A Novel Visualisation Technique for Dissolved Gas Analysis Datasets

A case study

G. E. Lee*, S. N. Hettiwatte*
School of Engineering and Information Technology
Murdoch University
Murdoch, Australia
gareth.lee@murdoch.edu.au, s.hettiwatte@murdoch.edu.au

H. A. Fonseka
Transpower New Zealand Ltd
Auckland, New Zealand
hasitha.fonseka@transpower.co.nz

***Abstract*— Dissolved gas analysis is the most widely used diagnostic test in power transformers. There are established methods used in industry for interpreting DGA results. Among these are the IEEE Key Gas Method, Rogers' Ratios and the Duval Triangle. However, collectively these methods can lead to conflicting results or unclassifiable measurements. This paper presents a visualization technique for interpreting DGA results to mitigate these effects, based on Kernel Principal Component Analysis. DGA measurements from more than 200 power transformers are used to validate the approach.**

Keywords—power transformers; condition monitoring; dissolved gas analysis; key gas method; Rogers' ratios; Duval triangle; kernel principal component analysis

I. Introduction

Gases are generated in transformers in service due to various mechanisms happening within. These gases then get dissolved in transformer oil, which acts as an insulator and a coolant. Dissolved Gas Analysis (DGA) is a method for extracting these gases from transformer oil with the aim of diagnosing the condition of the transformer. DGA is the most widely used diagnostic test for power transformers in service, worldwide. Typically, dissolved gases found in transformer oil include hydrogen (H_2), oxygen (O_2), nitrogen (N_2), carbon monoxide (CO), carbon dioxide (CO_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), propane (C_3H_8) and propylene (C_3H_6). However, not all these gases are used in the diagnosis. Most available diagnostic methods use no more than seven gases in total: H_2, CO, CO_2, CH_4, C_2H_6, C_2H_4, and C_2H_2.

Among the established methods available for DGA, the popular ones are the Key Gas Method [1], Rogers' Ratios [2] and the Duval Triangle [3], [4]. All these methods are valid provided that the gas concentration levels exceed some threshold value. It is widely recognized that DGA is a method with significant diagnostic limitations, not least the discontinuity that results when transformer oil is degassed [5]. Moreover the established methods described above define fault categories differently and can disagree about the categorization of a specific sample.

These experiments are based on a set of data collected by Transpower New Zealand Ltd. consisting of about 7600 maintenance records of power transformers and other switchgear dating back to 1970. The results reported here focus on a subset of 592 power transformer tests conducted during 2011-2012. In each case the oil was tested using DGA. The subset contains many examples that characterise the variation between well-behaved transformers, but also provides a few examples of faulty devices. This imbalance is common since utility companies wish to detect errant behaviour at the earliest possible stage and schedule pre-emptive maintenance to rectify problems before a transformer develops a severe and disruptive fault. The paucity of faulty examples has encouraged us to avoid using automated pattern recognition as many previous researchers have attempted [6], [7], since we do not have access to a balanced training set. Rather we have focused on developing a decision support system that tracks the behaviour of a transformer over time using visualisation techniques to assist investigating engineers to collate evidence when a transformer strays outside the bounds of normal variability. The approach, based on Kernel Principal Component Analysis, is aimed at plotting a locus for each transformer over time and highlighting the change of direction that signals a sudden deterioration in health.

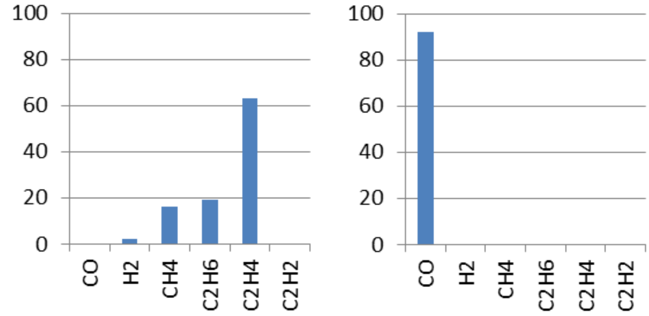
Despite the inherent limitations in interpreting DGA results, it is widely used in industry and others have examined the resulting data using Principal Component Analysis. PCA is frequently used as a feature extraction technique in pattern recognition systems [7], [8], [9]. No evidence has come to light of the application of kernel principal component analysis in this context previously or of principal component analysis used as part of a visualization system for DGA.

