

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Combination of multi-paradigm models for Power Transformer fault prediction

Francisco Aguiar



MESTRADO EM ENGENHARIA ELETROTÉCNICA E DE COMPUTADORES

Supervisor: Ricardo Jorge Teixeira de Sousa

Second Supervisor: António Manuel Lucas Soares

August 21, 2022

Abstract

Growing environmental concerns dictate how new products are designed. Power Transformers are a well-established device, making its optimization a challenge. Though, through the emerging technology of Machine Learning and combinations of the established and cutting edge models, new approaches to these machines are possible.

Fundamental services are capable of being integrated into newer products, offering innovative solutions to prominent issues. Predictive and prescriptive services are crucial, opening the possibility for anticipating faults and setting up the next best action to take. Thus, a particular focus will be given to transformer fault prediction using Dissolved Gas Analysis with combinations of established empirical and Machine Learning models.

Fault detection is tackled, to solve the problem of plentiful unlabelled data sources. Artificial generation and extension of data is also approached, so available Dissolved Gas Analysis databases are expanded to present more useful situations. This methodology will also pave the way to real world data.

A collection of models is suggested for a possible solution and its implementation, from empirical models, such as the Duval Triangle, to Machine Learning algorithms, such as Artificial Neural Networks. The attained data set is characterized, further expanded and labelled.

Subsequently, several Machine Learning models are trained and developed to provide future insight on the Power Transformer's behaviour and condition. A comparison to assess the performance of these algorithms is accomplished, revealing unexpected results, but ambitious prospects.

Acknowledgments

The task of writing this scientific document was not of a sole person, nor was this initiation to the scientific exploration. So, I would like to express my appreciation to the people who were there for me.

Firstly, I'm deeply indebted to Ricardo Sousa, whom without, this whole work would not be possible. I am very much grateful for my supervisor's guidance and good will.

I would also like to extend my gratitude to José Pinto, whose help cannot be overestimated. Particularly, during a time where matters were burdensome.

I am thankful for the valuable criticism that Professor António Lucas Soares gave me, while writing this document.

I also wish to thank INESC TEC, notably the Laboratório de Inteligência Artificial e Apoio à Decisão.

I would like to recognize the effort of Cristiano, Agostinho and the rest of the team at EFACEC.

I cannot finish without thanking my professors and colleagues at FEUP, who motivated me throughout my journey as an engineer student.

Por fim, quero agradecer verdadeiramente aos meus amigos e família. Sem a paciência, atenção e zelo deles, eu não alcançaria o meu propósito.

Obrigado,

Francisco Aguiar

“Nothing can have value without being an object of utility.”

Karl Marx

Contents

1	Introduction	1
1.1	Context	1
1.2	Motivation	2
1.3	Objectives	2
1.4	Structure of the document	3
2	State of the art review	5
2.1	Introduction	5
2.2	Diagnostic Techniques	6
2.3	Dissolved Gas Analysis	8
2.4	Machine Learning solutions	11
2.4.1	Supervised Learning	12
2.4.2	Unsupervised Learning	13
2.4.3	Semi-supervised Learning	14
2.5	Fault Prediction	15
2.6	Prediction Paradigms	17
2.6.1	Random Forest	17
2.6.2	Recurrent Neural Network	18
2.6.3	Long Short-Term Memory	19
2.7	Synthetic Data Generation	19
2.7.1	Auto-regressive Integrated Moving Average	20
2.8	Chapter Conclusion	20
3	Methodology	23
3.1	Overview	23
3.2	Datasets	24
3.3	Data Analysis	24
3.4	Artificial Labeling	25
3.5	Data augmentation	26
3.6	Modelization	26
3.7	Test and validation	29
3.8	Chapter Conclusion	30
4	Results	31
4.1	Data Analysis	31
4.2	Artificial Labeling	37
4.3	Data augmentation	41
4.4	Fault Prediction	43

4.4.1	Regression models	43
4.4.2	Auto-regressive Integrated Moving Average	47
4.4.3	Long Short-Term Memory	48
4.4.4	Curve fitting	51
4.5	Result Discussion	52
4.6	Chapter Conclusion	53
5	Conclusion	55
5.1	General conclusion	55
5.2	Contributions	56
5.3	Limitations	56
5.4	Future work	56
	References	59

List of Figures

2.1	Electrical Test Diagnostic Matrix [Fofana <i>et al.</i>].	6
2.2	Oil diagnostic techniques [N'Cho <i>et al.</i>].	7
2.3	Relative percentage of dissolved gas concentrations as function of temperature and fault type, according to the IEEE standard.	8
2.4	Typical faults, according to the IEC standard.	9
2.5	Duval and Low Energy Degradation Triangles.	10
2.6	Graphical representation of data clustering, [Dias <i>et al.</i>].	14
2.7	Degradation value (R) of a power transformer throughout the years, [Moodley]. .	16
2.8	Random Forest diagram.	17
3.1	Pipeline of the proposed methodology.	23
3.2	Schematic for one step ahead prediction using regression models.	27
3.3	Schematic for multi step ahead prediction using regression models.	27
3.4	Schematic for multi step ahead prediction using LSTM.	28
3.5	LSTM architecture diagram.	28
4.1	Plotting the Transformer C data frame.	32
4.2	Concentration histograms for (a) H ₂ , (b) CH ₄ , (c) C ₂ H ₂ , (d) C ₂ H ₄ , (e) C ₂ H ₆ and (f) CO gases.	32
4.2	Concentration histograms for (a) H ₂ , (b) CH ₄ , (c) C ₂ H ₂ , (d) C ₂ H ₄ , (e) C ₂ H ₆ and (f) CO gases.	33
4.3	Transformer C second region detailed.	34
4.4	Correlation through time between H ₂ and CO, with frame size 360 and step 1. . .	36
4.5	Correlation through time between CH ₄ and C ₂ H ₆ , with frame size 360 and step 1. .	36
4.6	Hydrocarbons in the second region of Transformer C.	37
4.7	Total Dissolved Combustible Gases in region 2 of Transformer C.	38
4.8	TDCG region plot of Transformers (a) H, (b) J, (c) L and (d) M.	39
4.9	Faulty type plot of Transformers (a) H, (b) J, (c) L and (d) M.	40
4.10	Augmented time series of region 2 of Transformer C.	41
4.11	Total Dissolved Combustible Gases in region 2 of Transformer C extended data set. .	42
4.12	Augmented biased time series of region 2 of Transformer C.	42
4.13	Total Dissolved Combustible Gases in region 2 of Transformer C extended and biased data set.	43
4.14	Transformer (a) J, (b) C region 2, (c) F region 1, (d) F region 2, (e) H, (f) L, (g) C region 1, (h) C region 4.	44
4.15	Regression and one time step prediction of Transformer C region 2.	45
4.16	Regression model 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7. . . .	45
4.16	Regression model 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7. . . .	46

4.17	ARIMA model 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7.	47
4.18	LSTM one time step ahead prediction of Transformer C region 2.	49
4.19	LSTM model 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7.	49
4.19	LSTM model 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7.	50
4.20	LSTM model 30 time steps prediction of data sets (a) 0 and (b) 2.	50
4.21	Curve fitting 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7.	51

List of Tables

3.1	Fault types in a Power Transformer.	25
3.2	Total Dissolved Combustible Gases Concentration.	25
4.1	Transformer C data frame.	31
4.2	Transformer C region 2 statistical analysis.	34
4.3	Transformer C region 2 correlation analysis, with the maximum values pointed out in bold.	35
4.4	Correlation analysis of the entire data set.	37
4.5	30 step ahead regression prediction performance.	46
4.6	30 step ahead ARIMA prediction performance.	48
4.7	Best 30 step ahead LSTM prediction performance parameters.	49
4.8	30 step ahead LSTM prediction performance.	50
4.9	30 step ahead Curve Fitting prediction performance.	52
4.10	Final 30 step ahead prediction performance.	52

Abbreviations

ANN	Artificial Neural Network
ARIMA	Auto-Regressive Integrated Moving Average
DGA	Dissolved Gas Analysis
DT	Digital Twin
FRA	Frequency Response Analysis
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
KNN	k Nearest Neighbor
LEDT	Low Energy Degradation Triangle
LSTM	Long Short-Term Memory
ML	Machine Learning
RF	Random Forest
RNN	Recurrent Neural Network
SVM	Support Vector Machine
TDCG	Total Dissolved Combustible Gas

Chapter 1

Introduction

Humanity has never before been so dependent on energy. The Power Transformer, an essential part of today's energy sector, is key to an effective and successful operation of the electrical system. This field of operation is, and will be in constant development, which presents us with great questions and challenges.

In this dissertation, an approach to Power Transformer fault prediction is achieved through analysis of in oil dissolved gases and the combination of different paradigms, such as established empirical models, but also Machine Learning algorithms.

1.1 Context

Nowadays an adequate device operation and preservation is vital, given the increasing concerns of energy efficiency and resource planning. Power transformers are no exception, considering that they are an essential part of the electrical grid and most industrial businesses, but also big and costly machines. Predicting its behaviors and diagnosing future problems is beneficial to the operators of these machines.

Since last century, the Power Transformer has been the universal solution for AC transmission systems. Although its development might seem that is settling down, there are still slight improvements to be made, both to the machine and its provided services.

Possible services that better the performance of any device are diagnostic, predictive, prescriptive, descriptive and visualization services. With these, closely monitoring the operation of a complex machine is more accessible, allowing responsible operators to act accordingly to the circumstances observed. Particularly, predicting Power Transformer's faults is pivotal. After all, anticipating failures allows preventive actions to be taken and maintains power supply.

These services can be combined into a Digital Twin (DT) system, encompassing every aspect of interest in one role. The Power Transformer is a well-known machine that indubitably can be modeled digitally, which would open new opportunities for its development and optimization.

Furthermore, most of the Power Transformers in service are now reaching their end of life. Thus, making sure that these machines accomplish their expected operation lifetime and that the new ones are built and installed with past issues in mind is substantial.

1.2 Motivation

Concerning the improvement of the Power Transformer and its provided services, to examine the machine's health and insulation degradation is crucial. Monitoring these aspects allows remaining useful life predictions and prognostic analysis, that is, estimating the type and the moment of the next fault.

This monitoring, however, must be approachable, in terms of cost and labor, but also accurate enough to provide substantial reports. Thus, advisedly predicting a faulty state in a Power Transformer is of the most importance, since it enables the possibility of thoughtful response and predictive maintenance.

The possibility of predictive maintenance allows the operation of a machine to its fullest, by means of reducing downtime and prolonging life. Upkeep costs also tend to decrease, since maintenance is carried only when necessary. For the same reasons, device's life is extended, resulting in an enhanced productivity.

1.3 Objectives

This work aims to integrate predictive maintenance in electrical grids and other uses of power transformers. The result of this implementation is a more reliable and energy efficient system, a legitimate concern in today's world.

Hence, the main objective of this project is to create a model of a Power Transformer, that combines different paradigms, capable of integrating a DT object. Although the DT will integrate services such as planning, monitoring and maintenance, the foremost development is fault prediction. The model will be useful for simulation and prediction of the transformer's behaviour and will combine several techniques from the literature and ML models.

A particular focus will be given to transformer fault prediction using a combination of different approaches. The most common method of diagnosing faults and its types in a transformer is through Dissolved Gas Analysis (DGA). There are many ways of analysing DGA, although the international standards provide general and specific guidance on the interpretation of dissolved and free gases on oil based equipments, such as Power Transformers. In addition to the already established empirical models, such as the Duval Triangle, Artificial Neural Networks (ANN) or other Machine Learning (ML) models can be used to build relationships between the data and the asset to provide future insight.

As a secondary objective, the thesis will also focus on solving the problem of unlabelled and scarce DGA data. To classify this data, well-developed methods proposed by researchers will be used. To further augment the data sets, ML models will be employed.

1.4 Structure of the document

Apart from this Introduction, this document includes 3 more chapters. Section 2 presents the background on fault identification and prediction in Power Transformers. Particular consideration is given to DGA, since it is the method of diagnosis of most interest to this work. This section also focuses on state of the art ML models for forecast and augmentation. Section 3 details the method of work and characterizes the available data and the approaches taken. Section 4 presents and interprets the results of the applied methodology. Section 5 concludes the topic in question, reviewing the work and setting future perspectives.

Chapter 2

State of the art review

In this chapter, the concept of fault in a Power Transformer is established. Fault diagnosing and prediction is addressed and the most practiced methods are introduced. The followed approach is to introduce the state of the art from a broader context leading to the specific DGA methodology. A collection of Machine Learning models used in the literature is also presented.

