Evaluation method for moisture content of oil-paper insulation based on segmented frequency domain spectroscopy: From curve fitting to machine learning

Huanmin Yao¹ | Haibao Mu¹ | Ning Ding¹ | Daning Zhang¹ | Zhao Jie Liang² | Jie Tian² | Guanjun Zhang¹

¹ State Key Laboratory of Electrical Insulation and Power Equipment, Xi’an Jiaotong University, 28 Xianning West Road, Xi’an, Shaanxi 710049, People’s Republic of China
² Shenzhen Power Supply Co., Ltd., Shenzhen, Guangdong, People’s Republic of China

Correspondence
Haibao Mu, State Key Laboratory of Electrical Insulation and Power Equipment, College of Electrical Engineering, Xi’an Jiaotong University, 28 Xianning West Road, Xi’an, Shaanxi, 710049, People’s Republic of China. Email: haibaomu@mail.xjtu.edu.cn

Abstract
In recent years, frequency domain spectroscopy (FDS) is often used to evaluate oil paper insulation state in power transformer bushing. But it is still very difficult to evaluate the moisture content accurately and quickly. In order to solve this problem, this paper proposes an intelligent algorithm based on random forest regression (RFR) to construct an efficient evaluation method through segmented FDS curves. Furthermore, the characteristics of FDS curves were studied and the intelligent method was compared with support vector regression (SVR) and deep neural networks (DNN). The results show that the dielectric loss, the real part and imaginary part of complex capacitance all move upward with the moisture increasing, so they can be used as the input feature of the evaluation model; The moisture content evaluation accuracy of the RFR model in the whole frequency band is higher than that of SVR and DNN models; With the increase of lower cut off frequency (FDS test stop frequency), the FDS test time is greatly shortened, and the accuracy of the RFR model can still meet the evaluation requirements. Therefore, the data in a compromise frequency band can be used to evaluate the moisture content of oil paper insulation accurately and quickly.

1 | INTRODUCTION

The key connection component between the power transformer and the grid is bushing. Among them, the capacitive oil-paper bushing plays a vital role in the safe and stable operation of power systems [1–3]. Moisture is one of the most important factors affecting the insulation state of bushing [4, 5]. In paper [6], it is found that the catalytic effect of moisture on cellulose degradation is much more serious than that of oxygen. Also, the increased moisture can cause the dielectric conductivity and the temperate of the oil-paper insulation to raise. Furthermore, the water vaporizes to generate bubbles, resulting in a decrease in the partial discharge inception voltage and breakdown voltage. Each time the moisture content doubles, the life of the insulation will be halved [7].

In recent years, the frequency domain spectroscopy based on dielectric response theory has attracted the attention of many scholars because of its low test voltage, simple wiring, and non-destruction to insulation [8, 9]. However, due to the long test time of low frequency point, the complete FDS test process usually takes several hours. The long test time seriously restricts the schedule of on-site maintenance. Therefore, how to accurately and quickly extract the moisture content information from FDS curves has become a hot topic.

In paper [10], it has been found that the FDS curve integral value $\int_{\tan \delta}^{\text{DC}}$ and DC conductivity $\sigma_{\text{DC}}$ have a non-linear relationship with the moisture content of oil-paper insulation; In paper [11, 12], FDS was used in high frequency band and PDC was used in low frequency band to obtain the response function $C(f)$ of oil-paper insulation of transformer, and the moisture content of oil-paper insulation was evaluated by comparing with database. The above results can be used to evaluate the moisture content of oil-paper insulation, but these methods are mostly empirical formulas based on curve fitting, which
have some problems such as complex calculation process, overfitting, and poor generalization performance.

With the development of artificial intelligence technology, some scholars have begun to use machine learning to evaluate the moisture content of oil-paper insulation equipment. Mousavi et al. [13] used genetic algorithm to find the equivalent circuit parameters of transformer and eliminated the influence of temperature on FDS curve through neural network. Ma et al. [14] used support vector machine (SVM) to evaluate the moisture content of transformers, where the complex permittivity in the FDS curves was the sample attribute, and the moisture content of oil-paper insulation was the mark. Fofana et al. [15] used FDS data to train a supervised neural network with 44 input layer neurons and 2 output layer neurons to evaluate the moisture content, and distinguished the influence of moisture and aging on FDS curve accurately, which is of great significance for evaluating the insulation states of oil-paper. Similarly, a method to diagnose the aging and moisture of bushing is proposed based on multiclass least square support vector machines optimized by cuckoo search algorithm [16]. However, it is usually difficult to find a suitable kernel function for SVM, resulting in poor evaluation performance. DNN also has many problems, such as complex network structure and slow learning speed. Also, these models all use the whole frequency band data to evaluate moisture content, which is not conducive to reducing the FDS test time.

