PATH: PREDICTING TRANSFORMER HEALTH

Maria Inês VERDELHO
EDP Distribuição
mariaines.verdelho@edp.pt

Cristina CARVALHO
EDP Distribuição
cristina.carvalho@edp.pt

Luís PINTO SÁ
EDP Distribuição
pinto.sa@edp.pt

João Vasco FERREIRA
EDP Distribuição
joaovasco.ferreira@edp.pt

Armando LEITÃO
INESC TEC (*)
afleitao@fe.up.pt

Luís Magalhães Dias
INESC TEC (*)
luis.m.dias@inesctec.pt

Xavier Andrade
INESC TEC (*)
xavier.andrade@fe.up.pt

Luís Guimarães
INESC TEC (*)
lguimaraes@fe.up.pt

(*) Faculdade de Engenharia, Universidade do Porto

ABSTRACT

This paper presents the project PATH, Predicting Transformer Health, whose purpose was the design of health indicators to classify the current and futures condition of all HV/MV and MV/MV Power Transformers (PT) of EDP Distribuição (EDPD).
We present the methodologies developed during the project to predict the condition of the transformer in a short and long term. Also, a model was developed that allows to calculate the probability of failure. In addition to automating and systematizing the process of transformer condition analysis, the models developed allow us to simultaneously incorporate the techniques already developed to signal the transformers with suspected potential faults or replacement in the near future. The models were refined using data provided by EDPD.

INTRODUCTION

Prior to the development of project PATH, EDPD had a traditional approach to diagnose the PTs condition based on some of the worldwide know condition assessment methods and on EDP engineering team knowledge and experience. This project's challenge was to integrate the existing knowledge and leverage the use of data analytics to perform the condition assessment of the PTs. PATH had three major objectives: *(1)* to provide a numerical evaluation of the PT condition in the short term and *(2)* long term, and *(3)* the identification of the internal and external factors that influence their condition. The models and algorithms that allow the PT condition calculation are described in the next sections. The first covers the short term approach which translates to a diagnose of the current PT condition based on the exploration of the DGA (Dissolved Gas Analysis) data. The second tackles the evaluation of the long term PT condition through the assessment of the paper insulation degradation and its correlation with internal and external factors.

METHODOLOGY

The chemical tests performed periodically to the oil (normally on a yearly base) monitor a fleet of 740 PTs. This data covers the years from 1989 to 2017 and is composed by the readings of furan compounds, gas concentrations and by physical, chemical and dielectric tests to the oil.
Additionally, equipment catastrophic failures were provided ranging from 2009 to 2016 (a total of 16 faults occurred). A short list of PTs inspected at the factory with major failures was also provided.
Furthermore, load measurements were also made available from 2004 to 2016. In these data, it is possible to find periodic measurements of electrical load discretized in intervals of 15 minutes.

Short term condition Evaluation

Based on the available data and acquired knowledge about the equipment, control methods were developed for the various test results in order to classify the PT condition. In addition, the authors pursued to identify some conditions and patterns recorded in the history to make the condition analysis of the PT more robust. The developed algorithm contributes encompasses five essential ideas:

1. To analyse whether the PT shows signs of faulty behaviour (Suspect);
2. To diagnose the condition of the PT based on the methods used in the DGA (Additional Analyses);
3. To understand the impact of past concentrations on the PT condition (Analyse the gas variation);
4. To quantify the PT condition based on diagnosis, historical records and absolute gas concentration (DGA condition);
5. To diagnose the degree of confidence in the respective PT's score (Diagnose fault)

To detect the PTs with signs of potential fault, two criteria were used. In the first criteria, the selection is based on the combined gas concentration (TDCG). If this concentration exceeds 720 (value obtained by the Key Gas method[1]) then the PT is flagged by the algorithm and is therefore subject to a more extensive analysis:

$$TDCG = [H_2] + [CH_4] + [C_2H_6] + [C_2H_4] + [C_2H_2] + [CO] \quad (1)$$

In the second criteria, the PT will be flagged as suspicious if any gas is above the limits defined by the user (Table 1). This first step allows to flag which PTs should be kept on watch. It is important to emphasize that meeting one of the

1 The Key Gas method diagnoses failures based on the proportion of combustible gases. The technique provides a number of "models" associated with standard failure conditions. For example, 63% ethylene with some ethane (19%) and methane (16%) indicate that the PT oil is overheated. If most of the gas is carbon monoxide (92%) then it indicates that the cellulose is overheating. The accuracy of this method is highly dependent on the researcher's experience and his ability to tune it.

criteria is enough for the PT to be flagged with a potential failure.

