due to the heterogeneity in feature dimensions, it is difficult to characterise trends consistently across scales. To address this issue, the coefficient of variation within each feature dimension is adopted to evaluate the relative dispersion of the sample set at different scales. The multidimensional evaluation metrics considered in this section include:

(1) coefficient of variation of time-domain features,

(2) coefficient of variation of frequency-domain features,

(3) coefficient of variation of damage-domain features,

(4) the error between the joint rotational-speed–torque probability distributions of sample and population data.

For the time-domain dimension, 18 time-domain features are computed for all segments corresponding to each window size. These features include typical time-series characteristics such as speed, acceleration, and torque, as well as statistical descriptors including mean, extreme values and variance [24]. Suppose that N sample fragments are obtained for a given window size w_k. The mean value of the *i-th* time-domain feature is calculated as:

$$\bar{F}_i^{\text{time}} = \frac{1}{N} \sum_{j=1}^N F_{ij}^{\text{time}} \tag{5}$$

Where F_{ij}^{time} denotes the *i-th* feature for the *j-th* sample fragment, $i = 1,2,\cdots,18$; $j = 1,2,\cdots,N$.

The standard deviation of feature F_{ij}^{time} across N sample fragments is given by:

$$\sigma_{F_i^{\text{time}}} = \sqrt{\frac{1}{N} \sum_{j=1}^N (F_{ij}^{\text{time}} - \bar{F}_i^{\text{time}})^2} \tag{6}$$

The coefficient of variation of feature $F_{ij}^{\text{time}}$is then computed as:

$$CV_{F_i^{\text{time}}} = \frac{\sigma_{F_i^{\text{time}}}}{\bar{F}_i^{\text{time}}} \tag{7}$$

The coefficient-of-variation matrix for time-domain features, representing the dispersion characteristics across all feature dimensions, is expressed as:

$$CV_\Omega^{\text{time}} = \{CV_{F_1^{\text{time}}}, CV_{F_2^{\text{time}}}, \cdots, CV_{F_{18}^{\text{time}}}\} \tag{8}$$

For a given window size w_k, the mean value $\mu_{CV_\Omega^{\text{time}}}$ of the time-domain feature coefficients reflects the overall level of the time-domain evaluation metric at that scale. Conversely, the standard deviation $\sigma_{CV_\Omega^{\text{time}}}$ of these coefficients represents the dispersion of the time-domain evaluation metric for the same window size. To compute the coefficients of variation for frequency-domain features across different sample segments, the frequency-domain characteristics must first be extracted. Twelve frequency-domain indicators are considered in this study, including the mean frequency, centre-of-gravity frequency, root-mean-square frequency and frequency standard deviation for each segment, covering vehicle speed, acceleration and torque. Similarly, the mean value $\mu_{CV_{\Omega}^{\text{freq}}}$ of the frequency-domain coefficients of variation characterises the average level of the frequency-domain evaluation metric,while the standard deviation $\sigma_{CV_{\Omega}^{\text{freq}}}$ quantifies the dispersion of the frequency-domain evaluation metric at that scale. For the damage-domain dimension, this study considers seven critical failure modes/components of the electric drive system, namely shaft torsional fatigue, tooth-surface contact fatigue, tooth-root bending fatigue, bearing A fatigue, bearing B fatigue, rotor magnetic-bridge fatigue, and stator winding thermal ageing [25-30]. For each sample segment, the damage corresponding to each failure mode/component is calculated using load-counting procedures, physics-based fatigue-life models, and linear damage accumulation rules. The damage rate of each segment is then obtained by dividing the calculated damage by the segment mileage, thereby generating seven mileage-normalized damage-rate indicators. The coefficients of variation of these seven damage-domain indicators are subsequently computed across all segments. As with the other dimensions, the mean value $\mu_{CV_{\Omega}^{\text{damage}}}$ represents the average level of the damage-domain evaluation metric , while the standard deviation $\sigma_{CV_{\Omega}^{\text{damage}}}$ represents its dispersion at the corresponding scale.

To compute the error in the joint probability distribution, let $P_j(S,T)$ denote the joint rotational-speed–torque probability distribution of the *j-th* sample segment for a given window size, and let $P_{All}(S,T)$ denote the corresponding distribution for the population data. The joint probability distribution error between the sample segment and the population is defined as:

$$E_j = \iint_D [P_j(S,T) - P_{All}(S,T)]^2 \, dSdT \tag{9}$$

The set of joint-distribution errors across all N sample segments is then expressed as:

$$E_\Omega = \{E_1, E_2, \cdots, E_N\} \tag{10}$$

Here, the integration domain D denotes the effective operating region of the electric drive system in the speed–torque space:

$$D = [n_{\min}, n_{\max}] \times [T_{\min}, T_{\max}] \tag{11}$$

where n and T represent rotational speed and torque, respectively. In implementation, the continuous integral is approximated by summation over discretized bins covering the valid operating range of the dataset.

Given the variability across multiple sample segments, the mean u_{E_Ω} and standard deviation σ_{E_Ω} of the squared-error set are used as key indicators to characterise the average level and dispersion of the joint probability-distribution errors between the samples and the population.

2.3. Relative recurrence fitting model

Based on the four-dimensional evaluation metrics defined in Section 2.2, a relative recurrence fitting model is introduced to comprehensively assess and quantify the trend behaviour of the evaluation metrics. The relative recurrence $\delta_{w_k}^m$ of the *m-th* evaluation metric within dimension w_k is defined as:

$$\delta_{w_k}^m = 1 - \frac{\mu_{w_k}^m / max(\boldsymbol{U})}{\sum_{w_1}^{w_p} (\mu_{w_k}^m / max(\boldsymbol{U}))} \tag{12}$$

where $\boldsymbol{U} = \{\mu_{w_1}^m, \mu_{w_2}^m, \cdots, \mu_{w_p}^m\}$ denotes the mean value of the *m-th* evaluation metric, with these values obtained across a range of different window sizes.

