A Dummy-Variable Model for Humidity-Influenced DC Film Capacitors Lifetime Estimation

Jigui Miao, Yang Liu, *Senior Member, IEEE,* Quan Yin, Shuai Zhao, *Member, IEEE,* Haichun Li and Huai Wang, *Senior Member, IEEE*

***Abstract*—Film capacitors are used in power electronic converters due to their higher reliability compared to aluminum electrolytic capacitors. Humidity is one of the critical factors affecting the aging of film capacitors. However, the existing methodologies for lifetime estimation of film capacitors rarely consider humidity. Besides, sometimes it is hard to obtain a large amount of data in the accelerated aging test of the device. To solve these problems, this paper proposes a lifetime estimation methodology that requires less data and testing time. In the proposed methodology, humidity dummy variables are introduced into the degradation model based on the multivariate linear regression (MLR) algorithm. The impact of introduced dummy variables with different structures on the degradation model is studied and the effect of the proposed lifetime estimation method is verified. The results show that the proposed scheme achieves a comparable performance and can be valid in reality.**

***Index Terms*—Dummy variable, film capacitors, relative humidity (RH), capacitance, lifetime estimation.**

I. Introduction

In power electronic converters, capacitors are utilized at DC-link to buffer energy, limit ripple, and balance power [1], [2]. Electrolytic capacitors are widely used at DC-link due to their high energy density and low cost, so have been extensively studied. However, they are susceptible to high temperature, overvoltage, and overcurrent, which leads to a short lifetime and requires regular replacement. Compared to electrolytic capacitors, film capacitors have higher reliability and longer lifetime [3]-[5], which are inevitably used as DC-link in high-frequency circuits and harsh environments. Therefore, film capacitors also need more attention from researchers in academe and industry.

According to statistics, capacitors are the most fragile components in power electronics with a failure rate of 30% [5]-[7]. Normally, related lifetime estimation studies are based on capacitor degradation indicators, the most commonly used are capacitance and equivalent series resistance (ESR) [3], [8], [9]. Lifetime estimation methods can be mainly divided into model-based methods and data-driven methods [10]. Model-based methods typically build a mathematical model for capacitor aging based on the physics-to-failure mechanism. In [11], a lifetime prediction method for metallized film capacitors considering harmonics and hot-spot temperature is proposed based on the Arrhenius model. In [12], particle filter is applied to estimate the remaining useful lifetime (RUL) of aluminum electrolytic capacitors. However, these model-based methods need to build an accurate mathematical model. Besides, their performance is susceptible to environmental conditions. By contrast, data-driven methods mine information from data without building complex mathematical models. Data-driven methods are divided into statistical methods and Artificial Intelligence (AI) methods [13]. In [14] and [15], the capacitor degradation is driven by a non-homogeneous Gamma stochastic process and a random-effect nonlinear Wiener process, respectively. AI methods are increasingly applied to capacitor lifetime estimation because of their powerful automatic ability, such as [13], [16], and [17]. In [13], an ensemble learning method based on chained support vector regression and a one-dimensional convolutional neural network is proposed for prognostics of aluminum electrolytic capacitors. In [16], the RUL of multi-layer ceramic capacitors is estimated using an artificial neural network (ANN). In [17], a bidirectional long short-term memory network is used to estimate the RUL of electrolytic capacitors. Model-based methods and data-driven methods are often combined as hybrid methods to take the advantages of both. [18] integrates bond graph model–based parameter estimation technique with the data–driven ANN technique for RUL estimation of capacitors in a dynamical system. All of these data–driven efforts achieve a good estimation performance.

Humidity is one of the main stress sources causing failures in power plants, second only to temperature and vibration [6]. As for film capacitors, since their aging is extremely sensitive to humidity, it is very necessary to research their health status affected by humidity. Although the above methods perform well without considering humidity, they may cause large errors and become no longer applicable in humidity conditions outside of the testing environment, such as the extremely high humidity condition. Besides, a considerable number of training data may be demanded in AI-based methods, especially the deep learning methods in [13] and [17].

At present, there are a few studies that focus on the effect of humidity on film capacitors. Failure modes and mechanisms of metallized film capacitors are discussed in [19]-[24], which indicates one of the main reasons for capacitance loss is electrochemical corrosion affected by operating voltage, temperature, humidity, and electrode material. Studies also suggest a standard testing condition for rapid capacitor reliability

Manuscript received Month xx, 2xxx; revised Month xx, xxxx; accepted Month x, xxxx. This work was supported in part by the ...Department of xxx under Grant xxxx. (Corresponding author: Yang Liu).

J.-G. Miao, Y. liu, Q. Yin, and H.-C. Li are with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, 430074 China (e-mail: miaojigui@hust.edu.cn; yangliu30@hust.edu.cn; yinquans@163.com; lihaich@163.com).

S. Zhao, H. Wang are with the Department of Energy Technology, Aalborg University, Aalborg DK-9220, Denmark (e-mail: szh@et.aau.dk, hwa@et.aau.dk).

evaluation is the temperature-humidity-bias test, with levels of 85°C and 85% relative humidity (RH) under an applied voltage [19], [21], [25]. In [26], Peck's model is derived based on the Arrhenius and inverse power law. In [23], the aging acceleration factors voltage, temperature, and humidity are considered in Peck's model to provide a good estimation of the capacitor lifetime. In [24], Peck's model is selected to describe the relationship of moisture ingress time with temperature and humidity, so as to determine the moisture ingress time in capacitance loss prediction. In these methods, lifetime estimation relies on the lifetime information under the reference environment. Sometimes it is inevitable to study the failure mechanism or inner structure of capacitors. In [27], degradation testing and failure analysis of DC film capacitors under 85 °C and 85% RH, 70% RH, and 55% RH are conducted. [28] analyzes capacitor lifetime in [27] using Weibull distribution and proposes a humidity-dependent lifetime derating factor based on Peck's model. However, the proposed lifetime model with humidity factor is unable to track the capacitor degradation with time. In [29], a variable-parameter Weibull distribution is proposed to predict capacitor lifetime based on 2.5% capacitance loss with reduced data. In [30], the percentage of capacitance loss of DC film capacitors under 85°C and 85% RH is driven by a nonlinear Gamma stochastic process. This model is also proved to achieve lifetime estimation with reduced degradation data. However, no humidity factor is introduced to the models proposed in [29] and [30], making them not consider the humidity impact. In addition, the trade-off between the amount of data and accuracy is not discussed in most studies, which is of great significance because it is usually difficult to obtain run-to-failure data, particularly for new systems because running systems to failure could be a lengthy as well as costly process [31]. This paper further studies the deficiencies of the aforementioned methods.

The MLR algorithm is a suitable method for small data with various categories. In the MLR model, categorical variables can be transformed into dummy variables to build estimation models properly [32]-[37]. Various methods that can handle dummy variables are described and compared in [32], which concludes that the most effective method is different for datasets and the stratification and dummy variables-based method should be tried at least. Besides, according to [32], although the stratification is flexible which divides the dataset into subsets and builds models for each subset, its estimation accuracy may be low because of the small subset, especially when there are many independent variables. In comparison, models with dummy variables may achieve higher estimation accuracy but require less data. If a dataset has p non-categorical variables and q categorical variables, it would need $5(p+q-1)$ data points for dummy variables while $5pq$ data points for the stratification [32]. Moreover, interaction terms made by multiplying independent variables and dummy variables also need to be introduced to the MLR algorithm if necessary [32]. [33]-[35] also covers some theoretical knowledge about dummy variable models. In [34], the impact of dummy variables on regression coefficients and canonical correlation indices is investigated. [35] discusses various strategies for converting the qualitative data to numerical information and manifests the introduction of dummy variables as a successful approach when the categorical variables are used as input variables or factors. [36] and [37] describe the modeling process in detail. In [36], the significance test is conducted to delete variables that have no significant effect on the dependent variable. Eliminations of autocorrelation and heteroscedasticity are also conducted to make the model more accurate. In [37], the significance test of explanatory variables and the statistical test of the model are carried out to optimize the proposed method.

In this paper, the dummy variable-based MLR methodology is selected to propose a novel lifetime estimation model for DC film capacitors under various humidity conditions. Compared with the previous practices, the main advantages of this method are summarized as follows.

1) The proposed method is data-driven without establishing a complex mathematical model based on the physics-of-failure mechanism. Compared to Peck's model, it can track the capacitor degradation with time. Since this method uses small samples to infer the population distribution, it does not require a large amount of data.
2) The proposed method introduces the humidity accelerating factor into the lifetime estimation model utilizing dummy variables. It mixes all the humidity data and only does one regression, thus increasing the degrees of freedom and improving the estimation accuracy of the regression coefficients.
3) The trade-off between the amount of data and estimation accuracy is discussed and it indicates that the proposed lifetime estimation model can achieve comparable accuracy with reduced data and testing time.
4) The proposed model has wide applicability. Since it is data-driven, the aging data with the same characteristics can also be evaluated with it. The degradation indicator in this study follows an exponential distribution, which is very common in humidity-influenced aging of capacitors and other electronic components.

The rest of this paper is as follows: Section II describes the experimental platform and data collection process. Section III presents the capacitance loss model-based lifetime estimation procedure. In Section IV, the proposed method is evaluated and analyzed based on case studies, followed by the conclusion in Section V.

II. Accelerated Aging Test for DC Film Capacitors Under Humidity Conditions

Metalized polypropylene film (MPPF) capacitor is one of the promising film capacitors for DC-link of converters due to their excellent electrical characteristics, high stability, and long lifetime. When the metalized polypropylene film suffers from high humidity and high temperature, the metal will react with oxygen and moisture at a fast rate to cause electrochemical corrosion. Reference [9] indicates the capacitance is more frequently used as the failure indicator for MPPF capacitors than ESR because the ESR is too small to be precisely measured. Therefore, the capacitance is used to be the failure indicator in this research.