II. Established Methods and Their Drawbacks

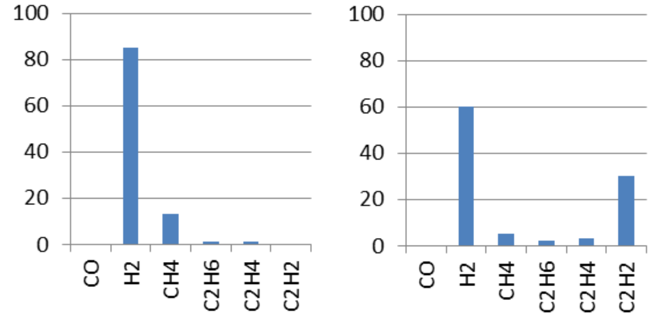
A. *Key Gas Method*

The Key Gas Method (KGM) is given in one of the IEEE Standards [1], as a method of interpreting faults in power transformers using the dissolved gases. It makes use of the dependence on temperature of the types of oil and cellulose decomposition gases. There are certain typical gases generated at certain temperatures, and if these gases are predominant, they can be traced back to the temperatures within a transformer and to the typical fault [1]. These predominant gases are called the 'key gases' and the KGM shows the typical percentages of key gases for four types of faults. These are shown in Fig. 1 [1].

The main drawback with the KGM is that it is highly unlikely for gas concentrations to exactly match the given signatures in Fig. 1. Therefore, based on the experience, the investigator has to make a judgment to correlate the data with the type of fault.



(a) Overheated oil (C_2H_4) (b) Overheated cellulose (CO)



(c) Partial discharge in oil (H_2) (d) Arcing in oil (C_2H_2)

Fig. 1. Typical fault gas percentages (with their key gases) [1]

B. Rogers' Ratios

Rogers [2] has found that under faulty conditions, the ratios between certain gas concentrations are bounded by specific values. These gas ratios are called Rogers' Ratios [1], [2] and the values they take for certain types of faults are categorized in Table I. Unlike with other diagnostic techniques this method also gives typical gas ratios when the unit is in normal operation (Case 0 in Table I). The major drawback with this method is certain values of ratios can fall outside the ranges given in Table I, and therefore, the fault could be indeterminate.

C. Duval Triangle

The Duval triangle provides a graphical method of identifying a fault. It uses a three-axis coordinate system with concentrations of CH_4, C_2H_4 and C_2H_2 on each axis. The area within the equilateral triangle is divided into different regions corresponding to the type of fault, as shown in Fig. 2 [3], [4]. The various regions indicated within the Duval Triangle are given in Table II. These regions and their boundaries have been decided empirically by visual inspection of a large number of faulty transformer measurements.

TABLE I. ROGERS' RATIOS [1]

Case	$\frac{C_2H_2}{C_2H_4}$	$\frac{CH_4}{H_2}$	$\frac{C_2H_4}{C_2H_6}$	**Suggested fault diagnosis**
0	<0.1	>0.1 to <1.0	<1.0	Unit normal
1	<0.1	<0.1	<1.0	Low-energy density arcing - PD
2	0.1 to 3.0	0.1 to 1.0	>3.0	Arcing-high-energy discharge
3	<0.1	>0.1 to <1.0	1.0 to 3.0	Low temperature thermal
4	<0.1	>1.0	1.0 to 3.0	Thermal<700^{O}C
5	<0.1	>1.0	>3.0	Thermal>700^{O}C

One advantage of using the Duval Triangle is that it always provides a diagnosis. There will always be a point within the triangle for known concentrations of CH_4, C_2H_4 and C_2H_2. The drawback with the Duval Triangle method is, sometimes wrong diagnosis may occur when data is in proximity to a boundary.

