Failure Rate Estimation for Transformer Population based on Health Index through Markov Model Approach

Nor Shafiqin Shariffuddin\textsuperscript{1,2}, Norhafiz Azis\textsuperscript{1,3,*}, Amran Mohd Selva\textsuperscript{1}, Muhammad Sharil Yahaya\textsuperscript{4}, Jasronita Jasni\textsuperscript{1}, Mohd Zainal Abidin Ab Kadir\textsuperscript{1} and Mohd Aizam Talib\textsuperscript{5}

\textsuperscript{1}Advanced Lightning, Power and Energy Research Centre (ALPER), Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
\textsuperscript{2}Electrical Technology Section, Universiti Kuala Lumpur British Malaysian Institute, 53100 Gombak, Selangor Malaysia
\textsuperscript{3}Institute of Advanced Technology (ITMA), Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
\textsuperscript{4}Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
\textsuperscript{5}TNB Research Sdn. Bhd., No. 1, Lorong Ayer Itam, Kawasan Institut Penyelidikan, 43000 Kajang, Selangor, Malaysia

ABSTRACT

This work examines the failure rate of the transformer population through the application of the Markov Model (MM) and Health Index (HI). Overall, the condition parameters data extracted from 3,192 oil samples were analysed in this study. The samples were from 370 transformers with the age range between 1 and 25 years. First, both HIs and failure rates of transformers were determined based on the condition parameters data of the oil samples known as Oil Quality Analysis (OQA), Dissolved Gas Analysis (DGA), Furanic Compounds Analysis (FCA) and age. A two-parameter exponential function model was applied to represent the relationship between the HI and failure rate. Once the failure rate state was obtained, the non-linear optimisation was used to determine the transition probability for each age band. Next, the future failure rate of the transformer population was computed through the MM prediction model. The goodness-of-fit test and Mean Absolute Percentage Error (MAPE) were utilised to determine the performance of the predicted failure rate. The current study reveals that the future state of the transformer population and failure rate could
be predicted through MM based on updated transition probabilities. It is observed that the MAPE between predicted and computed failure rates is 7.3%.

**Keywords:** Condition parameters data, failure rate, health index, Markov model, transformer

**INTRODUCTION**

One of the most crucial components in the electrical power network is a transformer. It can be subjected to various types of stresses during its operational lifetime. These events could lead to degradation, whereby without mitigation, it could lead to irreparable damages. As the population of in-service transformers are getting old, much attention should be highlighted to its reliability (CIGRE WG A2.18, 2003; Emsley et al., 2000; Lundgaard et al., 2004). Condition assessment is essential to optimise the transformer lifetime by an active maintenance and replacement strategies.

Most utilities have implemented Condition-based management to evaluate the health of transformers. This approach could increase the efficiency of asset management practices, which subsequently reduces the related costs through advanced analysis that acts as a preventive measure against underlying failures (Jahromi et al., 2009). Health Index (HI) is a concept under condition-based management whereby a single measurable index is proposed to classify the condition of a transformer. HI considers multiple condition parameters data and utilises various criteria to determine the ageing condition of transformers, which may not be assessable through individual measurement techniques. HI is computed through a scoring method whereby the condition parameters data are classified through weighting and ranking approaches. Based on HI, the condition is graded into several categories defined by Naderian et al. (2008).

Apart from HI, the failure rate is also one of the important aspects that can be used to analyse the reliability, which can be used to drive the optimal maintenance planning (Jürgensen et al., 2016a). Conventionally, the failure rate can be used to evaluate the effectiveness of maintenance strategies (Jürgensen, 2018). However, due to the long lifetimes of transformers, in-service failures data are quite difficult to be obtained (Zaidey et al., 2015). Many studies such as the proportional hazard model, Bayesian updating scheme and linear interpolation between different inspections outcomes are proposed to estimate the failures data based on the available information (Brown et al., 2018; Jürgensen et al., 2016a; Lindquist et al., 2005). Among the difficulties for failure rate modelling is the limitation of historical failures record (Jürgensen et al., 2018; Lindquist et al., 2005). One of the unique approaches to determine failure rate can be carried out based on the condition of the assets (Jürgensen, 2016). Currently, the study to predict the failure rate of the transformers is still limited. Other studies by Jürgensen (2016) and Jürgensen et al. (2016b) mainly focus on
single time condition parameters data whereby statistical data-driven approaches are used to model the failure rate. Markov model (MM) has been identified as one of the methods used to predict the failure rate of transformers.