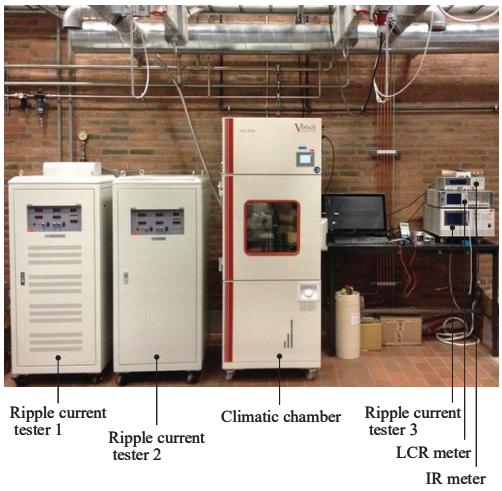


Fig. 1. Experiment setup for the accelerated aging test of capacitors.

The aging test and data collection were conducted at Aalborg University, which is the same as [35]. 1100V DC/40μF MPPF capacitors from EPCOS/TDK B32778G series are investigated, and the experimental setup for the accelerated aging test is shown in Fig. 1. It consists of a 100 L VCL 7010 climatic chamber, an Agilent E4980 LCR meter, a Chroma 11200 IR meter, and three ripple current testers. One of the ripple current testers is capable to supply DC voltage up to 500V and ripple current up to 30A, the other two supply DC voltage up to 2000V and ripple current up to 50A and 100A. In this study, 1100V DC voltage and 21.5Arms/100Hz ripple current is imposed on the capacitors. To distinguish the failure behavior caused by humidity from the effects of other stresses, a constant temperature and various humidity conditions are studied. According to standards IEC60068-2-67 and JESD22-A101-B, the reliability evaluation with levels of 85°C and 85% RH is the main standard for testing the limits the capacitor can withstand under the harsh environment of high-temperature moisture [25], [38]. Therefore, this testing situation must be considered. Because the leakage current of the capacitor may increase when it is stored at a temperature higher than 35°C and a relative humidity higher than 70%, 70% RH is selected as a testing condition. To reflect the capacitance degradation rate at 70% RH and 85% RH, we need to choose a low humidity level that has little effect on capacitor aging. Based on the experimental environment and practical application, we consider 35%∼55% RH as the low humidity [22], [26]. In view of the shortest possible testing time, 55% RH is chosen as a testing group.

We number the three humidity conditions 55% RH, 70% RH, and 85% RH as Group 1, Group 2, and Group 3, respectively. For each group, 10 MPPF capacitors are tested. During the degradation process, the testing proceeds until most or all samples fail. Therefore, different humidity stresses determines unequal testing time. In this study, the testing under 55% RH, 70% RH, and 85% RH last for 3850 hours, 2700 hours, and 2160 hours, respectively. As a result, the dataset consists of a total of 30 time-series data from the aging tests of three groups. The dataset details are summarized in Table I.

TABLE I
THE DATASET FORM THE ACCELERATED AGING EXPERIMENT OF MPPF CAPACITORS

Group number	1	2	3
Relative humidity and temperature condition	85°C and 55% RH	85°C and 70% RH	85°C and 85% RH
Number of capacitor samples	10	10	10
Number of time points of one capacitor	20	10	19

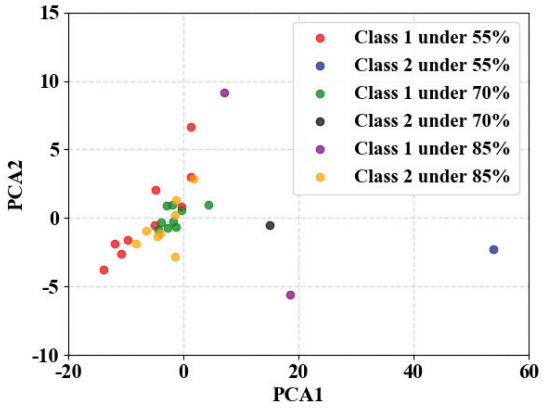


Fig. 2. Clustering results of samples under 55% RH, 70% RH, and 85% RH

It can be noted that the time-series data under 55% RH, 70% RH, and 85% RH has 20, 10, and 19 time points, respectively. Since we mainly study the severity of capacitance loss under 85% RH compared to 55% RH, the testing time interval under 70% RH is set to be relatively large for convenience, so the number of time points is less than the other two groups, but it can still reflect the degradation tendency under 70% RH.

Subsequently, the k-means unsupervised clustering is used for data cleaning to find and eliminate outliers that deviate from most of the data in the normal range. In the clustering, Euclidean distance is selected to measure the difference between samples, and generates 2 clusters. Then, the Principal Component Analysis (PCA) dimensionality reduction is performed on the clustering results for visualization. As shown in Fig. 2, the abscissa PCA1 and the ordinate PCA2 are two principal dimensions of samples after PCA dimensionality reduction. From Fig. 2, we can see that the mean distance between the two classes is very large under 55% RH, thus the sample in Class 2 is judged to be abnormal. In reality, the capacitance loss trend of this sample differs greatly from the samples in Class 1 and its capacitance decreases rapidly to 0, which exhibits unusual behavior. After removing 20 time points of the abnormal capacitor sample in Group 1, the remaining time-series data in three groups are used as training data.

To study the degradation performance under various humidity conditions, the mean normalized capacitance of capacitor samples in each group is learned, as shown in Fig. 3. It can be seen that the capacitance loss can be regarded as an exponential reduction with time. It decreases smoothly at the beginning, until a significant performance occurs, i.e., a

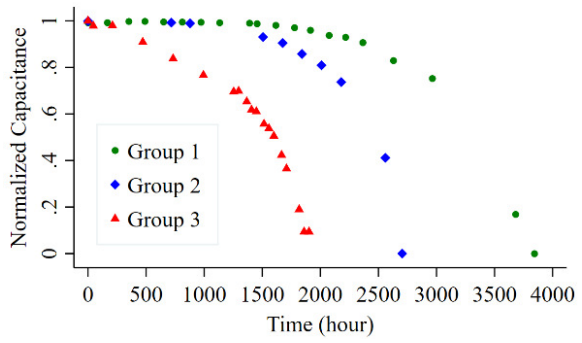


Fig. 3. The mean normalized capacitance of three humidity conditions.

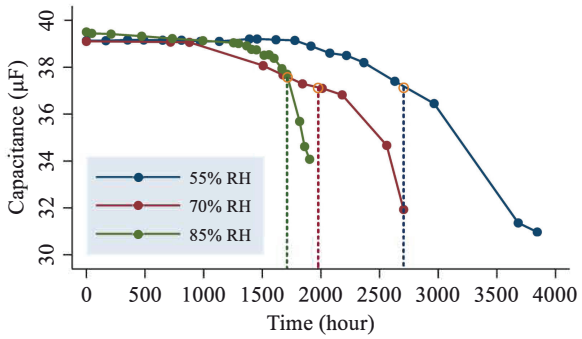


Fig. 4. Time-to-failure of capacitors under different humidity conditions.

data point in the degradation curve, after which the rate of capacitance loss becomes faster and the capacitor reaches the end-of-life quickly. As the humidity increases, the slope of the capacitor deterioration is steeper.

As an illustration, the determined failure time for three randomly selected capacitors subjected to various humidity conditions is shown in Fig. 4. It can be seen that the capacitance under 55% RH keeps in a small range until 1800 hours, then starts to decrease with an increasing speed, the capacitance under 70% RH and 85% RH starts to decrease with an increasing speed after 1250 hours and 900 hours, respectively. The dots circled in orange indicate a 5% capacitance drop, which means the capacitors reach the end-of-life at this time. We can give a rough estimation that the capacitor under 55% RH, 70% RH, and 85% RH runs to failure at 2710, 1980 and 1720 hours, respectively. These are lifetime estimations for a single capacitor under specific conditions. It is necessary to establish a general lifetime estimation method.

III. Lifetime Estimation Methodology Based on Dummy Variables

A dummy variables-based method is proposed to explicitly explore the effect of humidity on MPPF capacitors in this section. The procedure of the proposed methodology is given in Fig. 5. In the first part, a degradation model with the humidity-dependent factor is developed. In the second part, the unknown parameters of the proposed degradation model are estimated by the MLR method. In the third part, the evaluation of the degradation model is conducted from several perspectives. According to the evaluation results, the model is

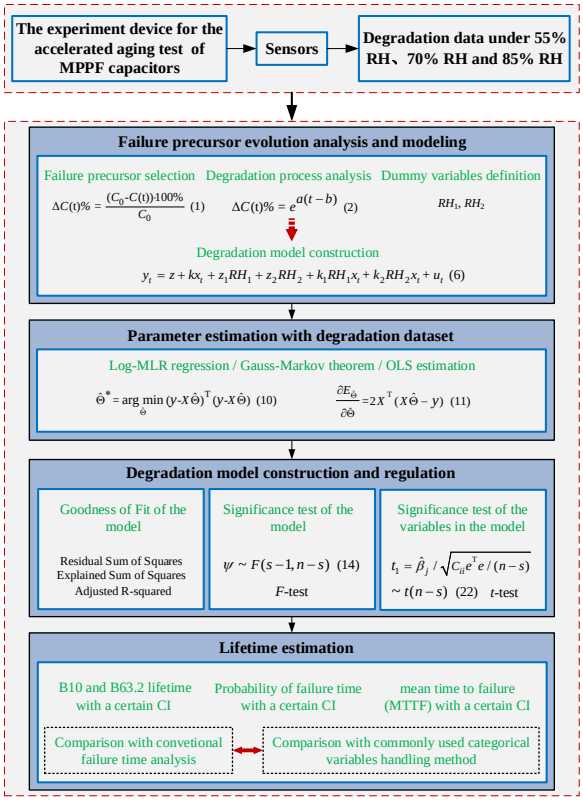


Fig. 5. The frame of the proposed method for DC film capacitors subject to humidity conditions.

adjusted to be more suitable for practical applications. In the last part, the failure time for MPPF capacitors under different humidity conditions is determined.