Since the random forest based on the decision tree introduces data disturbance and attribute disturbance, it has a lot of advantages, such as simple structure, no need to do feature selection, fast learning speed, and not easy to fall into local extremum [17–19]. It has been widely used in biology, medicine, information engineering and other fields, and can be used as an intelligent algorithm to identify the moisture content of oil-paper insulation [20, 21].

In order to solve the problem of long time to obtain input data and low accuracy when evaluating the moisture content of oil-paper insulation, this paper used segmented FDS curves as the input data set and selected RFR to build the evaluation model. At the same time, the evaluation accuracy of these three models applied to different lower-cut-off frequencies data were compared, which showed that RFR can evaluate the moisture content of oil-paper insulation accurately and quickly.

## EXPERIMENTS AND RESULTS

### 2.1 Basic theory of frequency-domain spectroscopy

In time domain, the step response of the total current density, $J(t)$, within an insulation system generated by the electric field strength, $E(t)$, is given by:

$$J(t) = \sigma_0 E(t) + \varepsilon_0 \frac{dE(t)}{dt} + \varepsilon_0 \int_0^t f(t - \tau) E(\tau) d\tau$$  \hspace{1cm} (1)

where $\sigma_0$ is the DC conductivity, $\varepsilon_0$ is vacuum dielectric constant.

Taking Laplace transform of Equation (1), which can write

$$f^* (\omega) = \{[\sigma_0 + \varepsilon_0 \chi''(\omega)] + j\omega \varepsilon_0 [1 + \chi'(\omega)]\} E^* (\omega)$$  \hspace{1cm} (2)

where $\chi'(\omega)$ and $\chi''(\omega)$ are the real and imaginary parts of complex susceptibility.

From Equation (2), the complex permittivity is:

$$\varepsilon_r(\omega) = \varepsilon'_r(\omega) - j\omega\varepsilon''_r(\omega) = [1 + \chi'(\omega)] - j \left[\chi''(\omega) + \frac{\sigma_0}{\omega\varepsilon_0}\right]$$  \hspace{1cm} (3)

Hence, the dielectric dissipation factor in frequency domain is:

$$\tan \delta = \frac{C''}{C'} = \frac{\varepsilon''_r(\omega)}{\varepsilon'_r(\omega)} = \frac{\chi''(\omega) + \frac{\sigma_0}{\omega\varepsilon_0}}{1 + \chi'(\omega)}$$  \hspace{1cm} (4)

Losses caused by conductance and polarization are included in $C''$. At the same time, the complex permittivity and the dielectric dissipation factor both depend on the frequency.

### 2.2 Preparation of oil-paper bushing samples

In order to simulate the real structure of field capacitive oil-paper bushing, insulation paper units with thickness of 1.04 mm, length of 600 mm and width of 100 mm were prepared [22], as shown in Figure 1.

In this section, the glassware was wiped clean with alcohol and placed in a vacuum oven. It was dried continuously at 105 °C/100 Pa for five hours to ensure that there was no residual moisture. The insulation paper units were put into the oven for dispersive arrangement and dried at 105 °C/100 Pa for 48 hours. After drying, the moisture content of the paper was measured by Karl Fisher Coulometer KFT831, which was less than 0.5%.

The new 25# Karamay transformer oil was dried at 105 °C/100 Pa with the same process for 72 hours. It should be noted that transformer oil and insulation paper units cannot
be dried in the same oven, because the transformer oil or oil stains would volatilize into oil vapour and be adsorbed on the surface of insulation paper when heated, thereby reducing the exchange ability of insulation paper units with water in air. It not only affects the dryness of insulation paper units, but also is difficult to accurately control the moisture content when naturally absorbing water.

