Health Assessment of Underground Power Cables: A Data-driven Approach based on One-Sample Maximum Mean Discrepancy

Qiaoqiao Li, *Member, IEEE*, Yan Xu, *Senior Member, IEEE*, Sungin Cho, *Member, IEEE*, Changyou Suo, *Member, IEEE*, and Terence Tik Lam Yip

***Abstract*— Effective health assessment and management of underground power cable system is essential for ensuring the cost-efficient operation of power grids. Existing practices usually monitor condition of the overall health of the power cable circuits but overlook its intricate structural complexity. This letter proposes a data-driven health assessment method considering both failure-related structural characteristics of each cable circuit and condition measurement data. Firstly, distinguishing failure indicators are identified by integrating in-depth component analysis with circuit condition measurements. Based on these features, a novel integrated health index is proposed, named as One-sample Maximum Mean Discrepancy (O-MMD), which evaluates the health condition of a circuit by quantifying the disparity between un-faulted assets and typical faulted groups in a high-dimensional feature space. This approach is able to assess whether new or modified circuits might exhibit similar fault characteristics as previously observed cases. Based on the O-MMD index, the probability of failure can be estimated for each cable circuit, and the maintenance plan can be prioritized. The proposed method is demonstrated on both seen and unseen real-world underground cable system data in Singapore.**

***Index Terms*— Underground power cable, Health assessment, Health index, Cable segment, Data-driven.**

I. Introduction

RELIABLE operation of power grid is crucial for the society and the economy. The backbone of an urban power grid is composed of underground power cables. For example, the Singapore power grid has more than eight thousand kilometers of 22 kV cables and three thousand kilometers of 6.6 kV cables. In practice, due to factors such as cable insulation aging and degradation, manufacturing defects, and installation damage, unexpected failures may occur in underground cables [1], leading to power outages and substantial repair costs.

In this respect, identifying potentially problematic cable assets is essential for predictive maintenance of cables. Existing methods can be primarily divided into statistical approaches and condition monitoring techniques. Statistical models analyze the survival or failure rates of individual cable components by leveraging life data and historical failure records. For example, Weibull analysis has been applied to develop cable aging models [2], while the Proportional Hazards Model (PHM) has been used to assess the influence of factors such as manufacturing and installation practices on cable segments [3]. However, these statistical approaches often rely heavily on detailed life-cycle data, which may not always be available. Moreover, they are unable to provide real-time insights into the current health status of cables, limiting their effectiveness for proactive maintenance planning.

Condition monitoring techniques include Insulation Resistance (IR), Partial Discharge (PD), VLF Tangent Delta, etc. In [4], PD characteristics under various typical defects were studied, revealing that most PD sites in highly service-aged cables are associated with extensive water trees. Similarly, [5] performed Tan δ experiments on field-aged and non-aged MV cable samples, providing guidance for operation and criteria establishment of 0.1 Hz Tan δ testing. Condition monitoring-based methods are valuable for identifying critical degradation patterns and assessing the actual status of cable assets. However, in practical applications, obtaining highly accurate condition data, particularly at the component level, is often impractical due to the significant resources required for locating, excavation, and reinstallation.

In realistic environment, regular condition tests, such as IR tests, are periodically conducted at substation terminations. Such tests provide measurements that reflect only the aggregated condition of all components within a circuit, rather than the health of individual segments or joints. As a result, maintenance decisions based solely on circuit-level measurements often overlook the structural complexity and variability of real-world cable circuits, reducing the effectiveness of predictive maintenance strategies.

This letter proposes a non-parametric data-driven approach for non-intrusive cable health assessment, combining the merits of statistical analysis and condition-based methods. By leveraging historical operational data in conjunction with routine IR readings from incidents and assets, the proposed method effectively captures the unique structural failure characteristics and health condition of each circuit. The proposed O-MMD index distinguishes the high-risk circuits by measuring disparity with typical faulted profiles. Validated using real-world data from the Singapore power grid, the method demonstrates high level of effectiveness. This innovative approach enables precise identification and prioritization of cables, paving the way for more effective and efficient predictive maintenance planning.