A. Dummy Variables-based Method

Dummy variables are usually used to transform qualitative variables with common attributes into numerical variables. Generally, the basic type is quantified as 0, the comparison type is 1. In the regression problem, the introduction of dummy variables can make the model more complicated and closer to reality. The precision and accuracy of the model are improved and the model can more intuitively reflect the influence of different categories of the attribute on the dependent variable. Besides, dummy variables-based regression not only learns parameters but also estimates the parameter distribution. The rules of setting dummy variables are as follows.

1) In a model with an intercept term, if the qualitative factor has n mutually exclusive types, only $n-1$ dummy variables can be introduced. Otherwise, it will fall into the ”dummy variable trap”, resulting in complete multicollinearity.
2) Avoid that the number of dummy variables introducing to the model exceeds the observations.
3) Dummy variables can be introduced to the regression model in two ways: addition and multiplication. Addition and multiplication change the intercept and slope of the model, respectively.

In the MLR model with dummy variables, the reference group setting rules are as follows.

1) We call the group with the dummy variable 0 as the reference group and the group with value 1 as the experimental group. Generally, 1 means that it appears or has a certain attribute, 0 means that it does not appear or does not have a certain attribute.
2) If the reference group setting is different, the significance of the experimental group is inconsistent.
3) Generally, set the lowest value or highest value as the reference group for analysis, or choose the group that interests researchers or the group with special meaning.
4) The reference group is usually the first level. With the reference group, it can be determined whether some other variables have an effect on the experimental group that deviates from reality.
5) After the reference group is selected, the dummy variable settings of other groups can be interchanged, which has no effect on the experimental results.

Since the higher the relative humidity is, the more easily the capacitor will be damaged, 70% RH and 85% RH may need more explanatory variables to interpret the impact of high humidity on capacitor degradation. Based on design rules of the reference group and practical application scenarios, 55% RH is selected as the reference group. The dummy variable 0 in the reference group means a certain explanatory variable does not work at this humidity, and the dummy variable 1 is to study the difference in the influence of higher humidity on capacitors compared with the reference group.

B. Degradation Model

5% capacitance drop from the initial capacitance is considered as the failure criteria in this paper. The percentage of the capacitance loss is defined as

$$\Delta C(t)\% = (C_0 - C(t))/C_0 \cdot 100\% \tag{1}$$

Where C_0 is the initial capacitance, $C(t)$ is the measured capacitance at time t.

As shown in Fig.3, the mean normalized capacitance in each group decreases exponentially with time. Therefore, the exponential and nonlinear characteristics of the deterioration process can be characterized by a general exponential model

$$\Delta C(t)\% \approx e^{a(t-b)} \tag{2}$$

Where a and b are the shape parameter and position parameter, respectively, t is the testing time. To apply the log transformation when $\Delta C(t)\%$ is around 0, we take the logarithm after adding 1 to $\Delta C(t)\%$. The sum of $\Delta C(t)\%$ and 1 can also be expressed in a general exponential function

$$\Delta C(t)\% + 1 \approx e^{a_1(t-b_1)} \tag{3}$$

Where a_1 and b_1 are the shape parameter and position parameter, respectively. Linearize (3), we can get

$$ln(\Delta C(t)\%+1) = a_1 t - a_1 b_1 \tag{4}$$

To reduce the aggregation of data points, make them symmetrically distributed on both sides of the linear model, and more extensible to mine the information contained in them, we perform twice logarithm operations, that is, the same logarithm operation is carried out on the left side of (4). The transformed data and testing time still satisfy a linear relationship. Let $ln(ln(\Delta C(t)\% + 1) + 1) = x_t$, $t = y_t$, we can obtain

$$y_t = kx_t + z \tag{5}$$

Where k and z are constants. It is worth noting that the end-of-life criteria for capacitors after modeling becomes x_t reaches 1.026672.

The dummy variables are used to find whether there are significant differences between the three groups and whether the differences come from the intercept term or the constant term. Based on essential definitions and setting rules of the dummy variables, three humidity levels are set by two dummy variables, named RH_1 and RH_2. The group under 55% RH is defined as the foundation type or reference group while the other two groups are comparison types or experimental groups. RH_1 and RH_2 under 55% RH, 70% RH and 85% RH are quantified as [0 0], [1 0] and [0 1], respectively. Considering the intercept and slope of (5) for each group may be different, RH_1 and RH_2 are introduced in the form of addition and multiplication simultaneously.

The degradation model of x_t, y_t, RH_1, RH_2 at monitoring time t is formulated as

$$y_t = z + kx_t + z_1 RH_1 + z_2 RH_2 + k_1 RH_1 x_t + k_2 RH_2 x_t + u_t \tag{6}$$

Where z and k denote deterministic constants and decide the characteristics of the foundation type. x_t is the general explanatory variable. RH_1 and RH_2 are dummy variables with values 0 or 1. k, z_1, z_2, k_1, k_2 are random coefficients to cover the heterogeneity of comparison types relative to the foundation type. u_t is the random error that follows a normal distribution with mean 0 and the same variance. The mean of y_t for 55% RH, 70% RH, and 85% RH can be formulated with (6) as

$$E(y_t|x_t, RH_1 = 0, RH_2 = 0) = z + kx_t \tag{7}$$

$$E(y_t|x_t, RH_1 = 1, RH_2 = 0) = z + z_1 + (k + k_1)x_t \tag{8}$$

$$E(y_t|x_t, RH_1 = 0, RH_2 = 1) = z + z_2 + (k + k_2)x_t \tag{9}$$

It is worth noting that RH_1 and RH_2 under 70% RH and 85% RH can also be set to [0 1] and [1 0]. In this case, the mean of y_t under 70% RH and 85% RH can be formulated as (9) and (8). Although variable terms are changed, the model structure remains the same. Therefore, the setting of experimental groups will not affect experiment results after the reference group is selected. The whole modeling process is based on exponential distribution. It can also be applied to other humidity-influenced components with exponential degradation characteristics.

C. Parameter Estimation with Training Dataset

The normalized capacitance of the three groups are used to estimate the parameters in (6). For the j^{th} sample in the i^{th} group, supposing there are m_i degradation measurements

$0 = t_{i,j,1} < t_{i,j,2} < ... < t_{i,j,k} < ... < t_{i,j,m_i}$, $i = 1, ..., 3$, $j = 1, ..., 10$, $k = 1, ..., m_i$. At monitoring time $t_{i,j,k}$, y_t and x_t are denoted as $y_{i,j,k} = y(t_{i,j,k})$ and $x_{i,j,k} = x(t_{i,j,k})$, respectively. According to the Gauss-Markov theorem, the least square estimator is the best linear unbiased estimator which has the smallest variance in the class of unbiased linear estimates [39]. Therefore, the ordinary least squares (OLS) is used to estimate the parameters in the degradation model.

Let $x_z = (x^z_{1,1,1}, ..., x^z_{1,10,m_1}, x^z_{2,1,1}, ..., x^z_{2,10,m_2}, x^z_{3,1,1}, ..., x^z_{3,10,m_3})^{\mathrm{T}}$ $(z = 1, ..., 5)$, Where $x^z_{i,j,k} = x^z(t_{i,j,k})$, $x^1 = x_t$, $x^2 = RH_1$, $x^3 = RH_2$, $x^4 = RH_1x_t$, $x^5 = RH_2x_t$. The input vector of the MLR model can be written as $X = (x_1, x_2, x_3, x_4, x_5, 1)$, where 1 is the unit vector with the same dimension as x_k. The output is written as $y = (y_{1,1,1}, ..., y_{1,10,m_1}, y_{2,1,1}, ..., y_{2,10,m_2}, y_{3,1,1}, ..., y_{3,10,m_3})^{\mathrm{T}}$ and the unknown weight of the input vector is denoted as $\hat{\Theta} = (\hat{k}, \hat{z}_1, \hat{z}_2, \hat{k}_1, \hat{k}_2, z)^{\mathrm{T}}$. With the OLS method, the parameters in the proposed degradation model can be estimated by

$$\hat{\Theta}^* = \underset{\hat{\Theta}}{\operatorname{argmin}} (y - X\hat{\Theta})^{\mathrm{T}}(y - X\hat{\Theta}) \tag{10}$$

Let $F_{\hat{\Theta}} = (y - X\hat{\Theta})^{\mathrm{T}}(y - X\hat{\Theta})$, the derivative of $F_{\hat{\Theta}}$ with respect to $\hat{\Theta}$ is calculated as

$$\frac{\partial F_{\hat{\Theta}}}{\partial \hat{\Theta}} = 2X^{\mathrm{T}}(X\hat{\Theta} - y) \tag{11}$$

Considering that the number of training data is larger than the number of variables, we can get the optimal solution by letting (11) be 0, the estimation parameters $\hat{\Theta}$ are obtained as

$$\hat{\Theta}^* = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y \tag{12}$$

In the established regression model, given an input $\hat{x}_{i,j,k} = (x^1_{i,j,k}, x^2_{i,j,k}, ..., x^5_{i,j,k}, 1)$, the corresponding estimation is $\hat{y}_{i,j,k} = \hat{x}^{\mathrm{T}}_{i,j,k}(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$.

D. The Metrics of Model Performance

The proposed degradation model is mainly evaluated in three aspects, including the goodness of fit (GF) of the model to the data, significance test of the model, and significance test of the variables in the model.

The GF of the model is characterized by the residual sum of squares (RSS) and the explained sum of squares (ESS). As shown in Fig. 6, the ESS reflects differences between groups while the RSS reflects differences within groups. The GF of the model is proportional to the ESS and inversely proportional to the RSS. With the values of RSS and ESS, the R-squared defined as the proportion of ESS in total sum of squares (TSS) is usually calculated to evaluate the regression effect. The larger R-squared means a better GF. To eliminate the influence of the number of variables on GF, the Adjusted R-squared is proposed by optimizing the R-squared considering the degree of freedom (DF).

