Prognostics & Health Management Oriented Data Analytics Suite for Transformer Health Monitoring

Jose I. Aizpurua[1], Brian G. Stewart[2] and Stephen D. J. McArthur[2]
[1]Mondragon University, Spain, [2]University of Strathclyde, Scotland, UK

An overview of machine learning based prognostics & health management applications

Prognostics and health management (PHM) solutions assist in transformer health monitoring activities through continuous health monitoring. A PHM-oriented data analytics suite is presented based on anomaly detection, diagnostics and prognostics applications implemented through machine learning techniques.

Abstract
Condition monitoring of power transformers is crucial for the reliable and cost-effective operation of the power grid. The unexpected failure of a transformer can lead to different consequences ranging from a lack of export capability, with the corresponding economic penalties, to catastrophic failure, with the associated health, safety, and economic effects. With the advance of machine learning techniques, it is possible to enhance traditional transformer health monitoring techniques with data-driven and expert-based prognostics and health management (PHM) applications. Accordingly, this paper reviews the experience of the authors in the implementation of machine learning methods for transformer condition monitoring.

Keywords: machine learning, data analytics, transformer health monitoring, anomaly detection, diagnostics, prognostics.

1. Introduction

The correct operation of a power transformer depends on the correct operation of multiple related components over a variety of conditions. When monitoring the transformer's health, the possible failure modes need to be examined. Traditional methods for transformer health assessment have been focused on the analysis of, e.g. gases dissolved in oil, temperature, or electrical parameters [1]. With the advance of machine learning (ML) techniques, traditional transformer health monitoring techniques can be enhanced with the development of prognostics and health management (PHM) applications including anomaly detection, diagnostics, and prognostics analysis modules [2]. For a consistent implementation of ML methods for transformer health monitoring, a multi-stage PHM-oriented methodology is needed. Fig. 1 shows the block diagram of the PHM-oriented design methodology.

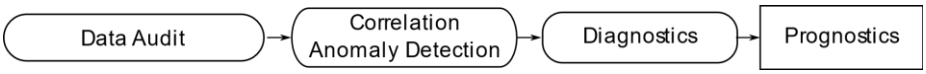

Fig. 1. PHM-oriented analytic modules.

The PHM-oriented analytics process starts from the *data audit* step by listing available datasets and identifying new variables that can be monitored to improve the health assessment process. Next, *correlation and anomaly detection are implemented* to identify abnormal data patterns. If an anomalous data trend is detected, then the *diagnostics* follows the process for the identification of failures. After diagnosing the current health, it is possible to implement *prognostics* methods to estimate the remaining useful life (RUL) through the application of future operation profiles. Fig. 2 shows prognostics and RUL prediction concepts, based on a set of collected health-state data samples, $\{y_1,\ldots,y_n\}$, with predictions from the prediction time instant t_p until crossing the failure threshold and reaching the end-of-life (EOL) limit. Note that the initial and final health states are expressed through probability density functions (PDF), which define the uncertainty associated with these states.

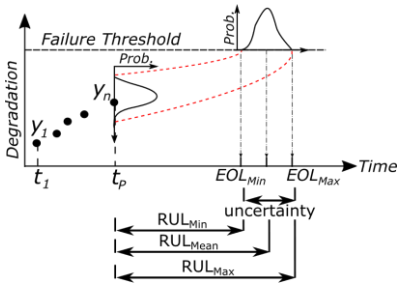

Fig. 2. Prognostics prediction and RUL estimation [3].

Fig. 3 shows the implementation of the PHM-oriented methodology for transformer health monitoring integrating different data sources, correlation between different parameters, diagnostics techniques, evidence combination of the analytic results, health index estimation and the final prognostics prediction. The PHM analytics suite defines different analytic layers from low-level data sources to high-level implementation levels, including diagnostics, prognostics, and health index applications.

