

# Improving the Accuracy of Transformer DGA Diagnosis in the Presence of Conflicting Evidence

Jose Ignacio Aizpurua, Victoria M. Catterson, Brian G. Stewart, and Stephen D. J. McArthur  
 Institute for Energy and Environment  
 University of Strathclyde  
 Glasgow, UK  
 jose.aizpurua@strath.ac.uk

Brandon Lambert, Bismark Ampofo  
 Bruce Power  
 Tiverton, Canada  
 Gavin Pereira, James G. Cross  
 Kinectrics Inc.  
 Toronto, Canada

**Abstract**— Transformers are critical assets for the reliable and cost-effective operation of the power grid. Transformers may fail if condition monitoring does not identify degraded conditions in time. Dissolved Gas Analysis (DGA) focuses on the examination of the dissolved gasses in the transformer oil and there exist different methods for transformer fault diagnosis based on different analyses of the gassing levels. However, these methods can give conflicting results, and it is not always clear which model is most accurate in a given situation. This paper presents a novel evidence combination framework for DGA based on Bayesian networks. Bayesian network models embed expert knowledge along with learned data patterns and evidence combination which aids in the consistency of analysis. The effectiveness of the proposed framework is validated using the IEC TC 10 dataset with a maximum diagnosis accuracy of 88.3%.

**Keywords**—DGA, transformer diagnosis, condition monitoring, Bayesian networks, evidence combination, ensembles

## I. INTRODUCTION

Dissolved Gas Analysis (DGA) is a very common method of assessing the health of transformers. There are a number of standard diagnostic techniques which are used to determine the type of problem causing gassing within the transformer, such as Duval's Triangle [1], Rogers' ratio [2], and Doernenburg's ratio [3]. When these techniques reach the same conclusion about the type of fault present, this can give increased confidence in the veracity of the diagnosis. However, in those cases where there is a disagreement between techniques, it is not always clear how to reach an appropriate conclusion.

Within the literature, there have been various approaches to solving this problem. The logic-based Evidential Reasoning technique has been shown to perform well for given case studies [4], and Dempster-Shafer Theory has been shown to give good accuracy on a larger database of samples [5]. However, a significant portion of DGA samples may be considered relatively easy to diagnose, since the techniques are in broad agreement. The accuracy of an evidence combination technique on a database of largely straightforward cases is not necessarily informative about the technique's applicability in the field. It is the performance of an evidence combination technique on cases where there is conflict in diagnosis that is crucial for the adoption of such a method by the industry.

This paper presents a Bayesian network framework for evidence combination which is particularly suited to combining the results of DGA diagnostic methods.

## II. RATIO-BASED DGA METHODS

Ratio-based DGA methods classify transformer faults through predefined gas ratio values. Table I shows Doernenburg's classification ratio values [3], where  $R_1=CH_4/H_2$ ,  $R_2=C_2H_2/C_2H_4$ ,  $R_3=C_2H_2/CH_4$  and  $R_4=C_2H_6/C_2H_2$ .

TABLE I. DOERNENBURG'S CLASSIFICATION RATIOS [3]

$R_1$	$R_2$	$R_3$	$R_4$	Diagnosis
>1	<0.75	<0.3	>0.4	Thermal
<0.1	N/A	<0.3	>0.4	PD
0.1-1	>0.75	>0.3	<0.4	Arcing

Duval's triangle evaluates the relative amount of three gasses ( $C_2H_2$ ,  $CH_4$ ,  $C_2H_4$ ) and classifies them within the coordinates of a triangle [1]. These coordinates define an area which can be transformed into a numerical classification method as displayed in Table II [6].