Since cellulose and mineral oil have strong hydrophilicity and hydrophobicity respectively, 97% of the water is mainly stored in the paper, while the water dissolved in the oil is very little. Therefore, only the insulation paper was damped in the experiment, and the transformer oil was kept dry. The paper samples were taken out and placed on the high-precision electronic balance immediately for weighting, and 110 dry insulation paper samples were divided into four groups, each containing 28, 28, 27, 27 samples respectively. These samples with the initial moisture content of 0.41–5.08% were prepared by controlling the time during absorbing water naturally. The ascending gradient of moisture content is 0.2%. The samples with moisture content greater than 6% were obtained by artificial humidification. The bushing in the field is usually not so severely damp. The purpose is to observe whether the FDS curves change rule at high moisture content is consistent with that at low moisture content, and provide more data for the models to improve the evaluation accuracy.

In order to eliminate the experimental error and ensure the accuracy of the measurement results, the Karl Fischer Coulometer KFT831 was used to titrate the moisture content of all samples three times, and the average of these results was taken as the final result.

After that, the prepared samples were placed in an oil-bearing sealed tank for 48 hours, so that the distribution of water was balanced naturally and the bubbles in the insulation paper overflowed.

2.3 FDS test and analysis

Oil-paper insulation samples with different moisture contents were obtained according to the above method. In order to obtain stable test results, the temperature is constant at 40 °C during the test process. Seven groups of samples with moisture content of 0.41%, 1.10%, 2.03%, 2.84%, 3.91%, 5.08% and 6.11% were selected for FDS test by DIRANA frequency domain dielectric response equipment. The test frequency band was 1 mHz–5 kHz and the peak voltage was 200 V. After that, we used the Karl Fischer Coulometer KFT831 to recalibrate the moisture contents. The FDS curves of insulation paper samples with different moisture contents are as follows.

As shown in Figure 2, tanδ–f curves move up with the increase of moisture content. With the moisture content of oil-impregnated paper increasing, the number of polar molecules in the medium increases, resulting in greater polarization loss. The dissociation rate of impurity ions increases with increasing in moisture content, which leads to the increase of carrier concentration, the overall conductivity of the medium and the conductivity loss. At the same time, the polarization establishment speed of oil-paper interface is accelerated and the polarization time constant is reduced. Hence, both the polarization loss in high frequency band and conductivity loss in low frequency band have increased, resulting in an upward trend of tanδ–f curves. The tanδ–f curves with moisture content of 0.41%, 1.10%, 2.03% and 2.84% show a tick shape. For the samples with moisture content of 3.91%, 5.08% and 6.11%, the tanδ–f curves have a slight increase in the low frequency band, while have a significant increase in the middle and high frequency band.

As shown in Figure 3, with the increase of moisture content, the C′–f curves remain consistent in the high frequency band (100 Hz–5 kHz). The C′–f curves with moisture content of 0.41%, 1.10%, 2.03% and 2.84% mainly change when the frequency is less than 10 Hz. The C′–f curves with moisture content of 5.08% and 6.11% change in the mid-range, and rise sharply with the increase of the moisture content.

As shown in Figure 4, the C″–f curves are similar to tanδ–f curves, and the overall trend is upward. The C″–f curves with moisture content of 0.41%, 1.10%, 2.03% and 2.84% show a tick shape in the whole frequency band. The C″–f curves with
moisture content of 3.91%, 5.08% and 6.11% are in a “straight line” shape, and the slope is approximately $-1$ in the frequency band between 1 mHz to 10 Hz. The slope of $C''$–f curves with moisture content of 0.41%, 1.10%, 2.03% and 2.84% is approximately $-1$ in the frequency band between 1 mHz to 1 Hz. The imaginary part of the complex capacitance $C''$ mainly characterizes the conductance loss and polarization loss process of oil-paper insulation. The conductivity loss is mainly reflected in the low frequency band. When the frequency is low enough, the polarization process has enough time to complete. Therefore, the conductivity process in oil-paper insulation is dominant, and the $C''$–f curves in low frequency band are straight lines whose slope is approximately $-1$.