Table 1 Gas individual threshold applied in the algorithm.

GAS	Concentration (ppm)
H_2	100 ppm
CH_4	80 ppm
C_2H_6	100 ppm
$C_2 H_4$	100 ppm
$C_2 H_2$	100 ppm

For all PTs which are not flagged by the algorithm a standard level condition score is assigned. Moreover, PTs with a standard condition score Nc are differentiated by absolute gas concentration, thus helping the tool used to understand which PTs are more likely to be flagged as suspect:

$$N_t = \frac{DGA_{score} \times N_c}{DGA_{score}^{max\,normal}} \quad (2)$$

The DGA_{score} is an indicator that aggregates the concentration of gases and is calculated as follows:

$$\boldsymbol{DGA_{score}} = \mathbf{80} \times [\boldsymbol{H_2}] + \mathbf{40} \times [\boldsymbol{CH_4}] + \mathbf{20} \times [\boldsymbol{C_2H_6}] + \mathbf{50} \times [\boldsymbol{C_2H_4}] + \mathbf{80} \times [\boldsymbol{C_2H_2}] \quad (3)$$

The $DGA_{\text{score}}^{\text{max}\,normal}$ is the worst analysis data for PT not flagged by the algorithm.

In the case of PTs with suspicious behaviour, the initial diagnosis of the PT condition relies on the combination of several literature techniques used for the DGA analysis, namely:

- Roger Ratios[2]
- IEC [3]
- Duval Triangle [4]
- Doernenburg [5]

All these methods enable the diagnosis of a potential fault in the PT. However, as already mentioned, the different techniques present some weaknesses in the diagnosis of certain failures. In order to obtain more robust results, a score Nt combining the diagnoses obtained by different methods is calculated. Each method considered in the score has an assigned weight, w_{method}.

$$N_t = \sum w_{method} \times f_{method} \quad (4)$$

For each fault detected f_{method} there is an assigned severity score, which reflects on a scale of [0; 10], where 0 (zero) is the best grade and 10 (ten) the worst, how bad is the failure.

The obtained score is adjusted by combining the different methods and the gas variation historical records. This variation ΔGt is calculated using the current combined gas concentration $DGA_{score}(t)$ and the PT's previous $DGA_{score}(t-1)$.

$$\Delta G^t = (DGA_t^{score} - DGA_{t-1}^{score}) \times \frac{365}{t-(t-1)} \quad (5)$$

Afterwards a condition degradation indicator, ΔH, aggregating past records, is calculated based on an exponential smoothing model that weights ($0 \leq w_g \leq 1$) the different variations of the gases:

$$\Delta H^t = w_g \times \Delta G^t + (1 - w_g) \times \Delta G^{t-1} \quad (6)$$

The degree of confidence score assigned by the algorithm is calculated by analyzing the level of agreement of the different used methods (IEC, Roger, Duval and Doernenburg).