Subsequently, polynomial fitting is applied to construct the relative recurrence evolution model. Assuming a fitting order of *i*, $i = 0,1,2,\cdots,n$, the polynomial function is expressed as:

$$F(\delta_{w_k}^m) = \sum_{i=0}^n a_i \cdot (\delta_{w_k}^m)^i \tag{13}$$

where $F(\delta_{w_k}^m)$ denotes the fitted function, a_i represents the polynomial coefficient of order *i*. These coefficients are estimated using the least-squares method:

$$\frac{\partial}{\partial a_i} \sum_{w_1}^{w_p} \left[F(\delta_{w_k}^m) - \delta_{w_k}^m\right]^2 = 0 \tag{14}$$

As the model order increases, the coefficient of determination $R^2 > 0.95$ is used as the fitting accuracy criterion. When the fitted recurrence reaches 0.95, the corresponding sample scale is regarded as the minimum required scale for the *m-th* evaluation metric, denoted $t_{0.95}^m$ .To ensure that all evaluation metrics meet the recurrence threshold of 0.95, the minimum user cycle period is defined as:

$$T_{MSS} = max\{t_{0.95}^1, t_{0.95}^2, t_{0.95}^3, t_{0.95}^4\} \tag{15}$$

In this study, polynomial fitting was adopted to describe the recurrence evolution of the multidimensional evaluation metrics because the recurrence values vary continuously with the window size but may still exhibit local fluctuations caused by sample discreteness. Compared with direct pointwise judgment based on the raw recurrence sequence, polynomial fitting provides a smoother and more analytically tractable representation of the recurrence trend, which facilitates stable identification of the window size corresponding to a prescribed recurrence level.

The recurrence threshold was set to 0.95 as a compromise between representativeness and efficiency. If the threshold is set too low, the resulting cycle period may be insufficient to ensure adequate consistency between the sample and the population. In contrast, an excessively high threshold would lead to unnecessarily long observation windows and increased data-acquisition cost. Therefore, the 0.95 threshold was adopted to ensure that the evaluation metrics had approached a sufficiently stable level while maintaining practical compactness of the minimum cycle period.

3. Method for establishing baseline operating conditions

3.1. Determination of the range of operating condition segments

To address the inconsistency in sample mileage across different segments under the minimum cycle period, the sample mileage range is defined as follows. Under the variable sliding-window method, when the window size equals the minimum cycle period T_{MSS} , the dataset is divided into N_T segments with corresponding mileage set $\Omega_{i=1}^{N_T}\{Mile_i\}$. The mileage range associated with the minimum cycle period is then given by:

$$Mile_{MSS} \in (min(\Omega_{i=1}^{N_T}\{Mile_i\}), max(\Omega_{i=1}^{N_T}\{Mile_i\})) \quad (16)$$

Assuming that the user has n running segments with mileage m_i , $i = 1,2,\cdots,n$, the number of small-sample operating segments required for filtering is calculated as:

$$n_{sample} \in \frac{Mile_{MSS}}{\sum_{i=1}^{n} m_i/n} \quad (17)$$

3.2. Classification random sampling

To ensure sampling uniformity, overcome the limitations of traditional methods, and maintain consistency in both load-distribution and damage-distribution characteristics between sampled segments and the population, a clustering-based random sampling strategy is adopted. First, all operating-condition fragments are grouped into distinct categories based on damage-rate characteristics. Then, random sampling is performed independently within each category. A sampling objective function is defined, and the number of sampling iterations is specified to obtain the optimal sample set.

The Davies–Bouldin (DB) index is used to determine the optimal number of clusters before clustering. This index accounts for both intra-cluster cohesion and inter-cluster separation.

$$DB = \frac{1}{n}\sum_{i=1}^{n} \max_{j\neq 1}\left(\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right) \quad (18)$$

where $c_i$is the *i-th* cluster centre, *n* is the number of cluster categories, and σ_i is the average distance from all sample points in category i to its respective cluster centre. $d(c_i, c_j)$ is the average distance between the centres of clusters *i* and *j*. A lower DB index value is indicative of a more satisfactory clustering outcome.

During the clustering process, the damage-rate characteristics of each component across all segments are used as input features. Specifically, seven damage-rate features were constructed for clustering, corresponding to the seven critical damage modes/components considered in the damage-domain analysis, including shaft torsional fatigue, tooth-surface contact fatigue, tooth-root bending fatigue, bearing A fatigue, bearing B fatigue, rotor magnetic-bridge fatigue, and stator winding thermal ageing. For each operating-condition fragment, the damage value of each failure mode was first calculated based on the corresponding load-history segment and physics-based damage model, and then normalized by the fragment mileage to obtain the damage rate. Accordingly, each fragment was represented by a seven-dimensional damage-rate feature vector for KFCM clustering.

By introducing these seven damage-rate features, the clustering process reflects not only the load similarity among fragments, but also the multi-component degradation consistency of the electric drive system. This makes the obtained operating-condition categories more physically relevant for reliability-oriented baseline operating condition construction. Kernel Fuzzy C-Means (KFCM) is employed to enhance linear separability in high-dimensional feature space by reconstructing the fuzzy C-means objective function using a Gaussian radial basis kernel. Let feature vector $\boldsymbol{Z} = \{z_1, z_2, \cdots, z_n\} \in R^s$ be mapped into feature space $\boldsymbol{H}$ by mapping$\Phi: \boldsymbol{Z} \to \boldsymbol{H}: \Phi(z) = y$. Let the cluster centre matrix be $V = [v_1, v_2, \cdots, v_c]_{s\times c}$. Let $L_{kfcm}(\boldsymbol{U}, \boldsymbol{V})$ denote the objective function of the KFCM clustering algorithm.

$$L_{kfcm}(\boldsymbol{U},\boldsymbol{V}) = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji}^{k}\left\|\Phi(z_i) - \Phi(v_j)\right\|_{\boldsymbol{H}}^{2} = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji}^{k}\,(2 - 2K(z_i, v_j)) \quad (19)$$

Where $\boldsymbol{U} = (u_{ji})_{c\times n}$is the fuzzy membership matrix of *n* samples to c categories, $k$is the fuzzy weighting exponent, and $K(z_i, v_j)$ is the Gaussian radial basis function, which can be expressed as:

$$K(z_i, v_j) = exp(\frac{-\|z_i - v_j\|^2}{2\sigma^2}) \quad (20)$$

The number of segments extracted for each operating condition is calculated based on the proportion of mileage for each condition relative to the total number n_{sample} . Two objective functions are used:the load-distribution objective, defined by the joint speed–torque probability error;the damage-distribution objective, defined by deviations in damage-rate parameters between the sample and the population. It is assumed that the joint probability-distribution error must be below threshold δ_1 and the deviation in damage-rate parameters must be below thresholdδ_2. The objective functions shall satisfy the following requirements:

$$\begin{aligned} &\iint_D [P_j(S,T) - P_{All}(S,T)]^2 \, dSdT < \delta_1 \\ &|F(D_s) - F(D_{all})| < \delta_2 \end{aligned} \quad (21)$$

Where $F(D_s)$ is the damage rate distribution parameter for the sample under consideration, and $F(D_{all})$ is the damage rate distribution parameter for the population sample.