2.1 Introduction

According to the IEC 60599, a fault in a Power Transformer is described as "an unplanned occurrence or defect in an item which may result in one or more failures of the item itself or of other associated equipment" and it can be either a non-damage or damage fault, the hindmost requiring some sort of action (repair or replacement) [1].

A fault usually leads to disturbances in the power supply, damage to susceptible electrical devices and sometimes even blackouts. Its causes range from core defects to insulation breakdowns. The latter being the most common motive, since throughout the machine's life, its insulation is continuously in exertion.

Therefore, a Power Transformer's age is often related to its internal insulation condition. As the machine is exposed to a wide range of stresses, such as electrical, mechanical and thermal, its insulation deteriorates. This degradation in the dielectric material is often followed by an increase in occurrence and severity of faults.

In the interest of preserving the transformer and its provided services, to check the device's health, but also to anticipate its defects, is vital. The monitoring must be accessible, in terms of cost and labor, but also accurate enough to provide meaningful results. Reasonably predicting a faulty state in a Power Transformer is beneficial, since it enables the possibility of thoughtful response and predictive maintenance.

In the next section, some of the tools used to monitor the condition of the Power Transformer are defined.

2.2 Diagnostic Techniques

Fofona and Hadjadj propose that the health evaluation of the machine must incorporate both electrical and physicochemical techniques. These usually involve measuring electrical parameters, physical and chemical properties, such as resistances, temperature, gas concentrations, etc. [2, 3]. Figure 2.1 shows which problems are able to be detected using only electrical analysis.

		Type of Problem					
		Magnetic Circuit Integrity	Magnetic Circuit Insulation	Winding Geometry	Winding/Bushing/OLTC Continuity	Winding/Bushing Insulation	Winding Turn to Turn Insulation
		Diagnostic Technique					
Basic Electrical	Winding ratio	•					
	Winding Resistance			•			
	Magnetisation current	•					•
	Capacitance and DF/PF		•		•	•	•
	Leakage Reactance				•		
	Insulation Resistance		•			•	
	Core Ground Test					•	
Advanced Electrical	Frequency Response of Stray Losses			•	•		
	Frequency Response Analysis	•			•	•	•
	Polarisation/ Depolarisation		•				
	Frequency Domain Spectroscopy		•				
	Recovery Voltage Method		•				
	Electrical Detection of PD	•	•				
	Acoustical Detection of PD	•	•				
	UHF Detection of PD	•	•				

Figure 2.1: Electrical Test Diagnostic Matrix [Fofana *et al.*].

According to Figure 2.1, not one single diagnostic technique is enough to manage all the possible issues. Nevertheless, the Frequency Response Analysis (FRA) seems to be the approach with the wider view.

FRA is a non-evasive, non-destructive method of detecting winding and core deficiencies. It inputs an electrical signal into the Power Transformer and reads its response. FRA tests can be employed with several connections, such as open-circuit and short-circuit, and may use all available terminals of the machine. Anomalies in the magnitude and frequency response indicate changes in the equivalent RLC circuit of the machine, thus evaluating the state of its electromagnetic components [4].

Although FRA is an adequate technique for evaluating Power Transformers mechanical condition, its state-of-the-art solutions are still transitioning to ML and computer intelligence inter-

pretations. Some researchers developed models for advanced analysis, which don't require human expertise for evaluation, but as of yet are only capable of detecting individual faults [5, 6].

Problems associated with the insulation are able to be observed through physicochemical techniques. These often require a sample from the transformer's oil to examine its properties by means of quantitative and qualitative methods, such as oil color, moisture content and dissolved gases. Figure 2.2 summarizes the designations of oil properties and possible methods for evaluation.

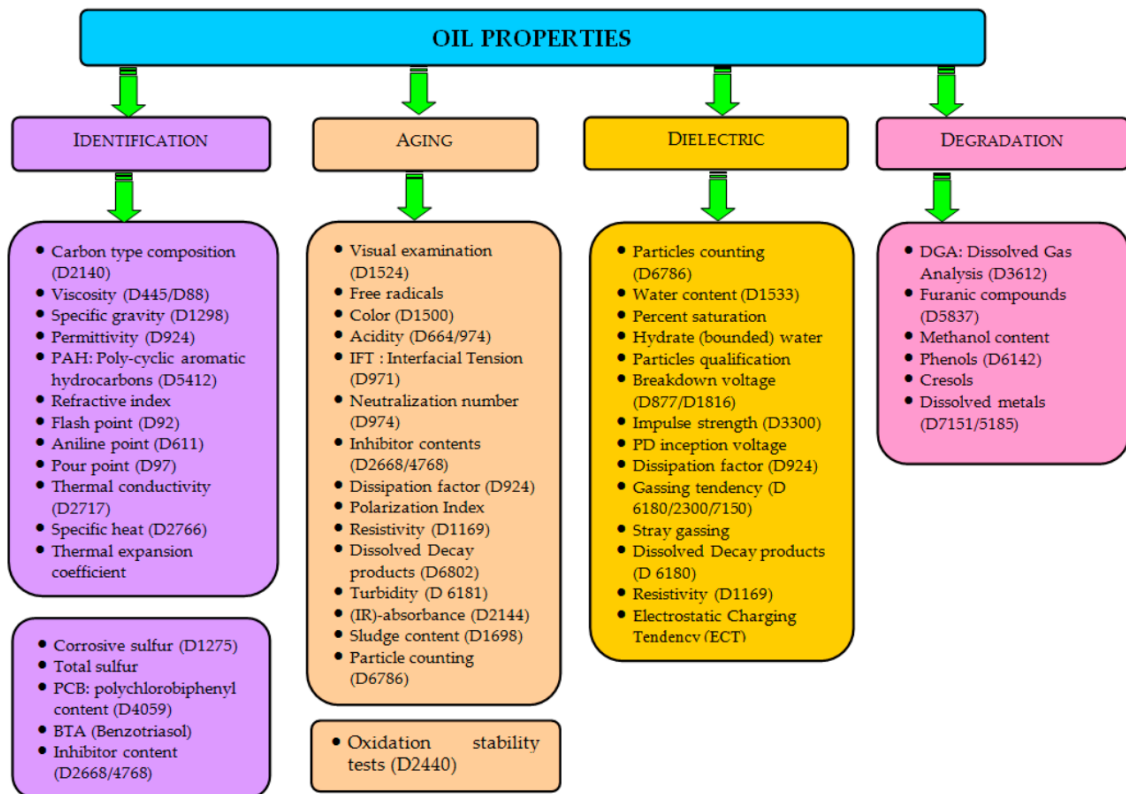


Figure 2.2: Oil diagnostic techniques [N'Cho *et al.*].

Throughout the transformer's life, it is exposed to electrical, mechanical and thermal stresses. This constant exertion during its operation deteriorates the oil, reducing the effectiveness of the insulation and cooling means. The decline in the oil quality is often linked to an increase of the concentration of gases such as hydrogen (H_2), oxygen (O_2), carbon dioxide (CO_2), methane (CH_4), ethane (C_2H_6), acetylene (C_2H_2), ethylene (C_2H_4) and others [3].

Thereby, dissolved gases in Power Transformer's oil are result of chemical reactions when the insulation is under stress. Concentration of these gases can be linked to faulty states, by what is known as Dissolved Gas Analysis (DGA). The next section focuses on this diagnostic method.

2.3 Dissolved Gas Analysis

Continuously necessities for proper condition diagnosing include reliable and non invasive techniques. DGA comes across as non-destructive and solid technique for oil-based transformer diagnosis and fault identification.

DGA is a monitoring method used to identify and quantify gases, through chromatographic analysis, produced by stresses which the transformer and its insulating oil are exposed to. Although DGA is an aging evaluation method, it does not have a single established approach, since it provides several tools, techniques and methods for fault identification and diagnosing. Thus, DGA is in constant development, also due to being based on statistics and empirical knowledge instead of science and physics-based models for understanding the meaning of gases quantities and concentrations.

In the past, correct assessment of DGA required technical know-how and skilled personnel. Nowadays, in order to reduce the irregularities of the DGA results, Machine Learning (ML) models can be used to suffice the inherent flaws in the techniques employed. Nevertheless, various techniques for DGA are widely used, such as the Total Dissolved Combustible Gas (TDCG) procedure, Rogers Ratios, IEC 60599 methods etc [7].

TDCG procedure is a method for transformer condition evaluation referenced in the IEEE C57.104 standard. TDCG is the sum of the concentrations of all combustible gases in the oil, i.e. H_2 , CH_4 , C_2H_6 , C_2H_4 , C_2H_2 , and CO . The IEEE proposes a four-level criterion, classifying risk of Power Transformer failure in 4 conditions, the first being the normal or healthy state and the fourth indicating presence of faults in the machine that could lead to failure. This procedure not only uses TDCG concentration but also individual gases concentrations for discerning between the criteria [8].

The IEEE standard also provides a wide view on the generation of gases in the oil and associates each gas formation and development of concentration to the transformer's internal temperature. Standard C57 relates gases' relative concentrations both to temperature and nature of the fault, as Figure 2.3 depicts.

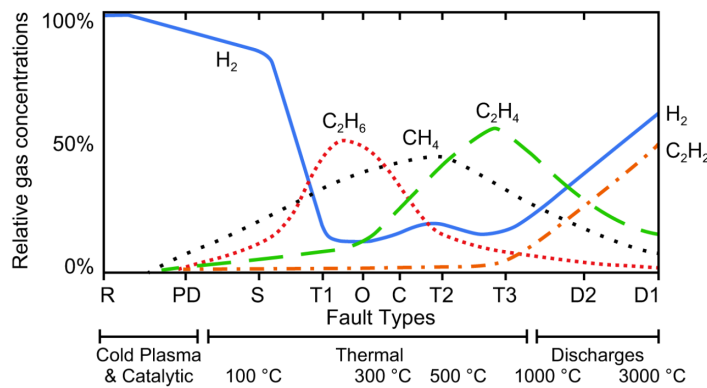


Figure 2.3: Relative percentage of dissolved gas concentrations as function of temperature and fault type, according to the IEEE standard.

The IEC 60599 norm provides an ample guide for analysis, although it lacks techniques that are contemplated in other reference standards. Standard 60599 also contemplates typical gas concentration values, and characterizes faults in six types, as Figure 2.4 displays. Accordingly, power transformers have seven states that can be classified either as faulty (thermal or electrical fault) or not faulty (healthy) [1]. However, Duval states that another faulty state, a mix of thermal and electrical faults, exists [9].

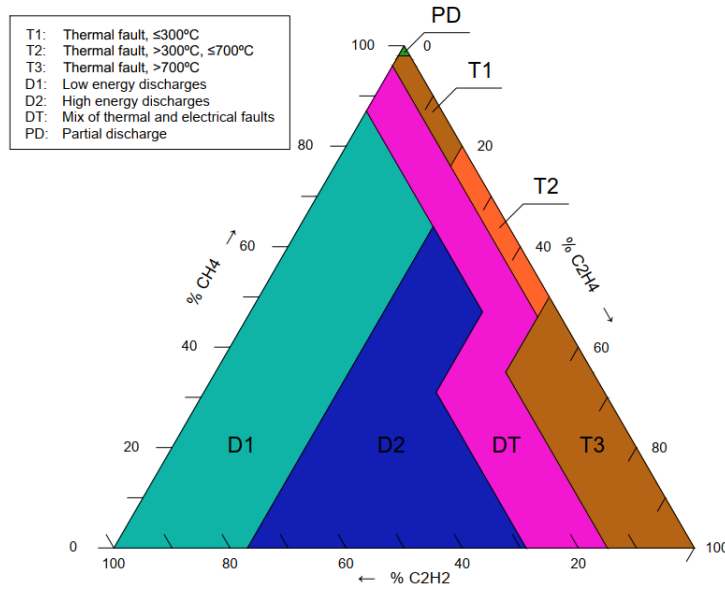
Type	Fault	Examples
PD	Partial discharges	Discharges in gas-filled cavities resulting from incomplete impregnation, high-humidity in paper, oil supersaturation or cavitation, and leading to X-wax formation
D1	Discharges of low energy	Sparkling or arcing between bad connections of different or floating potential, from shielding rings, toroids, adjacent disks or conductors of winding, broken brazing or closed loops in the core Discharges between clamping parts, bushing and tank, high voltage and ground within windings, on tank walls Tracking in wooden blocks, glue of insulating beam, winding spacers. Breakdown of oil, selector breaking current
D2	Discharges of high energy	Flashover, tracking, or arcing of high local energy or with power follow-through Short circuits between low voltage and ground, connectors, windings, bushings and tank, copper bus and tank, windings and core, in oil duct, turret. Closed loops between two adjacent conductors around the main magnetic flux, insulated bolts of core, metal rings holding core legs
T1	Thermal fault $t < 300\text{ }^{\circ}\text{C}$	Overloading of the transformer in emergency situations Blocked item restricting oil flow in windings Stray flux in damping beams of yokes
T2	Thermal fault $300\text{ }^{\circ}\text{C} < t < 700\text{ }^{\circ}\text{C}$	Defective contacts between bolted connections (particularly between aluminium busbar), gliding contacts, contacts within selector switch (pyrolitic carbon formation), connections from cable and draw-rod of bushings Circulating currents between yoke clamps and bolts, clamps and laminations, in ground wiring, defective welds or clamps in magnetic shields Abraded insulation between adjacent parallel conductors in windings
T3	Thermal fault $t > 700\text{ }^{\circ}\text{C}$	Large circulating currents in tank and core Minor currents in tank walls created by a high uncompensated magnetic field Shorting links in core steel laminations

Figure 2.4: Typical faults, according to the IEC standard.