II. Proposed Methodology

A. Overall Framework

The overall idea of the proposed method is illustrated in Fig. 1. The process starts with data pre-processing on the raw database, where component-level information for each segment and joint are recorded individually. The structured features of cable circuits are firstly constructed from the raw data. Based on the ability to distinguish faulted circuits from un-faulted ones, distinctive failure indicators are selected from these features for further analysis. These indicators are then merged with the IR condition data to form a new integrated database.

This research is supported by SP Group – NTU Joint Lab, and the Energy Market Authority (EMA), Singapore, under its Energy Programme (EMA-EP010-SNJL-001). Q. Li, Y. Xu and C. Suo are with Nanyang Technological University, Singapore. S. Cho was with SP Group, Singapore and now with University of Cambridge, UK. T. Yip is with SP Group, Singapore.

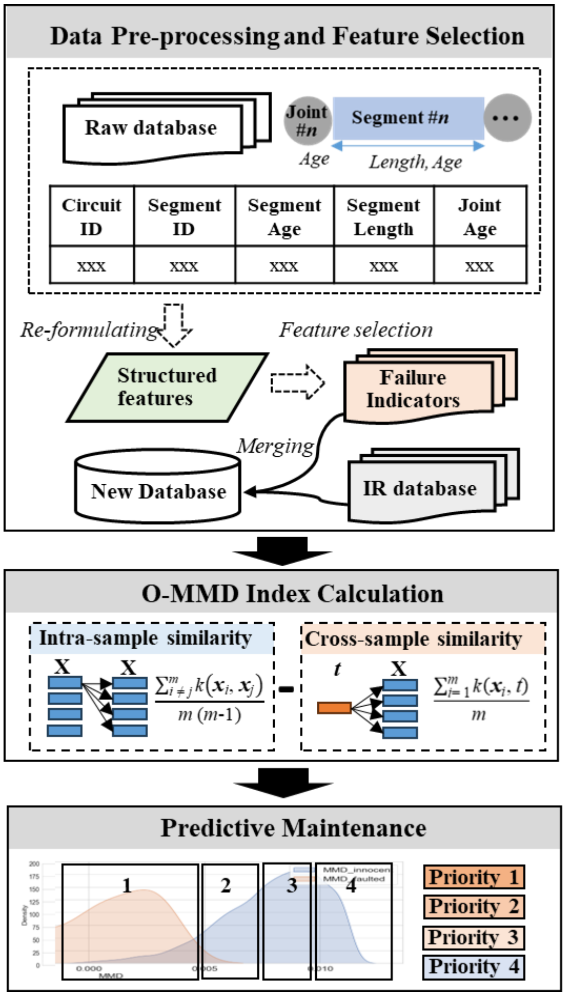


Fig. 1. Process of the proposed method for cable circuit health assessment

Given this combined dataset, an integrated health index is formulated using One-sample Maximum Mean Discrepancy (O-MMD) to quantify the distance of un-faulted circuits from a typical faulted group. Finally, maintenance priority can be made based on the estimated condition probability to the faulted and un-faulted circuit groups.

B. Feature analysis

The statistical analysis is conducted on faulted and un-faulted cable groups to identify specific patterns or trends associated with cable faults. The raw data in this study is collected from the sub-transmission and distribution underground cable systems in Singapore. Singapore's power grid includes an extensive network of underground cables, consisting of tens of thousands of individual segments and several thousand circuits. The datasets encompass a comprehensive asset information including basic operation data of *commission date* and *cable length of segments*, the *IR reading data*, and an incident database with *the historical failures records of circuits* during a 20-year period between 2002 and 2022.

In Table I, structural cable circuit features constructed from the raw database are listed with the condition measurements data. Specifically, the average segment age is calculated as:

TABLE I
COMPARATIVE ANALYSIS OF FAULTED AND UN-FAULTED GROUPS

Key Indicators of cable circuits (Unified value)	Un-faulted Circuits (M1)	Faulted Circuits (M2)	Difference % $(\frac{\text{M2-M1}}{\text{M1}})$
Structural Features			
No of Segment	0.14	0.24	**44.31%**
Length of Circuit	0.11	0.21	**47.57%**
Oldest Segment Age	0.64	0.71	10.12%
Average Segment Length	0.15	0.16	5.80%
Average Segment Age	0.04	0.02	**-172.18%**
Oldest joint age	0.50	0.58	14.38%
No of Newly Installed Segments	0.08	0.19	**55.36%**
Length of Newly Installed Segments	0.06	0.16	**61.20%**
Ratio of Newly Installed Segments	0.45	0.67	**32.82%**
No of Newly Installed joints	0.10	0.20	**51.45%**
Ratio of Newly Installed joints	0.74	0.81	8.24%
Conditional Features			
Min of 3-phase IR Reading	0.45	0.37	**-25.40%**
Max-Min Ratio of IR Reading	0.06	0.05	**-23.95%**
IR Reading- Blue Phase	0.20	0.15	**-29.84%**