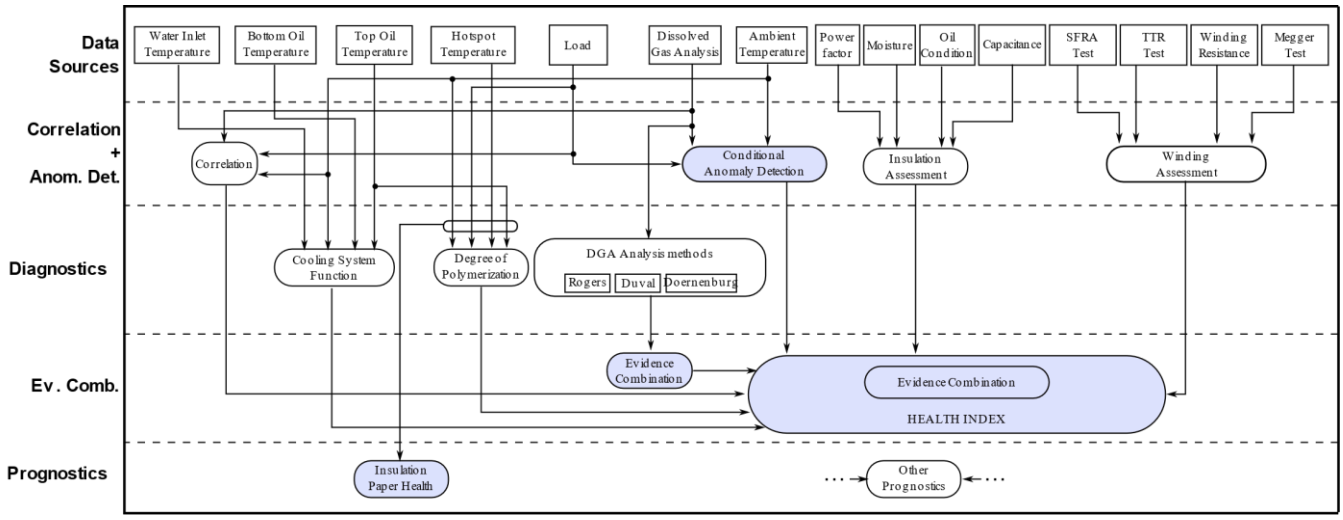

Fig. 3. PHM-oriented data analytic suite – adapted from [4].

From the high-level transformer operation point of view, there are two types of data: operational and environmental data, e.g., ambient temperature, load, and transformer condition-related data. Transformer condition-related data can be further divided into off-line and on-line parameters, which may require different analytics. Off-line parameter extraction requires the transformer to be de-energised, whereas data is sampled on an ongoing basis for on-line parameters. Listed data sources in Fig. 3 are relevant for transformer health monitoring, however, some of these datasets may not be easily collected, such as transformer hottest-spot temperature (HST), and sampled variables may depend on the type of transformer. Anomaly detection, diagnostics and prognostics modules highlighted in Fig. 3 will be addressed in the following sections.

2. Anomaly Detection

Anomaly detection focuses on the identification of abnormal patterns. Patterns or signatures define the correct operation, and they can be defined through ML and statistical learning models that learn to express the normal behaviour of the asset under study.

Many anomaly detection modules define the expected normality solely based on internal condition-related data. However, it is important to note that some systems are also influenced by external factors, such as ambient temperature and load. In this context, one alternative is to implement a conditional anomaly detection (CAD) model which correlates the internal transformer condition-related data and external operational data [5].

Fig. 4a shows the transformer CAD model concept. The goal of the CAD module is to distinguish situations where unusual operation conditions may be causing abnormal transformer behaviour from situations where the transformer condition is unusual under normal operation. The latter case is more likely to represent a true transformer health deterioration.

