Health index prediction of dissolved gases in transformer oil based on statistical distribution model

Xiaogang Li\textsuperscript{1,a}, Jiajia Liu\textsuperscript{1,b}, Xiaoguang Wang\textsuperscript{1,c}, Zhongyuan Wu\textsuperscript{1,d} and Chenyao Liu\textsuperscript{2}\textsuperscript{*}

\textsuperscript{1}Henan Jiuyu EPRI Electric Power Technology Co.,Ltd, Zhengzhou, Henan, China.
\textsuperscript{2}Guangxi Key Laboratory of Power System Optimization and Energy Technology, Nanning, Guangxi, China.
\textsuperscript{a}email: lxg622@163.com, \textsuperscript{b}email: as219860516@163.com, \textsuperscript{c}email: 13523088703@163.com, \textsuperscript{d}email: lxg622@163.com
\textsuperscript{*}Corresponding author’s email: lcy1421129@st.gxu.edu.cn

Abstract. In this paper, dissolved gas analysis (DGA) and statistical distribution model (SDM) were used to predict the health index (HI) of dissolved gas in transformer oil. First, the individual DGA data are classified according to transformer ages ranging from 1 to 4 years. Then, representative fitting models were selected and extrapolated from 5 to 25 years. The inverse cumulative distribution function (ICDF) of the selected distribution model was used to calculate the single conditional parameter data from 5 to 25 years. Finally, the traditional scoring method is used to estimate the future HI value. The results show that DGA parameters can be expressed by exponential equation based on statistical model. The predicted values of DGA health index of transformer oil from 1 to 7 years were basically consistent with the calculated values, and the DGA score was 100 points. By the 20th year, the DGA score had dropped to 75, requiring timely monitoring. The research results can provide powerful data support and theoretical reference for transformer life prediction.

1. Introduction

Power transformer is one of the expensive and key equipment in distribution system\cite{1}. Without appropriate operational intervention, this can result in power delivery failures, significant repair or replacement costs, and significant revenue losses for utilities due to the long lead time for mitigation actions. In recent years, state-based monitoring (CBM) has been paid more attention to in-service reliable operation\cite{2}. The health of the transformer can be evaluated by monitoring the performance parameters\cite{3}.

A health index (HI) is a known method of transforming multiple state information in a service into a target and a single computable index to provide the assets of a contained health transformer\cite{4}. At present, the existing transformer monitoring devices at home and abroad are generally special monitoring devices, which can only monitor the characteristic parameters in the transformer oil. This device has a single function, and can not realize the integration of multiple characteristic parameters to comprehensively evaluate the insulation state of the transformer, and it is difficult to make an accurate assessment of the operation state of the transformer\cite{5}.

The aim of this paper is to predict the health index of dissolved gases in transformer oil. In this paper, a statistical distribution model (SDM) is proposed to estimate the future HI values of dissolved gases in
distribution transformer population oil. Firstly, the representative distribution model of transformer parameters is identified. Next, SDM processing is carried out on the working condition parameter data, and finally, the HI value of dissolved gas in transformer oil in the future is estimated.

2. Method

2.1. Transformer operating condition index estimation model
SDM is used to estimate the future HI values of dissolved gases in transformer oil with limited historical parameter data. The overall framework for estimating the upcoming dissolved gas HI in transformer oil using the data parameters of individual working conditions is shown in figure 1.

![Technical flow chart of transformer health index life prediction based on statistical distribution model.](image)

The input parameter data in this paper are dissolved gases, including hydrogen (H2), methane (CH4), ethylene (C2H4), ethyne (C2H2), ethane (C2H6), carbon monoxide (CO), and carbon dioxide (CO2). This study does not consider any abnormal data generated by abnormal operating environment and electromagnetic interference during use.

In this study, two-parameter Weibull and normal distribution models are considered. Weibull distribution and normal distribution are widely used in industrial applications, especially in the field of reliability evaluation engineering. The probability density function (PDF) and cumulative distribution function (CDF) of the Weibull distribution model are shown in equations (1) and (2) below.

\[ f(t) = \frac{\beta}{\alpha} \left( \frac{t}{\alpha} \right)^{\beta-1} \exp \left( -\frac{t}{\alpha} \right)^\beta \]  

\[ F(t) = \int_0^t f(t) = 1 - \exp \left( -\frac{t}{\alpha} \right)^\beta \]
The probability density function (PDF) and cumulative distribution function (CDF) of the normal distribution model are shown in equations (3) and (4) below.

\[
f(t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] \tag{3}
\]

\[
F(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] dy \tag{4}
\]

2.2. Estimated distribution parameter

In this paper, the maximum likelihood estimation (MLE) method is used to estimate the population parameters of a distribution. This is because MLE is a parsing maximization process that applies to any form of data.