TABLE II. DUVAL'S CLASSIFICATION GASES [6]

$C_2H_2\%$	$CH_4\%$	$C_2H_4\%$	Diagnosis
0-0.02	0.98-1	0-0.02	PD
0-0.04	0.46-0.8	0.2-0.5	Thermal $300^\circ C < T < 700^\circ C$
	0.76-0.98	0.02-0.2	Thermal $T < 300^\circ C$
0-0.15	0-0.5	0.5-1	Thermal $T > 700^\circ C$
0.04-0.13	0.47-0.96	0-0.4	Mixture of thermal and electrical faults
0.13-0.29	0.21-0.56	0.4-0.5	
0.15-0.29	0-0.35	0.5-0.85	
0.13-0.29	0.31-0.64	0.23-0.4	High energy discharge (Arcing)
0.29-0.77	0-0.48	0.23-0.71	Low energy discharge (Arcing)

Table III displays Rogers' ratios ( $R_1=CH_4/H_2$ ,  $R_2=C_2H_2/C_2H_4$ ,  $R_5=C_2H_4/C_2H_6$ ) as proposed in [2].

The fault types diagnosed by each method are different. In this work all diagnoses are classified into four groups: *Thermal*, *Arcing* (including high energy discharges), *partial discharge (PD)*, and *Normal* degradation. These methods also assign a diagnosis with full confidence, regardless of proximity to a diagnostic boundary, and this may lead to conflicting situations. The proposed probabilistic diagnosis framework addresses this issue effectively through Bayesian networks.

TABLE III. ROGERS' CLASSIFICATION VALUES [2]

$R_1$	$R_2$	$R_5$	Diagnosis
>0.1-1	<0.1	<1	Normal degradation
<0.1	<0.1	<1	PD
0.1-1	0.1-3	>3	Arcing
>0.1-1	<0.1	1-3	Low temperature thermal
>1	<0.1	1-3	Thermal <700°C
>1	<0.1	>3	Thermal >700°C

### III. BASICS OF BAYESIAN NETWORKS

Bayesian networks (BN) use stochastic graphical models to represent dependencies among random variables [7]. These models are known as directed acyclic graphs (DAG). The structure of the BN model is interpretable such that each state can be mapped to the health condition of the modelled system. Let us assume that the DAG is comprised of  $p$  random variables, denoted  $X=\{X_1=x_1, \dots, X_p=x_p\}$ , which are linked through edges to reflect dependencies (see Fig. 1). In BN terminology a node  $x_1$  is said to be a parent of another node  $x_2$  if there exists an edge from  $x_1$  to  $x_2$ , and  $x_2$  is a child of  $x_1$ .

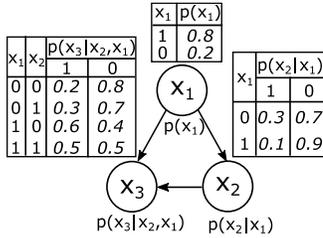


Fig. 1. Bayesian network example.

Statistically, dependencies are quantified through conditional probabilities. Bayes' theory states that the posterior probability,  $P(x_2|x_1)$ , can be estimated by multiplying the likelihood,  $P(x_1|x_2)$ , and the prior probability,  $P(x_2)$ , and normalizing with the probability of evidence,  $P(x_1)$  [7]:

$$P(x_2 | x_1) = P(x_1 | x_2)P(x_2) / P(x_1) \quad (1)$$

Bayesian networks are a compact representation of joint probability distributions [7]. 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 [7]. Accordingly the joint distribution of the set of random variables  $X=\{X_1=x_1, \dots, X_p=x_p\}$  is defined as follows:

$$P(X) = \prod_{i=1}^p P(x_i | x_1, \dots, x_{p-1}) \quad (2)$$

Using the information encoded in the DAG, (2) can be simplified to account only for parent nodes. Namely, if  $X$  is comprised of discrete random variables (assumed throughout the paper), the joint probability density function (the global distribution) is represented as a product of conditional probability distributions (the local distributions associated with each variable  $x_i \in X, 1 < i < p$ ) [7]:

$$P(X) = \prod_{i=1}^p P(x_i | x_{pa(i)}) \quad (3)$$

where  $x_{pa(i)}$  is the set of parents of  $x_i$  and  $P(x_i|x_{pa(i)})$  is the conditional probability distribution containing one distribution for each variable.