3 | RAPID EVALUATION MODEL

3.1 | Random forest regression

The random forest model obtained 55 training samples through Bootstrap and generated several regression trees \( h(x; \theta_n), n = 1, 2, \ldots, N \), details are as follows\[18, 23\].

\[
h(x) = \frac{1}{Q} \sum_{n=1}^{N} b(x; \theta_n) \tag{5}
\]

where \( Q \) is the number of samples, \( \theta_n \) is an independent and identically distributed random variable, and \( N \) is the number of regression trees.

If \( x_p \) is used as input data, the importance \( I_q \) of \( x_p \) on the tree is:

\[
I_q(x_p) = \frac{\sum_{i=1}^{Q_{OOB}} I(f'(x_{q,i})) - \sum_{i=1}^{Q_{OOB}} I(f''(x_{q,i}))}{Q_{OOB}} \tag{6}
\]

where \( x_{q,i} \) is the data measured at a certain moisture content condition, \( x_q \) is the data outside the package, \( Q_{OOB} \) is the number of samples outside the package, \( f'(x_{q,i}) \) is the nth sample value in the out of package data set, \( f''(x_{q,i}) \) is the estimated value of the nth sample in the out of package data after replacing variables randomly, and \( I \) is the discriminant function.

Hence, the importance of \( x_p \) in the random forest is:

\[
I(x_p) = \frac{\sum_{i=1}^{N} I_q(x_{q,i})}{N} \tag{7}
\]

3.2 | Feature subset selection

In high-dimensional space, problems such as sparse data samples, increased computational complexity, and complex calculations often occur. Selecting some important features as input features is an effective way to solve the “dimension disaster” problem.

Initial feature space was composed of dielectric loss tan\( \delta \), real part of complex permittivity \( \varepsilon' \), imaginary part of complex permittivity \( \varepsilon'' \), real part of complex capacitance \( C' \) and imaginary part of complex capacitance \( C'' \). In order to remove redundant features, the importance of each feature was obtained by using the evaluation criteria based on minimizing the correlation coefficient.

There was a bootstrap sample \( b = 1, 2, \ldots, B \). \( B \) is the number of training samples. We set the initial value of \( b \) to 1, and generated several regression trees \( h(x) \) and out of package data \( I_{OOB} \) according to Equation (5). Then, we used \( h(x) \) to regres the out-of-package data, perturbed the value of the feature \( X_{j,g} \) \( g = 1, 2, \ldots, 5 \), and calculated the importance of each feature according to Equation (7). The results are shown in Figure 5.

The imaginary part of the complex capacitor \( C'' \) has the largest importance, which is 0.368. This is because compared with other curves, the difference between the maximum and minimum values of \( C'' \) at any frequency point is more than two orders of magnitude, and the curves with different moisture contents have the largest discrimination.
Because of the proportional relationship between $C''$ and $\varepsilon''$, the importance of them is relatively close. Therefore, only one of the complex capacitance or permittivity was selected to represent these characteristics. This paper selected $C''$, $C'$ and $\tan\delta$ as the input features and readjusted the data to obtain a new sample data set.

### 3.3 Parameters selection

Parameter is one of the most important factors affecting model performance. This paper used the improved grid search method based on rectangular expansion to find the optimal parameters of RFR, SVR and DNN. The parameters that affect the prediction performance of RFR are the number of decision trees, the maximum number of features and the maximum depth of trees. The parameters selection steps are as follows:

1. The objective function is to make the coefficient of determination $R^2$ as close as possible to 1.

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{a} (\hat{y} - \bar{y})^2}{\sum_{i=1}^{n} (y - \bar{y})^2}$$  

$$= \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$  

(8)

where SST is the total sum of squares, SSR is the sum of squares for regression, $y$ is the measured value of the moisture content of the insulation paper samples, as shown in the testing set in Table 2, its mean value is $\bar{y}$, $\hat{y}$ is the estimated value of the sample moisture content calculated by evaluation models. The coefficient of determination $R^2$ is used to measure the similarity between estimated and measured values. A value close to 1 indicates the best estimate.

2. Since the maximum number of features is 3, it is only necessary to determine the number of decision trees and the maximum depth of trees.