Long term evaluation condition

Remaining Useful Life, RUL

Based on the furan compounds data the project's team developed a model that estimates the condition of the insulation paper in 5 to 10 years projecting the equipment's RUL (Remaining Useful Life). Since the insulation paper's degradation depends on multiple factors, the project also studied their influence. Thus, the model developed by the project team is based on three fundamental points:

1. The way in which RUL is calculated;
2. The selection of factors with a significant impact on the PT condition;
3. Calculation of the PT paper condition on a horizon of 5-10 years;

Prior to the project, the calculation of the RUL was based on the comparison of the DP[6] collected with the DP of a PT at its beginning and end of life. This results in the following linear relation to calculate the paper condition:

$$Paper\ condition\ (\%) = \frac{DP_{new} - DP_{measured}}{DP_{new} - DP_{final}} \quad (7)$$

The developed model incorporates a calculation formula that allows discerning when the PT is expected to reach its functional life limit. To do so, the project team analysed the relationship between the age of the PT and the DP of

2 The Roger Ratios (RRM) method uses the concentrations of different gases to identify and classify faults that occur in a PT. The proportions of C2H2 / C2H4, CH4 / H2 and C2H4 / C2H6 are used in this method. However, over time some gases are naturally produced in the PT without a fault occurring which leads to certain wrong classifications.

3 This method is similar to Roger ratio method except that the ratios C2H6/CH4 is excluded as it indicates only a limited range of decomposition.

4 The Duval Triangle Method (DTM) is a three-axis coordinate graphical method where axes represent percentages of CH4, C2H4 or C2H2 from 0 to 100%. Due to its accuracy and ability to detect a large number of failures, it is widely used by the community.

5 Doernenburg is a method based on four main gases (H2, CH4, C2H2, C2H4 and C2H6) that can detect potential internal faults. The accuracy of this method is high in the presence of large concentrations of gases.

6 The degree of polymerization (DP) is one of the most robust methods to evaluate the condition of paper insulation when compared to other methods [2]. The insulating paper of the PTs is a complex compound of carbon, hydrogen and oxygen, where molecules of the glucose monomer (C5H10O5) contribute to the cellulose formation. As the PT paper ages, the glucose rings react and as a consequence, the chain begins to break. To measure the degree of polymerization, paper samples are collected from various areas of the PT.

the paper. The project team also considered studies already carried out by EDPD that successfully related the concentration of furan compounds with the DP measured in the PTs. From this study, a logarithmic type curve was obtained, which allows the measurements of 2FAL[7] to obtain the DP of the paper. However, this curve is only valid in oil samples for PTs to which no oil maintenance action took place since the concentration of furans are significantly impacted whenever there are oil filtrations or replacements.

In order to maintain the same logarithmic relationship, the following expression was used to calculate the paper DP by using as input the PT age:

$$DP(age) = DP_{new} \times e^{-\frac{age}{k}} \quad (8)$$

Data of PTs that were put out of service (in these PTs the condition of the paper was effectively measured) was used to adjust the parameter k. This curve serves as a baseline for the RUL calculation model. This study resulted in a curve that will serve as the basis for the RUL model. According to the results obtained and assuming that a PT with new paper has a DP = 1200 and reaches the end of life with a DP = 120[8], the RUL calculation is as follows:

$$RUL = 60 \times \frac{1200 - DP_{theorical} \times \Delta d}{1200 - 120} \quad (9)$$

It is expected that the PT paper will theoretically last up to 60 years (results from the interception of the theoretical curve in the value of DP equal to 120).

During the development of this project, several brainstorming sessions were carried out in order to survey the factors that may influence the PT paper condition. These sessions resulted in the following factors that potentially impact the PT condition, namely: technical specifications; load regime; year of manufacture; oil quality; and electrical tests results.

External factors were not considered in this model since they are included in the PT failure probability model. Moreover, several factors were not considered in this model due to the scarcity of information.

To study the impact of the different factors on the PT condition analysis, the project team used logistic regression. To identify situations where there is accelerated or delayed ageing the project team related the $DP_{measured}$ obtained through analysis to furans with the $DP_{theorical}$ which is calculated with the age of the PT. The above approach results in the following equations:

$$DP_{measured} = DP_{theorical} \times e^{\beta_1 f_1 + \cdots + \beta_n f_n} \quad (10)$$

The selected factors help explaining the different rates of ageing making the estimation of the RUL more robust:

$$\Delta d = \ln\left(\frac{DP_{measured}}{DP_{theorical}}\right) = \beta_1 f_1 + \cdots + \beta_n f_n \quad (11)$$

Based on this relationship, high-performance analytical techniques were subsequently used to select the factors that present the best results in the modelling process and to explain the paper degradation rate Δd. The analytical technique that obtained the best results was *random forests*, which agrees with the several bibliographical references on the application of this technique [3]. This technique allowed to select the factors that best explain the coefficient of determination (*R^2 = 84%* of the variance of the results) and those that have the most impact on the final output.