* Note that all data presented in this table are not actual but normalized values

$A_{weighted} = \sum_{i=1}^{n} \left(\frac{L_i}{L_{total}} \times A_i \right)$. L_i and A_i is the length and age of segment i, and L_{total} is length of the circuit with n segments in total. Note that the oldest segment or joint of each circuit serves as the benchmark for identifying newer additions. Any segment or joint with an installation age younger than this benchmark is classified as a new installment. Such definition distinguishes between original infrastructure and newly added components. In general, these indicators are designed to provide structural insights into the circuit, such as historical replacements and the aging status.

To assess the distinguishing ability of these features between un-faulted and faulted circuits, the mean values of each feature across the two groups and their percentage difference are presented in Table I. Finally, features with absolute percentage difference that larger than 15% are identified as a distinctive indicator for the construction of the health index.

C. O-MMD-based health index

The above preliminary analysis highlights the importance of considering not just the measured condition data but also the integration of newer segments when assessing the health status of cable circuit. Using the identified distinctive failure indicators, this paper proposes a novel O-MMD-based health index. Unlike existing work which usually calculate the health index by simply adding or using a weighted sum of the features [6], the proposed index aims to capture the unique characteristics of each circuit and quantify its distance with failure profiles in the high-dimensional feature space.

Firstly, a database of circuits which historically experienced cable or joint failures is built as the faulted circuit profiles. Each faulted circuit profile is described by the failure indicators selected in *Section II.B.* Then, an O-MMD index is proposed to measure the disparity between un-faulted circuits to the faulted group. As a variant of the Maximum Mean Discrepancy (MMD) metric commonly used in transfer learning theory [7], the O-MMD index quantifies the difference between a sample (i.e., a

This article has been accepted for publication in IEEE Transactions on Power Delivery. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2025.3579927

TABLE II
MAINTENANCE RESULTS OF PRIORITY GROUPS

Key Indicators	Priority 1 0-20%	Priority 2 20-40%	Priority 3 40-60%	Priority 4 60-100%
Structural Features on different priority groups (on unseen cases)				
S1: No of Segment	1.00	0.90	0.71	0.40
S2: Length of Circuit	1.00	0.84	0.68	0.42
S3: Average Segment Age	1.00	1.27	2.05	4.04
S4: No of New Segments	1.00	0.72	0.49	0.16
S5: Length of New Segments	1.00	0.64	0.41	0.14
S6: Ratio of New Segments	1.00	0.77	0.64	0.25
S7: No of Newly joints	1.00	0.86	0.62	0.25
Conditional Features on different priority groups (on unseen cases)				
C1: Min of 3-phase IR Reading	1.00	1.23	1.43	2.38
C2: Ratio of IR Reading	1.00	0.98	0.71	1.76
C3: IR Reading- Blue Phase	1.00	1.28	1.35	2.22

new circuit) and a reference group (i.e., the faulted circuit) by evaluating the mean discrepancy in a high-dimensional feature space. This is achieved through kernel technique, which effectively capture non-linear relationships between high-dimensional vectors, ensuring a more precise and accurate assessment of whether new or modified circuits might exhibit similar fault characteristics as previously observed cases.

Mathematically, the calculation of O-MMD involves two main components: 1) **The intra-sample similarity**, computed as the average of the kernel evaluations between all pairs of points within the target group (the faulted group) itself. 2) **The cross-sample similarity**, computed as the average of the kernel evaluations between each point in the faulted group and a test sample point t in the database. Finally, the squared value of O-MMD is obtained by the difference of the two similarity measurements as follows:

$$\text{O-MMD}^2\,(\boldsymbol{X}, \boldsymbol{t}) = \frac{1}{m\,(m\text{-}1)}\sum_{i \neq j}^{m} k(\boldsymbol{x}_i, \boldsymbol{x}_j) - \frac{1}{m}\sum_{i=1}^{m} k(\boldsymbol{x}_i, t) \quad (2)$$