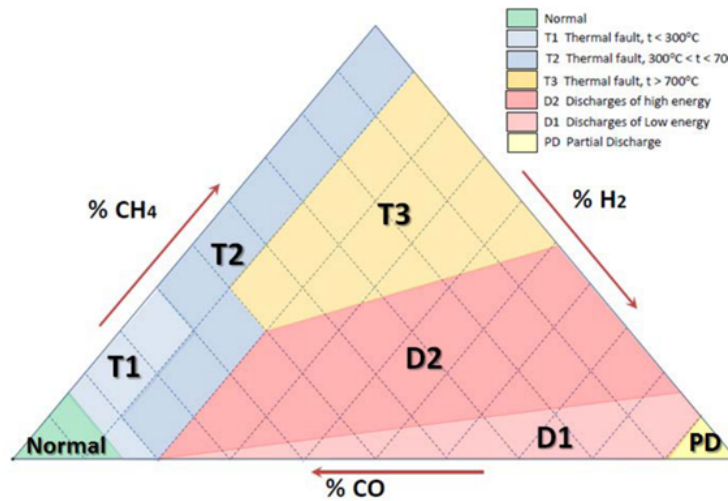
The IEC recommends the Basic Gas Ratios and the Duval Triangle (Figure 2.5a) for gas analysis and incipient fault identification. According to 60599 the basic gas ratios are the following: $\frac{C_2H_2}{C_2H_4}$, $\frac{CH_4}{H_2}$ and $\frac{C_2H_4}{C_2H_6}$. For the Duval Triangle the ratios adopted are the percentages considering only these three gases: CH_4 , C_2H_2 and C_2H_4 . These methods are used for graphic fault identification,

dividing the faulty states into well-defined zones, though sometimes other DGA tools are used for complementing the analysis [1].

Complementary gas ratios are $\frac{CO_2}{CO}$, $\frac{O_2}{N_2}$ and $\frac{C_2H_2}{H_2}$ and indicate the involvement of paper, excessive heating or oil tank contamination. Other ratios such as $\frac{C_2H_2}{CH_4}$ and $\frac{C_2H_6}{C_2H_2}$ are used in the Dörnenburg's Method, although it fell out of use, by evolving into simpler methods such as the Rogers Ratios or the Basic Gas Ratios [10].



(a) Duval Triangle.



(b) Low Energy Degradation Triangle, [Moodley].

Figure 2.5: Duval and Low Energy Degradation Triangles.

Moodley and Gaunt propose a method for identifying low energy faults and healthy states in a Power Transformer. This is the Low Energy Degradation Triangle (LEDT), and its construction is related to the Duval Triangle, as Figure 2.5 displays. Similarly, the triangle's components are percentages of three dissolved gases, being its summation 100%. These gases are H_2 , CO and

CH_4 , since they are associated with low energy faults, such as low temperature thermal faults [11].

Although the LEDT method is useful for presenting the natural progression of a Transformer's overtime degradation in a graphic and intuitive way, it falls prey to abnormal levels of carbon monoxide concentration. Inaccurate labelling of states might occur if CO levels are too high. For that reason, Moodley and Gaunt propose a limit of 500 ppm in the carbon monoxide concentration, when applying the LEDT method [12].

In addition, Pattanadech and Wattakapaiboon propose a graphic method that can be more consistent, taking into account, not only fault types but also aging of the insulation. This method is the Duval Pentagon and it considers two more gases against the typical Duval Triangle analysis. These are hydrogen (H_2) and ethane (C_2H_6). The Duval Pentagon method allows the ability "to distinguish the normal aging of the insulation condition of the transformers" by considering the associated "occurrence of gases under normal operation of the transformers" [13].

The aforementioned DGA methods can be defined as traditional methods, considering that they employ techniques derived from empirical observations. The next section will focus on advanced methods for power transformer fault identification by means of DGA.

2.4 Machine Learning solutions

Identifying faults in power transformers is necessary for lowering costs of maintenance, but correct assessment of the types of fault is not always guaranteed. As already stated, data-driven models can improve diagnosis through DGA. In this subsection, some of the ML algorithms present in the literature are exposed.

ML models were first introduced into fault identification to be combined with existing models and to provide an extra layer for an agent-based system, for the developed work in [7]. The purpose of this intelligent system was to combine acquired data from all the selected DGA techniques plus the Artificial Neural Network (ANN) method of identification. Hereby, a combined method with greater accuracy could be achieved. In such a way, with great amounts of data and some contradictions between DGA techniques, data-driven models were more and more adopted [10].

Thus, at first, ML was used to improve the accuracy of prevalent DGA methods, building upon established gas ratios or relative gas concentrations [14, 15]. This approach yields better results, comparing to traditional DGA methods, since its based in combinations of several of these methods, providing appropriate results in case of conflicting evidence or poor outcomes.

Some ML models used in Power Transformer fault identification by means of DGA data are ANNs [7], Decision Trees [16], Support Vector Machines (SVM) [17], k Nearest Neighbor (KNN) [18], Bayesian networks [19] and Markov models [20]. More and more often, these data driven models entirely replace traditional DGA methods, such as the ones presented in Section 2.3. The goal of researchers in this field is to "develop an alternative approach to DGA data interpretation", using advanced DGA methods based on ML [10]. Most of the ML algorithms used for this purpose, i.e. classification of transformer's incipient faults through DGA data, employ what is known as supervised learning techniques.

2.4.1 Supervised Learning

Supervised learning is a ML technique that learns from historic data. It requires correctly labeled data from data sets, such as the IEC TC10 [21], for algorithm training. This data set is consequence of direct inspection of transformers all over the world, throughout the years, providing a solid set for DGA ML appliances. Hence, the common use of this data set in supervised ML solutions in the literature.

Subsequently, to better the performance of supervised learning algorithms, feature extraction is crucial, since it's directly related to the quality of the training data. In this regard, redundant, pointless or noisy data features should be removed from the training set. Thus, increasing the trained algorithm performance, through reducing the computing time and bettering the value of the extracted knowledge.

Neural Networks are widely used throughout the literature for transformer DGA incipient fault diagnosis [7, 22, 23, 24, 25, 26, 27, 28, 29, 30]. Applying a simple ANN to DGA betters the diagnosis performance, improving fault identification when compared to traditional DGA methods. Still, it remains sub-optimal, as it can easily converge in local minima.

To overcome this, hybrid data-driven models [25, 29, 30] and fuzzy theory [22, 27, 31] have also been implemented, revealing high success both in fault identification accuracy and classifying missing data.

In [25], an extra algorithm, titled Cuckoo Search, is used to optimize the weights and biases of the Back-propagation Neural Network. Feature extraction follows the IEC basic gas ratios, following a supervised learning approach with real power transformer DGA data. As the hybrid Neural Network adjusts itself, due to the Cuckoo Search, it avoids local convergence and finds the best global solution.

Yang *et al.* use a Probability Neural Network enhanced by way of a custom developed called Multi-verse optimizer algorithm. Similarly, the auxiliary algorithm is adopted for optimizing the parameters of the Neural Network resulting in an improved convergence. The optimizer algorithm is fed an error signal that is direct result of the output of the Neural Network and the desired output [29].

ANNs can also be enhanced through a modified Evolutionary Particle Swarm Optimisation-time Varying Acceleration Coefficient. Once more, the ANN parameters are optimized by the auxiliary methods, finding the optimal values for the ANN's learning rate and momentum constant. Illias *et al.* compare their hybrid algorithm with other previous reported methods and conclude that the "performance of the ANN was successfully optimised" [30].

Fuzzy number theory is used in this context to embody the uncertain nature of DGA data. This uncertainty, associated with the formation of dissolved gases, the oil sampling or even the chromatographic analysis, can be modelled by fuzzifying discrete DGA data. Castro uses a Multilayer Feedforward ANN and basic gas ratios as in input to attempt a fuzzy rule extraction of power transformer's faults identification [22].

This knowledge extraction is further developed in [27], although it goes beyond the scope of this thesis. Miranda and Castro improve the IEC fault identification table with their results. The ANN used classifies power transformer faults by means of intervals rather than discrete categories or values.

Fuzzy logic comes across as an advantage, since it can output multiple faults and the weight of each diagnosis. The combination of fuzziness with ANNs provides a more accurate representation of particular transformers, by combining the intelligent and mutable characteristics of the ANN with fuzzy logic data processing [31].

Despite the widespread use of the IEC TC10 database for supervised training of ML models, a fault identification algorithm can be trained with unlabelled data. In the next subsection this approach will be considered.

2.4.2 Unsupervised Learning

Frequently, DGA data is not registered with the power transformer's condition in the moment of analysis. *A priori* this would be a problem, since labelled data is needed for "teaching" a data-driven model. Unsupervised learning is a mean of training an ML algorithm with non-categorized or unlabelled data, i.e. DGA data not diagnosed by an expert.

In [32], a unsupervised method to the fault identification problem in power transformers is undertaken. Dias *et al.* approach the problem of unlabelled data with an entirely data-driven approach, taking advantage of traditional DGA methods, such as the Duval Triangle, for cluster classification. They developed a ML algorithm capable of creating a data-driven model from DGA data and save it for future classifying and diagnosis use.

This algorithm's solution for feature extraction is based on the entropy method. The concept of entropy in ML refers to the meaningfulness of a certain variable, in this case a gas concentration, and is connected to its level of information content. If the concentration value is almost constant, it is identified as having lower significance. Alternatively, the more diverse the values are for a given gas, the more valuable information it provides. So, with entropy values assigned to each gas, significance weights are defined for each. Finally, those gases with an assigned weight below the threshold can be removed, simplifying and processing the input data.

After a data-driven feature extraction, information can be clustered for subsequent classification. Dias *et al.* propose two clustering alternatives, Principal Component Analysis and K-means Algorithm, both combining entropy metrics within their clustering algorithms. Subsequently, clusters are categorized by the TDCG procedure, ranking them by condition and effectively dividing the clusters in healthy or normal state and faulty state by severity. The definitive labelling is accomplished when the faulty state is known. This means resorting to the Duval Triangle method for transformer fault identification.

An example of the result of clustering is presented in Figure 2.6. Nine distinct clusters are presented, which seems to oppose the six types of faults reported in IEC 60599 as Figure 2.4 presents. However, as aforementioned, the model developed by Dias *et al.* besides categorizing faulty and normal states, also provides a ranking by severity level, following IEEE standard guidelines. Thus,

justifying the appearance of more than 6 states, being high energy faults associated with high level of risk and low severity with normal or healthy state.