The strength of relations among dependent nodes are synthesized through conditional probability tables (CPT). CPTs are defined for each node  $x_i \in X$  of the BN model expressing the conditional probability distributions (CPDs) for all the parent nodes  $x_{pa(i)}$ . If the nodes are discrete random variables, the CPD can be expressed as a multinomial distribution [7]. If a node does not have parents it will have a marginal probability table (e.g., node  $x_1$  in Fig. 1). If a node does have parents, each cell of the CPT will specify the conditional probability for the node being in a specific state given a specific configuration of the states of its parent nodes (nodes  $x_2$  and  $x_3$  in Fig. 1).

A BN model is completely defined with the DAG and the conditional probabilities between the nodes, i.e.  $BN=(DAG, \theta)$ , where  $\theta$  denotes the parameters of the CPT. The process of estimating the conditional probabilities between nodes is called *parameter learning* and the process of estimating the posterior distributions (i.e. diagnosis in the presence of specific data) is called *probabilistic inference*.

CPT values can be estimated via parameter estimation methods. This paper focuses on maximum likelihood estimation (MLE) which maximizes the likelihood of making the observations given the parameters. Given a training dataset  $D=\{D_1, \dots, D_{|train|}\}$ , first it is necessary to estimate the likelihood  $L$  that the dataset was generated by the model  $BN=(DAG, \theta)$ , and then find the maximum likelihood estimator,  $\hat{\theta}$ :

$$\hat{\theta} = \arg \max \{ L(\theta | D) \} \quad (4)$$

After determining the structure of the BN and learning the parameters, the BN model can be used to make inferences. The goal of this work is to perform diagnosis or causal analysis, reasoning from effects (measured DGA values) to causes (transformer faults). That is, measured gas values are given as evidence to the Bayesian network model and through the DAG structure the posterior probability of possible causes is evaluated,  $P(fault | gas \text{ data})$ . The posterior probability of each node enables inferences about the status of unobserved parameters and the most likely status for the node.

### IV. EVIDENCE COMBINATION THROUGH BAYESIAN NETWORKS

Fig. 2 shows the proposed evidence combination framework based on BN models created using Rogers' ratios, Doernenburg's ratios, and Duval's gases. The proposed approach is divided into cross-validation, data pre-processing, learning, inference and evidence combination stages.

The goal of the cross-validation stage is to validate the results and assess how they will generalize to an independent dataset. To this end, Monte Carlo cross-validations are implemented as follows [8]: (i) initialize the trial counter,  $trials=0$ ; (ii) randomly shuffle the dataset and execute preprocessing, learning, inference, and evidence combination steps, and store the results; (iii) if  $trials < Max\_trials$  iterate from the previous step and increase the trial counter by 1; (iv) otherwise extract mean and standard deviation values of the stored diagnosis results. For each trial, the random shuffle and the train/test steps generate different training and testing datasets, and therefore, this process evaluates the framework

with  $Max\_trials$  different training and testing datasets ( $Max\_trials=10^3$ ). As a result this process generates repeatable and consistent diagnosis results.

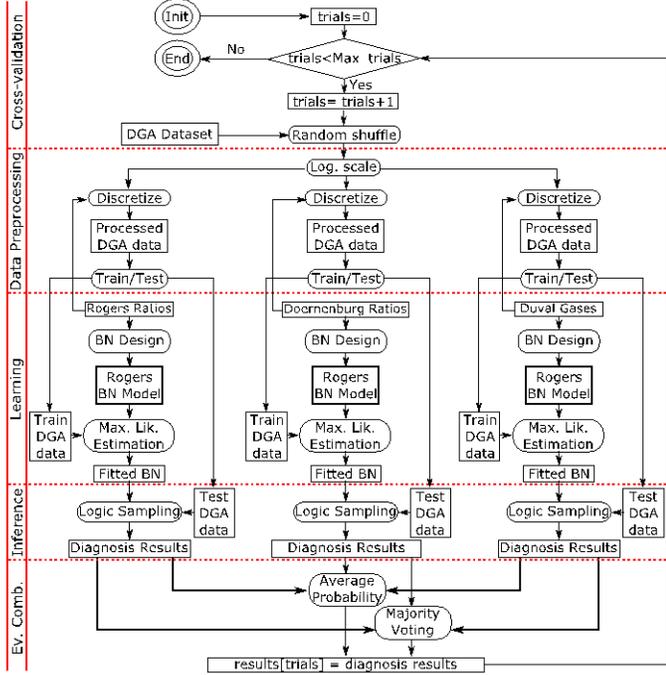


Fig. 2. Proposed Bayesian evidence combination framework.