3.3. Consistency test analysis

Given the multifaceted nature of the sampled dataset, which contains numerous combinations of operating-condition segments, sampling validity is evaluated using hypothesis tests, including the non-parametric K-S test and the parametric Z- and F-tests, which are used to analyse the consistency between the sample and the population.

The two-sample K–S test is used to evaluate the consistency between the sample and population distributions based on their cumulative distribution functions. The formula for calculating the test statistic is as follows:

$$D_{KS} = max(\,|F_z(x) - F_s(x)|) \quad (22)$$

Where $F_z(x)$ denotes the theoretical distribution function of the population, and $F_s(x)$ denotes the empirical distribution function of the sampled data.

The Z-test is used to examine the difference between the sample mean and the population mean. The test statistic for the *Z-test* is as follows:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}} \quad (23)$$

Where$\bar{x}$ denotes the sample mean, μdenotes the population mean, σ denotes the population standard deviation, and Ndenotes the sample size.

The F-test is used to determine whether a significant difference exists between the variances of two samples. The test statistic for the *F-test* is:

$$F = \frac{s_1^2}{s_2^2} \quad (24)$$

where s_1^2 denotes the sample variance, s_2^2 denotes the population variance, and the test statistic follows an F-distribution with $(n_1 - 1, n_2 - 1)$ degrees of freedom.

A significance level $\alpha = 0.05$ is specified for the hypothesis tests. If the asymptotic p -value is less than t α, the null hypothesis is rejected. If the p -value is greater than or equal to α, the null hypothesis is not rejected.

4. Case implementation and outcome analysis

4.1. User data analysis

The user data employed in this study includes operational records from 30 vehicles of the same model, reflecting the real-world driving behaviours and road conditions of different users over their recorded operating periods. Key statistical metrics, including operating time, accumulated mileage, overall damage contribution, and damage occurrence rate, are presented in Fig. 3. The overall damage contribution is calculated as a weighted average by considering the damage levels of different components. The damage occurrence rate is defined as the ratio of the user's total damage contribution to the corresponding accumulated mileage.

A goodness-of-fit analysis was carried out using a log-normal distribution model to characterize the statistical patterns of operating time, accumulated mileage, overall damage contribution, and damage occurrence rate across users. The corresponding probability density functions and cumulative distribution functions are shown in Figs. 4 and 5, respectively.

The results in Figs. 4 and 5 indicate clear inter-user differences in damage-rate characteristics. To determine the minimum cycle period and construct the baseline operating conditions, four representative users were selected according to cumulative damage-rate (DR-CDF) levels of 25%, 50%, 75%, and 95%. To improve robustness, the representative users were identified from the fitted DR-CDF curve rather than by direct pointwise extraction from the empirical cumulative distribution.

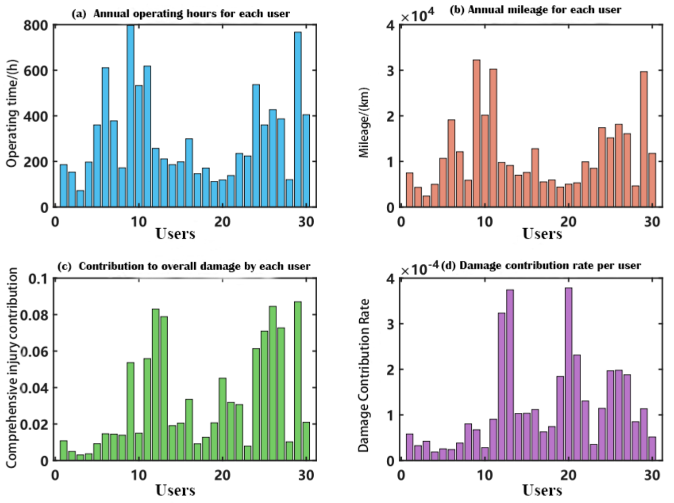
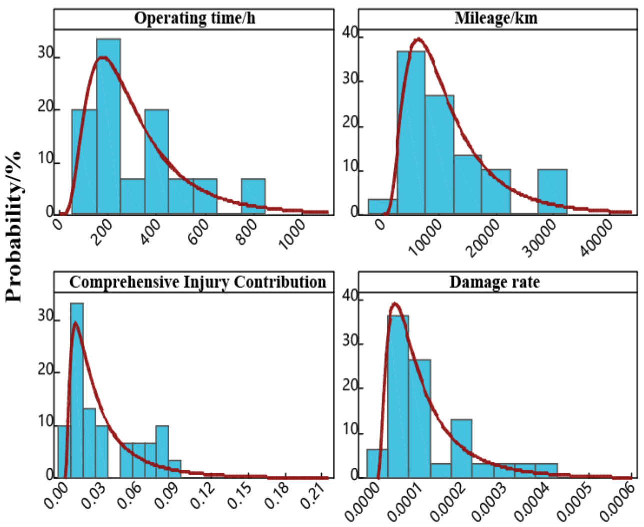


Figure 3. Statistics on different user characteristics.



Fiure 4. Probability distribution of various statistical characteristics among users.

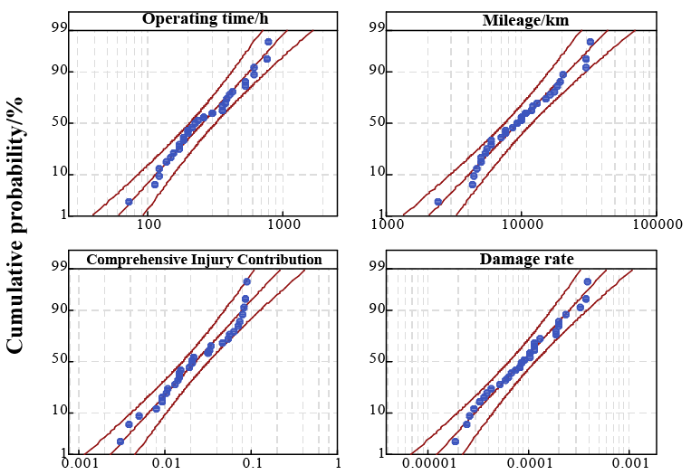


Figure 5. Cumulative probability distribution of various statistical characteristics for users.