In order to implement the CAD model, Gaussian Mixture Models are used which generate multivariate Gaussian distributions that embed condition-related data within a transformer model, *P(Transf)*, and operating environment related data within an environment model, *P(Env)*. The set of Gaussian components and their mixing proportions constitute a probabilistic model of the dataset. Then a correlation algorithm, named Expectation Maximization (EM) [5], is used to generate a conditional probability model that correlates the environment and transformer models, *P(Transf | Env)*. EM converges towards a locally optimal set of values that maximize the likelihood of the model *P(Transf | Env).*

The CAD module can take multiple different input variables to the transformer and environment models. As an illustrative example, load and ambient temperature have been used in the environmental model, and methane (CH_4) and hydrogen (H_2) for the transformer model. The development of the CAD model requires training environment and transformer models, and then generating a correlation model. Both models have been trained based on normal operation data, which is determined by a period of stable gas levels, and the rest of the data is used for testing. Fig. 4b shows the trained environment model, where the axes in the horizontal plane denote normalized training values of true power and temperature and the vertical axis identifies their joint probability density, e.g., when the true power is very high, it is likely that the temperature will be low. Fig. 4c shows the trained transformer model where horizontal plane axes denote methane and hydrogen, and the vertical axis denotes the probability density, e.g., it is very likely that when hydrogen is small methane is also small. These figures model the expected independent normal behaviour of the transformer and environment.

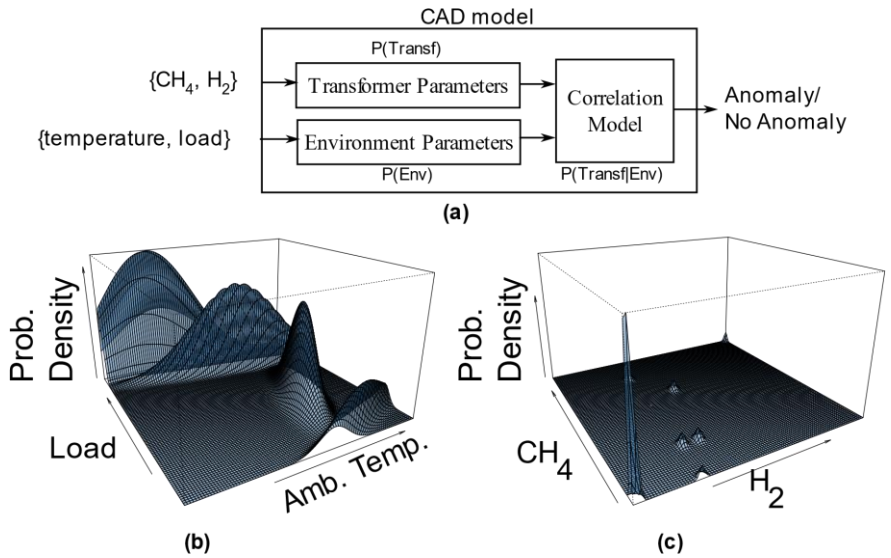

Fig. 4. **(a)** Transformer conditional anomaly detection approach. Examples of probabilistic **(b)** transformer model, P(Transf), and **(c)** environment model, P(Env).

After learning the environment and transformer models, the correlation models are tested. Fig. 5a shows the normalized test data and Fig. 5b shows the CAD outcome, where the red horizontal line identifies the failure threshold, and the vertical dashed line indicates the division between training and testing data.

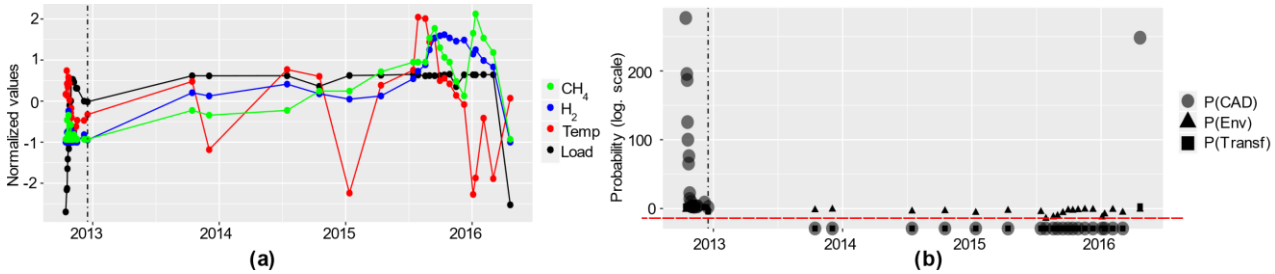

Fig. 5. **(a)** CAD training and testing datasets; **(b)** CAD results.