The *data preprocessing* stage starts by applying a *log-scale* step because diagnostic information does not reside in absolute gas values but instead in the order of magnitude [9]. Firstly the logarithm of every gas sample in the dataset is taken and then each variable in the dataset is scaled to mean zero and standard deviation one. This is done for each gas within the dataset, by subtracting the mean value and dividing by the standard deviation, for each sample of the variable.

Then a *discretization* step is applied particular to each specific ratio-based DGA method (Fig. 2). Table IV shows an example for the Doernenburg case according to the ratio values in Table I. The same discretization procedure is applied to the ratios defined in Tables II and III. The coding for the occurrence of failure modes is specified with binary coding (e.g. PD fault:  $PD=1$ ,  $Arc=0$ ,  $Thermal=0$ ,  $Normal=0$ ). Subsequently, the discretized dataset is divided into *train* and *test* datasets using 80% and 20% of the randomly shuffled dataset, respectively.

TABLE IV. DOERNENBURG'S CODING VALUES

Ratio	$R_1$			$R_2$		$R_3$		$R_4$	
Code	0	1	2	0	1	0	1	0	1
Range	$\leq 0.1$	0.1-1	$> 1$	$\leq 0.75$	$> 0.75$	$\leq 0.3$	$> 0.3$	$\leq 0.4$	$> 0.4$

#### A. Parameter Learning and Inference through BN

The DGA diagnosis through BN starts by *designing the BN*. To this end, expert knowledge is elicited from industry standards [3]. For instance, in the Doernenburg case (Table I) the *Thermal* fault depends on  $R_1$ ,  $R_2$ ,  $R_3$ , and  $R_4$ , whereas the *PD* fault depends only on  $R_1$ ,  $R_3$ , and  $R_4$ . Although the

classification ratios in Table I do not include *Normal* degradation values, when designing the BN model it is assumed that the *Normal* condition is indicated by the remainder values of all the ratios (i.e. when there is no diagnosis of *Thermal*, *Arc*, or *PD*). This allows the model to be trained for normal degradation data. Fig. 3 shows the BN model for Doernenburg's ratios.

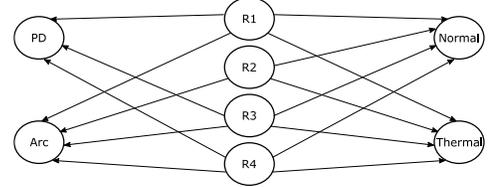


Fig. 3. Doernenburg's Bayesian network.

The same process is applied to extract BN models for Duval and Rogers' ratios. After designing the BN model, the next step is the implementation of the parameter learning through *MLE* using the training dataset. This step quantifies the CPT and Table V shows the CPT for a specific configuration of the Doernenburg BN model. When there is a *Thermal* fault, the value of  $R_1$  is likely (88%) to be in the range determined by the discretized number 2 ( $R_1 > 1$ , see Table IV). This value matches with the *Thermal* fault displayed in Table I. This node supplies one piece of probabilistic evidence for diagnosis. By utilizing the evidence from all nodes for a given sample, all the possible failure types can be evaluated probabilistically.

TABLE V. SUBSET OF THE CPT LEARNED FROM DATA

$R_1$	$PD=0, Arc=0, Thermal=1, Normal=0$
0	0.04
1	0.08
2	0.88

The next step is the *inference* which quantifies posterior probabilities through the *logic sampling* (LS) algorithm [7]. Given the test data and the BN model, LS infers the probability of each fault,  $P(fault | gas\ data)$ . The fault with the highest likelihood is the final diagnosis of the model. The probabilistic diagnosis framework generates more information about the strength of belief in a given diagnosis compared with the ratio technique, and this is combined into the overall decision. The same data processing, learning and inference steps are applied for Duval and Rogers ratios.