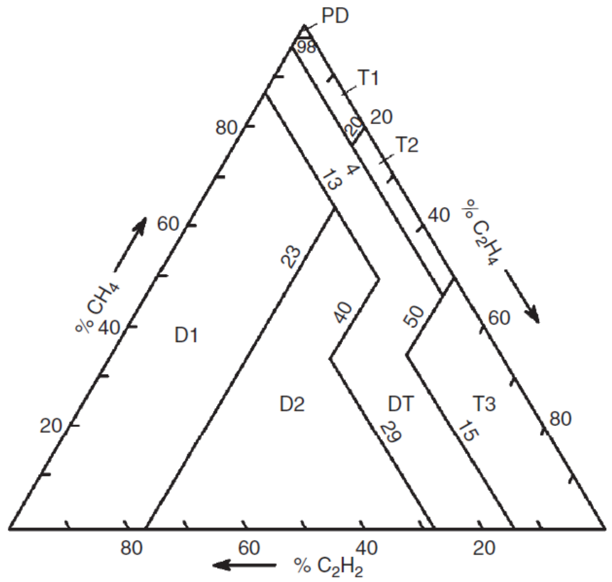


Fig. 2. Duval Triangle with fault regions [3], [4]

TABLE II. REGIONS WITHIN DUVAL TRIANGLE [4]

Region	Fault
PD	Partial discharge
D1	Discharges of low energy
D2	Discharges of high energy
T1	Thermal fault, $t < 300^{O}C$
T2	Thermal fault, $300^{O}C < t < 700^{O}C$
T3	Thermal fault, $t > 700^{O}C$
DT	Mixtures of thermal and electrical faults

III. DATASETS

The dataset consists of 592 DGA measurements taken over the period January 2011 until December 2012 on 224 separate power transformers. It is assumed that these measurements are representative of well-behaved transformers, defined as transformers which are not known to be faulty and operate within Transpower's agreed performance criteria [10]. Each measurement consists of concentrations for the six gases used by the Key Gas Method, taken from one transformer on a known date. Each measurement is treated as a 6 dimensional vector. In addition measurements were available from two faulty transformers that chart their deteriorating health. These are compared to the well-behaved dataset. Moreover during the research a cluster of 14 measurements were discovered corresponding to a further faulty transformer that is used as a third case study. These 14 points were quarantined from the dataset leaving 578 measurements. Finally signatures were generated and overlaid on the visualisation for each of the four KGM signatures in Fig. 1.

IV. EXPERIMENTAL METHODOLOGY

The DGA measurements were processed using Kernel Principal Component Analysis [11] (KPCA) as a way of seeking a new visualisation of the data that differentiates between the normal variation associated with well-behaved transformers and those that are developing or displaying faulty characteristics. KPCA is an extension to the established statistical method of principal component analysis (PCA) [12] that also takes advantage of the kernel trick [13], by using a multivariate Gaussian kernel function to map the six dimensional measurements into a reproducing kernel Hilbert space H of infinite dimension. The 578 well-behaved transformer measurements span a 124 dimensional subspace in H in which PCA is performed, seeking the principal basis vectors with greatest variance. This is determined by finding the eigenvectors and eigenvalues of the covariance matrix of the centred measurements [11]. The dataset gave rise to 124 non-zero eigenvalues. A five dimensional eigenspace was established, corresponding to the largest eigenvalues. The fifth eigenvalue is less than 10% the magnitude of the first which was deemed an adequate cut-off threshold. The eigenspace can be visualised by projecting onto selected pairs of eigenvectors. These KPCA projections are non-linear since they treat the kernel function as a non-linear distance metric between points.

Several parameter values were chosen when implementing the method using MATLAB: the eigenspace dimension was set as previously discussed; the DGA measurements were normalised prior to processing by finding the maximum concentration of each gas and scaling each concentration in the range [0, 1]; and a kernel variance σ was set for the Gaussian kernel function in the range [0.6, 0.9] as shown in later figures. The variance loosely sets the degree of non-linearity used in determining distances. Small values of σ make the representation highly non-linear and larger values approach the behaviour of conventional PCA. The final decision when visualising the data is which eigenvectors to project onto. The eigenvectors applying to each figure are stated in the captions. This is a matter of choosing the best vantage point to see the internal structure of the projection, to avoid coincidentally superimposed points. For example, we usually look from the front to back of a fish tank, but can sometimes see more by looking along the tank from one end or down into the tank from above. So it is with the projections resulting from KPCA.

V. RESULTS

Three fault conditions are considered in this section. Two of them arose in a previous publication of the authors [10] and are re-examined in the context of this new visualization approach. An additional fault was spotted and excised from the well-behaved transformer data as this paper was prepared. This fault will be discussed by way of an introduction.

A. A cluster of faulty data

When the data was visualized using the new KPCA-based projection it became apparent there was a previously unseen cluster of points. It is impossible to be confident from a single projection that points form a cluster, just as two stars that appear nearby in the night sky may be coincidentally superimposed along the same line of sight. In the case of KPCA it is possible to change the vantage point by selecting different eigenvectors, allowing the cluster to be verified from several different directions.