In Fig. 5b it can be observed that almost all the test data is classified as anomalous because the environment probability is high (triangles above the red line: normal environment), but the probability of the transformer model is low (squares below the red line), which indicates a true anomaly (circles below the red line). Note also that the last sample is classified as healthy (circle in the top right) because both transformer and environment models match with the trained models (Fig. 4b and Fig. 4c). In this case, as the gassing behaviour was being carefully monitored, the gas levels from the test period were known to be significantly different from those during the training period,. These changes were caused from a scheduled change in operation of the cooling of the transformer,. Therefore, this case shows the power of the CAD technique to recognize changing operation conditions, such as the change in cooling, and the existence of any anomalous gas behaviour.

3. Diagnostics

The overall health of the transformer can be evaluated looking at different parameters such as electrical, gas, thermal, or mechanical characteristics. Particularly, the analysis of the gassing behaviour is an important indicator of an early malfunction of the transformer.

Operational and fault events generate gases which are dissolved in the oil that circulates through a transformer for cooling and insulation purposes. Dissolved Gas Analysis (DGA) is a mature and industry-standard method that focuses on the study of these gases [6]. There are different industry-accepted DGA interpretation methods including Roger's ratios, Doernenburg's ratios, Duval's triangle and pentagon, and many more [7]. These techniques classify transformer faults based on the predefined range of specific fault gas ratios. However, their accuracy is limited because they assume crisp, deterministic decision bounds [7]. This leads to a decreased diagnostics accuracy and conflicting diagnostics outcomes among methods, which can be confusing and do not help engineers in the decision-making process.

In order to improve the classification accuracy a number of ML models have been proposed focused on solving a supervised learning problem. That is, using a DGA dataset comprised of *n* samples, $DGA = \{x_i, y_i\}_{i=1}^{n}$, where the pair $\{x_i, y_i\}$ contains the data related to the *i*-th observation, $x_i \in X$ and $y_i \in Y$. The matrix $X \in \mathbb{R}^{nxp}$ contains the information $X = \{x_1, \dots, x_n\}$ for p fault gases, and the vector $Y \in \mathbb{R}^{nx1}$ contains the information about the health state of the transformer. In a binary classification problem, the set of possible states of y_i are limited to normal and fault states. In contrast, multiclass classification problems are more challenging than binary classification problems, but they also generate more useful information for maintenance planning.

In this section, black-box (BB) and white box (WB) source classifiers are presented for multiclass transformer diagnostics, including normal degradation, thermal fault, arcing fault, and partial discharge (PD), based on the publicly available IEC TC 10 dataset that includes DGA samples and the corresponding forensic transformer health-state information [8]. BB models generally show a high accuracy, but their usability for resolving misclassified data samples is limited because their diagnostics is deterministic, and they do not generate uncertainty information. WB models capture expert-knowledge either as a causal model or through first-principle models. They generate the uncertainty associated with the decision-making process by quantifying the PDF of the likelihood of different diagnostics states. Artificial Neural Networks (ANN) and Support Vector Machines (SVM) models are used as BB classification models, as they have shown high accuracy on DGA data [7]. For WB modelling Gaussian Bayesian networks (GBNs) are used because they are able to capture the causality among random variables and infer uncertainty information.

3.1 Artificial Neural Networks

ANNs are BB models are widely used for classification and regression [7]. The multilayer perceptron (MLP) feedforward model was used, which is a three-layer network (input, hidden, and output) comprised of fully connected neurons. Each neuron performs a weighted sum of its inputs and passes the results through an activation function. All the designed ANN models use a sigmoid activation function for hidden and output nodes.