In the first traversal, the value range of the number of decision trees is $e \in [E_{\min}, E_{\max}]$ and the maximum depth of the trees is $f \in [F_{\min}, F_{\max}]$, which form a two-dimensional space area. We divided the area into $S \times T$ grids, then the parameter combination of each grid point is:

$$\begin{align*}
    e &= E_{\min} + (s - 1)(E_{\max} - E_{\min})/S \\
    f &= F_{\min} + (t - 1)(F_{\max} - F_{\min})/T
\end{align*}$$  

(9)

where $s = 1, 2, 3, \ldots, S; t = 1, 2, 3, \ldots, T$.

3. As shown in Figure 6, the objective function value of each grid point was calculated separately, and the rectangular area with the largest objective function value was selected as the new parameter value range. Further, the mesh was redivided into smaller steps, where $a = (E_{\max} - E_{\min})/S$, $b = (F_{\max} - F_{\min})/T$.

After cross-validation, the optimal parameters of the three models are shown in Table 1.

This method can ensure that all possible regions around the approximate optimal combination can be searched, and it is easier to obtain the global optimal solution. Moreover, the selection of the step size is not directly related to the results, which enhances the adaptability of the grid method.

### 3.4 Model building

Dielectric loss $\tan\delta$, complex capacitance real part $C'$ and complex capacitance imaginary part $C''$ of FDS curves at 40 °C/1 mHz–5 kHz were selected as input attributes, and moisture content of oil-paper insulation samples was taken as output mark. There are 55 sets of sample data, and each set contains 17 frequency points. In Section 2.2, 110 samples were divided into four groups. According to the stratified sampling method, in the first and second group, 28 paper samples with moisture content interval close to 0.2% were selected as the training set, and in the third and fourth group, 27 paper samples with little difference in corresponding moisture content from the training set were selected as the testing set. The data are shown in Table 2.
YAO ET AL.

### TABLE 2 Experimental data

| Order | Moisture content (%) | Training data | Testing data |
|-------|----------------------|---------------|--------------|
| 1     | 0.410                | 0.475         |              |
| 2     | 0.540                | 0.648         |              |
| 3     | 0.755                | 0.783         |              |
| 4     | 0.810                | 0.945         |              |
| 5     | 1.080                | 1.090         |              |
| 6     | 1.100                | 1.308         |              |
| 7     | 1.515                | 1.540         |              |
| 8     | 1.565                | 1.758         |              |
| 9     | 1.950                | 1.990         |              |
| 10    | 2.030                | 2.233         |              |
| 11    | 2.435                | 2.520         |              |
| 12    | 2.605                | 2.723         |              |
| 13    | 2.840                | 3.050         |              |
| 14    | 3.260                | 3.317         |              |
| 15    | 3.375                | 3.608         |              |
| 16    | 3.840                | 3.875         |              |
| 17    | 3.910                | 4.165         |              |
| 18    | 4.420                | 4.458         |              |
| 19    | 4.495                | 4.783         |              |
| 20    | 5.070                | 5.075         |              |
| 21    | 5.080                | 5.338         |              |
| 22    | 5.595                | 5.658         |              |
| 23    | 5.720                | 5.840         |              |
| 24    | 5.960                | 6.035         |              |
| 25    | 6.110                | 6.155         |              |
| 26    | 6.200                | 6.355         |              |
| 27    | 6.510                | 6.665         |              |
| 28    | 6.820                | /             |              |

3. For each node of 600 decision trees, a subset containing two attributes was selected randomly from the attribute set, and then the optimal attribute was selected from this subset. The optimal partition attribute was obtained by the Gini index, and the expression is as follows:

\[
\text{Gini} = 1 - \sum_{k=1}^{|y|} p_k^2
\]

where \( p_k \) (\( k = 1, 2, \ldots, |y| \)) is the proportion of class \( k \) in the sample set. The regression results of 600 decision trees are arithmetic averaged to get the final model output.

4. There are several data points under each moisture content, the predicted results are averaged and compared with the real value. We used the coefficient of determination \( R^2 \) to measure the similarity between estimated and measured values, which is shown in Equation (8).

### 4 QUANTITATIVE EVALUATION OF MOISTURE CONTENT

In order to evaluate the moisture content of oil-paper insulation more quickly, this paper used RFR to identify the moisture content of FDS curves with different lower-cut-off frequencies, and compared the results with SVR and DNN models. The divided frequency bands are shown in Figure 8.