#### B. Evidence Combination

When ratio-based DGA methods agree on the diagnosis it increases confidence in its veracity, but sometimes there are conflicts among their outcomes. The application of BN models for DGA increases the diagnosis accuracy and also provides mechanisms to combine the outcomes of different methods in a logical manner.

For example, given the gas values ( $R_1=0.25$ ,  $R_2=0.15$ ,  $R_3=0.6$ ,  $R_4=1.6$ ,  $R_5=4$ ,  $\%CH_4=0.1$ ,  $\%C_2H_2=0.1$ ,  $C_2H_4=0.71$ ), if ratio-based DGA methods are used, Rogers will classify it as an *Arc* fault, Duval as a *Thermal* fault, and Doernenburg will not classify it because it is not in the bounds of Table I. After the implementation of the proposed BN framework for Rogers,

Doernenburg and Duval (Fig. 2), the classification results are shown in Fig. 4. Doernenburg’s BN classifies it as *Normal* degradation, Rogers’ BN classifies it as an *Arc* fault, and Duval’s BN classifies it as *Normal* degradation. The inclusion of *Normal* values in the BN model enabled Doernenburg’s BN model to learn the normal degradation pattern (Fig. 3).

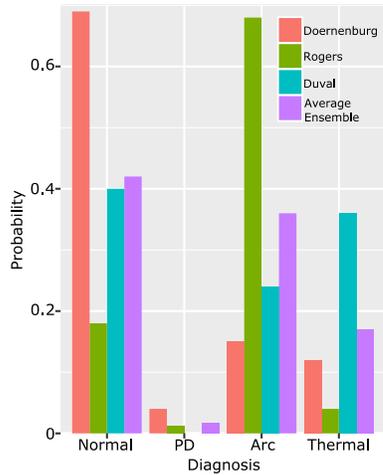


Fig. 4. Classification results example.

This conflict situation can be resolved using evidential reasoning methods [10]. This work focuses on arithmetic-based evidence combination methods:

(i) *Majority voting*: taking into account the outcome of each technique, the votes of each technique are counted, and the fault with the highest number of votes is assigned. In case of equal votes, the fault with the highest likelihood is selected.

(ii) *Average ensemble*: after generating the probabilistic outputs of each technique, the mean probability is quantified for each fault considering all the outcomes of all the techniques and assigns the corresponding fault to the highest likelihood fault.

Majority voting diagnoses the case above as *Normal* degradation with full confidence, whereas the average ensemble diagnoses *Normal* degradation with probabilistic values as shown in Fig. 4. Both results match with the actual transformer’s health state.

## V. RESULTS

Ratio-based DGA methods, BN models, and evidence combination methods were tested on the IEC TC 10 dataset [11] through the procedure introduced in Section IV. This dataset is comprised of 167 samples of seven different gasses ( $C_2H_6$ ,  $C_2H_4$ ,  $H_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $CO$ ,  $CO_2$ ), labelled with their corresponding failure causes as follows: 5.3% *PD*, 44.4% *Arc*, 20.4% *Thermal*, 29.9% *Normal*. Table VI displays the results for each of these cases. The performance was measured through the accuracy of the classifier with respect to the labelled failure causes in the IEC TC10 database.

The highest accuracy is obtained with the average probability model with a maximum accuracy of 88.3%. As for the accuracy for each failure mode, the highest accuracy for the thermal and PD faults is obtained with the Duval’s triangle

(87.7%, 96.9%), the arcing fault is best captured with the average probability ensemble, and normal degradation with the Doernenburg’s Bayesian network model (64%).