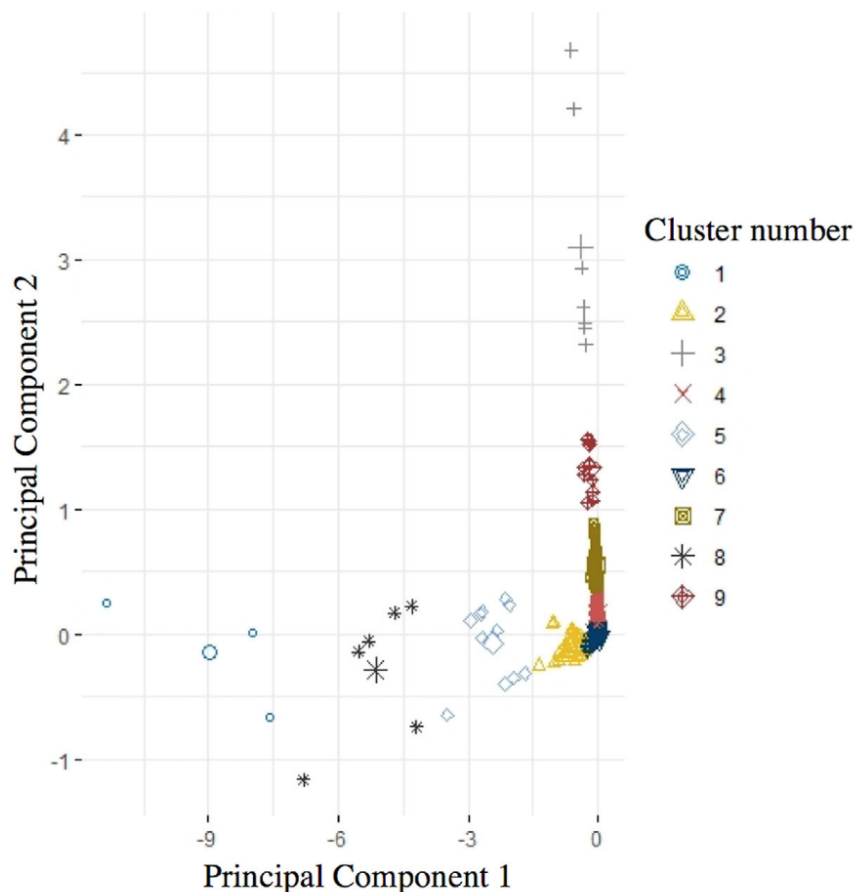


Figure 2.6: Graphical representation of data clustering, [Dias *et al.*].

The performance of the proposed model is then tested with the IEC TC10 database, in the interest of comparing to other methods reported in the literature. It revealed high accuracy, directly comparable to state of the art supervised methods and "proving the generalization potential of the novel methodology".

Besides this unsupervised method, a third approach to ML training algorithms is possible, combining the two already reviewed concepts. In the next section, semi-supervised learning is presented.

2.4.3 Semi-supervised Learning

Semi-supervised learning is a ML training technique that uses both labelled and unlabelled data. Essentially, a model that uses a semi-supervised approach benefits from the information contained in unlabeled data to improve its prediction. In the literature, a few examples of application were found.

In [33], a semi-supervised autoencoder with an auxiliary task is proposed. An autoencoder is an unsupervised neural network, composed of coder and encoder parts and hidden layers. First, this algorithm learns, by unsupervised means, the characteristics of the data. Then it learns, by supervised auxiliary means, to detect normal and faulty states and to identify the faulty types.

Kim *et al.* divide the data in four distinct groups: an unlabeled DGA data set, a labeled DGA data, and two information data sets for the supervised learning tasks. After the training, the health features are represented in a two dimensional space, facilitating the visualization of the data. Boundaries, based on the auxiliary fault identification algorithms, are also traced to provide an artificial labelling of the unlabeled data.

The next section addresses the prediction methods and models for Power Transformers using DGA.

2.5 Fault Prediction

As it was illustrated, classifying and identifying faults by means of DGA is a well studied topic. Hence, the next step is predicting future faults and their types in power transformers.

To achieve this Jiang *et al.* developed "a dynamic early prediction and warning model combined Gaussian mixture model with hidden Markov model" capable of identifying sub-healthy states (state before fault), which conventional DGA methods would consider not faulty [20].

Predicting faults in power transformers requires first an extraction of the features of the machine's states. Jiang *et al.* also introduced an extra state, labeled as sub-healthy, to "enrich the health index and aging stages of power transformers". With the static features validated, a dynamic (time series) technique for state prediction is tested using an hidden Markov model. This model offers a probability based solution for the transitions between states, effectively predicting the next faulty state. Jiang *et al.* also focused more on thermal faults, which opens the possibility of testing the presented solution in every type of fault.

The LEDT can also act like a predicting model for transformer faults, since it assesses insulation degradation associated with low energy events. In [34], Moodley proposes a R-value trend derived from plotting the triangle with polar coordinates, being R the magnitude of the vector starting in the left corner of the LEDT. Hence, and as Figure 2.5 alludes to, when R is zero or under a certain value, the transformer condition can be treated as normal or healthy. According to Moodley's analysis, "a R value of 0.13 is considered the outer limits of normal [state]".

Hence, Moodley does a time-based analysis of several transformers employing the LEDT R-trend. Accordingly, Figure 2.7 shows the development of the degradation value of a power transformer throughout its years of operation. Most of the R values are in the normal region, although it changes when a trigger event occurs. Thereafter, the transformer's condition does not return to normal. Although the transformer's oil got changed in 2004, the analysis still categorizes its state as not normal. In the end the analysis is proven right, considering the transformer failed in 2005, revealing internal high energy discharges.

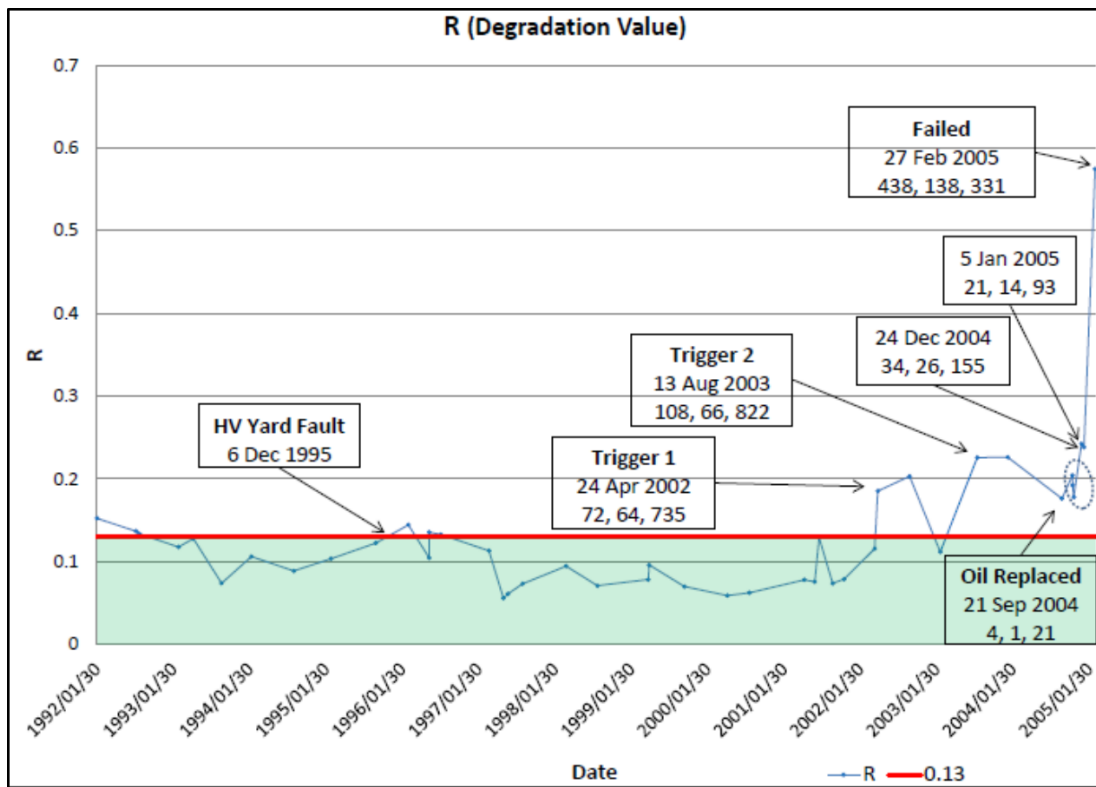


Figure 2.7: Degradation value (R) of a power transformer throughout the years, [Moodley].

This method can be considered effective for "identifying trends of insulation degradation". Therefore, the LEDT is capable of detecting future incipient faults even if the insulation oil is replaced.

Other ML models such as Long Short-term Memory (LSTM), Random Forest (RF) and Auto-regressive Integrated Moving Average (ARIMA) are also applied to Power Transformer DGA predictions. These approaches are associated with the prediction of the dissolved gases' behaviour throughout the transformer's life, which then can be interpreted and analysed to identify future faults [35, 36].

Mahrukh *et al.* employ and compare LSTM and RF approaches to DGA data prediction. They propose, by regression square and correlation analysis, that the input of the prediction model is an array of the gases' concentrations and that the predicted output is the total hydrocarbons, i.e. total sum of gases containing hydrogen and carbon. Mahrukh *et al.* obtain a timeseries forecast of total hydrocarbon concentrations with great accuracy and conclude that the RF based solution shows the best performance [35].

In [36] Nandagopan *et al.* use a LSTM combined with a vector auto regression model and an ARIMA model for DGA data prediction. The LSTM model employed can only generate single-step predictions, therefore they combined the model with other approaches to achieve multiple time steps predictions. The data used, besides the typical gas concentrations, also included transformer load and winding temperature.

The ARIMA model utilized predicts a single variable with good results, although a multi-variable approach would be superior. This model also tends to behave like a linear regression the further the prediction, leaving Nandagopan *et al.* to propose that the ARIMA model "is unreliable for prediction of future trends" [36].

In the next section, a review of some of the ML models used for time series prediction is given.

2.6 Prediction Paradigms

Accurately predicting a time series requires a proper model to be built and to adjust it to the problem at hands. The best model for prediction depends on the characteristics of the time series. Setting the correct parameters of an algorithm may also involve other ML models, for they can be used for optimization of those same parameters. Hence, prediction methods are usually not a single algorithm but a combination of different paradigms.

Nonetheless, in this Section simple prediction algorithms are described and presented.

2.6.1 Random Forest

RF or Random Forest is a collection of multiple decision trees, a ML algorithm used for classification and regression. Each decision tree is generated, trained and randomly displayed in what is called an ensemble classifier. The RF classifier is an ensemble of decision trees.

Each individual tree yields a prediction which is aggregated with other trees' predictions, following a particular criterion (averaging or majority vote), and combining the results in to one single prediction.

This method of combining prediction presents the advantage of reducing the variance of the prediction error. Figure 2.8 depicts a simple ensemble classifier method.

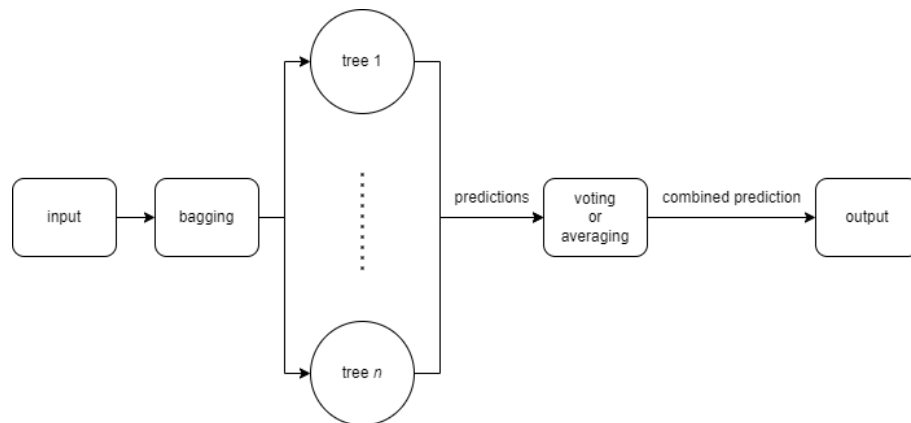


Figure 2.8: Random Forest diagram.

In order to build the RF, n individual decision trees are created with random feature variables and trained with subsets of the original training set. Given $D_{training}$ as the training set, through bootstrap aggregation or bagging, $D_{training}$ is able to be sampled and to generate n new training

sets, such as $D_{training} = \{D_1, D_2, \dots, D_n\}$. Each new data set is unique and may present duplicate instances.

Random Forest benefits from the technique of bagging, that consists of selecting a set of examples randomly with replacement. For each set, a tree is produced which is able to provide a prediction. The aggregation is then applied to determine a single prediction that is representative of the models. The quality (degree of uncertainty) of the prediction is measured by the dispersion of the trees prediction. The main advantage of this model is the simplicity and explainability. Despite being state of the art, this algorithm can not model well the relations between time, since it only outputs a single value.

The ensemble of trees is then assembled. After the classifications of each tree are made, the voting occurs. The outcome with most expression, i.e. the popular vote, gets to determine the output of the forest. In case of an ensemble RF for regression, the final output is determined by the average of all the decisions from the trees [35].

2.6.2 Recurrent Neural Network

RNN or Recurrent Neural Network is a type of ANN that makes use of sequential information by means of its "memory". This is possible due to the feedback cycle that artificial neurons present, directly influencing the current output with information from previous inputs.