This treatment reduces the influence of local fluctuations and provides a smoother percentile-based mapping between cumulative probability level and representative user position. In addition, the 95% DR-CDF level was selected as the upper representative group instead of using the extreme tail of the distribution, because values too close to 100% are more sensitive to tail randomness and individual outliers. Therefore, the 95% DR-CDF level can better represent high-damage users while maintaining statistical robustness.

4.2. Determination of the minimum cycle period

To avoid repetitive presentation, only the representative results for the lower-damage case (25% DR-CDF) and the higher-damage case (95% DR-CDF) are shown in detail. The quantitative results for all four representative user groups are summarized in Table 1.

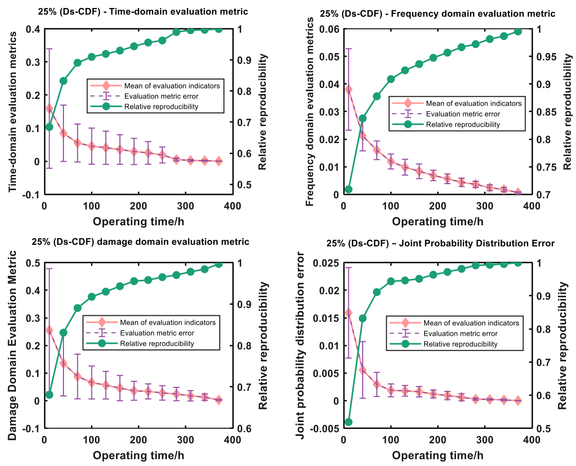


Figure 6. Trend chart of multi-dimensional evaluation indicators under 25% DR-CDF.

For the representative user corresponding to a cumulative damage rate of 25%, the total recorded operating time was 378 h and the accumulated mileage reached 12,124 km. Window sizes ranging from 10 h to 370 h, with increments of 30 h, were applied sequentially. For each window size, the dataset was segmented into operating-condition fragments using a sliding step of 1 h. Multidimensional evaluation metrics were then established from the time-domain, frequency-domain, damage-domain, and joint distribution-error perspectives. The mean values, errors, and relative recurrence of these evaluation metrics at different time scales are shown in Fig. 6.

As illustrated in Fig. 6, with increasing operating duration, the evaluation metrics in all dimensions gradually decrease and stabilize, while their corresponding errors also diminish. To further quantify the minimum cycle period, relative recurrence fitting curves were constructed for each dimension, and the operating duration corresponding to 95% recurrence was identified, as shown in Fig. 7.

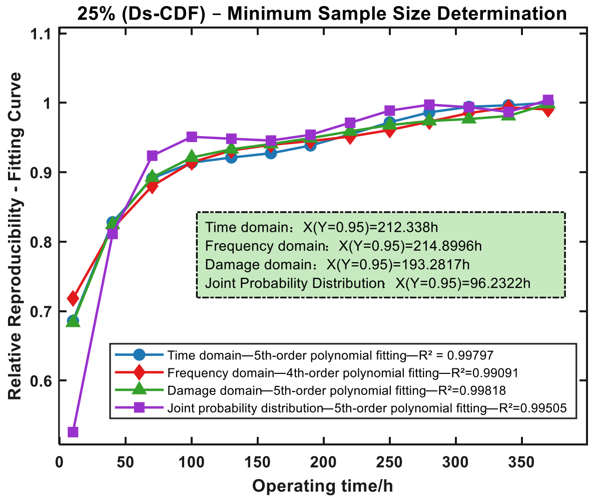


Figure 7. Determination of the minimum cycle period under 25% DR-CDF.

According to Fig. 7 and Equation (15), the minimum cycle period for this user was determined to be 214 h. The mileage statistics of all operating-condition fragments at this scale were then calculated, yielding an equivalent mileage mean of 6,928 km according to Equation (16), with a corresponding mileage range of 6,810–7,026 km.

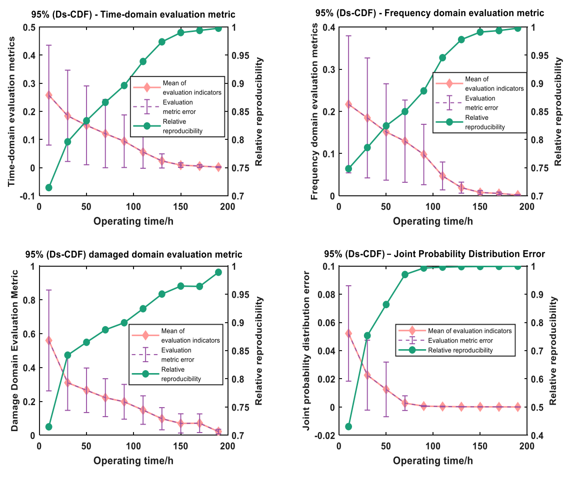


Figure 8. Trend Chart of Multi-dimensional Evaluation Indicators under 95% DR-CDF.

By contrast, for the representative user corresponding to a cumulative damage rate of 95%, the total recorded operating time was 210 h and the accumulated mileage was 9,124 km. Window sizes from 10 h to 200 h were applied sequentially, and operating-condition fragments were generated using a sliding step of 1 h. The mean values, errors, and recurrence levels of the multidimensional evaluation metrics at different time scales are shown in Fig. 8.

As shown in Fig. 9, the recurrence-fitting equations for all four dimensions achieve coefficients of determination greater than 0.99. When all dimensional evaluation metrics exceed the 95% recurrence threshold, the minimum cycle period for this user is identified as 129 h. Statistical analysis of the fragment mileage at this time scale yields an equivalent mileage mean of 5,525 km and a mileage range of 5,344–5,691 km.