MM is a stochastic model whereby the conditional distribution of future states of the process given present and previous states depends only on the present state (memoryless property) (Chatfield, 1973). In principle, MM is used as a model to predict the performance of assets by identifying discrete conditions through utilisation of transition state probability over multiple discrete time intervals, where \( P_{ij} \) is the probability of a failure rate to transit from state \( i \) to state \( j \) within a certain year of interval (McDonald, 2004; Borovkov, 2003). MM is widely implemented in civil engineering for various applications (Edirisinghe et al., 2015; Setunge & Hasan, 2011; Agrawal et al., 2006; Camahan et al., 1987; Micevski et al., 2002). It is also utilised in electrical equipment such as switchgear and transformers (Hoskins et al., 1999; Hamoud, 2011; Hamoud, 2012). Recently, many studies have been carried out to employ the MM to determine the transition probabilities of the transformer population for condition state prediction (Selva et al., 2018; Yahaya et al., 2017).

In this paper, the estimation of the future failure rate for the transformer population-based on HI is carried out by MM. The contribution of this work is the utilisation of MM and HI to predict the future failure rate of the transformer population. First, the HI s and failure rates are computed. Next, the transition probability is obtained, and subsequently, the failure rate prediction is carried out. Finally, the predicted failure rate is compared with the computed failure rate.

METHODS
The model was developed based on oil samples data of 33/11 kV distribution transformers with ages ranging from 1 to 25 years retrieval from one of Malaysia’s utilities (Company A, personal communication, November 13, 2016). Oil Quality Analysis (OQA) parameters used in the current study were AC breakdown voltage, acidity, moisture in oil, colour and interfacial tension. The Dissolved Gas Analysis (DGA) included hydrogen (H\(_2\)), carbon monoxide (CO), carbon dioxide (CO\(_2\)), methane (CH\(_4\)), ethane (C\(_2\)H\(_6\)), ethylene (C\(_2\)H\(_4\)) and acetylene (C\(_2\)H\(_2\)). The Furanic Compound Analysis (FCA) consisted of 2-furaldehyde (2FAL), while age was also considered in this case study. The condition parameters data were used to calculate both HI and Individual Failure Rate (IFR). The workflow of this study can be seen in Figure 1.

The HI was computed through the scoring and weighting method utilising the corresponding scales given by Jahromi et al. (2009) and Naderian et al. (2008). Both oil quality and dissolved gasses factors were determined based on Equation 1 (Jahromi et al., 2009; Naderian et al., 2008):
Figure 1. Workflow of the research

\[ DGF = OQA = \frac{\sum_{i=1}^{n} S_j \times W_j}{\sum_{i=1}^{n} W_j} \]  

where \( W_j \) represents the parameter’s weighting factor, \( n \) represents the number of parameters in each factor, and \( S_j \) represents the parameter’s score. All parameters rating codes were identified based on Jahromi et al. (2009); Naderian et al. (2008). The final HI was determined based on Equation 2 (Yahaya et al., 2017):

\[ HI = \frac{K_{DGA}HIF_{DGA} + K_{OQA}HIF_{OQA} + K_{FCA}HIF_{FCA} + K_{AG}HIF_{AG}}{4K_{DGA} + 4K_{OQA} + 4K_{FCA} + 4K_{AG}} \]  

where \( K \) represents the rating assigned to each factor, and \( HIF \) represents each factor’s score.

The failure rate of the individual transformer was determined based on the relative risk method (Jürgensen et al., 2016b). The average failure rate, failure location and confidence level were determined as per Jürgensen et al. (2016b). Based on the condition parameters data of the transformer population, the average condition parameter was computed according to Equation 3 (Jürgensen et al., 2016b):

\[ \theta_{j,N} = \frac{1}{m} \sum_{k=1}^{m} \theta_{j,k} \]
where $m$ represents the number of transformer population, $\theta_{j,N}$ represents the average of all obtained values $\theta_{j,k}$ in each parameter. In this study, all condition parameters data were assumed to correspond to winding failures (Jürgensen et al., 2016b). Therefore, when multiple condition parameters data was associated with winding failures, each parameters’ weighting was required to demonstrate its significance to indicate the condition in one failure location. The assigned scores and weights of all condition parameters data were determined based on Equation 4 (Jahromi et al., 2009; Jürgensen et al., 2016b):