In the significance test of the model, we mainly test whether the linear relationship between the explanatory variables and the explained variable is significant. We assume there is no linear relationship between them, which means all coefficients

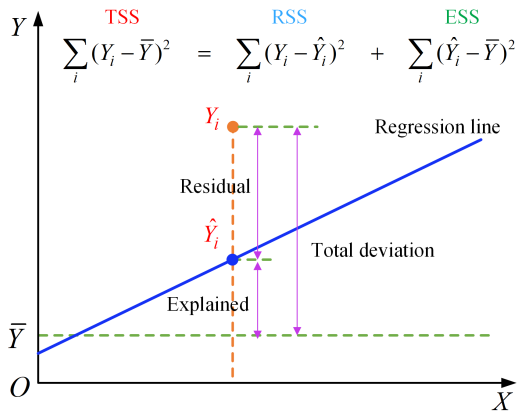


Fig. 6. Schematic diagram of the relationship between RSS, ESS, and TSS

of model (6) are 0, then test whether this hypothesis is true. The original hypothesis is written as

$$H_0 : z = k = z_1 = z_2 = k_1 = k_2 = 0 \tag{13}$$

It is known the sum of squares of several random variables with the standard normal distribution obeys the chi-square distribution. The ratio of two independent random variables with the chi-square distribution divided by respective DF obeys the F distribution. Under hypothesis (13), the explained variable y_t in (6) follows the normal distribution. In Fig. 6, Y_i can be regarded as samples taken from the normal population y_t, and they are independent of each other. Therefore, both RSS and ESS follow the chi-square distribution. We denote the number of model parameters including the intercept term as s, the number of samples in MLR as n, the DFs of ESS and RSS can be calculated as s-1 and n-s, respectively. At the same time, the mean square error (MSE) can be obtained by calculating the ratio of RSS to its DF. After dividing by respective DF, RSS and ESS can be used to construct F-statistics to test the hypothesis (13). With the transformation expression $\psi = (ESS/(s-1))/(RSS/(n-s))$, ψ is distributed as

$$\psi \sim F(s - 1, n - s) \tag{14}$$

Given a significance level α, we can reject or accept the original hypothesis in (13) by F-test and judge whether the linearity of the model is significant. When the F value locates in the rejection range determined by the significance level, the hypothesis is rejected, otherwise, the hypothesis is accepted.

It is also necessary to consider the effect of each explanatory variable on the explained variable and decide whether the explanatory variable should be retained in the model. We assume the explanatory variable has no significant influence on the explained variable, then test whether this hypothesis is true. Let $\beta_j (j = 0, 1, ..., 5)$ denote the parameters in the model, the original hypothesis is written as

$$H_0 : \beta_j = 0 \tag{15}$$

Since $\hat{\beta}_j$ obeys the normal distribution with mean β_j and unknown variance, it can be used to construct t-statistics to test the hypothesis (15). We write model (6) in matrix form

$y = X\Theta + u$, the estimation difference between $\hat{\Theta}$ and Θ can be obtained with (12) as

$$\hat{\Theta} - \Theta = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}u \tag{16}$$

Thus, $\hat{\Theta}$ is the unbiased estimation of Θ and its mean is Θ. We denote the variance of u as σ_u^2, the covariance of $\hat{\Theta}$ can be calculated as

$$E((\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^{\mathrm{T}}) = {\sigma_u}^2(X^{\mathrm{T}}X)^{-1} \tag{17}$$

Let $(X^{\mathrm{T}}X)^{-1}$=C_{ij} $(i, j = 0, 1, 2..., k)$, the variance of $\hat{\beta}_j$ is estimated as

$$Var(\hat{\beta}_j) = {\sigma_u}^2 C_{ii} \tag{18}$$

Next we derive an unbiased estimation of ${\sigma_u}^2$. The OLS estimation of y can be written as

$$\hat{y} = X\hat{\Theta} \tag{19}$$

Let $e = y - \hat{y}$, the RSS is derived as

$$e^{\mathrm{T}}e = u^{\mathrm{T}}(I_n - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}})u \tag{20}$$

Where I_n denotes the n order unit matrix. Then, the mean of RSS is obtained as

$$E(e^{\mathrm{T}}e) = {\sigma_u}^2(n - s) \tag{21}$$

Thus, an unbiased estimation of ${\sigma_u}^2$ can be obtained from (21). Substituting it into (18), the variance estimation of $\hat{\beta}_j$ can also be obtained. By subtracting the mean and dividing by the standard deviation estimation, $\hat{\beta}_j$ can be constructed as $\tau = (\hat{\beta}_j - \beta_j)/\sqrt{C_{ii}e^{\mathrm{T}}e/(n-s)}$, τ is distributed as

$$\tau \sim t(n - s) \tag{22}$$

Under hypothesis (15), $\tau_1 = \hat{\beta}_j/\sqrt{C_{ii}e^{\mathrm{T}}e/(n-s)} \sim t(n-s)$. Given a significance level α, we can accept or reject the original hypothesis in (15) by the t-test and judge whether this parameter has a significant effect on the explained variable. When the value locates in the rejection range determined by the significance level, the hypothesis is rejected, otherwise, the hypothesis is accepted. Meanwhile, we can construct the interval estimation of parameters. The probability that t locates in the receptive domain is 1-α, from which we can get the confidence interval (CI) of β_j

$$\left(\hat{\beta}_j - t_{\alpha/2}\sqrt{C_{ii}e^{\mathrm{T}}e/(n-s)}, \hat{\beta}_j + t_{\alpha/2} \cdot \sqrt{C_{ii}e^{\mathrm{T}}e/(n-s)}\right) \tag{23}$$

That is to say, we believe the CI covers the true value of β_j with 1-α certainty.

E. Lifetime Estimation Method

In addition to the absolute value of lifetime, the CI of the estimated lifetime need to be calculated to take into account the uncertainty which is inherent to failure prognostics [40]. In model (6), the random error u_t is assumed to satisfy a normal distribution with mean 0 and the same variance, which is described as

$$u_t \sim N(0, \sigma_u^2) \tag{24}$$

Model (6) can be written in matrix form as

$$y = X\Theta + u \tag{25}$$

Where Θ is the coefficient matrix, X is the explanatory variable matrix. According to (24), y obeys a normal distribution. Given a set of observations of explanatory variables X_0, the OLS estimation of y is written as

$$\hat{y}_0 = X_0\hat{\Theta} \tag{26}$$

Based on the property of OLS, $\hat{y}_0$ obeys a normal distribution. Thus, the estimation error of (26) also follows a normal distribution, which is expressed as

$$\hat{y}_0 - y \sim N(0, \sigma_u^2(1 + X_0(X^{\mathrm{T}}X)^{-1}X_0^{\mathrm{T}})) \tag{27}$$

Therefore, a t-statistic can be constructed as

$$\frac{\hat{y}_0 - y}{\sqrt{\hat{\sigma}_u^2(1 + X_0(X^{\mathrm{T}}X)^{-1}X_0^{\mathrm{T}})}} \sim t(n - s) \tag{28}$$

Where s denotes the number of explanatory variables in (6). Given a significance level α, the CI of y_0 is obtained as

$$(\hat{y}_0 - l_{in}^*, \hat{y}_0 + l_{in}^*) \tag{29}$$

Where $l_{in}^* = t_{\alpha/2}\sqrt{\hat{\sigma}_u^2(1 + X_0(X^{\mathrm{T}}X)^{-1}X_0^{\mathrm{T}})}$.

Under explanatory variable X_0, the conditional expectation of y is described as

$$E(y|X_0) = X_0\Theta \tag{30}$$

Similarly, the difference between $\hat{y}_0$ and $E(y|X_0)$ follows a normal distribution expressed as

$$\hat{y}_0 - E(y|X_0) \sim N(0, \sigma_u^2 X_0(X^{\mathrm{T}}X)^{-1}X_0^{\mathrm{T}}) \tag{31}$$

The CI of $E(y|X_0)$ under the significance level α can also be obtained as

$$(\hat{y}_0 - l_{me}^*, \hat{y}_0 + l_{me}^*) \tag{32}$$

Where $l_{me}^* = t_{\alpha/2}\sqrt{\hat{\sigma}_u^2 X_0(X^{\mathrm{T}}X)^{-1}X_0^{\mathrm{T}}}$.

Moreover, the normal distribution of y under X_0 can be described as

$$f(y|X_0, \Theta, u) \sim N(X_0\Theta, \sigma_u^2) \tag{33}$$

Where σ_u^2 are estimated by the OLS method, refer to (21). In this way, the probability density function of y under X_0 is written as

$$f(y|X_0, \Theta, u) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_u}\exp(-\frac{(y - X_0\Theta)^2}{2\hat{\sigma}_u^2}) \tag{34}$$

The corresponding cumulative distribution function (CDF) is expressed as

$$\begin{aligned} F_L(y|X_0, \Theta, u) &= P(Y \le y|X_0, \Theta, u) \\ &= \int_{-\infty}^{y} f(y|X_0, \Theta, u)dy \end{aligned} \tag{35}$$

Given the set of explanatory variables when the failure criterion is reached, the lifetime with a certain CI can be estimated based on (29) and (32).

The commonly used lifetime standard is BX (X=1, 10, or 63.2) lifetime, which represents the time when 1%, 10%, or

63.2% of the components run to failure. In this paper, B10 and B63.2 lifetime are also studied. With the CDF of failure time in (35), B10 lifetime can be calculated as

$$F_Y(y_{\mathrm{B10}}|X_0^*,\Theta,u) = P(Y<y_{\mathrm{B10}}|X_0^*,\Theta,u) = 10\% \quad (36)$$

and B63.2 lifetime can be similarly calculated as

$$F_Y(y_{\mathrm{B63.2}}|X_0^*,\Theta,u) = P(Y<y_{\mathrm{B63.2}}|X_0^*,\Theta,u) = 63.2\% \quad (37)$$

Where X_0^* denotes the explanatory variables reaching the failure criterion.