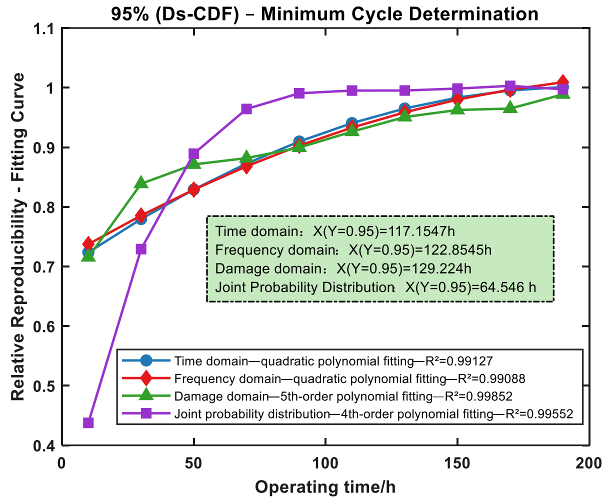


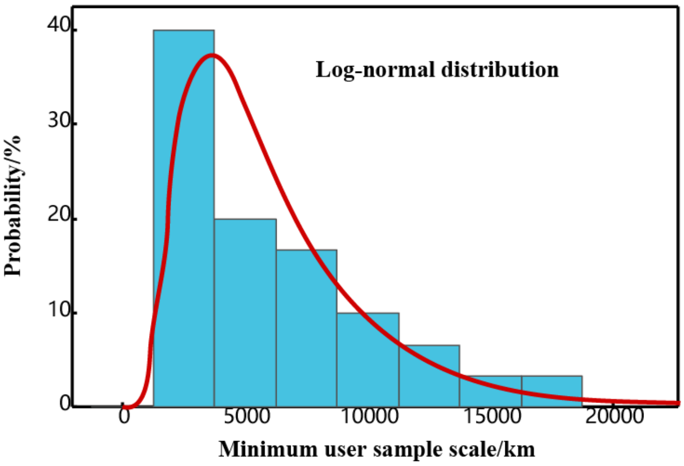
Figure 9. Determination of the minimum cycle period under 95% DR-CDF.

The minimum cycle period results for all four representative DR-CDF groups are summarized in Table 1. Although the convergence rates differ among user groups, all cases exhibit a consistent tendency toward stabilization as the observation-window length increases. These results indicate that the identified minimum cycle periods provide compact yet sufficiently representative observation horizons for subsequent sampling and baseline operating-condition construction.

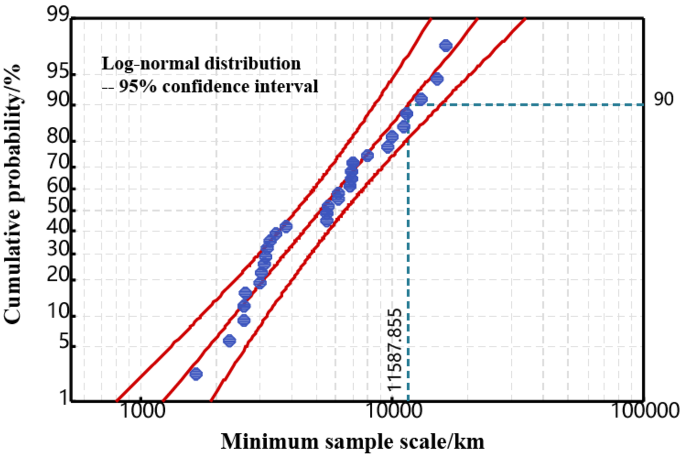
To further characterize inter-user variability, the minimum cycle period was extracted for all users, and a log-normal distribution model was established to characterize the overall distribution. The corresponding probability density and cumulative distribution curves are shown in Fig. 10.

Table 1. Summary of minimum cycle period results for the four representative DR-CDF groups.

DR-CDF group	Total operating time (h)	Accumulated mileage (km)	Minimum cycle period (h)	Equivalent mileage mean (km)	Mileage range (km)
25%	378	12124	214	6928	6810–7026
50%	185	6980	79	3026	2844–3189
75%	387	16090	164	6796	6648–6940
95%	210	9124	129	5525	5344–5691



(a)



(b)

Figure 10. Distribution model of minimum cycle periods for different users.

As illustrated in Fig. 10, 90% of users exhibit a minimum cycle period shorter than 11,587 km. This result provides a useful engineering reference for determining the minimum cycle period in practical applications and for defining the minimum data-acquisition mileage during load collection.

Overall, the results of the four representative DR-CDF groups show a consistent convergence pattern of the multidimensional recurrence metrics with increasing observation-window length. Although the convergence rates differ among user groups, all cases exhibit a clear tendency toward stabilization. The identified minimum cycle periods therefore provide a compact yet sufficiently representative observation horizon for subsequent sampling and baseline operating-condition construction.

4.3. Establishment of baseline operating conditions for electric drive systems

The number of sample segments was determined according to Equation (17) to construct the baseline operating conditions for the electric drive system. The KFCM clustering algorithm was then applied to cluster all operating-condition fragments, using as input a feature matrix composed of the damage-rate characteristics of each electric-drive component. During clustering, the fuzzy weighting exponent was set to 2, and the iteration termination threshold was set to 1×10−4. The Davies–Bouldin (DB) index computed according to Equation (18) is shown in Fig. 11, from which the optimal clustering result was obtained with six clusters.

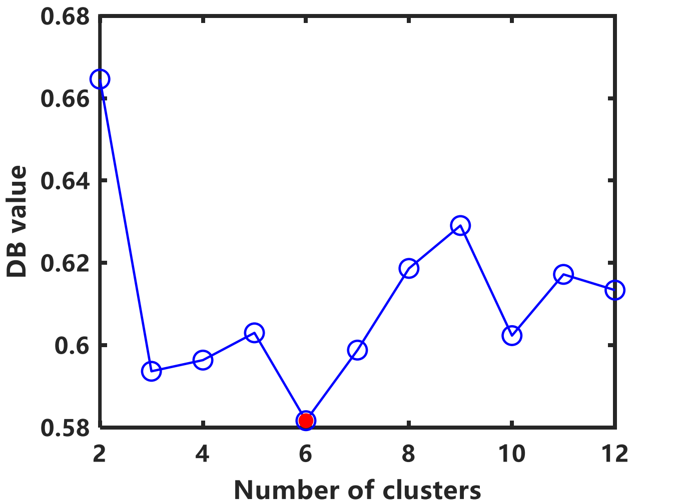


Figure 11. Number of clusters evaluation metric.

The cluster cloud diagram shown in Fig. 12 is used to illustrate the distribution of fragments across different categories after clustering. Owing to the high-dimensional feature mapping of the KFCM method, fragments within the same category are tightly concentrated, while clear separations are observed between categories. The mileage proportions for the six categories are 48.3%, 5.7%, 12.6%, 0.5%, 2.1% and

30.9%, respectively. These proportions are recommended as a reference for users when selecting category-specific operating-condition segments during the construction of baseline operating conditions.