The prediction of the PT condition in a long-term perspective, 5 to 10 years, is made through the historical evolution of the ageing of the paper conjugated with the factors that influence the degradation. Fundamentally the developed model calculates the Δd of paper degradation and uses *a*' historical records of that degradation to make the forecast in the established period t, see equation (8). Subsequently, the calculation of the RUL is made based on the theoretical paper DP curve. The final calculation formula that allows to make the prediction of the RUL is calculated as (9).

$$\Delta \mathrm{d} = \Delta d_k(t) = \alpha \times \Delta d_{t-k} + \sum \beta_i \times f_i(t) \quad (12)$$

As the data is collected and entered into the database for a given PT, the model adjusts the theoretical base curve in order to evaluate the condition of the PT from a long-term perspective.

Random failures

With the objective of improving the risk analysis of EDP Distribuição, a model encompassing different condition factors was developed in order to obtain the PT failure. Thus, a model capable of adjusting the probability of failure, according to the PT condition, was proposed. The model developed by the team project is based on two fundamental points:

1. The selection of factors with impact on the probability of failure;
2. Calculation of the failure probability based on the factors influencing it;

The condition indicators of the PTs calculated in this project, the DGA condition score and the RUL result, are combined with the atmospheric events to create the set of factors used to estimate the failure probability.

The use of the DGA and RUL factors allows to understand the impact of the PT condition in the short and long term on the failure probability, while the atmospheric events incorporate the meteorological effects.

Several factors were not considered in this model due to the same reason already mentioned in the previous section. An example is the recording of the number of alarms and activation of the PT protections which started to be collected at the beginning of 2011. This type of information might be important to include in the model, however it is necessary to increase the number/period of historical records so that the information has the desired technical validity. Furthermore, it was decided that this type of information should be introduced and studied in

7 2-furfural (2-FAL).

8 Value used by the paper analytical expertise.

future projects.
There are alternative methods that can be used to model the reliability function through a set of faults. The PHM (Proportional Hazards Model) model was chosen because: (i) it requires, by principle, less assumptions, (ii) it easily incorporates censored data (also usually referred to as incomplete or suspended times), (iii) it is able to estimate both factors that affect reliability and a base reliability curve $R(t)$ - a function that represents the probability that a system will survive without failures until a given time *t*.
The PHM describes the reliability function through the equation:

$$R(t,z) = R_0(t)e^{z_1\beta_1+\cdots+z_n\beta_n} = R_0(t)e^{z\beta} \quad (13)$$

where *R* is the reliability, *t* is the variable that measures the interval between occurrences (typically it is time), z_i is a variable that indicates the value of factor *i* and β_i is the coefficient associated with factor *i*. Thus, positive coefficients mean that the factor has a negative impact on reliability, while negative coefficients imply that the factor has a positive impact on reliability.
The baseline risk curve $R_0(t)$ is obtained through the Kaplan-Meier model since it does not need to assume a specific distribution for the time between failures. The absence of parameterization of the baseline curve represents, in complex systems a significant advantage.
After calculating the factors coefficients and the baseline risk curve it is possible to obtain, for each PT, the adjusted failure probability using the following equation:

$$P_f = R(t,z) \quad (14)$$

In case the PT does not have sufficient data to calculate the RUL indicator and/or the DGA indicator, the failure probability is calculated using the baseline risk curve $R0$ (t).

$$P_f = R_0(t) \quad (15)$$

RESULTS

Short term condition

For the short-term model six gases are used to analyse the PT condition - given that they are the ones that are most prevalent in the detection of internal faults:

H_2 (Hydrogen)	C_2H_4 (Ethylene)
C_2H_2 (Acetylene)	CO (carbon monoxide)
CH_4 (Methane)	C_2H_6 (Ethane)

By analysing recent data and taking into account the historical records of each PT, it was possible to highlight PTs in worst condition. Considering the worst score that a PT can have is 10, it was possible to conclude that the PT in the worst condition, for this model, had a score of 7.5.
Figure 1 shows the current distribution of PTs condition. In this Figure, only the most recent results are shown.