Similar to a Feedforward ANN, a RNN can be of the multi-layer perceptron type, that is it can incorporate more than one hidden layer besides the input and output layers, and can be mapped in one-to-one, one-to-many, many-to-one and many-to-many ways. It also utilizes the typical neuron activation function, the sigmoid function, described in equation 2.1, but can also use the tanh or relu functions.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

A cell from a RNN has a hidden state, that is responsible for its "memory". This internal state is computed at every time step and can be described by equation 2.2.

$$h_t = \sigma(W_{hh} \cdot h_{t-1} + W_{xh} \cdot X_t + b_h) \quad (2.2)$$

Where h_t is the hidden internal state, and W_{hh} , W_{xh} and b_h are the cell parameters, weights and bias, respectively.

The output can then be calculated, according to equation 2.3.

$$y_t = \sigma(W_y \cdot h_t + b_y) \quad (2.3)$$

RNN neuron's parameters can be joined in a set, such as $\theta = \{W_{hh}, W_{xh}, W_y, b_h, b_y\}$, and when training the network, attempt to attribute the optimal values to these parameters.

A recurrent ANN can be trained with the help of a special variation of the back propagation training algorithm, the back propagation through time method. Though, with this training algorithm two hurdles arise: the vanishing and exploding gradient problem. The exploding gradient problem is easily solved by limiting the gradient value at each cell and the vanishing problem can be solved by using variants of the RNN, such as the LSTM [37].

2.6.3 Long Short-Term Memory

LSTM or Long Short-Term Memory Neural Network is a variant of a RNN introduced by Hochreiter and Schmidhuber to tackle the problems associated with training a traditional RNN [38].

The LSTM cells differ from the RNN units by having an additional three gates. These are the input gate, forget gate and output gate, and their values are described by equations 2.4, 2.5 and 2.6.

$$i_t = \sigma(W_{hi} \cdot h_{t-1} + W_{xi} \cdot X_t + b_i) \quad (2.4)$$

$$f_t = \sigma(W_{hf} \cdot h_{t-1} + W_{xf} \cdot X_t + b_f) \quad (2.5)$$

$$o_t = \sigma(W_{ho} \cdot h_{t-1} + W_{xo} \cdot X_t + b_o) \quad (2.6)$$

The internal cell state is given by equation 2.7, which then, along with its gates' information, compute the unit's memory, as equation 2.8 demonstrates.

$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma(W_{hc} \cdot h_{t-1} + W_{xc} \cdot X_t + b_c) \quad (2.7)$$

$$h_t = o_t \odot \sigma(c_t) \quad (2.8)$$

Where $\odot$ is the Hadamard product of matrices or element-wise multiplication.

Thus, the LSTM cells are able to recall short term information without jeopardizing long term trends, as well as bypassing the vanishing gradient problem.

In the next section, an overview of artificial time series data generation is presented.

2.7 Synthetic Data Generation

In the context of ML, synthetic data generation is a way of expanding a data set by following or reproducing the trends present in the original data. Data sets might require extension, because of inadequate, not representative or scarce data, mainly unlabeled data, and to achieve better model training. This section focuses primarily on augmenting time series data sets.

A time series can be reproduced by ML models, as it was implied in Section 2.6, but also by simpler methods, such as slicing, rearranging, scaling, magnitude and time warping, adding noise

(jittering) or by statistical means. In spite of the simplicity of these methods, they often fail to consider the trends or cyclic and seasonal information of the time series [39].

2.7.1 Auto-regressive Integrated Moving Average

ARIMA or Auto-regressive Integrated Moving Average is a time series forecasting algorithm that utilizes real past data to predict future values. This model is characterized by three independent terms: (p,d,q) . Respectively, these terms equate to the order of the auto-regression, how often the time series is differenced and the order of the moving average.

The auto-regression, of order p , makes predictions using a linear combination of historical data, as equation 2.9 presents:

$$Y_t = \alpha + \beta_1 \cdot Y_{t-1} + \dots + \beta_p \cdot Y_{t-p} + \varepsilon_t \quad (2.9)$$

Where Y_t is the value to be predicted at the t point in time, Y_p are the p previous or lag values, α is the offset or intercept value, β_p is the coefficient of the p lag value and ε_t is the error term.

The moving average, of order q , makes predictions using a linear combination of residual errors, as equation 2.10 presents:

$$Y_t = \alpha + \theta_1 \cdot Y_{t-1} + \dots + \theta_q \cdot Y_{t-q} + \varepsilon_t \quad (2.10)$$

Where θ_q is the coefficient of the q error value.

The differencing of order d is how many times the consecutive data is subtracted. Equation 2.11 shows a differencing of order 1:

$$Y'_t = Y_t - Y_{t-1} \quad (2.11)$$

Seasonality is also able to be integrated in the ARIMA model through the difference operator. Equation 2.12 shows the mathematical implementation of seasonal differencing:

$$Y'_t = Y_t - Y_{t-m} \quad (2.12)$$

Where m is the duration of the season.

The ARIMA model is a simple algorithm capable of tackling dynamic data, such as time series data. The difference operator, although "theoretically insufficient" is easy to implement and "satisfactory for some specific problems in engineering" [40].

2.8 Chapter Conclusion

This chapter reviewed important definitions and concepts for power transformer fault prediction. Emphasis was given to Dissolved Gas Analysis, since it is the method with the widest use for fault identification, prediction and ML applications.

The main objective of this work is to develop a model capable of predicting future power transformer incipient faults, but also, as a secondary objective, to tackle the problem of abundant unlabelled and scarce time series DGA data. Thus, the attention given to fault identification models, unsupervised and semi-supervised learning alternatives, and predicting and data augmentation algorithms. Many techniques were considered to give the reader a view of all possible paths to be taken.

Several potential solutions were raised for solving the prediction problem, from analytical approaches, such as the LEDT, to ML predicting models, such as LSTM ANN. Even with an unlabelled and limited data set, capable of being artificially labeled and expanded by the methods raised in this chapter, the objective of predicting future faults is achievable.

Chapter 3

Methodology

This chapter focuses on the development of the procedure to solve the proposed problems. ML algorithms, empirical models and public databases are presented, detailed and further established. Methods for data analysis, processing and augmentation are demonstrated. The algorithm modeling for fault prediction and its validation and performance tests are also enacted. A pipeline for implementation of the solution and its results evaluation is proposed.

3.1 Overview

As stated, in this chapter the methodology for this work is proposed. An overview of the several stages for building, training and evaluating the predicting models is performed. The groundwork for processing and analysis of the data is tackled, as well as the fault identification algorithms. A method for extending DGA time series is also proposed.

In Figure 3.1 a diagram of the whole planned approach is visible. This is the pipeline detailed in this chapter and followed throughout the work.

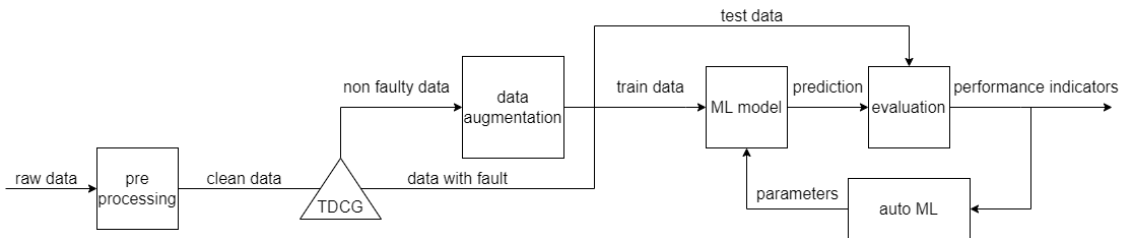


Figure 3.1: Pipeline of the proposed methodology.

The main framework for data processing was implemented in Python and composed by packages for statistical analysis (statistics), numerical manipulation (numpy and pandas) and visualization (matplotlib). The adoption of this implemented framework is justified by the fact of being rich in terms of functionalities and easy to install software.

3.2 Datasets

For proper and adequate model training, an accessible database with plenty and meaningful data is key. Databases from credible sources in the field of Power Systems, such as IEEE and IEC, have been used and studied in many scientific works. Particularly, the IEC TC10 database is openly available in [21].

However, this database does not have time series DGA records. This time based information is needed for an adequate prediction model training. Contact with EFACEC provided a time series data set, although not large and detailed enough for the purpose of this work. An alternative was found in [41]. This data set has almost 5 years of DGA data from 13 unique Power Transformers from the UK. The data includes concentrations from 8 different gases and humidity content, but it does not offer information about faults or any health condition.

This data set provides three readings of each gas concentration (in parts per million, ppm), each reading associated with one of three phases of the Power Transformer. The exception being two transformers that are only characterized by its main readings, and one that has missing oxygen readings in one of its phases. Gases present in this data set are: H_2 , CH_4 , C_2H_2 , C_2H_4 , C_2H_6 , CO , CO_2 , O_2 and H_2O .

3.3 Data Analysis

For a thorough examination of all gathered data a collection of statistical analysis is carried out. These range from histograms to correlation analysis. Although the data set holds information about 9 gases, only six of them are considered. These are H_2 , CH_4 , C_2H_2 , C_2H_4 , C_2H_6 and CO , which are the six combustible gases found in Power Transformer insulating oil.

Histograms of all considered gases are plotted with a logarithmic scale, to ensure that outliers are visible. These bar graphs provide a straightforward visualization of the distribution of the data. Additionally, correlation analysis is important to verify relations between components of the TDCG gases which might be important to the values predictions and data expansion. Hence, three types of correlation analysis are enacted, namely linear relation (Pearson correlation coefficient), monotonous relation (Spearman's rank correlation coefficient) and variation relation (phi coefficient or mean square contingency coefficient).

Particularly, the phi coefficient is applied to the binarized differentiation. This is, when the value increases from one value to the next, the value 1 is assigned. If the value decreases or is equal, the value -1 is assigned. In order to verify local in time relationships, short-term analysis is performed using a sliding window on the paired time series, where the above mentioned correlation metrics are applied. The variation of the metrics can be observed and localized through time.

3.4 Artificial Labeling

In order to obtain labels with meaning, data categorization was employed. The goal is not to operate with factual faults but to use a concentration threshold that is associated with a fault warning. Hence, to artificially label the data, first, categories must be defined.

Power Transformer's faulty labels can be described with 8 states, ranging from 0 to 7, as Table 3.1 displays. To achieve this, two traditional methods found in the literature for fault identification using DGA are employed.

State	Type	Fault
0	Healthy	None
1	PD	Partial discharge
2	T1	Thermal fault, $t < 300^{\circ}\text{C}$
3	T2	Thermal fault, $300^{\circ}\text{C} < t < 700^{\circ}\text{C}$
4	T3	Thermal fault, $t > 700^{\circ}\text{C}$
5	D1	Discharges of low energy
6	D2	Discharges of high energy
7	DT	Mix of electric and thermal faults

Table 3.1: Fault types in a Power Transformer.

For labelling the data as healthy or faulty, the elementary form of the TDCG procedure is chosen, this is, considering only the total concentration of all combustible gases. In Table 3.2, the four criterion of this procedure are detailed. For the sake of simplicity, condition 2 and above are considered faulty states in Power Transformer's condition evaluation. Therefore, at any given moment, if the TDCG is larger than 720 ppm, a fault identification method may be conducted. This value is defined as an established limit for warning alarm, effectively labeling data as "Fault warning" when associated to the condition $TDCG > 720\text{ppm}$.

Condition	TDCG	Faulty State
1	< 720	No fault
2	721 - 1920	Fault may be present
3	1921 - 4630	Fault probably present
4	> 4630	Could lead to failure

Table 3.2: Total Dissolved Combustible Gases Concentration.

Likewise, for discerning between faulty states, the Duval Triangle, as presented in Figure 2.5a, is applied. After this procedure, the data frames contain an extra variable or column with its fault state information, according to the Table 3.1.

Considering that not all available time series reach the determined fault warning state, a background truth must be determined for those. In the next section, the algorithm for augmenting the available data is posed.

3.5 Data augmentation

To those time series that do not reach the considered threshold for it to be considered faulty, a process of expanding the data set is carried. The results of this data augmentation are then considered as a background truth, since they are the available data to work with.

At first, an ARIMA model for data augmentation was considered. However, it tended to produce constant values, providing no valuable forecasts and failing to reproduce the growing tendencies of gases' concentrations. Thus, an alternative model for data augmentation is considered. This is a deep learning algorithm for time series forecast.

N-BEATS, or Neural Basis Expansion Analysis for Time Series is an ensemble of deep feed-forward ANN with backward connections. The name deep, derives from the large stacks of hidden neuron layers. This neural network design enables interpretation, is simple and easy to use and train [42].