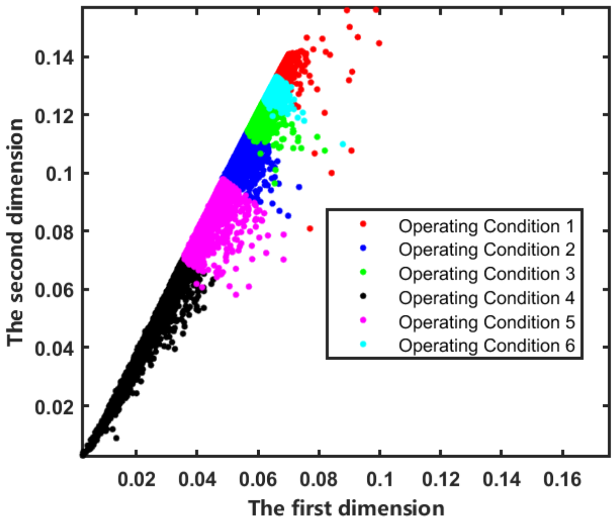


Figure 12. Two-dimensional cluster cloud diagram.

Using the representative users corresponding to cumulative damage rates of 25%, 50%, 75%, and 95%, and following the constraints defined in Equation (21), the thresholds δ_1= 0.01 and δ_2= 5% were adopted. A total of 10,000 sampling trials were carried out, with random sampling conducted independently within each user category. Consistency-testing methods were then employed to evaluate the selected operating-condition fragments by assessing the consistency between the sampled and population datasets in terms of distribution, mean, and variance.

The explicit consistency-test results are summarized in Table 2. For all four representative DR-CDF user groups, the p-values of the K-S test, Z-test, and F-test are all greater than 0.05. Therefore, the null hypotheses are not rejected for any of the three tests, indicating that no statistically significant differences exist between the sampled datasets and the corresponding overall datasets in terms of distribution, mean, or variance.

Table 2. Explicit consistency-test results for the four representative DR-CDF user groups.

DR-CDF group	K-S statistic	K-S p-value	Decision	Z statistic	Z p-value	Decision	F statistic	F p-value	Decision
25%	0.0187	0.8929	Not rejected	0.3457	0.7296	Not rejected	1.0296	0.6822	Not rejected
50%	0.0197	0.9197	Not rejected	0.8723	0.3824	Not rejected	0.9891	0.7534	Not rejected
75%	0.0368	0.7891	Not rejected	1.2309	0.2185	Not rejected	1.0543	0.6262	Not rejected
95%	0.0446	0.6993	Not rejected	1.5735	0.1150	Not rejected	0.9671	0.8164	Not rejected

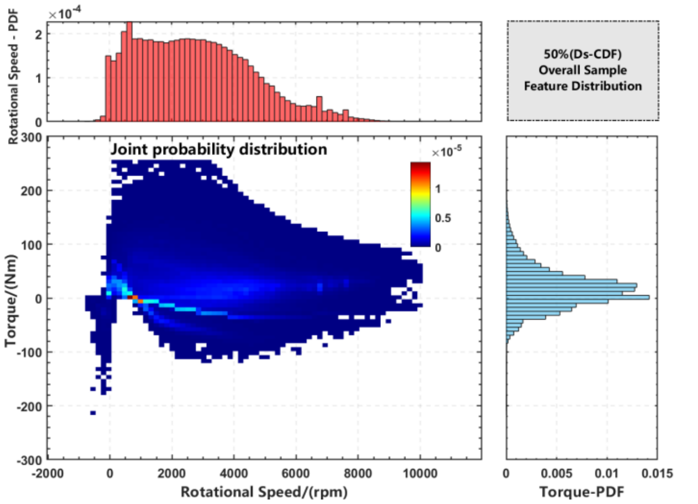


Figure 13. Distribution of overall user sample characteristics at 50% DR-CDF.

Because the distribution plots for different DR-CDF groups exhibit highly similar visual patterns, only the median case (50% DR-CDF) and the high-damage case (95% DR-CDF) are presented as representative examples in Figs. 13–16. Figs. 13 and 14 show the overall and sampled speed–torque joint probability distributions, together with the corresponding speed and torque probability density distributions, for the 50% DR-CDF case.

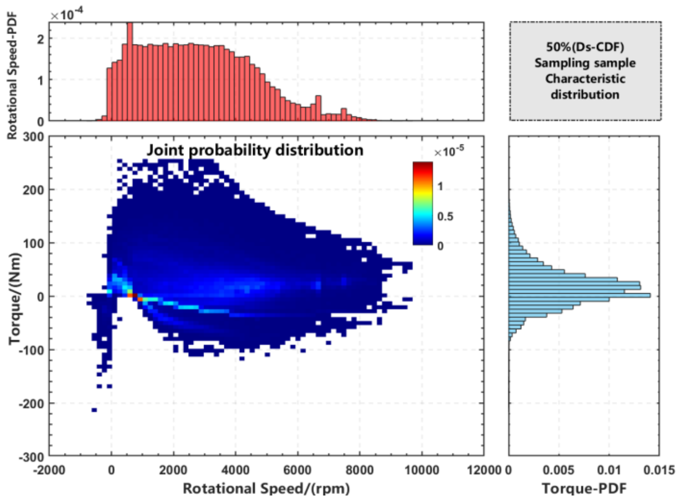


Figure 14. Distribution of user sample characteristics at 50% DR-CDF.

Similarly, Figs. 15 and 16 present the corresponding results for the 95% DR-CDF case. In both representative cases, the sampled datasets reproduce the main characteristics of the overall operating-condition distributions with good agreement.

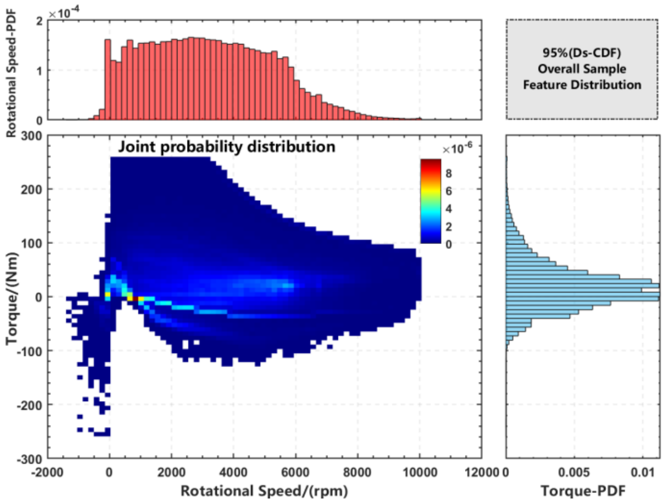


Figure 15. Distribution of overall user sample characteristics at 95% DR-CDF.