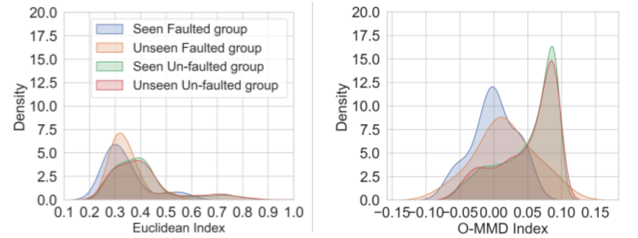
where $k\,(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(\frac{-\|x_i\text{-}x_j\|^2}{2\sigma^2}\right)$ is the Gaussian kernel function. $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ represent the feature vectors corresponding to the i-th and j-th samples in the faulted group, respectively. $\boldsymbol{X}$ is the group of faulted cases, described by the identified indicators. The O-MMD value can leverage a quantitative measurement of how compatible a new circuit data is with the existing distribution of faulted group. A higher O-MMD value indicates greater dissimilarity, suggesting that the new data deviates from typical faulted cases, potentially indicating less risk or different fault characteristics. Other widely used distance measures like Euclidean distances typically compare individual data points or vectors [8]. In contrast, the O-MMD index assesses the entire distribution of datasets, providing a more holistic comparison.

D. Preventive maintenance and dynamic updating

For preventive maintenance, four priority groups are formulated using the 0.2, 0.4, and 0.6 quantiles of the O-MMD index of the prioritized circuits. Higher-priority groups are subjected to more in-depth condition monitoring techniques, such as VLF tangent delta testing and PD analysis. These techniques provide more precise and effective detection of problematic cable and joint but are more costly and typically conducted less frequently than routine IR recordings. When available, these advanced condition monitoring data serve as stronger evidence for detecting potential early fault indications.

Further, three updating mechanisms are developed to renew the O-MMD index and the corresponding priority levels, including event-triggered, periodic, and preventive updates. Firstly, event-triggered structural and operational changes, such as circuit reconfigurations, segment replacements, or joint repairs, will be reflected in the O-MMD index. When a new fault occurs on a previously un-faulted circuit, it is added to the faulted dataset, triggering recalculation of the index. Secondly, the O-MMD index will be periodically renewed by the routine condition monitoring data, such as IR test results. The periodic recalibration ensures the index captures the evolving health status of circuits. Thirdly, early faults detected by the proactive maintenance strategy in this method will prompt preventive repairs. These cases are then added to the faulted dataset.

Overall, the application of the O-MMD index facilitates more targeted monitoring and preventive strategies. By identifying circuits whose characteristics deviate significantly from known fault patterns, the method allows operators to preventively address areas of concern. Dynamic and periodic updates enable the method to track changes in circuit health over time, ensuring that maintenance planning remains timely and effective.



(a) M1: Euclidean distance (b) M2: Proposed O-MMD index

Fig. 2. Testing results of the proposed method and the baseline approach

TABLE III
CONFUSION METRICS OF PROPOSED METHOD

Ratio (Count)	Priority 1 0-20%	Priority 2 20-40%	Priority 3 40-60%	Priority 4 60-100%
Un-faulted	20.64% (187)	19.75% (179)	20.64% (187)	38.96% (353)
Faulted	**57.14% (8)**	**28.57% (5)**	**7.14% (1)**	**7.14% (1)**

TABLE IV
COMPARATIVE ANALYSIS OF PROPOSED METHOD

Methods	Incident coverage rate			
	0-0.2 quantile	0-0.4 quantile	0-0.6 quantile	0-1.0 quantile
M1 (Baseline)	35.71%	71.42%	85.71%	100.00%
M2 (Proposed)	**57.14%**	**85.71%**	**92.85%**	**100.00%**

III. HEALTH ASSESSMENT RESULTS

Based on the feature analysis results in Section II-B, cable health assessment and prioritization are conducted using the O-MMD index introduced in Section II-C. Firstly, both faulted and un-faulted data are split into seen and unseen sets in a 7:3 ratio, ensuring that the method is tested on data it has not previously encountered. Specifically, the seen faulted data are used to construct the faulted circuit profiles. The seen un-faulted data are used to fine-tune the kernel parameters for the O-MMD index calculation and to establish the criteria for determining priority groups. The unseen datasets are used for evaluating the performance of the proposed method and its impact on maintenance planning.