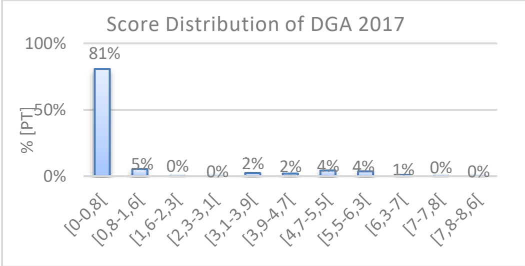

Figure 1 Current overview of PTs condition.

Long term condition

In the RUL model, to determine the coefficients associated with each factor considered, the team used estimators of the model parameters obtained by the least squares method. The final model, where the statistically significant variables for RUL modelling are identified, was obtained using the multiple regression technique called *Random Forest* [3].
For the analysis of the paper condition, we considered an annual time metric, the list of significant factors and their relative importance for the model using the Mean Standard Error Percentage (%MSE) [3]. The higher the %MSE is, the greater the factor weight in the model. The model was generated using 1737 samples covering 230 TP. Table shows the selected factors, the %MSE value and the respective impact on the response variable $\Delta d(t)$.

Table 2 Critical Factors to RUL model and its impact.

Factor	%MSE	Impact
Density	18%	Negative
PC1[9] (oil quality)	27%	Negative
PC2[10] (oil humidity)	15%	Negative
Total Load	8%	Positive
Delta tangent	11%	Positive
Connection group	44%	Positive/Negative
Manufacture year	66%	Negative

With the obtained results it is possible to conclude that a set of PT must be replaced in a time horizon of 5 to 10 years since they are expected to reach their considered end of life.
Figure 2 shows the probability of the paper reaching the end of life for a given number of PTs by 2020 (grey curve) or by 2023 (red curve) if we consider the most conservative scenario made by the RUL model (with a 95% confidence interval).

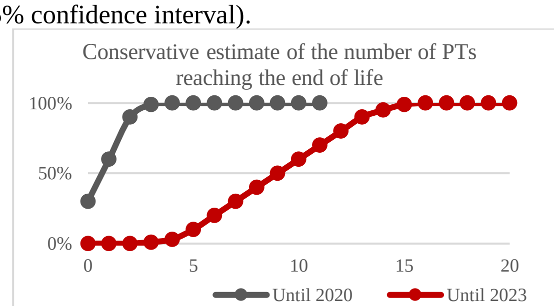

Figure 2 Conservative estimate of the number of PTs reaching the end of life.

9 PC1 – Number depended on neutralization test and interfacial strength

10 PC2 – Number depended on water content and the breakdown voltage.

This way and using the grey curve as an example, the probability of having at most one PT in which the insulation paper reaches its end of life until 2020 is around 65%. Hence, for this case within two years the estimation point to at most 3 failures since the remaining scenarios have very low probabilities. However, looking at the same scenario on the red curve we conclude that within 5 years the failures in PTs is very likely to exceed 4.

Random failures

In the *PHM* model, to determine the coefficients associated with each of the three factors, the team used the maximum likelihood method.

The failure mode considered implies the complete replacement of the PT. The time metric used is the calendar age. Table 3 presents the mean factor for all recorded faults - μ, the resulting coefficient - β - and the test value – p for all the factors. The test value represents the probability of the data being compatible with the initial hypothesis, which states that the factor does not affect reliability ($\beta = 0$), and its value must be compared with the chosen level of significance. That is, the closer to zero is the test value, the greater is the probability that the factor has a significant impact on the equipment reliability. The test value should be compared with a predefined significance level, e.g. *5%*, to determine whether a factor is statistically significant.