$$W_{jp} = \frac{C_p}{\sum_{i=1}^{n} C_i}$$  \hspace{1cm} [4]

where $C_p$ represents the weight score assigned to each parameter, the 2FAL and age scores were obtained based on Jahromi et al. (2009). Next, the condition parameter factor, $S_{j,k}$ was identified as either a positive or negative linear function (Jürgensen et al., 2016b). Finally, the positive or negative linear function was determined based on Equations 5 and 6 (Jürgensen et al., 2016b):

$$S_{j,k} = \frac{1}{\theta_{j,N}} \theta_{j,k}$$  \hspace{1cm} [5]

$$S_{j,k} = \frac{\theta_{pc,j} - \theta_{j,k}}{\theta_{pc,j} - \theta_{j,N}}$$  \hspace{1cm} [6]

where $\theta_{pc,j}$ is described as a perfect condition or as a new condition. Next, the condition parameter factors of all condition parameters for each of the transformers were computed by Equation 7 (Jürgensen et al., 2016b):

$$S_{j,k} = W_{jp}^T S_{j,p,k}$$  \hspace{1cm} [7]

Finally, the IFR of the transformer was determined based on Equation 8 (Jürgensen et al., 2016b):

$$\lambda_{IFR,k} = \lambda_N \sum_{j=1}^{n} \left( S_{j,k} \rho_j \alpha_i + (1 - \rho_j) \alpha_i \right)$$  \hspace{1cm} [8]

where $\lambda_N$ represents the average failure rate, $S_{j,k}$ represents the condition parameter factor for each transformer, $\alpha_i$ represents failure location, and $\rho_j$ represents the probability of condition parameters.

Once the average HI and failure rate of the individual transformer was obtained, the corresponding relationship was modelled based on a two-parameter exponential function model as shown in Equation 9:

$$y = ab^x$$  \hspace{1cm} [9]
where \( y \) is the failure rate, \( a \) is the initial failure rate, \( b \) is the accumulation factor, and \( x \) is the HI. Based on Equation 9, the failure rate indicator scales and states were determined with reference to the condition of HI from 100% - 0%, corresponding to very good until very poor as defined by Jahromi et al. (2009); Naderian et al. (2008). The failure rate scales and states used for the MM are shown in Table 1.

MM was implemented to predict the future failure rate of the transformer population. Two assumptions were made to simplify the model in this study. Firstly, natural and monotonic distributions were considered for the future failure rate model. Secondly, the probability summation in each of the rows for the MM transition matrix was set to one. In total, five \( P_{ij} \) terms were required to formulate the transition matrix of the MM as seen in Equation 10:

\[
\begin{bmatrix}
P_{11} & 1 - P_{11} & 0 & 0 & 0 \\
0 & P_{22} & 1 - P_{22} & 0 & 0 \\
0 & 0 & P_{33} & 1 - P_{33} & 0 \\
0 & 0 & 0 & P_{44} & 1 - P_{44} \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]  

[10]

The final state, \( P_{55} \), was set to 1 to simplify the MM process, assuming that all transformers would reach a deplorable state. As represented in Equation 11, the nonlinear optimisation technique was implemented to determine the transition probabilities matrix, which could minimise the overall difference between computed and predicted failure rates for every age zone (Selva et al., 2018; Yahaya et al., 2017).

\[
\min \sum_{t=1}^{N} |A(t) - B(t,P) |
\]  

[11]

where \( N \) represents the number of years in each zone, \( P \) represents the transition probabilities, \( P_{11}, P_{22}, P_{33}, P_{44}, A(t) \) represents the computed failure rate at time \( t \), and \( B(t,P) \) represents the predicted failure rate by MM at time \( t \). Next, the prediction of the future failure rate state in a year, \( t \) can be represented by Equation 12:

\[ F_t = F_0 \times P^t \]  

[12]

where \( t \) represents the interval number, \( F_0 \) represents the initial state, and \( P \) represents the transition probability matrix. In this study, the \( F_0 \) was set as \( F_0 = [10000] \) based on the
assumption for an initial state for a newly installed transformer. Finally, the future failure rate of the transformer population at \( n \) year was computed based on Equation 13:

\[
F_{n+1} = F_n \times P \times R^T
\]  

[13]

where \( F_{n+1} \) is the next failure rate at the specific interval, \( F_n \) is the current failure rate, and \( R^T \) is the matrix transform of the failure rate state scales where \( R = \begin{bmatrix} 1.59 & 1.99 & 2.69 & 3.59 & 5.60 \end{bmatrix} \) obtained from Table 1.