To provide a more direct comparison, the statistical characteristics of the overall datasets and the sampled datasets for the four representative DR-CDF groups are summarized in Table 3. As shown in Table 3, the sampled datasets closely match the overall datasets across all four representative groups. The differences in mean damage rate are all within 1%, and the damage-rate variances remain highly consistent between the overall and sampled datasets. In addition, the joint probability distribution errors are all lower than 0.001, indicating that the sampled datasets preserve the principal load-distribution characteristics of the original data with high fidelity.

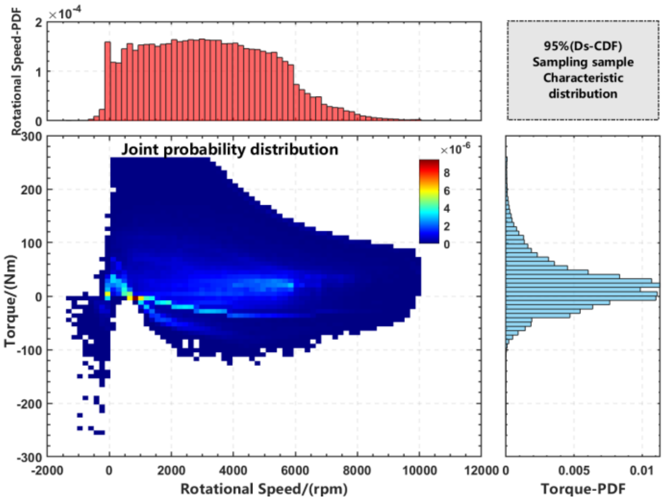


Figure 16. Distribution of user sample characteristics at 95% DR-CDF.

Table 3. Comparison between the overall dataset and the sampled dataset for the four representative DR-CDF groups.

DR-CDF group	Number of fragments (overall / sampled)	Time (h) (overall / sampled)	Mileage (km) (overall / sampled)	Mean damage rate (overall / sampled)	Damage rate variance (overall / sampled)	Joint probability distribution error
25%	11637 / 6649	378 / 216	12124 / 6958	0.0217 / 0.0216	2.63E-04 / 2.82E-04	7.54E-06
50%	4217 / 1813	185 / 79	6980 / 2999	0.0243 / 0.0242	1.20E-03 / 1.12E-03	4.42E-04
75%	10334 / 4366	387 / 163	16090 / 6761	0.0192 / 0.0191	7.49E-04 / 7.16E-04	5.96E-05
95%	6601 / 3900	210 / 126	9124 / 5499	0.0268 / 0.0267	1.71E-03 / 1.72E-03	1.69E-05

Overall, the results of this section demonstrate that the baseline operating conditions constructed from the identified minimum cycle periods can preserve the principal statistical properties of the original large-scale operational data. The clustering analysis and consistency tests confirm that the proposed framework can effectively retain representative operating characteristics while substantially reducing data volume. This provides a feasible basis for reliability-oriented operating-condition reconstruction and subsequent engineering application of electric drive systems.

5. Conclusion

This study proposes a methodology for establishing baseline operating conditions of electric drive systems to support reliability assessment. The main findings are summarised as follows:

(1) A variable sliding-window method was developed to segment multi-scale operating conditions of electric drive systems. This method enables the identification of trend behaviours in multidimensional evaluation metrics across different time scales by analysing variance coefficients and joint probability distribution errors in the time, frequency and damage domains. The proposed relative recurrence index effectively quantifies the minimum cycle period for each user. Statistical analysis of the cumulative distribution shows that 90% of users exhibit minimum cycle periods shorter than 11,587 km.

(2) Considering the physical characteristics of multi-component degradation in electric drive systems, the KFCM clustering method was applied to group operating conditions with similar degradation rates. The Davies–Bouldin (DB) index identifies six as the optimal number

of clusters. After high-dimensional feature mapping, operating-condition segments within the same cluster become densely concentrated, while clear separations between clusters are achieved.

(3) A methodology was developed to construct baseline operating conditions for electric drive systems based on minimum cycle period identification and category-based random sampling. By minimizing both the joint speed-torque distribution error and the deviation in damage-rate characteristics, the proposed method preserves the principal statistical characteristics of the original dataset. For the representative users considered in this study, the differences in damage-rate means between sampled and overall datasets remain within 1%, while the joint probability distribution errors remain below 0.001. In addition, the K-S, Z-test, and F-test results indicate that no statistically significant differences are observed between the sampled datasets and the corresponding overall datasets.

In summary, the proposed framework provides a practical approach for extracting compact, representative, and reliability-relevant baseline operating conditions from large-scale user operational data. However, the effects of threshold selection, fitting-model choice, rare-event representation, and cross-platform generalization still require further study. Future work will therefore focus on strengthening the statistical foundation and extending the method to broader engineering scenarios.