Using the prioritization criteria in Section II D, Table II presents the statistical characteristics of each prioritized group on unseen data. Firstly, the un-faulted cases closely follow the pre-determined quantile distribution, with a 0.2:0.2:0.2:0.4 split across the four priority groups. Advanced monitoring methods, such as PD or VLF tangent delta testing, are suggested to monitor these high-priority circuits to detect early fault indications. Table II highlights the trends in structural and conditional features across priority groups, with higher-priority circuits showing more segments, longer total circuit lengths, and lower IR readings. Further, Kernel Density Estimation (KDE) is used to visualize the probability density of data points across seen and unseen, faulted and un-faulted groups. Fig. 2 compares the KDE curves of the O-MMD index with those based on Euclidean distance. Notably, Fig. 2(b) shows that the O-MMD index achieves much clearer separation between the groups. The un-faulted group (green and red) generally has higher O-MMD values, while the faulted group (blue and yellow) clusters around lower values, demonstrating the superior discriminative capability of the proposed method.

Table III showing the proportion and count of faulted and un-faulted circuits assigned to each maintenance priority level. The unseen faulted cases are primarily concentrated in Priority 1 and 2, demonstrating the effectiveness of the proposed method in prioritizing circuits with higher failure risks. Then, a new evaluation metric called the ***incident coverage rate*** **(ICR)** is proposed. The **ICR** metric quantifies the percentage of faults being covered within the total unseen fault set at each pre-defined monitoring quantile, as presented in Table IV. The results demonstrate that the proposed method significantly outperforms the baseline Euclidean distance-based approach. By monitoring the top 40% of prioritized circuits, over 85% of potential faults can be effectively covered, validating the efficiency and practicality of the O-MMD index for preventive maintenance planning.

The findings suggest that the O-MMD is an effective quantitative measure of how distinct the circuit conditions are relative to typical profiles of faulted cases. Based on the O-MMD index, we can further prioritize those cables that shows significant similarity with faulted case profiles and develop a more targeted and efficient maintenance strategy to address issues before actual failures occurs. In the future, the proposed health index can be integrated with Weibull distribution analysis for cable survival modeling and laboratory-based water tree studies for post-mortem validation. This combination will allow the development of a comprehensive framework to assess the health and reliability of the cable system.

REFERENCES

[1] J. Luo, Y. Qiu, and L. Yang, "Operation fault analysis of CLPE power cable above 10 kV," *High Voltage Eng*., vol. 29, no. 6, pp. 14–16, 2003

[2] Z. Tang et al., "Comparison of the Weibull and the Crow-AMSAA Model in Prediction of Early Cable Joint Failures," *IEEE Trans. Power Delivery*, vol. 30, no. 6, pp. 2410-2418, Dec. 2015

[3] M. H. P. Klerx, J. Morren and H. Slootweg, "Analyzing Parameters That Affect the Reliability of Low-Voltage Cable Grids and Their Applicability in Asset Management," *IEEE Trans. Power Delivery*, vol. 34, no. 4, pp. 1432-1441, Aug. 2019.

[4] M. S. Mashikian, and A. Szarkowski. "Medium voltage cable defects revealed by off-line partial discharge testing at power frequency." *IEEE Electrical Insulation Magazine*, vol.22, no.4: pp: 24-32,2006.

[5] J. C. Hernandez-Mejia, et al. "Characterization of ageing for MV power cables using low frequency tan δ diagnostic measurements." *IEEE Trans. Dielectrics and Electrical Insulation*, vol.16, no.3, pp: 862-870, 2009.

[6] T. Somsak, T. Suwanasri, C. Suwanasri, "Lifetime estimation-based health index and conditional factor for underground cable system", *Energies*, vol.14, no.23, pp: 8113, 2021.

[7] F. Zhuang et al., "A Comprehensive Survey on Transfer Learning," *in Proceedings of the IEEE*, vol. 109, no. 1, pp. 43-76, Jan. 2021.

[8] C. Shen, J. Kim, L. Wang, "Scalable large-margin Mahalanobis distance metric learning," *IEEE Trans. Neural Networks*, vol.21, no.9, pp: 1524-1530, Sept. 2010.