The 14 points forming the cluster were examined and it became apparent that they were all observations made of the same transformer. Moreover when these measurements were tested using the Duval triangle they were all classified as category DT.

The faults are shown as x crosses (in red) in Fig. 3. (Diagrams in this paper are in colour to assist in discerning the detail). This is generated using a kernel variance $\sigma = 0.6$. This is in contrast to the well-behaved transformer measurements portrayed as small (black) diamonds. In addition the four faults of the KGM are overlaid as follows: (a) overheated oil is shown as a + cross; (b) overheated cellulose as a hollow circle; (c) partial discharge in oil as a hollow square; and (d) arcing in oil as a hollow diamond.

The origin of the original linear space is shown near coordinate (0.4, -0.2), translated due to the centring of data in the Hilbert space. The six axes of the original linear representation are shown emerging from the origin. These show the directions within the space corresponding to increases in concentration of each of the key gases and are colour coded for easy comprehension. The order of the axes is (from top to bottom) carbon monoxide (blue), ethylene (cyan), methane (green), hydrogen (red), acetylene (black) and ethane (magenta). Note that the carbon monoxide axis terminates in proximity to the hollow circle since this is the appropriate fault signature for high CO (cf. Fig. 1 (b)). Likewise the ethylene axis ends near the + cross and hydrogen axis near the hollow square. The final fault signature (hollow diamond) is less clearly located as it consists of a mixture of hydrogen and acetylene.

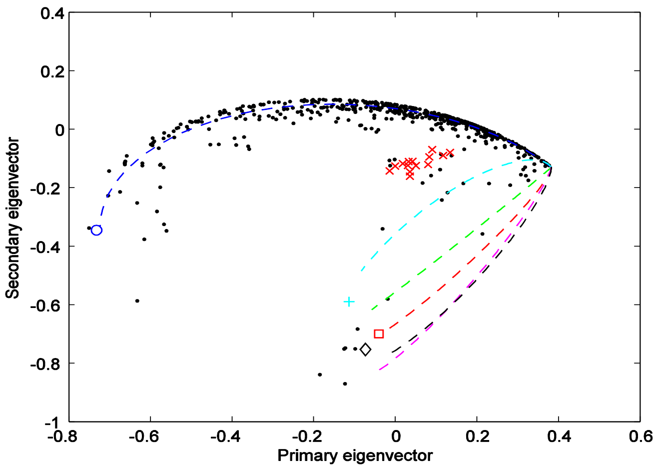


Fig. 3. KPCA projection of a discovered fault cluster

B. *Transformer B Case Study*

In an earlier publication [10] the authors highlighted a transformer that had failed in service and labelled it transformer B. It was also of interest for this project in that it had a long history of DGA data collection, with 12 measurements made between March 2003 and the time of failure in November 2010. The final measurement can be classified as a T3 fault using the Duval triangle method [10]. It was expected that the locus that resulted from this sequence of measurements could provide information about the time at which the transformer deteriorated.

Fig. 4 shows the locus superimposed on the elements from the previous figure. This plot required kernel variance σ = 0.9. Each of the measurements is shown by an x cross and these are joined by green line segments in chronological order, starting at the origin which is located near (0.3, 0.0) on this projection (March 2003) and ending at the top middle of the graph (November 2010). The transformer starts near the origin since it was already displaying errant behaviour prior to March 2003 and so its oil was degassed at this point so that it could be more clearly diagnosed.

It can be seen from the figure above that this locus moves along in the subspace occupied densely by well-behaved transformers (small black diamonds) for the first three measurements. In the 4th measurement (taken in November 2005) the locus has changed direction by about 90^{o} and risen vertically above the densely populated region. Thereafter it rises up through an area of the space which clearly has no well-behaved transformers.