Therefore, for Weibull distribution, the \((\hat{\alpha}, \hat{\beta})\) MLE estimation of \((\alpha, \beta)\) solution can be obtained as follows (5) and (6).

\[
\frac{n}{\hat{\alpha}} = \sum_{i=1}^{n} x_i^{\hat{\beta}} \tag{5}
\]

\[
\frac{n}{\hat{\beta}} = \sum_{i=1}^{n} \ln x_i = \hat{\alpha} \sum_{i=1}^{n} x_i^{\hat{\beta}} \ln x_i \tag{6}
\]

Similarly, for normal distribution, the \((\hat{\alpha}, \hat{\beta})\) MLE estimation of \((\alpha, \beta)\) solution can be obtained as follows (7) and (8).

\[
\hat{\mu} = \hat{\alpha} \tag{7}
\]

\[
\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2} \tag{8}
\]

2.3. State data estimation

The conditional parameter data represented by the Weibull distribution model is calculated at the appropriate place in the sample. The data were evaluated by probabilistic values using ICDF with scale parameter \(\alpha\) and shape parameter \(\beta\). ICDF of Weibull distribution model can be expressed by Equation (9) below.

\[
x = F^{-1}(\alpha, \beta) = -\alpha \left[ \ln(1 - p) \right]^{1/\beta} I_{[0,1]}(p) \tag{9}
\]

Similarly, ICDF of the normal distribution model can be expressed by equation (10) below.

\[
x = F^{-1}(\mu, \sigma) = \{x : F(\mu, \sigma) = p\} \tag{10}
\]

3. Simulation

3.1. A statistical distribution model is applied to CBM data

The conditional data used in this study were obtained from oil samples of distribution transformers, with data ages ranging from 1 to 4 years. Prioritize, condition parameter data and draw on the x axis, the X-axis represents the variables interval, y axis weibull or normal cumulative probability distribution (CDF), red dot line according to the analysis data of mean and variance to build normal or weibull distribution line, blue point for the analysis of sample points, the closer it gets to the straight line shows the sample normal feature, the better, figure 2 is the normal cumulative distribution probability diagram of H2. According to the figure, the sample distribution is highly fitted to the line, so the normal distribution model can be selected for the data of H2.
Figure 2. Normal distribution probability diagram of H₂ (Figure a is year 1, Figure b is year 2, Figure c is year 3, and Figure d is year 4).

3.2. Estimation of distribution parameters and state data

The distribution parameters of normal and Weibull were calculated according to (5) ~ (8), and then the curve fitting process of optimal algorithm was used to fit and extrapolate the distribution parameters of 5–25 years. Then the estimated distribution parameters in (9) and (10) are calculated by ICDF to verify the single conditional parameter data in the next 10 years. The predicted values of individual conditional parameters for H₂, CH₄, CO, CO₂, C₂H₂, C₂H₄ and C₂H₆ are shown in figure 3 below. The corresponding equations of the principal curve are listed in table 1.

| Parameter | Distribution model | curve equation | R²     |
|-----------|-------------------|----------------|--------|
| H₂        | Normal            | y=3.617exp(0.04x) | 0.9498 |
| CH₄       | Weibull           | y=1.948exp(0.2319x) | 0.9423 |
| CO        | Normal            | y=73.74exp(0.1372x) | 0.9876 |
| CO₂       | Weibull           | y=462.7exp(0.06087x) | 0.9792 |
| C₂H₄      | Weibull           | y=0.5047exp(0.01301x) | 0.9963 |
| C₂H₆      | Normal            | y=0.4772exp(0.01287x) | 0.9963 |
Figure 3. Comparison of DGA calculation results and prediction results (Figure a is H$_2$, Figure b is CH$_4$, Figure c is CO$_2$, Figure d is CO, Figure e is C$_2$H$_6$ Figure f is C$_2$H$_4$).

3.3. Health index prediction and verification

Next, according to the main curve equation of each parameter in DGA obtained in the previous section, the HI value of dissolved gas in transformer oil for 5 ~ 25 years is predicted by using the traditional scoring algorithm and equations (11) and (12). The results show that the predicted HI value of dissolved gas in transformer oil is basically consistent with the calculated HI value. The transformer DGA health index prediction diagram is shown in Figure 4.

\[
SWHI = \sum_{i=1}^{n} S_i \cdot W_i
\]  

(11)

\[
SF_j = \sum_{i=1}^{n} 4 \times W_j
\]

(12)

In Formula (11), $S_i$ represents the score of each parameter in DGA, $W_i$ represents the weight of each parameter in DGA, In Formula (12), $SF_j$ represents the score of DGA, and $W_j$ represents the weight of DGA.
4. Conclusion
In this study, dissolved gas analysis (DGA) and statistical distribution model (SDM) are used to predict health index (HI) of the dissolved gases in transformer oil. Firstly, the single condition parameter data of dissolved gas in transformer oil are calculated. Finally, the future HI value is estimated by traditional scoring method. The results showed that the predicted values of DGA health index of transformer oil from 1 to 7 years were basically consistent with the calculated values, and the DGA score was 100 points. By the 20th year, the DGA score had dropped to 75, requiring timely monitoring. The research results can provide strong data support and theoretical reference for transformer life prediction.

Acknowledgments
This work was supported in part by the National Natural Science Foundation of China under Grant 51907034, and in part by the Specific Research Project of Guangxi for Research Bases and Talents under Grant 2020AC19010.

References
[1] A. Azmi, J. Jasni, N. Azis, M.Z.A.A. Kadir, Evolution of transformer health index in the form of mathematical equation, Renew. Sustain. Energy Rev. 76: 687–700.
[2] R. Murugan, R. Ramasamy, (2019) Understanding the power transformer component failures for health index-based maintenance planning in electric utilities, Eng. Fail. Anal. 96: 274–288.
[3] S. Li, H. Ma, T. Saha, G. Wu, (2018) Bayesian information fusion for probabilistic health index of power transformer, IET Gener. Transm. Distrib. 12: 279–287.
[4] S.S.M. Ghoneim, I.B.M. Taha, (2020) Comparative study of full and reduced feature scenarios for health index computation of power transformers, IEEE Access. 8: 181326–181339.
[5] W.R. Tamma, R.A. Prasojo, (2021) High voltage power transformer condition assessment considering the health index value and its decreasing rate, High Volt. 6: 314–327.