To verify the dimension of the impact of the factor in the reliability, it must be taken into account the multiplication of the coefficient by the factor mean ($\beta \times \mu$).

Table 3 Significant factors for the PTs.

Factor	μ	β	p
DGA score	0.1	0.55	0.7%
RUL	0.36	-11.84	1.0%
Meteorology effects	$3.25e^{-4}$	1656	$< 0{,}1\%$

Final estimations

The final reliability estimation of the PTs merges the results from the paper degradation model and the random failure model. Table shows the number of PTs estimated to fail from 2019 to 2021, 2023, 2024 and 2025, It considers two different scenarios: *(i)* "Most likely scenario" and *(ii)* "Conservative scenario".

Table 4 Estimation of PTs to fail from different periods.

Period	*Conservative scenario*	*Most Likely scenario*
2019-2021	8 PTs	5 PTs
2019-2023	16 PTs	12 PTs
2019-2024	22 PTs	17 PTs
2019-2025	30 PTs	23 PTs

The "Most likely scenario" points to the expected number of failures while the "Conservative scenario" indicates an estimate for the maximum number of faults in the next few years (considering confidence of 95%). This way, it is expected that 8 PTs will fail by 2021 if it is considered the "conservative scenario", and that in the "most likely scenario" 5 PTs will fail. If we analyze the next 5 years, it is possible that up to 16 PTs will break down, but it is expected that a total of 12 PTs will effectively break. The same type of reading is applied for the next 6/7 years.

CONCLUSION

The PATH project had the purpose of designing health indicators that integrated the power PTs HV/MV and MV/MV of EDPD with a view to three major purposes: the first and the second purpose consist of providing an evaluation of the PT condition in the short and long term and the third purpose is to identify the factors that influence PTs condition.

This paper describes the models that allow calculating the condition of the PT in a short and long term. Also, a model was developed that allows estimating the PT probability of a random failure. In addition to automating and systematizing the process of PT condition analysis, the models developed allow us to simultaneously incorporate the techniques already developed by the employees of EDPD to signal the PTs with suspected potential faults or replacement in the near future. The models were refined using data provided by EDPD. The results enable an improved decision-making process regarding O&M actions of PT and also to project investment needs on an ageing fleet.

REFERENCES

[1] Ali Naderian Jahromi, R. P., 2009, "An approach to Power Transformer Asset Management Using Health Index", *IEEE Electrical Insulation Magazine,* 15.

[2] Bogdan Gorgan, P.V., 2012, “Calculation of the Remaining Lifetime of Power Transformers Paper Insulation”, *IEEE*, 8.

[3] Breiman, L., 2001, “RANDOM FORESTS”, *Berkeley, CA 94720: University of California. ion*.

[4] Di He, Y. Z., 2016, “Failure Probability Model of Transmission and Transformation Equipment for Risk Assessment”, *IEEE*, 5.

[5] Douglas Bates, M. M. (s.d.). “Fitting Linear Mixed-Effects Models using lme4”, *Journal of Statistical Software*, 51.

[6] Gockenbach, X. Z., 2008, “Asset-Management of Transformers Based on Condition Monitoring and Standard Diagnosis”, *IEEE Electrical Insulation Magazine*, 15.

[7] GRAY, I. (s.d.)., *A GUIDE TO TRANSFORMER OIL ANALYSIS. A GUIDE TO TRANSFORMER OIL ANALYSIS*.

[8] IEEE., 2009, *IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers*.

[9] Md Mominul Islam, G. L., 2017, “A review of condition monitoring techniques and diagnostic tests”, *Springer*, 25.

[10] Power, A. A., 2008, “An Approach to Determine the Health Index of Power Transformers”, *IEEE*, 5.

[11] UK Power Networks, 2015, *DNO COMMON NETWORK ASSET INDICES METHODOLOGY*.

[12] Weisberg, J. F., 2011, *Cox Proportional-Hazards Regression for Survival Data in R*.