Since the proposed algorithm does not involve intensive variable storage and reading, the algorithm complexity mainly refers to time complexity, which comes from parameter estimation in (12), variance estimation in (20) and (21), and lifetime estimation in (29), (32) and (33). The algorithm complexity in these three parts is $O(n)$, $O(n^2)$, and $O(1)$, respectively. Therefore, the complexity of the proposed algorithm is $O(n^2)$.

IV. Performance Evaluation and Lifetime Estimation

In this section, the degradation model is adjusted by performance evaluation. Then, lifetime estimation is carried out and the effectiveness of the proposed method is verified. At last, the selection of the reference group is validated. The performance evaluation and lifetime estimation are conducted on a personal computer with Windows 10 operating system, 1.60 GHz Intel(R) Core(TM) i5-8250U CPU, and 8 GB RAM. The statistical analysis software Stata/MP 16.0 is adopted to perform the algorithm. During operation, the size of the temporary storage space occupied by the algorithm and software is about 40 MB. The entire process takes 0.85s. The algorithm can be executed on open-source free software at a low cost.

A. Residual Normality Test

As mentioned earlier, the premise of interpreting the proposed model is that the residuals follow a normal distribution. Therefore, a Quantile-Quantile plot of the residuals is created to assess whether the residuals are normally distributed, as shown in Fig. 7, the orange line is the 45-degree line, and the blue points are residuals. We can see the residuals fall along a straight line at the 45-degree angle and only deviate slightly on the tail end, thus the residuals are considered to be normally distributed.

B. Degradation Model Assessment

In Table II, the high F(5,465) declares the proposed model is significant. The p-value of the F-test shows the significance level of accepting the original hypothesis in (13) can reach as low as 1%, thus we can think the variable correlation of the proposed model is intense under the confidence level of more than 99%. The degrees of freedom of the ESS and RSS are 5 and 465, respectively, which are determined by the number of model variables and data samples. According to ESS, RSS and their degrees of freedom, the R-squared and Adjusted R-squared can be calculated, which reveals the overall goodness

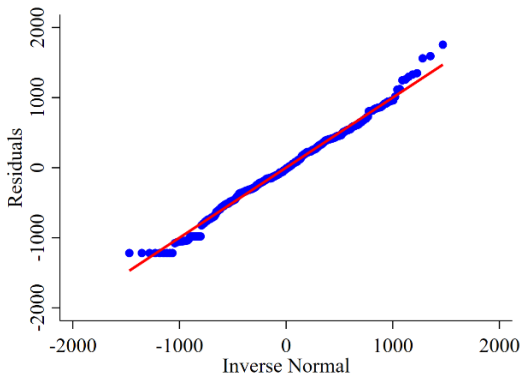


Fig. 7. Quantile-Quantile plot of the regression residuals.

TABLE II
The Goodness of Fit and Significance Test of The Proposed Model

Indicator	Value
F(5, 465)	179.06
p-value	6.2e-106
R-squared	0.6582
Adjusted R-squared	0.6545

of fit of the model. There exists a margin between 0.6545 and 1, which indicate a part of model residuals cannot be explained by the explanatory variables. This is mainly limited by the experimental data, which cannot deny the validity of the model. Usually, increasing the number of samples can enhance the R-squared. The low p-value of the F-test and the comparable R-squared fully demonstrates the feasibility of the proposed model.

Table III shows the estimation and t-test results of parameters. It can be seen the variables with coefficients k, z_2, and z have a significant effect on the explained variable, which can reject the null hypothesis in (15) under a 99.9% significance level, while the impact of the variables with coefficients z_1, k_1 and k_2 seems a bit inferior. This encourages us to study the influence of these variables on the degradation model further.

After obtaining the general degradation model for three humidity conditions, the respective degradation model applied to different humidity conditions can be given, as shown by the red curve in Fig. 8. It can be seen the degradation data is almost distributed on both sides of the regression curve. The deviation between the measured values and the regression curve reflect the heterogeneity between samples. It is in this deviation that we study the mean-time-to-failure of capacitors. Based on this distribution deviation, we can give the confidence interval of the mean lifetime at various confidence levels. The time when the percentage of the capacitance loss reaches the largest value decreases with the humiduty increases, which represents the capacitor lifetime is shorter when the humudity is higher. Besides, the initial points are concentrated at the degradation ratio of about 0 under 55% RH, indicating the capacitance fluctuates around the initial value for a long time when the humidity is low.

In the regression analysis, the variables that are not im-

TABLE III
THE ESTIMATION AND SIGNIFICANCE TEST OF PARAMETERS

Coefficient	Nominal	Standard deviation	t	p-value	95% CI
k	1679.664	76.19643	22.04	-3.9e-57	[1529.932,1829.396]
z_1	-237.8624	85.54381	-2.78	5.8e-3	[-405.9628,-69.76211]
z_2	-1066.952	94.07881	-11.34	4.8e-25	[-1251.824,-882.0793]
k_1	-366.0364	-366.0364	-2.71	7.1e-3	[-631.1389,-100.934]
k_2	-365.724	120.472	-3.04	2.6e-3	[-602.4609,-128.9871]
z	1218.296	42.70882	28.53	-3.9e-57	[1134.37,1302.222]

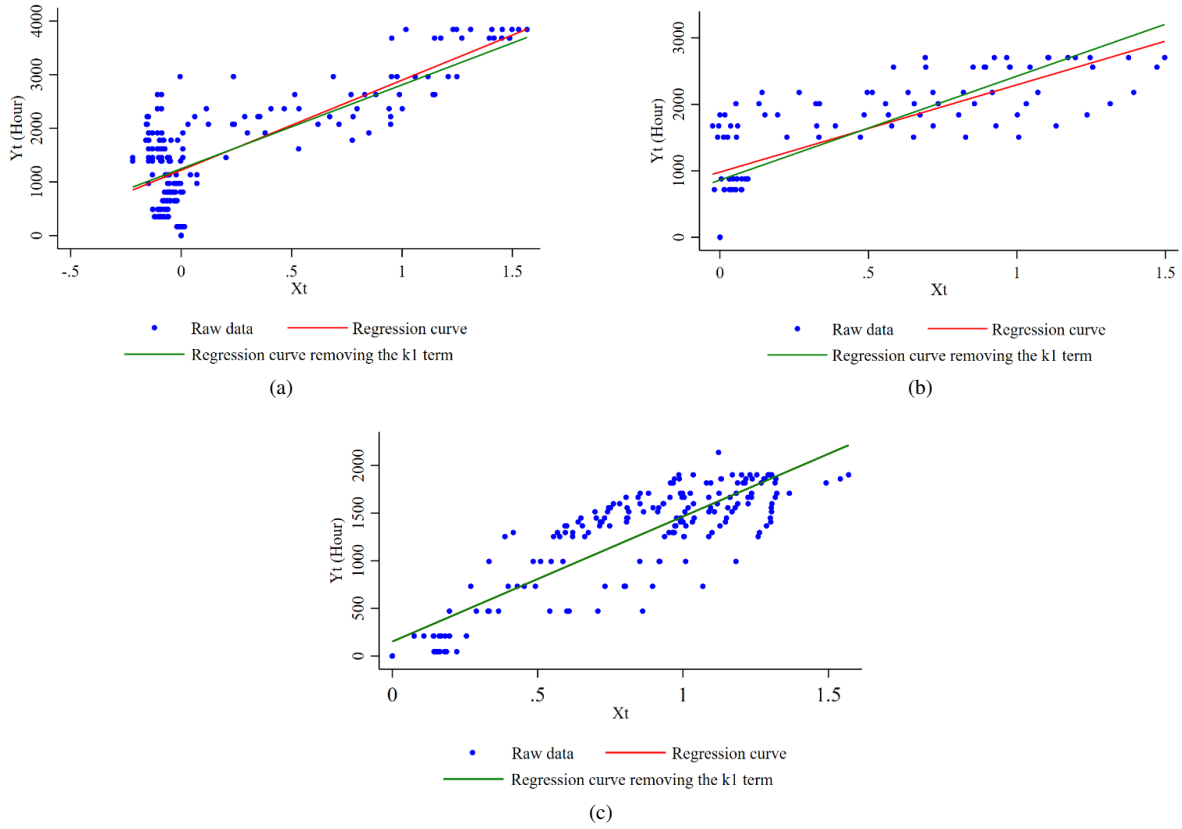


Fig. 8. Regression results of the MLR model under various humidity conditions ((a), (b), and (c) are corresponding to 55%, 70%, and 85% RH, respectively).

portant can be removed from the model by diagnosing their multicollinearity and p-value, to obtain the simplified model [36]. To learn more about the influence of the z_1, k_1, and k_2 term on the explained variable, the variable is removed and the regression is performed again. Here, only the k_1 item is studied. The corresponding results are plotted by the green curve in Fig. 8, we can see the slope and intercept of the regression curve differ from the red curve under 55% RH and 70% RH. The mean of residuals from the observation point to the red regression line and green regression line under each group are both 0. The residual standard deviation of the red line under 55% RH, 70% RH, and 85% RH are 664.528, 537.387, and 295.223, respectively, while are 667.160, 549.830, and 295.223, respectively for the green line. The residual standard deviation under 55% RH and 70% RH both increase, which indicates the model without the k_1 item is insufficiently adaptable to the data. Therefore, this item can not be omitted. The residual standard deviation under 85% RH remains the same, this is because removing the k_1 term does not change the model structure at this humidity. Besides, the mean and standard deviation of the standardized residuals are calculated as 0 and 1, further verifying the assumption that the residual follows a normal distribution with mean 0 and the rationality of the proposed model. Moreover, considering the nature of the degradation ratio, we can modify the cross term in (6) and get the following

$$y_t = z' + k' x_t + z'_1 RH_1 + z'_2 RH_2 + k'_1 RH_1 \Delta C(t)\% + k'_2 RH_2 \Delta C(t)\% + u_t \quad (38)$$

Where z' and k' denote the deterministic constants and decide the characteristics of the foundation type. $k', z'_1, z'_2, k'_1, k'_2$ are random coefficients to cover the heterogeneity of comparison types relative to the foundation type.