With the support of packages such as Pytorch and Darts, N-BEATS is implemented in the Python environment. This deep learning model training is done repeatedly. This is, it not only trains once and computes a set of a hundred forecasts, but also trains on itself again with the added information. This process is done independently with each gas at a time and extending the time series up to then thousand steps more.

Since gas relationships are not always kept throughout the forecast, there is the possibility to reintroduce these in the form of a bias. This inserted bias is a simple linear model, that predicts a individual gas value from the others.

Timestamps are also augmented, by way of reproducing the sampling rate present in the original data sets. The average time between readings in the non extended data sets is calculated, and then added to expanded points. This process enables a time based approach for prediction.

3.6 Modelization

For modeling the solution to the fault prediction problem, several different ML algorithms are used for regression and forecast. These are implemented in Python with the support of compatible libraries, such as scikit-learn, statsmodels, pyramid-arima, tensorflow and keras. Also for convenience, the data sets utilized for training and testing are polished, that is, clearing them of obvious outliers and high frequency noise, using low pass filters, but also scaling them, in other words, setting the lowest occurring value as 0, and the maximum as 1. This process enables a more efficient trend capture by the ML models.

As shown previously in section 3.4, the problem of fault prediction can be boiled down to a uni-variable time series forecast. Hereupon, simple and efficient models that satisfy the established requirements may perform well.

Simpler algorithms such as linear or exponential regressions, but also polynomial and logarithmic regression are employed as a trivial approach. Though not all are implemented successfully, a linear combination framework is constructed, allowing further transformations. Models using

regression algorithms are trained with a variable window of consecutive TDCG values to predict a future one. The parameter to use for regression models training optimization is R^2 . As Figure 3.2 demonstrates, a regression model is trained for one step prediction with an input window of 3 values. The model is then tested with previously unknown data to estimate the last instances of the time series.

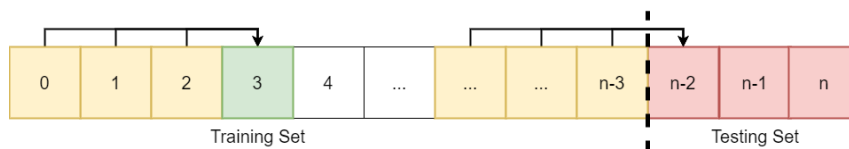


Figure 3.2: Schematic for one step ahead prediction using regression models.

However, with these models, to achieve a multi step ahead prediction, an ensemble is necessary. In Figure 3.3 a case for forecasting 3 time steps ahead is depicted. In this example, three individual models are trained, each one for predicting a unique shifted value. With the concentration values on the testing set estimated, results can be evaluated. More, with the achieved forecasts, the time for fault warning can be obtained, unless the threshold is not reached. In this case, a fit approach can be computed to project the tendency and obtain the desired time for warning.

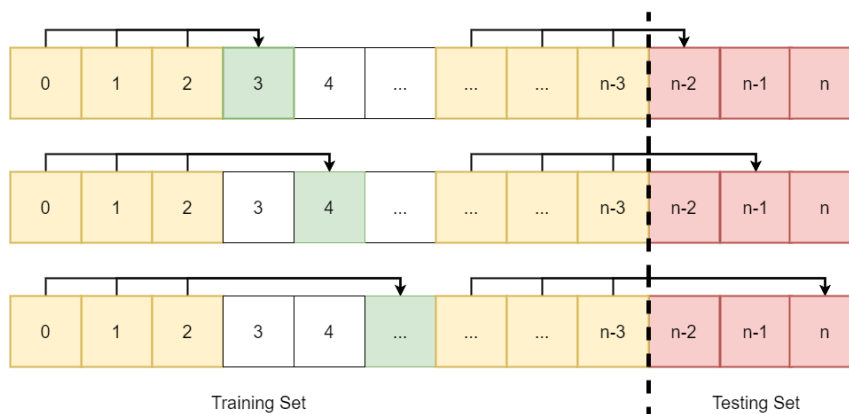


Figure 3.3: Schematic for multi step ahead prediction using regression models.

Regression analysis are done independently, but they also can be combined, averaging their forecasts to obtain improved predictions. These models' results will serve as groundwork for comparison. Indeed, after the observation of the trends in the time series, simple functions such exponential and third degree polynomial typically explain physical phenomena.

For the ARIMA model predictions to be effective, the data must be first converted to stationary, that is, removing the positive trend. This is achieved by the differencing method. In addition, seasonality is not employed, since none was observed in the data. ARIMA forecasting is done by lagging the datapoints in the time series, and can be done to any extent.

Being ANNs more complex and flexible algorithms, they provide better generalizations and may perform better. But they also require a careful approach, since their training parameters are plenty and prone to optimization. The employed LSTM is trained for varying forecast windows,

and the leading model is saved, considering the minimization of the training root-mean-squared error metric.

In fact, for multi step ahead prediction, just one LSTM model is necessary, provided that the neural network architecture is changed accordingly. Figure 3.4 presents a diagram of the time series prediction algorithm for a 3 step ahead forecast.

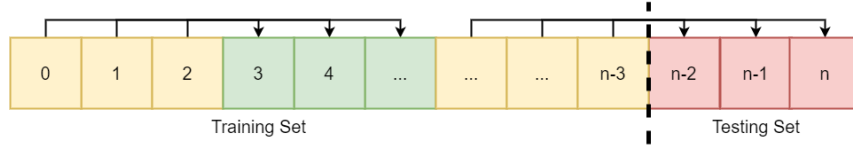


Figure 3.4: Schematic for multi step ahead prediction using LSTM.

The architecture of the network is sequential and comprised of one LSTM layer and two final dense layers of feed-forward artificial neurons. The number of units per layer is dependent on sizes of the input and forecast windows, such as equations 3.1, 3.2 and 3.3 present.

$$Input_{units} = LSTM_{units} = n \quad (3.1)$$

$$Output_{units} = m \quad (3.2)$$

$$HiddenDenseLayer_{units} = \lfloor \frac{n+m}{50} \rfloor \quad (3.3)$$

Where n is the number of consecutive TDCG input values, and m the number of points to predict. The number of units in the hidden layer is equal to the flooring operation, that is, the greatest integer less than or equal to $\frac{n+m}{50}$. In Figure 3.5 is present a schematic representation of the LSTM architecture.

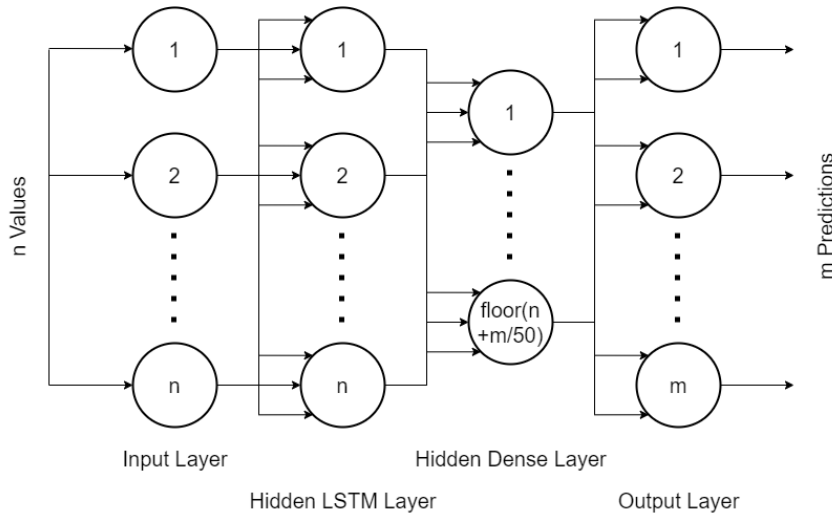


Figure 3.5: LSTM architecture diagram.

The LSTM layer is also capable of dropout, i.e. the probabilistic rejection of input and recurrent connections to LSTM cells during training, effectively disabling them momentarily. This process reduces overfitting and may improve overall performance, since it forces the network to rely on every unit to generalize.

A final approach to the predicting models is the curve fitting. This method simply fits the best polynomial curve to the given data. For the most part, this is used to continue forecasts that have positive trends, but that do not reach the desired output, although curve fitting can also be used solely.

Approaches, such as auto ML and the method for evaluating the performance of a given prediction model is described in the next section.

3.7 Test and validation

To better determine the hyper-parameters of the models, an Auto ML approach is executed. This is done with the help of parameter tuning frameworks, such as grid search and random search, available in the scikit-learn and keras tuner libraries, but also auto ARIMA, using the pyramid-arima package. Advanced methods are not employed due to the simplicity of the problem, since few optimization of monotonous parameters are needed. Default optimization criteria were considered. The goal of these search algorithms is to minimize the performance metrics, eventually providing an optimal set of parameters for a given model.

After a model computes its predictions on each time series, a validation process is done to prove the efficacy of the forecasts. Models, and their capability of predicting distant faults, are also tested against each other.

Validation sets are used to confirm the captured trends, through a process of optimization of statistical parameters, such as R^2 coefficient of determination, root-mean-squared error (RMSE 3.4) and mean absolute percentage error (MAPE 3.5). The coefficient of determination is implemented with sklearn metrics R^2 score and used only for the simple regressions and linear models. RMSE is used for comparisons between models or architectures, but for the same time series. MAPE is used for an overall inter-model and inter-set testing. These metrics measure how well the model predicts responses, ensuring that the training is efficient and provides an accurate ML model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_i - y_i)^2}{N}} \quad (3.4)$$

$$MAPE(\%) = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i - y_i}{Y_i} \right| \cdot 100 \quad (3.5)$$

Sizes of these sets are usually around 20% of the original data set size, but vary according to the forecast windows. The validation datapoints are also randomly chosen, according to a set seed, when training the model.

The testing data set is the final faulty segment of the original data set with the addition of the augmented regions. This is the background truth to work with. Testing data sets sizes vary according to the forecast time. Contact with industrial companies, such as EFACEC, suggested that predicting failures in advance with a determined time window is of great value. Thus, models can be tested for their inner parameters and overall ability to reproduce tendencies, but also with different window spans of the training and testing set.

3.8 Chapter Conclusion

In this chapter, the groundwork for fault prediction is presented. Methodology for managing unlabelled and limited DGA data is determined. Moreover, procedures for model building and assessment are detailed.

Chapter 4

Results

In this chapter the product of the application of the proposed methodology is analyzed. Results are evaluated, compared and discussed. Relevant insights are extracted and used to substantiate the conclusion.

4.1 Data Analysis

Data was treated as proposed in the methodology. Table 4.1 and Figure 4.1 illustrate the results of the procedures applied to the data for preparation and analysis of ML tasks.

Timestamp	H2	CH4	C2H2	C2H4	C2H6	CO	CO2	O2	H2O
2010-09-14 17:14:00	2.0	7.7	0.0	11.6	9.1	58.5	817.0	8498.0	3.5
2010-09-14 21:00:00	2.2	7.5	0.1	11.5	8.3	55.9	800.0	8538.7	3.6
...			...						...
2015-07-08 22:00:00	7.8	7.4	0.5	6.8	3.4	336.8	2352.0	24081.0	3.5
2015-07-09 06:00:00	7.5	7.3	0.3	6.7	3.4	335.6	2345.0	24026.6	3.6

Table 4.1: Transformer C data frame.

Observing Table 4.1, the data from Transformer C is organized by ascending time and properly titled. Approximately, each reading is obtained with a 4 hour interval. To the other Power Transformers the same method is applied, to obtain a comparable and homogeneous framework.

An ensemble of histograms was computed, resulting in Figure 4.2. These are presented in a logarithmic scale to provide an easier observation of outliers. Promising distributions of hydrocarbons with the exception of C_2H_2 are shown. This particular gas' distribution is centered in a value of concentration below 5 ppm, in other words, it carries little to no information. Regardless, this behavior is expected, considering that C_2H_2 is usually associated with high temperature faults, as validated by Figure 2.3. The data relates to the normal operation of a transformer, although it may also resemble a pre-faulty state, as C_2H_6 distribution hints.

The gas C_2H_6 , followed by carbon monoxide and hydrogen, is the gas with the widest distribution by value of concentration, reaching up to 600 ppm. These gases show the highest absolute

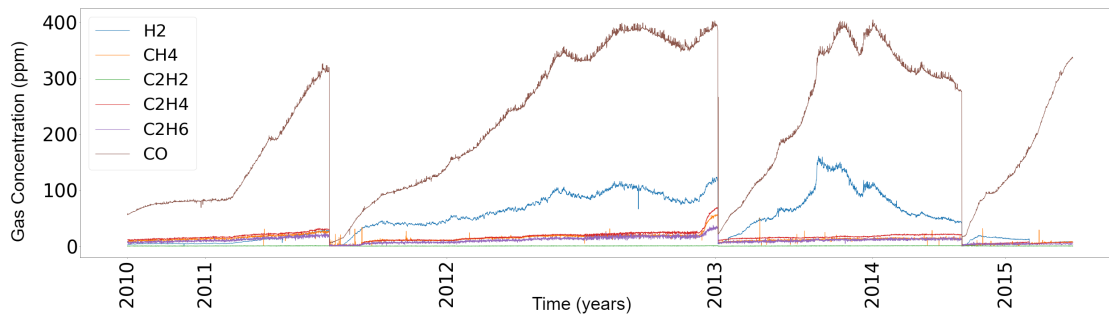


Figure 4.1: Plotting the Transformer C data frame.