Model training is performed using a back-propagation algorithm. The goal is to learn the neuron weights to estimate the transformer health-state (network output) from DGA values (sample input), which minimize the estimation error with respect to the target transformer health state. Best results were obtained with 20 hidden nodes with C_2H_6, C_2H_4, H_2, CH_4, and C_2H_2 as inputs as shown in Fig. 6.

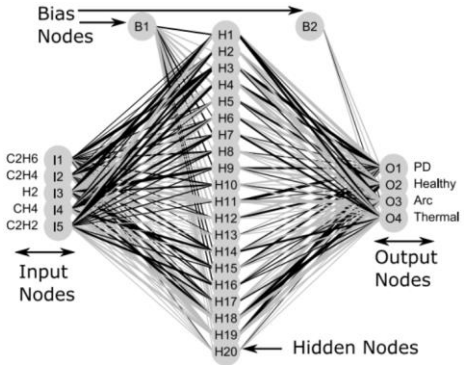

Fig. 6. ANN configuration for DGA diagnostics [7].

3.2 Support Vector Machines

The SVM maps input data into a space using a kernel function [7]. The SVM learns the boundary separating one transformer health state from another with maximum distance. The kernel function aims to translate a problem that is nonlinearly separable into a feature space, which is linearly separable by a hyperplane. The hyperplane represents the transformer health-state classification boundary. The SVM training consists of calculating the kernel hyper-parameters and a cost variable, which is a trade-off between separation variables and obtaining a large margin for the SVM. Grid search was used to find the optimal parameters. A number of configurations were trained using all different gases and their ratios as input to the SVM. Of the trained SVMs, the one with the highest accuracy from the test data was selected as the choice for that output, which matches with the input data used for the ANN model: C_2H_6, C_2H_4, H_2, CH_4 and C_2H_2.

3.3 Gaussian Bayesian Networks

Bayesian networks (BN) are statistical models that capture probabilistic dependencies among random variables [9]. Graphically, these variables are represented through nodes linked through edges that reflect dependencies between variables. Statistically, dependencies are quantified through conditional probabilities. BNs are a compact representation of joint probability distributions. In probability theory, the chain rule permits the calculation of any member of the joint distribution of a set of random variables using conditional probabilities. When a BN is comprised of continuous random variables Gaussian BNs capture dependencies through linear Gaussian distributions and variable distributions are modelled through Normal random variables. Fig. 7 shows the GBN model.

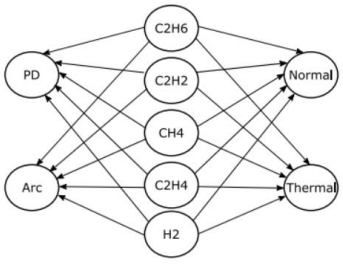

Fig. 7. GBN configuration for DGA diagnostics [9].

The parameter estimation for GBN models is based on the maximum likelihood algorithm which estimates the corresponding parameters for each node in the BN model, e.g. for the PD node (Fig. 7): $Pr(PD|C_2H_6, C_2H_2, CH_4, C_2H_4, H_2) \sim N(\beta_0 + \beta_1 C_2H_6 + \beta_2 C_2H_2 + \beta_3 CH_4 + \beta_4 C_2H_4 + \beta_5 H_2;\ \sigma^2)$. After learning the parameters, the estimation of the conditional probability of nodes is based on inferences. In this case, the likelihood weighting algorithm is implemented, which fixes the test DGA gas samples (evidence) and uses the likelihood of the evidence to weight samples [10]. When applied to the DGA dataset, for each of the analysed faults the outcome of the inference is a set of random samples from the conditional distribution of the fault node given the evidence. Density values of the inference outcomes can be calculated through Kernel density estimates.

3.4 Performance Results

After training and testing all the classifiers through Monte-Carlo cross validation, thus enabling classifier statistical evaluations, Fig. 8 shows the overall classification accuracy results for the different source classifiers.