We name the three models: the degradation model (6), the degradation model (6) removing the k_1 item, the degradation model (38) as Model I, Model II, and Model III

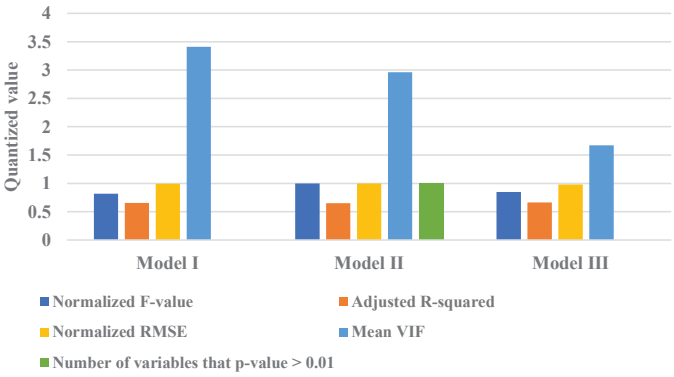


Fig. 9. The comparison of four characteristics for three models.

successively. Four characteristics including the normalized F-value, Adjusted R-squared, normalized Root Mean Square Error (RMSE), multicollinearity, and the number of variables whose p-value is greater than 0.01 are compared for these models. Multicollinearity refers to the linear correlation between independent variables and may lead to large variance and low accuracy for OLS parameter estimation. In this paper, the variance inflation factor (VIF) is calculated to diagnose multicollinearity. The VIF is defined as

$$\mathrm{VIF}_i = \frac{1}{1 - R_i^2} \tag{39}$$

Where R_i^2 denotes the determination coefficient of variable x_i. The closer the VIF is to 1, the lighter the multicollinearity is, and a serious linear relationship exists between explanatory variables when the VIF exceeds 10.

As shown in Fig. 9, the mean VIF of each model is within 5, representing the multicollinearity between explanatory variables is weak and the OLS estimation results are credible. It exists an explanatory variable whose p-value is larger than 0.01 in Model II, which demonstrates the k_1 item in (6) can not be ignored. Since Model III performs better than Model I and Model II, it be used for lifetime estimation later.

C. Lifetime estimation

Fig. 10 shows the estimated degradation time with 95% CI corresponding to various capacitance loss ratios under three RH environments. We can see that it takes a longer time for the capacitance under lower RH to decrease to the same value.

In Table IV, the nominal, mean confidence interval (Mean CI) and individual confidence interval (Individual CI) of the mean time to failure (MTTF) lifetime estimated by the proposed method are given for three cases. The nominal lifetime is determined by model (38) based on the OLS estimation, given the observed values of the explanatory variables when x_t reaches 1.026672. We can see the Individual CI is greater than the Mean CI, which accords with the common sense. As the confidence level increases, the Mean CI and Individual CI are extended to a larger range. It is obvious that the deterioration speed of capacitors within 70% RH and 85% RH is larger than that less than 70% RH, the nominal lifetime decreases almost 860 hours from 70% RH to 85% RH while 500 hours from 55% to 70%.

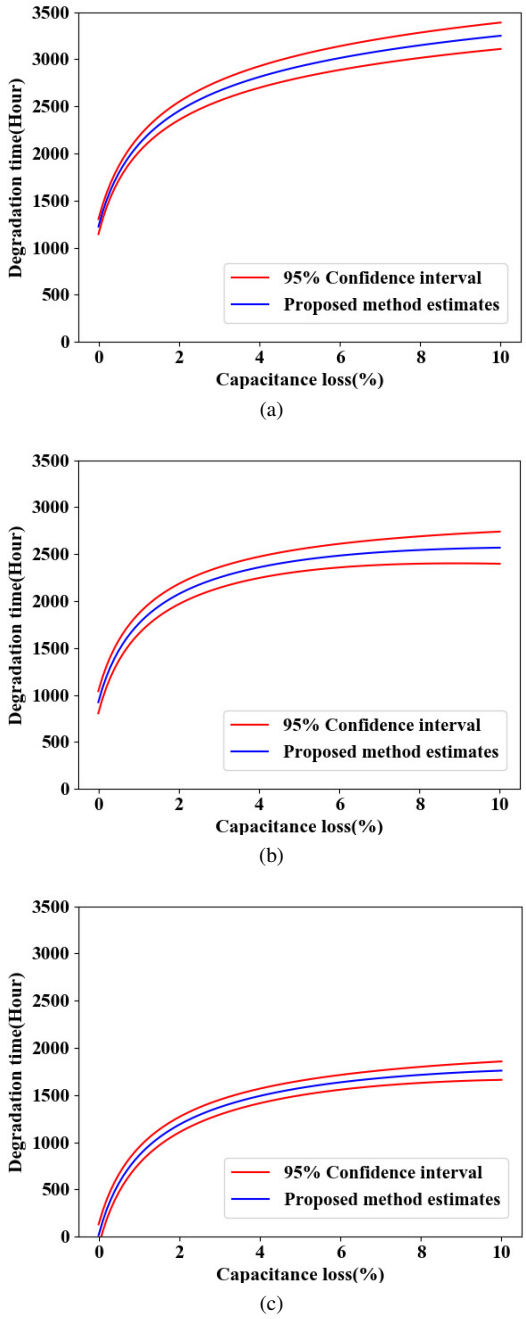


Fig. 10. The estimated degradation time with 95% CI corresponding to capacitance loss under various humidity conditions ((a), (b), and (c) are corresponding to 55%, 70%, and 85% RH, respectively).

The details of the estimated B10 and B63.2 lifetime with various Mean CI are given in Table V. The nominal B10 and B63.2 lifetime are estimated by (36) and (37), given the set of explanatory variables when the capacitors reach the nominal lifetime. It can be also found that the deterioration between 70% RH and 85% RH is more severe than that between 55% RH and 70% RH.

The lifetime calculated by Weibull distribution is applied to be a standard to verify the proposed method. The Weibull failure distribution and corresponding MTTF lifetime are given

TABLE IV
MTTF LIFETIME UNDER VARIOUS HUMIDITY CONDITIONS AND CONFIDENCE LEVEL

Humidity condition	Confidence level	Nominal	Mean CI	Individual CI
85°C and 55% RH (Hours)	90%	2924	[2823, 3026]	[2079, 3769]
	95%		[2803, 3045]	[1917,3931]
	99%		[2765, 3084]	[1601, 4248]
85°C and 70% RH (Hours)	90%	2433	[2333, 2533]	[1601, 4248]
	95%		[2314, 2552]	[1426,3440]
	99%		[2277, 2590]	[1110, 3756]
85°C and 85% RH (Hours)	90%	1575	[1510, 1639]	[733, 2416]
	95%		[1497, 1652]	[572, 2577]
	99%		[1473, 1676]	[257, 2892]

TABLE V
THE ESTIMATED BX LIFETIME USING THE PROPOSED METHOD

Humidity condition	Lifetime	Nominal	90% Mean CI	95% Mean CI	99% Mean CI
85°C and 55% RH (Hours)	B10	2271	[2169, 2372]	[2150, 2392]	[2111, 2430]
	B63.2	3096	[2995, 3198]	[2975, 3217]	[2937, 3255]
85°C and 70% RH (Hours)	B10	1780	[1680, 1880]	[1661, 1899]	[1623, 1936]
	B63.2	2605	[2505, 2705]	[2486, 2724]	[2448, 2762]
85°C and 85% RH (Hours)	B10	924	[861, 988]	[849, 1000]	[825, 1024]
	B63.2	1748	[1684, 1812]	[1671, 1825]	[1648, 1849]

TABLE VI
THE ESTIMATED LIFETIME USING WEIBULL DISTRIBUTION

Lifetime	85°C and 55% RH (Hours)	85°C and 70% RH (Hours)	85°C and 85% RH (Hours)
MTTF	3001	2441	1536
B10	2319	1710	850
B63.2	3210	2655	1715

as

$$F(y) = 1 - \exp(-\left(\frac{y}{\eta}\right)^{\beta}) \tag{40}$$

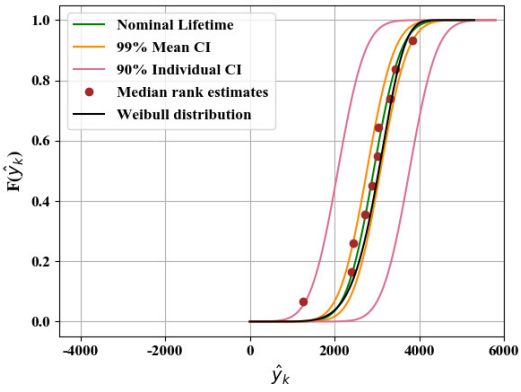
$$\text{MTTF} = \eta \cdot \Gamma(1 + \frac{1}{\beta}) \tag{41}$$

Where y denotes time, $\Gamma(\ )$ denotes gamma function, η and β represent the scale parameter and shape parameter, respectively. η is the time when 63.2% of the samples fail, which are 3210, 2655, and 1715 for Group 1, Group 2, and Group 3, respectively. β can be obtained as 6.9, 5.1, and 3.2, respectively. The parameter estimation is based on linear regression, the algorithm complexity is $O(n)$, n is the number of samples. The estimation results are given in Table VI. Compared with the values in Table IV and Table V, the estimation difference between the proposed method and Weibull analysis is acceptable, which proves the effectiveness of the proposed methodology.