The data to be tested contain five different frequency bands, which are 1 mHz–5 kHz, 0.01 Hz–5 kHz, 0.1 Hz–5 kHz, 1 Hz–5 kHz, 10 Hz–5 kHz. For the convenience of expression, they are numbered 1–5. As a standard practice, at least two cycles of the sinusoidal test voltage are recorded to accurately ascertain the amplitude and phase differences between the voltage and current signals. On the other hand, under field conditions, testers usually determine the test time according to the trend and amplitude of the initial stage (100 mHz–5 kHz) curves of the FDS. When the initial amplitude is small, the measurement is stopped at a higher frequency, otherwise, the full frequency test is carried out. Here we only considered the condition of laboratory test. Therefore, the time required to obtain the data of these frequency bands are 51, 23, 5, 1 and 1 min respectively.

For the above data of five frequency bands, the evaluation accuracy and time of the RFR, SVR and DNN models are shown in Table 3.

The model building process is shown in Figure 7.

The construction process of the RFR is more complicated than that of the SVR and DNN. The characteristics of the three models have been marked in the flowchart. Among them, RFR introduces data disturbance and attribute disturbance, which can enhance the generalization performance of the model and prevent overfitting. At the same time, by combining the results of multiple decision trees, the evaluation accuracy of the model can be improved.

The RFR model building process is as follows.

1. The data sets with the same number of the training samples were extracted by using Bootstrap sampling method.
2. The optimal parameters of the model were obtained by using the improved grid search method based on rectangular expansion. And the generalization performance of the learner was evaluated through tenfold cross-validation.

The model building process is shown in Figure 7.
FIGURE 7  Flow chart
training time depends on the complexity of each decision tree. The RFR model constructed by us contains 600 decision trees, and the training data contains 28 instances and 3 attributes, so the training time is about 5.3 s.

In order to further analyse the evaluation accuracy of the model, the accuracy comparison of each frequency band with $R^2$ as the index is shown in Figure 9.

The predicted results with the frequency band from 1 mHz to 5 kHz are that RFR has the highest evaluation accuracy, which is 0.9137; the $R^2$ of SVR model is 0.8969; the $R^2$ of DNN model is 0.9012.

When the frequency band of test data is 0.01 Hz–5 KHz, the $R^2$ of RFR, SVR, DNN are 0.8969, 0.8718, 0.9012 respectively. Compared with the results from 1mHz to 5 kHz, the $R^2$ of RFR is only reduced by 0.0419. In addition, the acquisition time of 0.01 Hz–5 kHz data is shorter than that of the whole frequency band data by 28 min. On the basis of sacrificing the evaluation accuracy properly, the speed of model evaluation is greatly improved. The moisture content of oil-paper insulation can still be identified with high accuracy.

At the same time, with the increase of lower-cut-off frequency, the number of input sample data is gradually reduced, resulting in a continuous decrease in the accuracy of the model. When the lower-cut-off frequency is greater than 0.1 Hz, the accuracy of DNN is higher than that of SVR and RFR, which is due to the small amount of data, resulting in over-fitting of all, so the evaluation accuracy is less reliable in these two frequency bands. Although the test time has been reduced accordingly, it is impossible to evaluate the moisture content of oil-paper accurately.

Therefore, the data in 0.01 Hz–5 kHz frequency band can be used to evaluate the moisture content accurately and quickly, while the data in other frequency bands cannot be used due to their low accuracy.

As shown in Figure 10, the evaluation accuracy of RFR from 0.01 Hz to 5 kHz is higher than that of SVR and DNN. This is due to the fact that using a frequency band containing fewer frequency points obtains data less than using the whole frequency band. It is equivalent to the missing value in the input data, which will make the model unable to get accurate results.

### TABLE 3  The evaluation accuracy of three models

| Frequency band number | SVR    | RFR    | DNN    |
|-----------------------|--------|--------|--------|
| $R^2$                 |        |        |        |
| 1                     | 0.8969 | 0.9137 | 0.9012 |
| 2                     | 0.8581 | 0.8718 | 0.8305 |
| 3                     | 0.7796 | 0.8201 | 0.7669 |
| 4                     | 0.6749 | 0.7232 | 0.7838 |
| 5                     | 0.5363 | 0.5945 | 0.6625 |