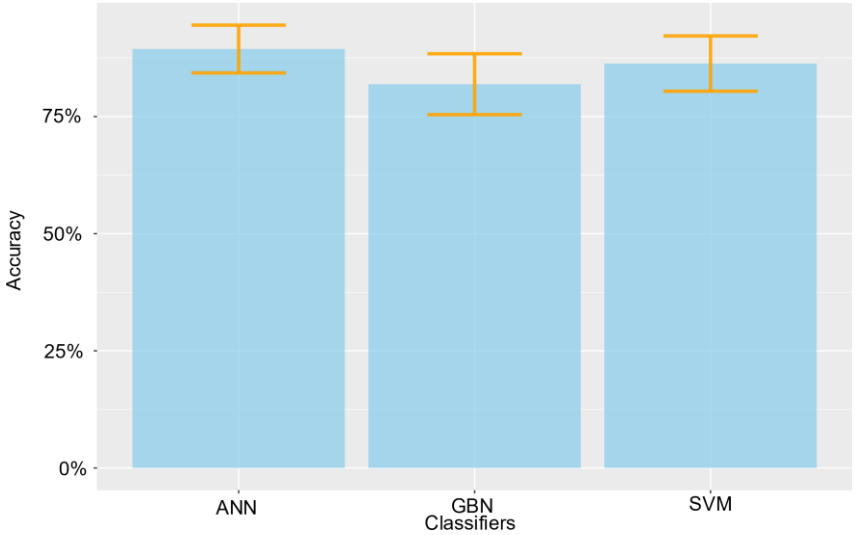

Fig. 8. DGA diagnostics results tested on the IEC TC 10 dataset.

It can be seen that the ANN model is the more accurate diagnostics method. It is interesting to observe that the accuracy of the GBN model is not as high as SVM and ANN models, but the information inferred from the outcome of GBN models is more useful. For instance, consider that after training the source classifiers, they are tested for the following absolute gas values [9]: H_2 = 26788 ppm, C_2H_4 = 27 ppm, C_2H_6 = 2111 ppm, C_2H_2 = 1 ppm and CH_4 = 18342 ppm and the observed fault type is partial discharge. Table I displays probabilistic results for the different classifiers and for the different possible health states of the transformer.

Table I. Classification accuracy results example [9].

Method	Pr(Normal)	Pr(Thermal)	Pr(Arc)	Pr(PD)
GBN	0.23	0.28	0.18	**0.31**
SVM	0.08	**0.45**	0.07	0.4
ANN	3.9E-2	**0.5**	4.8E-6	0.46

ANN and SVM models do not generate more information other than shown in Table I. However, GBN models generate PDFs with uncertainty information as shown in Fig. 9. If the designed GBN model is confident about the diagnostics output it will generate a PDF with a narrow standard deviation, *i.e.,* PD fault. In contrast, for uncertain diagnostics outcomes the generated PDFs are wider with greater standard deviation values, e.g., Normal degradation.

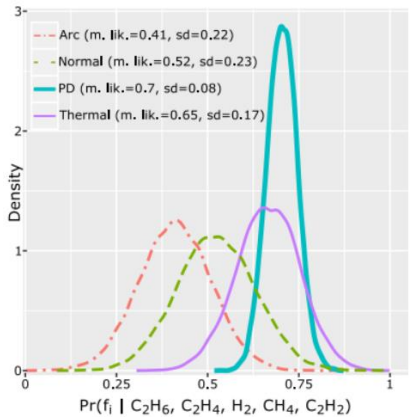

Fig. 9. GBN output example [9].

So long as the test data is comprised of gas samples and faults, which are learned during the training stage, the GBN model should return a prediction with high confidence. However, if the model is tested on an unseen class of fault, the model should be able to quantify this with uncertainty levels which can convey information about the confidence

of the diagnosis of the model. This uncertainty information can be used to further post-process the results and improve the diagnostics accuracy [7].

4. Prognostics

Transformer insulating paper is subjected to high temperatures and undergoes a polymerization process causing the breakdown of the cellulose fibres. When the fibre size is reduced, the kraft paper losses its mechanical strength. By becoming brittle, the paper breaks due to vibration during transformer operation resulting in failure of electrical insulation. Therefore, high temperatures reduce transformer insulation life, and it is essential to monitor the temperature. Accordingly, the focus of this section is on the quantification of the transformer paper aging rate, and prediction of parameters to calculate the aging rate and remaining paper life.