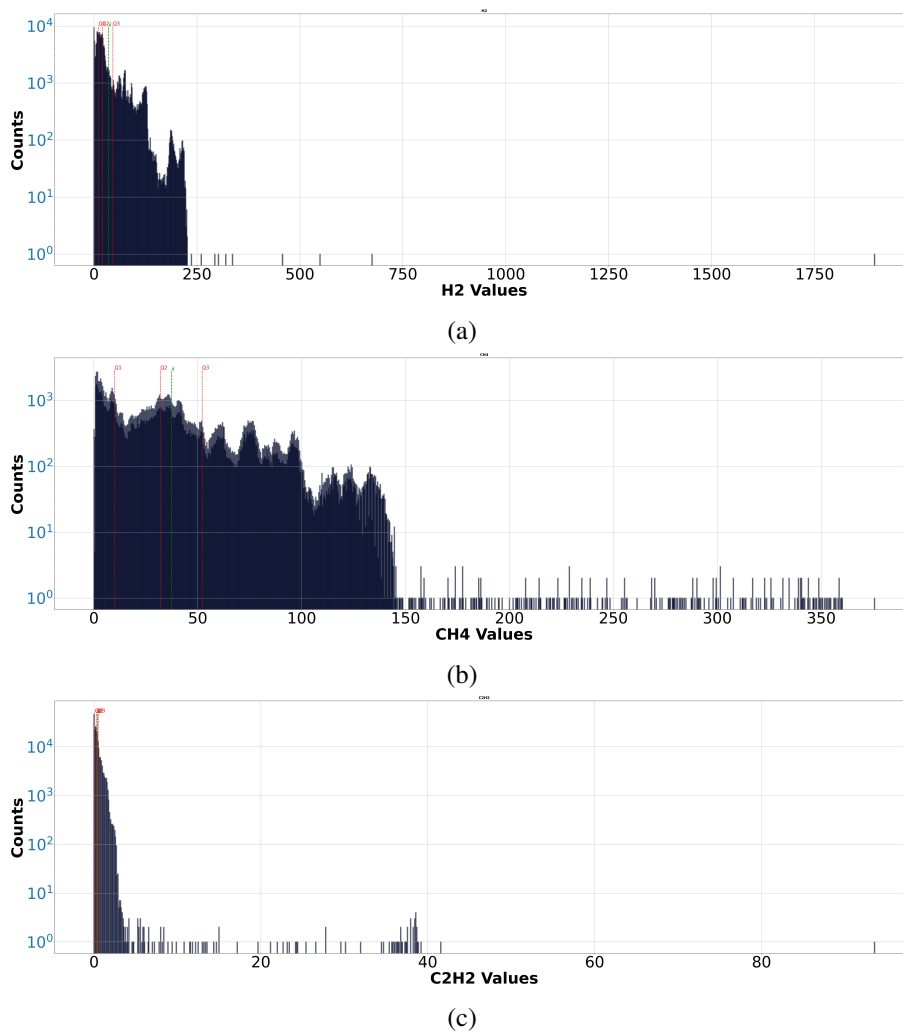


Figure 4.2: Concentration histograms for (a) H₂, (b) CH₄, (c) C₂H₂, (d) C₂H₄, (e) C₂H₆ and (f) CO gases.

concentrations in the data set, thus being the best candidates for prompting a faulty state, according to the TDCG procedure.

With this knowledge, Figure 4.1 was created to provide a deepened visualization of the data,

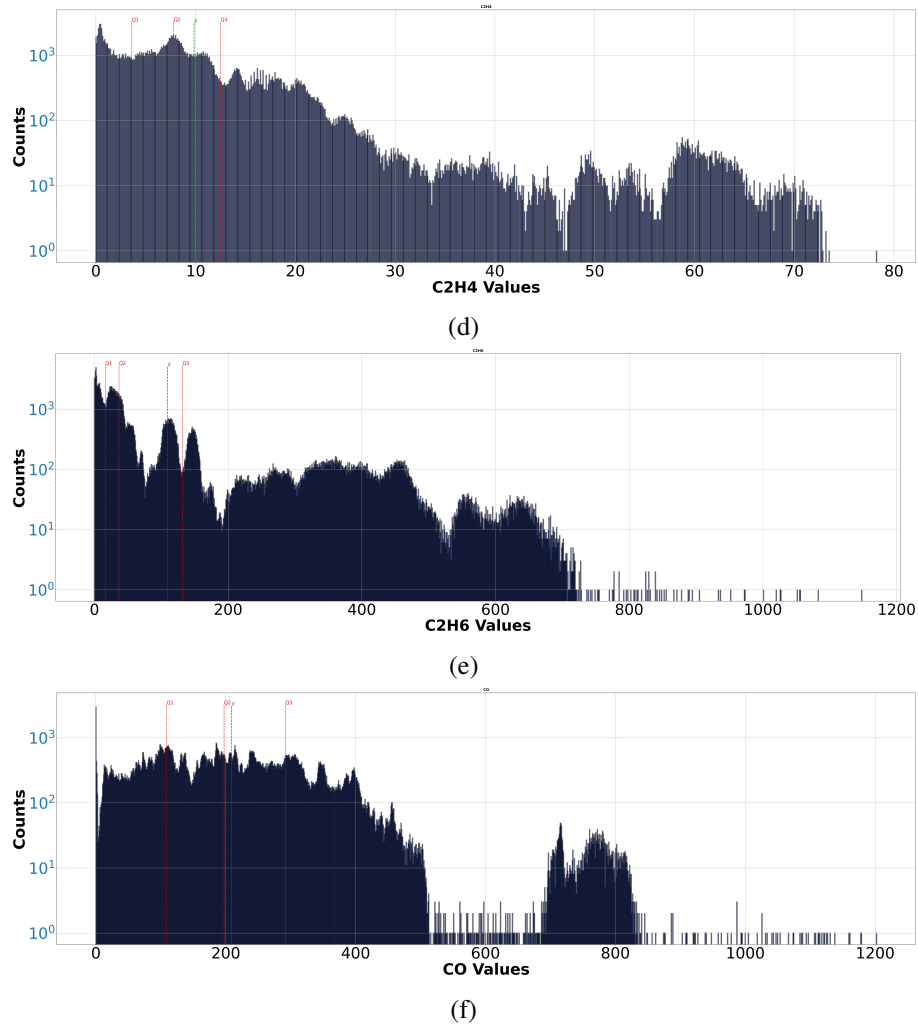


Figure 4.2: Concentration histograms for (a) H_2 , (b) CH_4 , (c) C_2H_2 , (d) C_2H_4 , (e) C_2H_6 and (f) CO gases.

in this study case, belonging to Transformer C. Again, only the combustible gases were considered. Examining deeper Figure 4.1, several abrupt changes of gas concentrations throughout the years of the transformer's operation are noticeable. These discontinuations are most likely related to oil changes, since most of the gases concentrations drop when this event occurs, though they can also be attributed to operation and maintenance breaks. When the transformer oil is replaced or renewed, a decrease in dissolved gas content is detected in the next analysis, hence this phenomenon.

Knowing that this event occurs in the data set, dividing the intervals that are continuous becomes of first concern due to the need of adequate processing of the data before model training. The employed model must not learn the discontinuities, but what leads to them and the overall process of dissolved in oil gas generation. After the partition of segments, various DGA regions are obtained. Figure 4.3 details the second region of the Transformer C, spanning from late 2011 to the beginning of 2013.

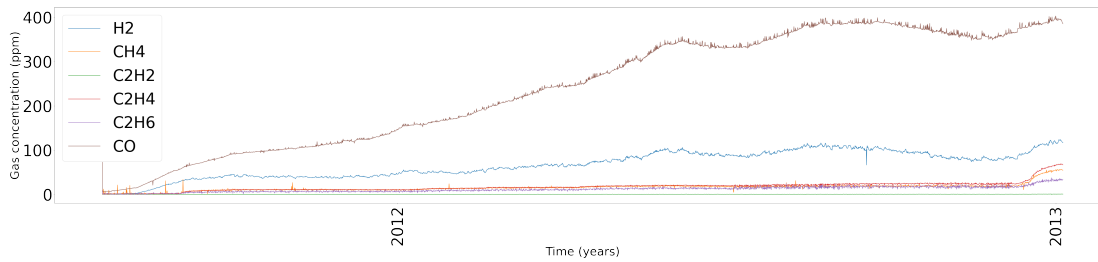


Figure 4.3: Transformer C second region detailed.

In addition, this region of the data is characterized statistically in Table 4.2.

Through examining Figure 4.3 is clear the increase of gas concentrations overtime. This increase is more prominent, at least with CH_4 , C_2H_4 and C_2H_6 , in the last instances of the time series, a few days before the end of 2012. This behavior is a potential trigger, and possibly an indication of future fault, since qualified actors performed a maintenance on the transformer, as Figure 4.1 corroborates.

Column Name	Mean	Std.Dev.	Coef.Var.	Q1	Median	Q3	Max.	Min.
H2	67.255	29.746	0.442	41.325	68.05	92.8	123.5	1.1
CH4	15.465	8.273	0.535	10.700	16.20	18.7	56.4	0.7
C2H2	0.421	0.172	0.409	0.300	0.40	0.5	1.1	0.0
C2H4	17.428	10.435	0.599	10.800	17.30	22.3	69.2	0.0
C2H6	11.604	6.322	0.545	6.800	11.40	15.9	37.6	0.0
CO	240.348	125.004	0.520	119.200	251.40	359.6	402.7	6.2

Table 4.2: Transformer C region 2 statistical analysis.

In Table 4.2, data associated with C_2H_2 is by no means significant, since its standard deviation is roughly null, confirming the already stated report. Also as anticipated, hydrogen and carbon monoxide show high deviations, a result of growing tendencies overtime. However, C_2H_6 does not follow exactly the distribution presented in Figure 4.2, considering its maximum value in this region from the Transformer C is about 40 ppm. This inconsistency may be attributed to the maintenance operated to the transformer, not letting the positive gradient of the gas develop.

In Table 4.3, a correlation analysis between the dissolved combustible gases is presented. As foretold, correlations between C_2H_2 and any other variable are weak, due to lack of information this gas provides. The most notable correlations, both Pearson and Spearman, are between the hydrocarbons (with the exception of C_2H_2) and between hydrogen and carbon monoxide. The Spearman's rank coefficient shows high positive correlation, but also a high Pearson correlation coefficient, which indicates an almost linear and monotonous relation between the variables, implying a simultaneous increase of gas' concentrations.

The Phi coefficient displays significant values in two unique relations, namely between H_2 and CO and between CH_4 and C_2H_6 . The relation between the two non hydrocarbon gases equals 0.54 and implies a positive relationship between them. As Figure 4.3 confirms, when carbon

Columns Pair	Pearson	Spearman	Phi Coef.
H2 CH4	0.781	0.892	0.057
H2 C2H2	0.663	0.663	0.019
H2 C2H4	0.810	0.907	0.066
H2 C2H6	0.888	0.905	0.021
H2 CO	0.972	0.960	0.540
CH4 C2H2	0.627	0.659	-0.017
CH4 C2H4	0.976	0.924	-0.027
CH4 C2H6	0.887	0.869	-0.427
CH4 CO	0.741	0.908	0.073
C2H2 C2H4	0.662	0.704	0.007
C2H2 C2H6	0.686	0.683	-0.005
C2H2 CO	0.681	0.693	0.003
C2H4 C2H6	0.942	0.965	0.232
C2H4 CO	0.785	0.961	0.053
C2H6 CO	0.889	0.945	-0.011

Table 4.3: Transformer C region 2 correlation analysis, with the maximum values pointed out in bold.

monoxide's concentration values fluctuate, so does hydrogen's. Contrariwise, the Phi coefficient of CH_4 and C_2H_6 expresses a somewhat opposing behavior. This relationship, however, is not so visible. Further analysis is required, to capture a the relationship at a short term level. Thus, a correlation through time is enacted.