RESULTS AND DISCUSSION

The distribution of the number of oil samples between 1 and 25 years is shown in Figure 2. It is observed that the highest and lowest numbers of oil samples are distributed at the age of 14 and 25 years, respectively. The computed failure rate of the transformers according to the zone and age is shown in Table 2. It is found that the failure rate increases with age. Once the transformer population reaches 6 years, the failure rate slightly exceeds the global average failure rate, which is 2%, as reported in Bossi et al. (1983). However, as the age increases from 7 to 13 years, the transformer population failure rate decreases slightly than the global average failure rate. Finally, at the age between 14 and 25 years, the transformer population failure rate once again exceeds the global average failure rate.

Next, based on computed failure rate values, the transition probabilities were determined by minimising the summation of absolute differences between the computed and predicted failure rates for each year based on Equation 11. The computed failure rates for zones 1 and 2 were used to determine the transition probability matrix for training and application. For zones 3 to 5, the computed failure rates data were used to verify the

![Figure 2. Distribution of oil sample data versus age](image)
predicted failure rate computed by the MM approach. Based on Equation 11 and Table 2, the transition probability matrix for zone 1 was computed as seen in Equation 14:

\[
P = \begin{bmatrix}
0.9348 & 0.0652 & 0 & 0 & 0 \\
0 & 0.9900 & 0.0100 & 0 & 0 \\
0 & 0 & 0.5493 & 0.4507 & 0 \\
0 & 0 & 0 & 0.9900 & 0.0100 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]  

[14]

Next, the failure rate state probability for years 1 to 5 was determined based on Equation 14 based on the product of the initial state and transition probability matrix. The computational process of failure rate state probability for each year can be seen in Table 3.

The last failure rate state probability computed for year 5 from Table 3 was used as the initial state for zone 2. The computed transition probability matrix for zone 2 is shown in Equation 15:
Table 3

Failure rate state for years 1 to 5

<table>
<thead>
<tr>
<th>Year</th>
<th>MM Process</th>
<th>Failure Rate State Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F₀ x P¹</td>
<td>[0.9348, 0.0652, 0.0000, 0.0000, 0.0000]</td>
</tr>
<tr>
<td>2</td>
<td>F₀ x P²</td>
<td>[0.8739, 0.1254, 0.0007, 0.0000, 0.0000]</td>
</tr>
<tr>
<td>3</td>
<td>F₀ x P³</td>
<td>[0.8170, 0.1811, 0.0016, 0.0003, 0.0000]</td>
</tr>
<tr>
<td>4</td>
<td>F₀ x P⁴</td>
<td>[0.7637, 0.2326, 0.0027, 0.0007, 0.0003]</td>
</tr>
<tr>
<td>5</td>
<td>F₀ x P⁵</td>
<td>[0.7139, 0.2800, 0.0038, 0.0012, 0.0010]</td>
</tr>
</tbody>
</table>

The computed transition probability matrix for zone 2 was used to calculate the failure rate state probability for years 6 to 10 as in Table 4.

Table 4

Failure rate state for years 6 to 10

<table>
<thead>
<tr>
<th>Year</th>
<th>MM Process</th>
<th>Failure Rate State Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>F₀ x P⁶</td>
<td>[0.5365, 0.4546, 0.0066, 0.0006, 0.0016]</td>
</tr>
<tr>
<td>7</td>
<td>F₀ x P⁷</td>
<td>[0.4032, 0.5834, 0.0111, 0.0004, 0.0020]</td>
</tr>
<tr>
<td>8</td>
<td>F₀ x P⁸</td>
<td>[0.3030, 0.6778, 0.0168, 0.0003, 0.0021]</td>
</tr>
<tr>
<td>9</td>
<td>F₀ x P⁹</td>
<td>[0.2277, 0.7463, 0.0234, 0.0003, 0.0023]</td>
</tr>
<tr>
<td>10</td>
<td>F₀ x P¹⁰</td>
<td>[0.1711, 0.7954, 0.0306, 0.0004, 0.0025]</td>
</tr>
</tbody>
</table>

These steps were repeated to determine the future failure rate state probability for zones 3 to 5, where the last failure rate state probability from the previous zone was used as the initial state for the next zone. The initial state for zones 1 to 8 are shown in Table 5.