The IEEE standard C57.91 defines the insulation paper aging acceleration factor at time t, $F_{AA}(t)$, as [10]:

$$F_{AA}(t) = e^{\frac{15000}{383} - \frac{15000}{273+\theta_H(t)}} \tag{3}$$

where $\theta_H(t)$ is the transformer winding's hottest-spot temperature at time instant t in ˚C which can be calculated from other measurements through [10]:

$$\theta_H(t) = \theta_{TO}(t) + \Delta\theta_{TO,H}(t) \tag{4}$$

where $\theta_{TO}(t)$ is the top oil (TO) temperature at time instant t and $\Delta\theta_{TO,H}(t)$ is the hottest-spot temperature rise over top oil temperature at time instant t.

4.1 Hottest-spot Temperature Prediction

The hottest-spot temperature in (4) is inferred from other measurements and these measurements may include measurement errors that may affect the hottest-spot temperature estimation accuracy. Assuming steady-state measurements, and including measurement errors, the hottest-spot temperature can be calculated as follows [10]:

$$\theta_H(t) = (\theta_{TO}(t) + \varphi_{TO}) + \Delta\theta_{H,R}[(i(t) + \varphi_i)/i_R]^{2m} \tag{5}$$

where $i(t)$ is the transformer load at time t, i_R is the rated load, $\Delta\theta_{H,R}$ is the hottest-spot temperature rise at rated load, m is a transformer parameter determined from a lookup table depending on the cooling system of the transformer, φ_{TO} denotes the top oil measurement error and φ_i designates the load measurement error.

Underestimated HST may lead to the pumps and fans running less and the transformer will be running hotter with an accelerated aging rate and significantly reduced lifetime. If HST measurements are available, it is possible to directly build HST forecasting models based on ML models. However, it is often the case that HST measurements are expensive and not available. In this case, the hottest-spot temperature may be inferred from indirect measurement, such as top-oil temperature, ambient temperature, and load, as defined in (5). Given an input load profile and other influencing parameters, it is possible to design a top-oil temperature forecasting model through ML techniques, which can then be used to predict the HST [3, 11].

The forecasting horizon will determine the application, e.g., short-term forecasting may be used for online health monitoring, mid-term forecasting for maintenance planning and long-term for strategic design decisions. For a given prediction horizon, the first step when designing a forecasting model, is to select and evaluate features that minimize the forecasting error. Depending on the forecasting model, delayed signals may be considered, and this may impact on the forecasting error due to the feedback configuration. There are different ML and statistical methods that can be used to forecast the top-oil temperature, e.g. [3, 11].

4.2 Prognostics through state-estimation

The degradation process of the insulation paper is not a deterministic process, and it needs to integrate sources of uncertainty corresponding to the degradation process, which can be represented through [3, 11]:

$$RUL(t) = RUL(t - \Delta t) + \omega_{RUL_{t-\Delta t}} - e^{(15000+\omega_t)\left(\frac{1}{383}-\frac{1}{273+\hat{\theta}_H(t)}\right)} \quad (6)$$

where ω_t denotes the degradation process uncertainty due to the lack of exact knowledge at time instant t, $\omega_{RUL_{t-\Delta t}}$ denotes the uncertainty of the lifetime estimation at time instant t- Δt, and $\hat{\theta}_H(t)$ is the HST estimate. Initially $\omega_{RUL_{t-\Delta t}}$ will denote the initial lifetime estimation error, ω_{RUL_0}. This error will be propagated and updated in subsequent iterations through the recurrence relation form of the RUL estimation in (4).

Using the designed hottest-spot temperature forecasting model it is possible to connect predictions with the RUL model in (6) and predict future RUL ageing profiles. Fig. 10 shows the potential integration in the transformer lifetime prediction framework.