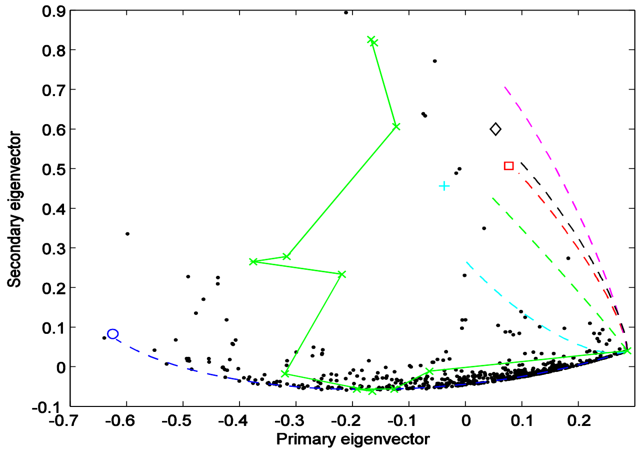


Fig. 4. KPCA projection of the locus of transformer B

C. *Transformer C Case Study*

As a further case study transformer C, also arising from [10], was examined. This unit was of interest because of its high acetylene levels which exceeded Transpower's criteria, warranting closer examination. The unit was partially degassed in 2009 providing just four measurements over the period June 2009 until July 2011. These resulted in a short locus that can be plotted on a KPCA projection, particularly since the final two values were very similar giving just three dispersed points.

The final projection, Fig. 5, shows the fourth and fifth most significant eigenvectors with kernel variance σ = 0.8. Viewed from this direction, most of the well-behaved transformers form a cluster in the middle of the projection. This case is interesting since transformer C's fault does not resemble any of the KGM signatures. Their locations can be seen bottom centre and on the right of the figure with the final circular signature near (-0.1, 0.2).

The magenta locus starts near (-0.1, 0.0) (June 2009) and moves to the left, ending with the final pair of points in proximity to (-0.4, 0.1) (July 2011). Again the projection makes it clear that the locus is moving away from the bulk of well-behaved transformers, but there are outliers in proximity to the termination point of the locus (and this has been confirmed by seeking other projections of the data). This may be accounted for by the range of variability of "normal" transformers, but could equally well indicate as yet undiscovered fault behaviour among the well-behaved examples. It reflects the fact that it is only possible to be confident in a fault diagnosis once a transformer has undergone an internal physical examination which is expensive to conduct. Due to the significant cost, maintenance tends to occur when the evidence is conclusive about the fault status.

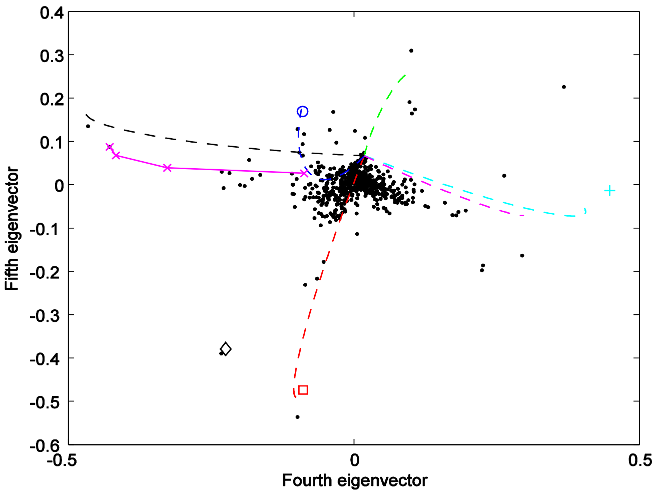


Fig. 5. KPCA projection of the locus of transformer C

Transformer C, not only does not resemble any KGM signature, but also cannot be diagnosed using Rogers' ratios. However the Duval triangle method diagnoses the final location of the locus as categories D1/D2.

VI. Discussion of Approach

There is other recent work [14], [15] providing graphical ways to diagnose fault behaviour of transformers based on DGA measurements, but much is derived from gas ratio combinations and relative content of fault gases. These methods seek out rules resulting in decision boundaries that partition a graph into regions associated with various fault categories.

In contrast the approach reported here has focused on a visualization technique that allows investigating engineers to decide when transformer behaviour has deviated away from the bulk of past measurements. The approach does not aspire to definitive classifications, but does include projections of the axes and key gas fault signatures to identify regions associated with various faults. The strength of this method is that it allows an investigator to visualize a large set of past measurements as a basis for comparison with a new one. Indeed the structure of the projection is derived from the span of past measurements. It also has the potential for chronological analysis, allowing a sequence of measurements to form a locus and this has been illustrated with a couple of case studies.