TABLE VI. ACCURACY RESULTS FOR DIFFERENT DIAGNOSIS METHODS

Approach	Accuracy				
	Overall	Thermal	PD	Arc	Normal
<b>Rogers Ratio</b>	42.39% ± 7.4%	58.5% ± 18.72%	12.7% ± 33.53%	66.2% ± 11.4%	3.7% ± 11%
<b>Doern. Ratio</b>	60.8% ± 6.5%	74.3% ± 16.9%	73.9% ± 35.2%	94% ± 6%	0%
<b>Duval Triangle</b>	64.79% ± 7.4%	<b>87.7% ± 12.6%</b>	<b>96.9% ± 17%</b>	70.7% ± 10%	0%
<b>Rogers BN</b>	73.76% ± 7.2%	70.2% ± 18%	34.08% ± 13.6%	93.72% ± 6.1%	61.5% ± 15.8%
<b>Doern* BN</b>	79.86% ± 6.6%	74.3% ± 16.9%	73.9% ± 35.2%	94% ± 6%	<b>64% ± 15%</b>
<b>Duval* BN</b>	72.8% ± 7%	49.9% ± 20.2%	96.3% ± 18.8%	96.3% ± 5%	49.7% ± 16.1%
<b>Majority Voting</b>	72.75% ± 7%	78.76% ± 15.2%	50.4% ± 39.8%	91.9% ± 6.5%	43.7% ± 14.8%
<b>Average Probability</b>	<b>81.2% ± 7.1%</b>	83.63% ± 14.3%	69.9% ± 37.25%	<b>98.28% ± 3.4%</b>	56.56% ± 42.5%

\* including normal states in the BN model

## VI. CONCLUSIONS

This paper presents an evidence combination approach based on Bayesian networks. The novel method takes into account expert knowledge embedded in industry standard DGA evaluation methods (Rogers, Duval, Doernenburg), increases the engineering confidence in the results by removing inconsistencies, and obtains a maximum diagnosis accuracy of 88.3% tested on the IEC TC10 dataset [10].

## REFERENCES

- [1] M. Duval, "A review of faults detectable by gas-in-oil analysis in transformers," IEEE Electr. Insul. Mag., vol. 18, pp. 8-17, 2002.
- [2] R. R. Rogers, "IEEE and IEC Codes to interpret incipient faults in transformers, using gas in oil analysis," IEEE Trans. Dielectr. Electr. Insul., vol. EI-13, pp. 349-354, 1978.
- [3] "IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers," IEEE Std C57.104-2008, pp. 1-36, 2009.
- [4] K. Spurgeon, W. H. Tang, Q. H. Wu, Z. J. Richardson, and G. Moss, "Dissolved gas analysis using evidential reasoning," IEE Proceedings - Science, Measurement and Technology, vol. 152, pp. 110-117, 2005.
- [5] M. Lee Hui and C. S. Chang, "Application of Dempster-Shafer's theory of evidence for transformer incipient fault diagnosis," In Proc. APSCOM 2009 Conference, 2009, pp. 1-6.
- [6] S. S. Desouky, A. E. Kalas, R. A. A. El-Aal, and A. M. M. Hassan, "Modification of Duval triangle for diagnostic transformer fault through a procedure of dissolved gases analysis," In IEEEIC 2016, 2016, pp. 1-5.
- [7] R. Nagarajan, M. Scutari, and S. Lbre, Bayesian networks in R: with Applications in Systems Biology: Springer, Incorporated, 2013.
- [8] Q.-S. Xu and Y.-Z. Liang, "Monte Carlo cross validation," Chemometrics and Intelligent Laboratory Systems, vol. 56, pp. 1-11, 2001.
- [9] P. Mirowski and Y. LeCun, "Statistical machine learning and dissolved gas analysis: a review," IEEE Trans. Power Del., vol. 27, pp. 1791-1799, 2012.
- [10] V. M. Catterson, S. D. J. McArthur, "Using evidence combination for transformer defect diagnosis," International Journal Innovations in Energy Systems and Power, vol. 1, no. 1, 2006
- [11] M. Duval and A. dePablo, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases," IEEE Electr. Insul. Mag., vol. 17, pp. 31-41, 2001.