Table 5

Initial state for zone 1 to 8

<table>
<thead>
<tr>
<th>Zone</th>
<th>Initial State</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.7139</td>
</tr>
<tr>
<td>3</td>
<td>0.1711</td>
</tr>
<tr>
<td>4</td>
<td>0.1627</td>
</tr>
<tr>
<td>5</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>0.0000</td>
</tr>
<tr>
<td>7</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
Figure 3 shows the predicted failure rate computed by MM based on Equation 13 and Table 2 within 40 years. Most of the predicted failure rates are quite close to the computed failure rates, as illustrated in Table 2. The predicted failure rates exhibit slight differences from computed failure rates at 6, 23, 24 and 25 years. The predicted failure rate of the transformer population is quite low during the first 13 years of service, and it is lower than the global average failure rate, which is 2%. The predicted failure rate of the transformer population begins to exceed the global average failure rate at the age of 14 years and remains high until 40 years. Based on the prediction, the high increment failure rate occurs at 23 to 40 years.

![Comparison between computed and predicted failure rates](image)

The Chi-squared test based on Equation 16 was used to evaluate the goodness-of-fit between the predicted and computed failure rates:

$$X^2 = \sum_{i=1}^{k} \frac{(R_i - E_i)^2}{E_i}$$  \[16\]

where $k$ represents the observation number, $E_i$ represents the computed value at $i$ year, $R_i$ represents the predicted value of the $i$ year and $X^2$ represents a Chi-squared distribution coefficient with $k - 1$ degree of freedom. The coefficient $X^2$ is 0.79 where it is lower than the significance value at the region of 0.05 with 24 degrees of freedom which is 36.42 based on the Chi-Square distribution table (Onchiri, 2013). This result indicates that there is no significant difference between the predicted and computed failure rates.

Mean absolute percentage error (MAPE) was used to calculate the difference between computed and predicted failure rates according to Equation 17:

$$MAPE \% = \frac{\sum_{i=1}^{25} \frac{|Y_i - X_i|}{|Y_i|} \times 100\%}{25}$$  \[17\]
where $Y_t$ represents the computed failure rates, $X_t$ represents the predicted failure rates, and $t$ represents the age of the transformer. The MAPE and accuracy level are tabulated in Table 6. The absolute errors between computed and predicted failure rates for each zone are shown in Figure 4. It is found that the highest and lowest percentage of errors in zone 1 is at the age of 4 and 5 years, respectively. For zone 2, the highest and lowest percentage of errors are between 6 and 9 years. The highest and lowest percentage of errors in zone 3 is at 13 and 15 years, respectively. Zone 4 has the highest and lowest percentage of errors at the age of 19 and 20 years. Finally, the highest and lowest percentage of errors in zone 5 is 23 and 22 years. A significant difference in the percentage of errors between computed and predicted failure rates are observed at the age of 23 to 25 years due to the low distribution of oil samples data used in failure rates computation for this zone.

Table 6
Mean absolute percentage error and accuracy level for failure rate

<table>
<thead>
<tr>
<th>Zone</th>
<th>MAPE (%)</th>
<th>Level of accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone 1 - 5</td>
<td>7.27</td>
<td>92.73</td>
</tr>
<tr>
<td>Zone 3 - 5</td>
<td>8.91</td>
<td>91.09</td>
</tr>
</tbody>
</table>

**CONCLUSION**

It is found that the MM can be utilised to predict future failure rates based on HI. The majority of the predicted failure rates are quite close to the computed failure rates, with an accuracy level of 91.10%. Notwithstanding, higher discrepancies between the computed and predicted failure rates in Zone 5 are observed due to insufficient oil samples data to compute the failure rates in the respective zone. Based on the MM prediction, the failure rate of the transformer population increases as the transformer’s age increases, and it begins.
to exceed the global average failure rates at 14 years. Overall, MM can be considered a viable approach to predict transformer failure rates, and it can be used as an alternative option to determine the forecasted failure data.

ACKNOWLEDGEMENTS

The authors would like to express sincere gratitude to the Ministry of Higher Education for the funding provided for this study under the FRGS scheme of FRGS/1/2019/TK07/UPM/02/3 (03-01-19-2071FR).

REFERENCES


Failure Rate Estimation for TX Based on HI through MM Approach