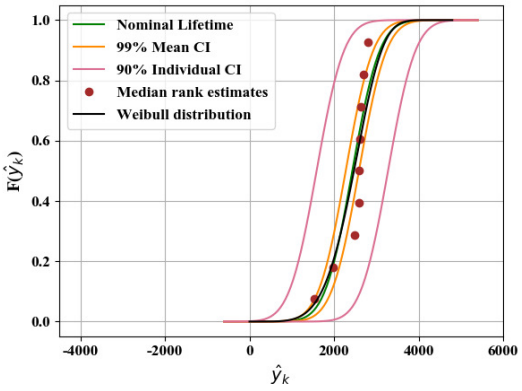
In order to verify the effectiveness of the proposed method more comprehensively, the median rank estimates are regarded as individuals to test whether the estimated CDF is consistent with the individual distribution. Fig. 11 shows the median rank estimates, Weibull distribution, and the estimated CDF of MTTF lifetime. Most of the median rank estimates distribute inside or tightly on both sides of the 99% Mean CI except for individual points. The range of 90% Individual CI is relatively wide and basically covers all possible individual lifetimes. Besides, the rising trend of the estimated CDF agrees well with the median rank estimates and Weibull distribution.

To evaluate the capability of the proposed lifetime estimation method, the capacitance loss data with progressively reduced time points is applied for model construction. Fig. 12 shows the comparison between the Weibull failure distribution and the evolution of the estimated CDF of MTTF failure time under 85% RH. It can be seen the estimated CDF keeps a narrow interval with the Weibull failure distribution until the fourth time point is reduced, which corresponds to the testing time of 1666 hours. Through calculation, the maximum error percentage of MTTF, B10, and B63.2 lifetime estimation compared to the Weibull analysis is 3.76%. Similarly, the testing time can be reduced by 162 hours and 144 hours at 55% RH and 70% RH, with maximum error percentages of 5.76% and 4.78%. It indicates the proposed method can achieve comparable accuracy with reduced data and testing time. Instead of running all capacitors failed, the testing can be finished when the first few samples fail, and the degradation data of samples monitored up to this moment are used for modeling.

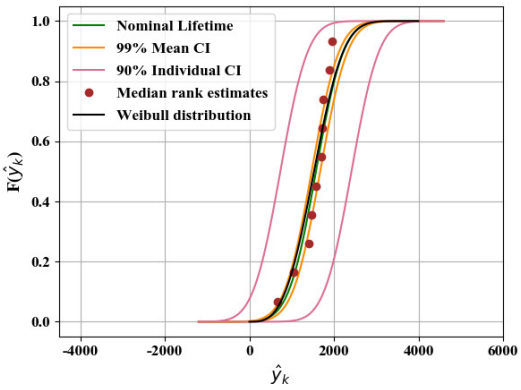
Different humidity conditions are seen as categorical variables in this paper. The commonly used categorical variables



(a)



(b)



(c)

Fig. 11. The estimated CDF of failure time under various humidity conditions ((a), (b), and (c) are corresponding to 55%, 70%, and 85% RH, respectively).

handling method is called stratification, which divides the dataset into subsets and builds multiple models for each subset. Compared with the usual stratification, the introduction of dummy variables increases regression data, which may bring a higher estimation accuracy [32]. To further investigate the effect of dummy variables, the comparison of the model regression effect and lifetime estimation accuracy between the proposed method and the stratification is conducted. As previously described, the time-series data of 29 testing samples in three groups are applied as training data for the proposed method. But in the stratification, the time-series data of 9, 10, and 10 testing samples are used for subset training of 55% RH, 70% RH, and 85% RH, respectively. The RMSE and

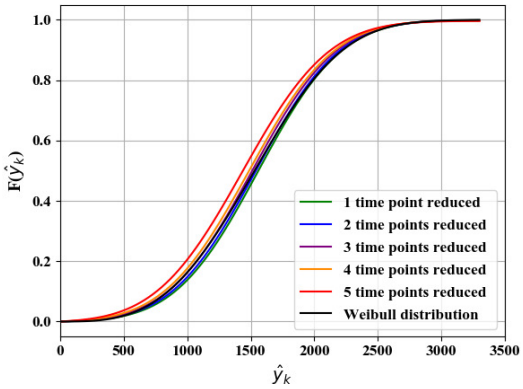


Fig. 12. Evolution of the estimated CDF when a different number of time points is reduced under 85% RH.

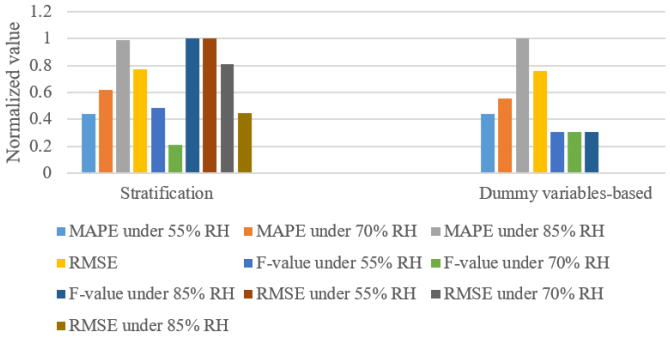


Fig. 13. The comparison of evaluation indicators between stratification and dummy variables-based method.

average absolute percentage error (MAPE) are used to evaluate the regression accuracy and lifetime estimation accuracy, respectively.

Fig. 13 shows the crucial normalized evaluation indicators in model regression and lifetime estimation, we can find the entire effects of model regression and lifetime estimation with dummy variables are better than the stratification. The MAPE, F-value, and RMSE under 70% RH are significantly improved after transforming dummy variables. The poor model performance under 70% RH with the stratification may due to its only 10 time points in time-series data. In contrast, the dummy variables-based method improves performance by increasing the number of training data. Therefore, the proposed model based on dummy variables performs better than the usual stratification method in terms of the dataset being studied.

D. Reference group selection validation

As mentioned earlier, if the reference group is different, the significance of the experimental group is inconsistent. In this part, the proposed model (6) and model (38) are used to illustrate the influence of the reference group selection on experimental results. It is known that modifying the dummy variable settings of the experimental groups will not change the model form, nor will change the experimental results. In this way, we can ignore the settings of dummy variables in

TABLE VII
SETTINGS OF DUMMY VARIABLES FOR DIFFERENT REFERENCE GROUPS

RH	55% RH reference group			70% RH reference group			85% RH reference group		
	55%RH	70%RH	85%RH	55%RH	70%RH	85%RH	55%RH	70%RH	85%RH
RH_1	0	1	0	1	0	0	1	0	0
RH_2	0	0	1	0	0	1	0	1	0

TABLE VIII
MODEL PERFORMANCE EVALUATION UNDER DIFFERENT REFERENCE GROUPS

Performance Evaluation		Reference group for model (6)			Reference group for model (38)		
		55%RH	70%RH	85%RH	55%RH	70%RH	85%RH
Evaluation indicators	F value	179.06	179.06	179.06	186.24	181.31	180.36
	p-value	6.2e-106	6.2e-106	6.2e-106	1.4e-108	8.8e-107	2.0e-106
	R-squared	0.6582	0.6582	0.6582	0.6670	0.6610	0.6598
	Adjusted R-squared	0.6545	0.6545	0.6545	0.6634	0.6573	0.6561
p-value of t-test	$k(k')$	-3.9e-57	3.8e-24	6.5e-35	-3.9e-57	2.7e-59	5.7e-61
	$z_1(z_1')$	5.8e-3	5.8e-3	4.8e-25	1.7e-5	7.1e-6	2.0e-42
	$z_2(z_2')$	4.8e-25	1.3e-12	1.3e-12	1.1e-52	3.0e-24	1.0e-27
	$k_1(k_1')$	7.1e-3	7.1e-3	2.6e-3	3.2e-4	0.0293	0.0153
	$k_2(k_2')$	2.6e-3	0.9981	0.9981	7.7e-5	0.0107	0.0265
	$z(z')$	-3.9e-57	8.7e-32	0.0713	-3.9e-57	5.0e-37	0.5024
MTTF lifetime estimation	55% RH	2943	2943	2943	2924	2817	2800
	70% RH	2329	2329	2329	2433	2416	2349
	85% RH	1500	1500	1500	1575	1534	1530
	$\hat{\sigma}_u^2$	267087	267087	267087	260217	264896	265814

experimental groups. The settings of dummy variables for different reference groups are shown in Table VII. Corresponding experimental results are in Table VIII.

As seen from Table VIII, for model (6), evaluation indicators and lifetime estimation results under different reference groups are the same, but the t-test results of each explanatory variable are significantly different. To improve model performance, it is necessary to remove explanatory variables whose p-value are greater than 0.05. In the experiment with 70% RH as the reference group, after removing the explanatory variable with coefficient k_2, model forms under 70% RH and 85% RH are the same, but different from 55% RH. In the experiment with 85% RH as the reference group, after removing the explanatory variable with coefficient k_2, model forms under 70% RH and 85% RH are the same, and different from 55% RH. This proves high humidity has a significant effect on the aging of DC film capacitors, 70% RH and 85% RH conditions should be experimental groups. In model (38), when 70% RH or 85% RH is used as the reference group, evaluation indicators and p-value of t-test for explanatory variables are worse than using 55% RH. Lifetime estimation results under 55% RH have a large deviation. Therefore, variable items reflecting the impact of high humidity are not needed in the model under 55% RH, which further verifies it is suitable to choose 55% RH as the reference group in our research.

V. CONCLUSION

This paper puts forward a lifetime estimation methodology for MPPF capacitors subject to various humidity levels. The capacitor degradation model is established based the percentage of capacitance loss. The humidity factor is neatly introduced to the degradation model in the shape of dummy variables. Statistical methods such as the t-test and F -test are applied to evaluate and modify the model. Then, the probability of failure time is constructed, and the B10, B63.2, and MTTF lifetime with a certain CI are obtained. It is worth noting that the proposed methodology is also applicable for other components whose humidity-influenced aging characteristics follow an exponential distribution, whether for electrolytic capacitors or film capacitors, or other electronic components.