| MSE                   |        |        |        |
|-----------------------|--------|--------|--------|
| 1                     | 0.4049 | 0.3458 | 0.3643 |
| 2                     | 0.5577 | 0.5045 | 0.7171 |
| 3                     | 0.8663 | 0.7154 | 0.9116 |
| 4                     | 1.2777 | 1.0730 | 0.8908 |
| 5                     | 1.8228 | 1.5889 | 1.2998 |

| RMSE                  |        |        |        |
|-----------------------|--------|--------|--------|
| 1                     | 0.6563 | 0.5880 | 0.6036 |
| 2                     | 0.7468 | 0.7103 | 0.8468 |
| 3                     | 0.9308 | 0.8458 | 0.9548 |
| 4                     | 1.1304 | 1.0359 | 0.9438 |
| 5                     | 1.3501 | 1.2605 | 1.1401 |

| MAE                   |        |        |        |
|-----------------------|--------|--------|--------|
| 1                     | 0.5198 | 0.4513 | 0.4672 |
| 2                     | 0.6427 | 0.5326 | 0.7270 |
| 3                     | 0.8108 | 0.6202 | 0.8120 |
| 4                     | 0.9690 | 0.7731 | 0.7593 |
| 5                     | 1.1170 | 0.9406 | 0.8979 |

| Time (s)              |        |        |        |
|-----------------------|--------|--------|--------|
| 1                     | 0.6882 | 5.4601 | 41.740 |
| 2                     | 0.6789 | 5.3930 | 41.145 |
| 3                     | 0.6770 | 5.4024 | 41.779 |
| 4                     | 0.6691 | 5.3447 | 41.370 |
| 5                     | 0.6625 | 5.3846 | 41.204 |
However, the RFR model is an ensemble learning method based on decision trees. Since the decision tree can divide samples with missing attribute values into different sub-nodes with different probabilities, RFR is less sensitive to data loss. On the contrary, SVR and DNN do not have strategy to deal with missing samples, resulting in lower evaluation accuracy than RFR.

At the same time, these three models all have higher evaluation values for samples with low moisture content. This is because the polarization process at low moisture content is weak, and the change degree of $\tan \delta$, $C'$, $C''$ are small. The $\tan \delta$ under low moisture content has little difference with higher moisture content, which leads to the overall higher evaluation results of these models.

5 | CONCLUSION

In this paper, FDS curves of oil-paper insulation samples with different moisture contents were studied. And a new method for evaluating the moisture content based on RFR was proposed.

The experiments show that the $\tan \delta/f$ curves move up with the increase of moisture content; the influence of moisture content on the real part of complex capacitance $C'$ is mainly reflected in the low frequency band; the imaginary part of complex capacitance $C''$ increases with the increase of moisture content in the whole frequency band.

Among the three moisture evaluation models, the RFR has the highest accuracy in evaluating the moisture content represented by the FDS curves at different lower-cut-off frequencies. Moreover, when the frequency band of the model input data is 0.01 Hz–5 KHz, the evaluation accuracy of RFR model does not decline significantly, but the test data acquisition time is reduced by 28 min. Thus, we can use the frequency band containing fewer frequency points to solve the problem of long time to obtain input data in the moisture content evaluation process, and provide a new method to measure the moisture content of oil-paper bushing on site accurately and quickly.

ACKNOWLEDGMENTS

This work was supported by the project (090000KK58190036). The authors thank for the experimental instruments provided by Shenzhen Power Supply Co., Ltd. The authors also thank the anonymous editor and the reviewers’ accurate and valuable comments on their paper.