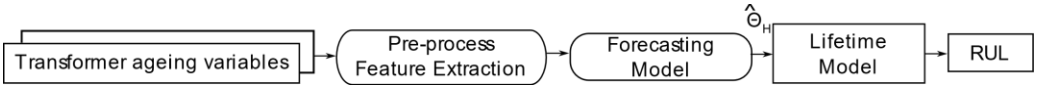

Fig. 10. Lifetime estimation framework.

For different examples in the implementation of transformer RUL prediction frameworks including temperature forecasting strategies, the reader is referred to [3] and [11].

5. Conclusion

This paper has presented a number of machine-learning methods and tools for transformer condition monitoring via prognostics and health management applications. Namely, the implementation details for the development of anomaly detection, diagnostics and prognostics applications have been described. Most of the presented techniques combine data-driven machine-learning methods with expert knowledge-based strategies with a special emphasis on uncertainty processing for an improved performance in operation contexts with various sources of uncertainty. This overview paper along with the data analytics suite, will assist engineers in formulating, developing, understanding and identifying potential applications for machine-learning informed prognostics and health management applications for power transformers.

Bibliography

[1] M.J. Heathcote, *J & P Transformer Book*, thirteenth edition, Newnes, 2007.

[2] G. J., Vachtsevanos, F. L, Lewis, M. Roemer, A. Hess and B. Wu, *Intelligent fault diagnosis and prognosis for engineering systems*, Hoboken, N.J.: Wiley, 2006.

[3] J.I. Aizpurua, B. G. Stewart, S.D.J. McArthur, M. Penalba, M. Barrenetxea, E. Muxika, and J. V. Ringwood, *Probabilistic forecasting informed failure prognostics framework for improved RUL prediction under uncertainty: A transformer case study*, Reliability Engineering & System Safety, Vol. 226, No. 108676, 2022.

[4] J.I. Aizpurua, B. G. Stewart, S.D.J. McArthur, B. Lambert, J. G. Cross, and V. M. Catterson, *Improved Power Transformer Condition Monitoring under Uncertainty through Soft Computing and Probabilistic Health Index*. Applied Soft Computing, Vol. 85, No. 105530, 2019.

[5] V. M. Catterson, S. D. J. McArthur, and G. Moss, *Online conditional anomaly detection in multivariate data for transformer monitoring*. IEEE Transactions on Power Delivery, Vol. 25, No. 4, pp. 2556-2564, 2010.

[6] IEEE Power & Energy Society, *IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers*, IEEE Standard C57.104-2019, 2019.

[7] J. I. Aizpurua, V. M. Catterson, B. G. Stewart, S.D.J. McArthur, B. Lambert, and J. G. Cross. *Uncertainty-Aware Fusion of Probabilistic Classifiers for Improved Transformer Diagnostics*. IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 51, No. 1, pp. 621–633, 2018.

[8] M. Duval, A. de Pablo, *Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases*, IEEE Electrical Insulation Magazine, Vol. 17, No. 2, 2001.

[9] J. I. Aizpurua, V. M. Catterson, B. G. Stewart, S.D.J. McArthur, B. Lambert, B. Ampofo, G. Pereira, and J. G. Cross, *Power Transformer Dissolved Gas Analysis through Bayesian Networks and Hypothesis Testing*. IEEE Transactions on Dielectrics and Electrical Insulation. Vol. 25, No. 2, pp. 494–506, 2018.

[10] IEEE Power and Energy Society, *IEEE Guide for Loading Mineral-Oil-Immersed Transformers and Step-Voltage Regulators*, IEEE Standard C57.91-2011, 2011.

[11] J.I. Aizpurua, S. D. J. McArthur, B. G. Stewart, B. Lambert, J. G. Cross, and V. M. Catterson. *Adaptive Power Transformer Lifetime Predictions through Machine Learning & Uncertainty Modelling in Nuclear Power Plants*. IEEE Transactions on Industrial Electronics, Vol. 66, No. 6, pp. 4726-4737, 2018.