References

1. Wang Z, Zhao L, Zhang D, Yan C. Study on dynamic prediction method for degradation state of electric drive system based on deep learning and uncertainty quantification. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* 2023; 237: 09544070231205063. doi:10.1177/09544070231205063.
2. Moussa A, Aoulmi Z. Improving electric vehicle maintenance by advanced prediction of failure modes using machine learning classifications. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2025; 27(3): 201372. doi:10.17531/ein/201372.
3. Zhao L, Wang Z, Feng J, Zheng S,Ning X. Construction method for reliability test driving cycle of electric vehicle drive system based on users' big data. *Journal of Mechanical Engineering* 2021; 57(14): 129-140. doi: 10.3901/JME.2021.14.129.
4. Wang Z, Zhao L, Kong Z, Yu J, Yan C. Development of accelerated reliability test cycle for electric drive system based on vehicle operating data. *Engineering Failure Analysis* 2022; 141: 106696. doi:10.1016/j.engfailanal.2022.106696.
5. Qian Q, Zhang J, Luo J, Qin Y. Integrated-dispersion manifold distance: a new distribution discrepancy metric for machine fault transfer diagnosis under time-varying conditions. *IEEE Transactions on Cybernetics* 2025. doi:10.1109/TCYB.2025.3630879.
6. Zhou J, Qin Y, Luo J, Zhu T. Remaining useful life prediction by distribution contact ratio health indicator and consolidated memory GRU. *IEEE Transactions on Industrial Informatics* 2023; 19(7): 8472-8483. doi:10.1109/TII.2022.3218665.
7. Zhang G, Wang Y, Li X, Qin Y, Tang B. Health indicator based on signal probability distribution measures for machinery condition monitoring. *Mechanical Systems and Signal Processing* 2023; 198: 110460. doi:10.1016/j.ymssp.2023.110460.
8. Guo J,Fu H. Reliability assessment method based on small sample accelerated life test data. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2025; 27(1): 192170. doi:10.17531/ein/192170.
9. Deng S, Zhao L, Wang Z, Zhou C,Zheng S. A method of rapid construction of reliability target load for electric drive systems. *Chinese Journal of Automotive Engineering* 2021; 11(5): 320-329. DOI:10.3969/j.issn.2095-1469.2021.05.02.
10. Pielecha I, Pielecha J. Simulation analysis of electric vehicles energy consumption in driving tests. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2020; 22(1): 130-137. doi:10.17531/ein.2020.1.15.
11. Puisa R, Montewka J, Krata P. A framework estimating the minimum sample size and margin of error for maritime quantitative risk analysis. *Reliability Engineering & System Safety* 2023; 235: 109221. doi:10.1016/j.ress.2023.109221.
12. Woo S, O'Neal DL, Pecht M. Improving the lifetime of mechanical systems during transit established on quantum/transport life-stress prototype and sample size. *Mechanical Systems and Signal Processing* 2023; 193: 110222. doi:10.1016/j.ymssp.2023.110222.
13. Gope P. Determination of sample size for estimation of fatigue life by using Weibull or log-normal distribution. *International Journal of Fatigue* 1999; 21: 745-752. doi:10.1016/S0142-1123(99)00048-1.
14. Xiong JJ, Shenoi RA. A reliability-based data treatment system for actual load history. *Fatigue & Fracture of Engineering Materials &*

Structures 2005; 28: 875-889. doi:10.1111/j.1460-2695.2005.00924.x.
15. Zheng G, Wang Q, Cai C. Criterion to determine the minimum sample size for load spectrum measurement and statistical extrapolation. *Measurement* 2021; 178: 109387. doi:10.1016/j.measurement.2021.109387.
16. Zheng G, Wan Y, Wei H, Zhao S. High-confidence minimum acquisition mileage of durability load spectrum of off-road special vehicle. *Acta Armamentarii* 2024; 45: 3017-3028. doi: 10. 12382 / bgxb. 2023. 0817.
17. Zheng L, Sayed T, Mannering F. Modeling traffic conflicts for use in road safety analysis: a review of analytic methods and future directions. *Analytic Methods in Accident Research* 2021; 29: 100142. doi:10.1016/j.amar.2020.100142.
18. Wang J, Wang N, Wang Z, Zhang Y, Liu L. Determination of the minimum sample size for the transmission load of a wheel loader based on multi-criteria decision-making technology. *Journal of Terramechanics* 2012; 49: 147-160. doi:10.1016/j.jterra.2012.02.001.
19. He C, Zhou H,Xiang J. Study on the method of road spectrum sample size determination in vehicle road testing. *Automobile Technology* 2003: 14-17.
20. Andráš I, Dolinský P, Michaeli L, Šaliga J. A time domain reconstruction method of randomly sampled frequency sparse signal. *Measurement* 2018; 127: 68-77. doi:10.1016/j.measurement.2018.05.065.
21. Chen D, Qin Y, Wang Y, Zhou J. Health indicator construction by quadratic function-based deep convolutional auto-encoder and its application into bearing RUL prediction. *ISA Transactions* 2021; 114: 44-56. doi:10.1016/j.isatra.2020.12.052.
22. Ansari S, Ayob A, Lipu MSH, Hussain A, Abdolrasol MGM, Zainuri MAAM,Saad MHM. Optimized data-driven approach for remaining useful life prediction of lithium-ion batteries based on sliding window and systematic sampling. *Journal of Energy Storage* 2023; 73: 109198. doi:10.1016/j.est.2023.109198.
23. Karolczuk A, Kurek M. Fatigue life uncertainty prediction using the Monte Carlo and Latin hypercube sampling techniques under uniaxial and multiaxial cyclic loading. *International Journal of Fatigue* 2022; 160: 106867. doi:10.1016/j.ijfatigue.2022.106867.
24. Zhao L, Wang Z, Feng J, Zheng S. Research on reliability condition characteristics of electric drive system based on user big data. *Automotive Engineering* 2020; 42(10): 1386-1396. doi:10.19562/j.chinasae.qcgc.2020.10.013.
25. Dukalski P, Będkowski B, Parczewski K, Wnęk H, Urbaś A, Augustynek K. Dynamics of the vehicle rear suspension system with electric motors mounted in wheels. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2019; 21(1): 125-136. doi:10.17531/ein.2019.1.14.
26. Hou N, Ding N, Qu S, Guo W, Liu L, Xu N, Tian L,Xu H,Chen X,Wu CM. Failure modes, mechanisms and causes of shafts in mechanical equipment. *Engineering Failure Analysis* 2022; 136: 106216. doi:10.1016/j.engfailanal.2022.106216.
27. Liang X, Zuo MJ, Feng Z. Dynamic modeling of gearbox faults: a review. *Mechanical Systems and Signal Processing* 2018; 98: 852-876. doi:10.1016/j.ymssp.2017.05.024.
28. Poplawski JV, Peters SM, Zaretsky EV. Effect of roller profile on cylindrical roller bearing life prediction—Part I: comparison of bearing life theories. *Tribology Transactions* 2001; 44: 339-350. doi:10.1080/10402000108982466.
29. Chai F, Li Y, Liang P, Pei Y. Calculation of the maximum mechanical stress on the rotor of interior permanent-magnet synchronous motors. *IEEE Transactions on Industrial Electronics* 2016; 63: 3420-3432. doi:10.1109/TIE.2016.2524410.
30. Brancato EL. Estimation of lifetime expectancies of motors. *IEEE Electrical Insulation Magazine* 1992; 8: 5-13. doi:10.1109/57.139066.