ORCID

Haibao Mu https://orcid.org/0000-0002-8790-6336

REFERENCES

1. Liao, W., et al.: A diagnostic method for moisture intrusion fault in oil bushing. IEEE Trans. Instrum. Meas. 69(9), 7072–7081 (2020).
2. Buchacz, J., Cichon, A., Skubis, J.: Detection of conductive layers short circuit in Hv condenser bushings using frequency domain spectroscopy. IEEE Trans. Dielectr. Electr. Insul. 24(1), 552–558 (2017).
3. Wang, D., et al.: Moisture estimation for oil-immersed bushing based on fds method. At a reference temperature. IET Gener. Transm. Distrih. 12(10), 2480–2486 (2018).
4. Liao, R., et al.: Understanding and analysis on frequency dielectric parameter for quantitative diagnosis of moisture content in paper–oil insulation system. IET Electr. Power Appl. 9(3), 213–222 (2015).
5. Smith, Dj., et al.: A dielectric frequency response model to evaluate the moisture content within an oil impregnated paper condenser bushing. IET Sci. Meas. Technol. 7(4), 223–231 (2013).
6. Fofana, I., Borsi, H., Gockenbach, E.: Results on aging of cellulose paper under selective conditions. In: 2001 Annual Report Conference on Electrical Insulation and Dielectric Phenomena (Cat. No.01CH37225). Kitchener, Ontario, pp. 205–208 (2001).
7. Manuel, A.V.R., Vanessa, M.L.J.: Bushing failure in power transformers and the influence of moisture with the spectroscopy test. Eng. Fail. Anal. 94, 300–312 (2018).
8. Fofana, I., et al.: On the frequency domain dielectric response of oil-paper insulation at low temperatures. IEEE Trans. Dielectr. Electr. Insul. 17(3), 799–807 (2010).
9. Dutta, S., et al.: Effect of measurement temperature on power transformer insulation diagnosis using frequency-domain spectroscopy. IET Sci. Meas. Technol. 11(6), 773–779 (2017).
10. Liao, R., et al.: Quantitative diagnosis of moisture content in oil-paper condenser bushing insulation based on frequency domain spectroscopy and polarization and depolarization current. IET Gener. Transm. Distrih. 11(6), 1420–1426 (2017).
11. Koch, M., Kruger, M.: A fast and reliable dielectric diagnostic method to determine moisture in power transformers. In: International Conference on Condition Monitoring & Diagnosis. Beijing, pp. 467–470 (2008).
12. Koch, M., Krueger, M., Tenbohlen, S.: On-site methods for reliable moisture determination in power transformers. In: IEEE PEDS TeD 2010. New Orleans, LA, pp. 1–6 (2010).
13. Mousavi, S.A., et al.: Artificial neural network based method for temperature correction in Fds measurement of transformer insulation. J. Phys. D: Appl. Phys. 53(14), 145103 (2020).
14. Ma, H., Saha, T.K., Ekanayake, C.: Machine learning techniques for power transformer insulation diagnosis. In: AUPEC 2011. Brisbane, pp. 1–6 (2011).
15. Betie, A., et al.: Neural network approach to separate aging and moisture from the dielectric response of oil impregnated paper insulation. IEEE Trans. Dielectr. Electr. Insul. 22(4), 2176–2184 (2015).
16. Wang, D., et al.: Insulation defect diagnostic method for Oil bushing based on multiclass LS-SVM and cuckoo search. IEEE Trans. Instrum. Meas. 69(1), 163–172 (2019).
17. Fernandez-Delgado, M., et al.: Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15, 3133–3181 (2014).
18. Ham, J., et al.: Investigation of the random forest framework for classification of hyperspectral data. IEEE Trans. Geosci. Remote Sens. 43(3), 492–501 (2005).
19. Senoussaoui, M.E.A., Brahami, M., Fofana, I.: Combining and comparing various machine-learning algorithms to improve dissolved gas analysis interpretation. IET Gener. Transm. Distrib. 12(15), 3673–3679 (2018)

20. Shah, A.M., Bhalja, B.R.: Fault discrimination scheme for power transformer using random forest technique. IET Gener. Transm. Distrib. 10(6), 1431–1439 (2016)

21. Jeong, M., Nam, J., Ko, B.C.: Lightweight multilayer random forests for monitoring driver emotional status. IEEE Access 8, 60344–60354 (2020)

22. Zhang, D., et al.: Study on Fds characteristics of oil-immersed paper insulation bushing under non-uniform moisture content. IET Sci. Meas. Technol. 12(5), 691–697 (2018)

23. Rastgou, M., et al.: Estimating the soil water retention curve: Comparison of multiple nonlinear regression approach and random forest data mining technique. Comput. Electron. Agric. 174, 105502 (2020)

How to cite this article: Yao, H., et al.: Evaluation method for moisture content of oil-paper insulation based on segmented frequency domain spectroscopy: From curve fitting to machine learning. IET Sci. Meas. Technol. 15, 517–526 (2021).
https://doi.org/10.1049/smt2.12052