It is possible to convert additional signatures (such as those from the KGM) and superimpose them on the projection. However portraying boundaries, such as those from Rogers' ratios and the Duval triangle, within the projection is difficult. For example a boundary such as that resulting from the Rogers' rule $C_2H_4 / C_2H_6 \geq 3$ defines a 5 dimensional hyperplane and its surface and structure would be complex to visualize in KPCA projections. Moreover, it is only easy to see whether points are above or below the surface when viewing it edge on. It is hoped to unify the method proposed here with the three alternatives presented in Section II, but this may require a new visualization approach such as including semi-opaque surfaces in the projections and fading objects depending on their distance behind the surface. This will form the basis for future research.

VII. Conclusions

This project was a case study to determine whether a non-linear visualization approach, based on kernel principal component analysis, would offer benefits for managing a large collection of DGA measurements. Through several examples it has been demonstrated that the method allows clusters of unusual (possibly faulty) measurements to be located and differentiated from well-behaved transformer measurements. It has also been possible to plot the temporal evolution of suspect transformers, as loci in the space, to find the time at which their behaviour deteriorates.

The results of this study are encouraging and will lead to future work automating the visualization to a greater extent, attempting to unify the approach further with established methods such as Duval's triangle and possibly developing a "polished" implementation to be trialled in industry.

Acknowledgment

The authors would like to thank Transpower NZ Ltd. for sharing their transformer data, and Mr. Jon Brown, Senior AC Stations Engineer, Asset Management Services, Transpower NZ, for his technical input.

References

[1] IEEE Standard C57.104-2008, "IEEE guide for the interpretation of gases generated in oil-immersed transformers," 2009.

[2] R. R. Rogers, "IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis," IEEE Trans. in Electrical Insulation, vol. EI-13, no. 5, October 1978.

[3] M. Duval, "A review of faults detectable by gas-in-oil analysis in transformers," IEEE Electrical Insulation Magazine, Vol. 18, No. 3, pp. 8-17, May/June 2002.

[4] IEC 60599, "Mineral oil-impregnated electrcial equipment in service – guide to the interpretation of dissolved and free gas analysis," Edition 2.1, International Electrotechnical Commission, Geneva, Switzerland, 2007.

[5] M. J. Heathcote, J & P Transformer Book, 13th ed., Newnes, 2007, pp.558-615.

[6] W. H. Tang and Q. H. Wu, Condition Monitoring and Assessment of Power Transformers Using Computational Intelligence, Springer, 2011, pp. 105-123.

[7] Z. Peng and B. Song, "Research on transformer fault diagnosis expert system based on DGA database," icic, vol. 2, pp.29-31, Second International Conference on Information and Computing Science, 2009.

[8] C. Zhang, Y. Luo and X. Wen, "Study on fault diagnosis of transformer based on principal component analysis of dissolved gas," J. High Voltage Engineering, CNKI, vol. 2005-08, 2005, in Chinese.

[9] C. Kocaman, E. Kilic, and O. Ozgonenel, "Detection of transformer internal faults by using dynamic principle component analysis," 14th IEEE Signal Processing and Communication Applications Conf., Antalya, Turkey, 2006, in Turkish.

[10] S. N. Hettiwatte and H. A. Fonseka, "Analysis and interpretation of dissolved gases in transformer oil," IEEE Intl. Conf. on Condition Monitoring and Diagnosis, Bali, Indonesia, 2012, pp. 35-38.

[11] B. Scholkopf and A. J. Smola, Learning with Kernels, MIT Press, 2002, pp. 429-437.

[12] H. Hotelling, "Analysis of complex statistical variables into components," J. Educ. Psychol, vol. 24, 1933, pp. 417.

[13] N. Cristianini and J. Shaw-Taylor, Support Vector Machines, Cambridge University Press, 2000, pp. 30-41.

[14] S. W. Kim, S. J. Kim, H. D. Seo, J. R. Jung, H. J. Yang and M. Duval, "New methods of DGA diagnosis using IEC TC10 and related databases part 1: application of gas-ratio combinations," IEEE Trans. on Dielectrics and Electrical Insulation, vol. 20, no. 2, pp. 685-690, April 2013.

[15] S. J. Lee, Y. M. Kim, H. D. Seo, J. R. Jung, H. J. Yang and M. Duval, "New methods of DGA diagnosis using IEC TC10 and related databases part 2: application of relative content of fault gases," IEEE Trans. on Dielectrics and Electrical Insulation, vol. 20, no. 2, pp. 691-696, April 2013.