The effectiveness of the presented methodology is validated by comparison with the conventional Weibull analysis. It indicates the proposed model can achieve comparable accuracy with reduced data and testing time. Besides, the lifetime estimation effects of the dummy variables-based scheme are proved to be better than the commonly used stratification method in terms of the dataset being studied. Although the complexity of the proposed algorithm is slightly increased than the Weibull analysis, the proposed algorithm considers humidity when estimating lifetime, which can easily study the distribution regularity of humidity and the distribution of capacitor lifetime which is conditionally dependent on humidity. This is significant because humidity is a key factor affecting the lifetime of film capacitors.

REFERENCES

[1] Y. Ohno and H. Haga, "Control method of electrolytic capacitor-less dual inverter for harmonic compensation under distorted grid voltage," in *2021 IEEE 12th Energy Conversion Congress Exposition - Asia (ECCE-Asia)*, 2021, pp. 1396–1401.

[2] Z. Liao, N. C. Brooks, and R. C. Pilawa-Podgurski, "Design constraints for series-stacked energy decoupling buffers in single-phase converters," *IEEE Transactions on Power Electronics*, vol. 33, no. 9, pp. 7305–7308, 2018.

[3] P. Sundararajan, M. H. M. Sathik, F. Sasongko, C. S. Tan, J. Pou, F. Blaabjerg, and A. K. Gupta, "Condition monitoring of dc-link capacitors using goertzel algorithm for failure precursor parameter and temperature estimation," *IEEE Transactions on Power Electronics*, vol. 35, no. 6, pp. 6386–6396, 2020.

[4] C. Zhang, L. Xu, X. Zhu, Y. Du, and L. Quan, "Torque ripple reduction of pmsm with small capacitor drive systems based on combined control method," *IEEE Access*, vol. 9, pp. 98 874–98 882, 2021.

[5] P. Sundararajan, M. H. M. Sathik, F. Sasongko, C. S. Tan, M. Tariq, and R. Simanjorang, "Online condition monitoring system for dc-link capacitor in industrial power converters," *IEEE Transactions on Industry Applications*, vol. 54, no. 5, pp. 4775–4785, 2018.

[6] S. S. Manohar, A. Sahoo, A. Subramaniam, and S. K. Panda, "Condition monitoring of power electronic converters in power plants — a review," in *2017 20th International Conference on Electrical Machines and Systems (ICEMS)*, 2017, pp. 1–5.

[7] Y. Wu and X. Du, "A ven condition monitoring method of dc-link capacitors for power converters," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1296–1306, 2019.

[8] C. Wu, J. Yue, J. Liu, and L. Wang, "An online proactive health monitoring method for output capacitors of vehicular auxiliary converter," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, pp. 1–1, 2021.

[9] H. Soliman, H. Wang, and F. Blaabjerg, "A review of the condition monitoring of capacitors in power electronic converters," *IEEE Transactions on Industry Applications*, vol. 52, no. 6, pp. 4976–4989, 2016.

[10] A. Gupta, O. P. Yadav, D. DeVoto, and J. Major, "A review of degradation behavior and modeling of capacitors," in *International Electronic Packaging Technical Conference and Exhibition*, vol. 51920. American Society of Mechanical Engineers, 2018, p. V001T04A004.

[11] C. Lv, J. Liu, Y. Zhang, W. Lei, and R. Cao, "An improved lifetime prediction method for metallized film capacitor considering harmonics and degradation process," *Microelectronics Reliability*, vol. 114, p. 113892, 2020.

[12] M. Rigamonti, P. Baraldi, E. Zio, D. Astigarraga, and A. Galarza, "Particle filter-based prognostics for an electrolytic capacitor working in variable operating conditions," *IEEE Transactions on Power Electronics*, vol. 31, no. 2, pp. 1567–1575, 2015.

[13] F. Wang, Y. Cai, H. Tang, Z. Lin, Y. Pei, and Y. Wu, "Prognostics of aluminum electrolytic capacitors based on chained-svr and 1d-cnn ensemble learning," *Arabian Journal for Science and Engineering*, pp. 1–18, 2022.

[14] A. Gupta, O. P. Yadav, A. Roy, D. DeVoto, and J. Major, "Degradation modeling and reliability assessment of capacitors," in *International Electronic Packaging Technical Conference and Exhibition*, vol. 59322. American Society of Mechanical Engineers, 2019, p. V001T06A018.

[15] X. Ye, Y. Hu, B. Zheng, C. Chen, R. Feng, S. Liu, and G. Zhai, "Reliability assessment of film capacitors oriented by dependent and nonlinear degradation considering three-source uncertainties," *Microelectronics Reliability*, vol. 126, p. 114277, 2021.

[16] C. Bhargava and P. K. Sharma, "Statistical and intelligent reliability analysis of multi-layer ceramic capacitor for ground mobile applications using taguchi's approach," *International Journal of Quality & Reliability Management*, 2021.

[17] D. K. B. Kulevome, H. Wang, and X. Wang, "A bidirectional lstm-based prognostication of electrolytic capacitor," *Progress In Electromagnetics Research C*, vol. 109, pp. 139–152, 2021.

[18] O. Prakash and A. K. Samantaray, "Prognosis of dynamical system components with varying degradation patterns using model–data–fusion," *Reliability Engineering & System Safety*, vol. 213, p. 107683, 2021.

[19] N. Valentine, M. H. Azarian, and M. Pecht, "Metallized film capacitors used for emi filtering: A reliability review," *Microelectronics Reliability*, vol. 92, pp. 123–135, 2019.

[20] Z. Li, H. Li, F. Lin, Z. Wang, G. Zhang, and L. Qi, "Capacitance loss analysis of metallized film capacitors under ac condition," in *2020 IEEE 3rd International Conference on Dielectrics (ICD)*. IEEE, 2020, pp. 173–176.

[21] H. K. Boshkova, "New kemet's miniaturized emi-suppression and dc-link power box unique designs for harsh environment in energy, industrial and automotive application."

[22] Q. Chen, H. Li, L. Li, H. Jiang, Y. Liu, Q. Zhang, F. Lin, and C. Zhang, "Moisture ingress of metallized film capacitor under high temperature and different humidity condition," in *2018 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP)*. IEEE, 2018, pp. 422–425.

[23] R. Gallay, "Metallized film capacitor lifetime evaluation and failure mode analysis," *arXiv preprint arXiv:1607.01540*, 2016.

[24] H. Li, Z. Li, F. Lin, Q. Chen, T. Qiu, Y. Liu, and Q. Zhang, "Capacitance loss mechanism and prediction based on electrochemical corrosion in metallized film capacitors," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 28, no. 2, pp. 654–662, 2021.

[25] Steady-State Temperature Humidity Bias Life Test, EIA/JEDEC Standard JESD22-A101-B, Apr. 1997.

[26] D. S. Peck, "Comprehensive model for humidity testing correlation," in *24th International Reliability Physics Symposium*. IEEE, 1986, pp. 44–50.

[27] H. Wang, D. A. Nielsen, and F. Blaabjerg, "Degradation testing and failure analysis of dc film capacitors under high humidity conditions," *Microelectronics Reliability*, vol. 55, no. 9-10, pp. 2007–2011, 2015.

[28] H. Wang, P. D. Reigosa, and F. Blaabjerg, "A humidity-dependent lifetime derating factor for dc film capacitors," in *2015 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, 2015, pp. 3064–3068.

[29] Y. Zhang, Z. Wang, S. Zhao, F. Blaabjerg, and H. Wang, "Lifetime prediction of the film capacitor based on early degradation information," in *2021 IEEE Applied Power Electronics Conference and Exposition (APEC)*. IEEE, 2021, pp. 407–412.

[30] S. Zhao, S. Chen, and H. Wang, "Degradation modeling for reliability estimation of dc film capacitors subject to humidity acceleration," *Microelectronics Reliability*, vol. 100, p. 113401, 2019.

[31] A. Kabir, C. Bailey, H. Lu, and S. Stoyanov, "A review of data-driven prognostics in power electronics," in *2012 35th International Spring Seminar on Electronics Technology*. IEEE, 2012, pp. 189–192.

[32] M. Tsunoda, S. Amasaki, and A. Monden, "Handling categorical variables in effort estimation," in *Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement*, 2012, pp. 99–102.

[33] Ö. İ. GÜNERİ and B. DURMUŞ, "Dependent dummy variable models: an application of logit, probit and tobit models on survey data," *International Journal of Computational and Experimental Science and Engineering (IJCESEN)*, vol. 6, no. 1, pp. 63–74, 2020.

[34] N. Ekong, I. Moffat, A. Usoro, and I. Matthew, "A comparative study of the impact of dummy variables on regression coefficients and canonical correlation indices: An empirical perspective," *International Journal of Analysis and Applications*, vol. 19, no. 4, pp. 576–586, 2021.

[35] H. Hernandez, "Quantitative analysis of categorical variables," *ForsChem Research Reports*, vol. 6, pp. 2021–04, 2021.

[36] Y. Huang, S. Wang, and J. Yan, "Research on influencing factors of chinese movie microblog marketing effect based on multiple linear regression model:——official microblog of ne zha as an example," in *2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, vol. 1. IEEE, 2020, pp. 79–83.

[37] R. Zhang and X. Lu, "Research on the influencing factors of package storage time in the parcel lockers based on user classification," in *2020 6th International Conference on Information Management (ICIM)*. IEEE, 2020, pp. 215–222.

[38] Environmental testing - Part 2: Tests - Test Cy: Damp heat, steady state, accelerated test primarily intended for components, IEC 60068-2-67, Dec. 1995.

[39] F. Li, S. Li, N. Tang, and T. Denœux, "Constrained interval-valued linear regression model," in *2017 20th International Conference on Information Fusion (Fusion)*. IEEE, 2017, pp. 1–8.

[40] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "A data-driven failure prognostics method based on mixture of gaussians hidden markov models," *IEEE Transactions on reliability*, vol. 61, no. 2, pp. 491–503, 2012.