Figures 4.4 and 4.5 present an evolution of relationships over time computed with optimal window size and step, obtained by a grid search algorithm. Displayed, are the three above mentioned correlations within limits. These limits, between magnitude 0.75 and 1, define the most correlated areas of the time series.

In Figure 4.4, the Phi coefficient increases through time, that is, the association between H_2 and CO increases overall. It averages the value of 0.54, as estimated in Table 4.3. Likewise, both gases display, at times, true linear and monotonous correlations. In Figure 4.5, the Phi coefficient is always negative, oscillating around -0.5, which would indicate a negative relationship. To some extent, this is true, although a near positive linear correlation confronts this hypothesis. This correlation is most strong on the tail of the windowed region.

In Figure 4.6, is possible to recognize the intrinsic correlation of hydrocarbon gases. Up close, the readings of these gases display notably amounts of variability. This inconsistency would explain the faint Phi coefficient association between the hydrocarbons. Nevertheless, the linear and monotonous correlations are unmistakable, providing a solid relationship between these gases.

Since just the behavior of this transformer was approached in depth, and considering that it might not represent the full picture of the data set, an average of all correlation analysis was calculated, present in Table 4.4.

As evidence showed, in the analysis of Transformer's C region, present in Figure 4.3, the Pearson and Spearman correlation analysis conferred similar results. The same fact stays true in

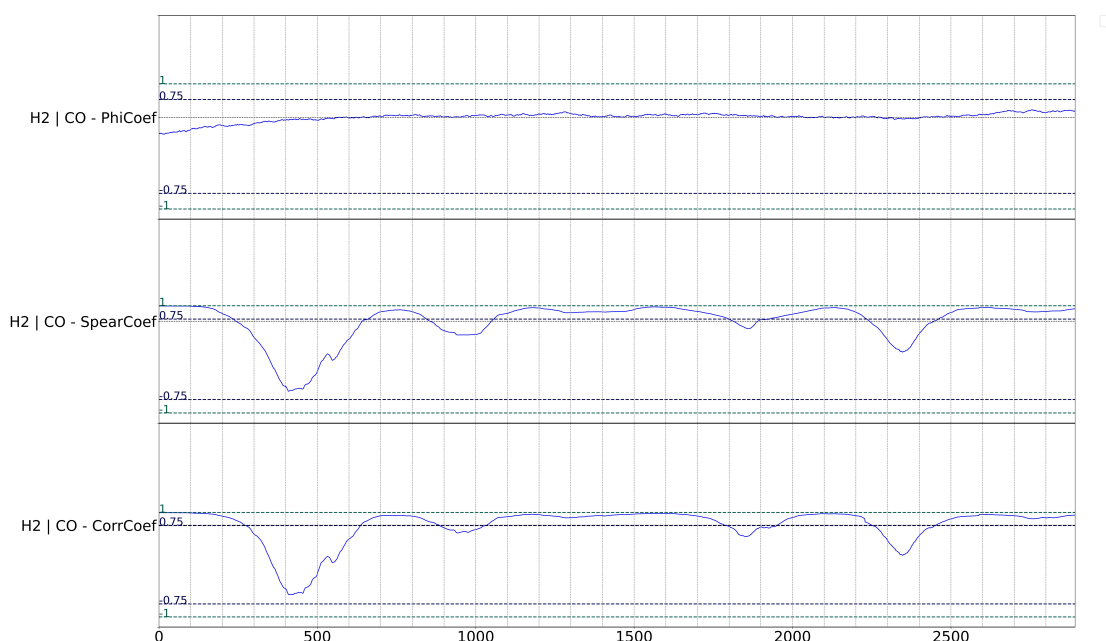


Figure 4.4: Correlation through time between H_2 and CO , with frame size 360 and step 1.

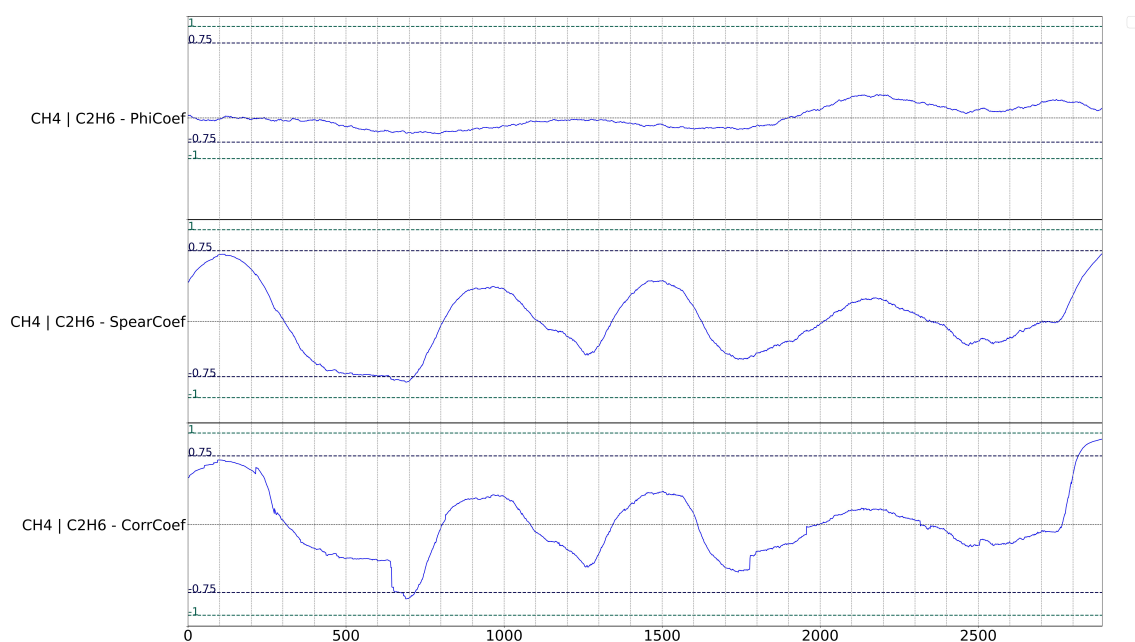


Figure 4.5: Correlation through time between CH_4 and C_2H_6 , with frame size 360 and step 1.

Table 4.4. Overall, the strongest relationships in the data are those that are characterized by a high absolute mean and a low coefficient of variation.

Monotonous and linearly, the strongest correlations are between the hydrocarbons, and between carbon monoxide and the other gases (with the exception of C_2H_2 , as aforementioned). Although, the averages of the correlation analysis values are not significantly high, in comparison with the other tabulated values and considering their low coefficient of variation, they indicate a

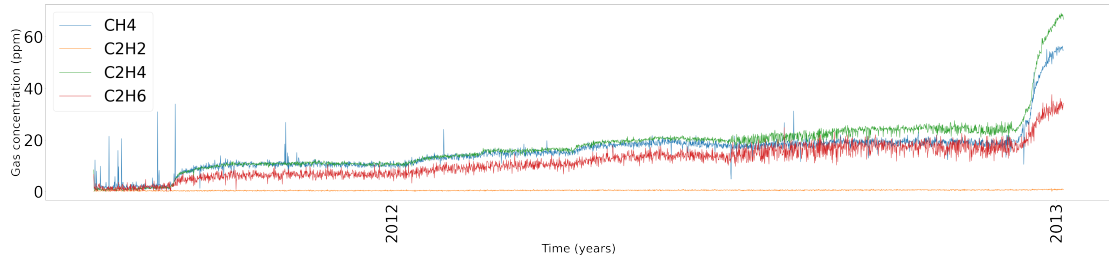


Figure 4.6: Hydrocarbons in the second region of Transformer C.

Columns Pair	Pearson Correlation		Spearman Correlation		Phi Coefficient	
	Mean	Coef. Var.	Mean	Coef. Var.	Mean	Coef. Var.
H2 CH4	0.381	1.321	0.388	1.230	0.081	1.439
H2 C2H2	0.122	2.994	0.096	3.185	0.081	1.107
H2 C2H4	0.338	1.247	0.334	1.237	0.022	8.950
H2 C2H6	0.321	1.538	0.342	1.398	0.027	3.990
H2 CO	0.549	0.879	0.532	0.895	0.527	0.341
CH4 C2H2	0.107	2.996	0.100	3.040	-0.010	4.049
CH4 C2H4	0.582	0.639	0.558	0.640	-0.071	1.274
CH4 C2H6	0.561	0.835	0.516	0.885	-0.304	0.701
CH4 CO	0.663	0.542	0.642	0.589	0.119	1.117
C2H2 C2H4	0.143	2.238	0.114	2.667	0.004	13.632
C2H2 C2H6	0.045	7.044	0.052	5.601	0.009	4.170
C2H2 CO	0.102	3.613	0.065	5.220	0.024	1.778
C2H4 C2H6	0.617	0.538	0.614	0.508	0.237	0.673
C2H4 CO	0.538	0.744	0.513	0.788	0.048	1.026
C2H6 CO	0.554	0.821	0.533	0.826	0.025	3.589

Table 4.4: Correlation analysis of the entire data set.

discernible correlation.

Regarding the Phi coefficient, the strongest association is between H_2 and CO , followed by CH_4 and C_2H_6 . Hydrogen and carbon monoxide, as recognized previously, have the strongest association throughout the whole data set, as it's low coefficient of variation corroborates. The preceding analysis on the Phi coefficient of CH_4 and C_2H_6 still holds true, and the global analysis confirms that this pattern is recognizable in the whole data set.

Generally, it seems that the preceding analysis on that one region conveys a broad aspect of the data. In the next section, the artificial labeling of the data set is undertaken.

4.2 Artificial Labeling

To demonstrate the method for fault detection and identification, the prior example is further depicted. In Figure 4.7 the TDCG is plotted, as well as a reference line for healthy performance, according to Table 3.2.

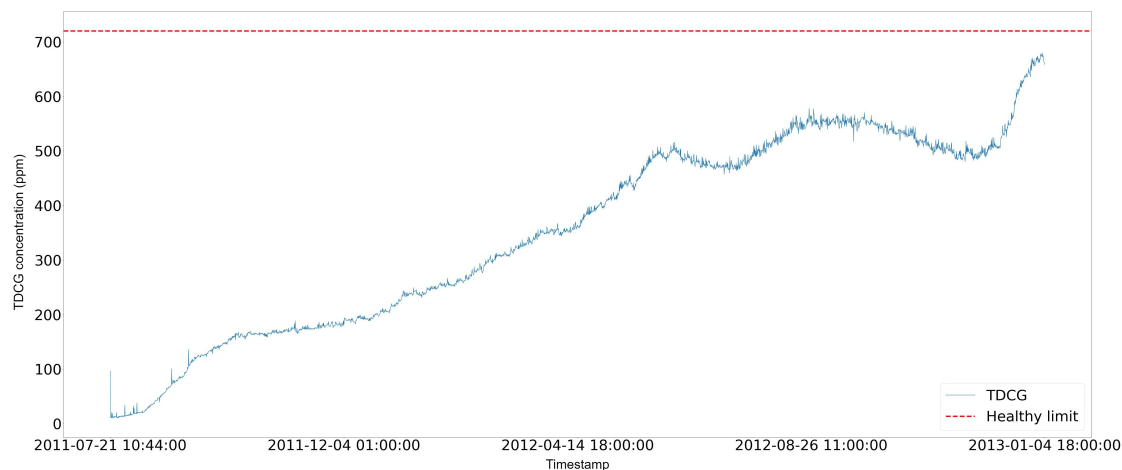


Figure 4.7: Total Dissolved Combustible Gases in region 2 of Transformer C.

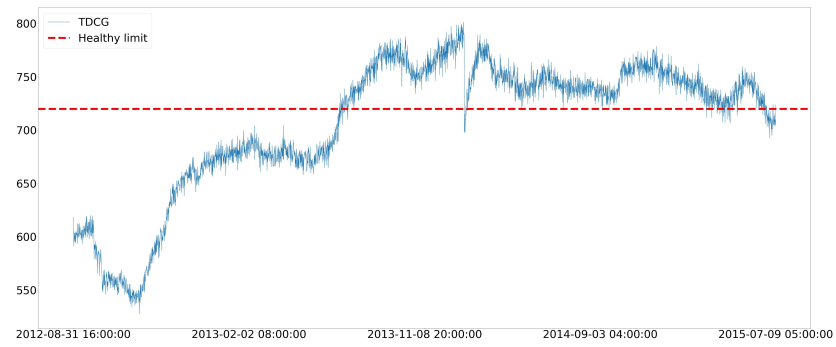
Although the TDCG increases overtime, the total of gases does not reach the threshold for faults to be present. This behaviour is not favored, since broad data with different events is preferable. Besides, it might also require means for augmenting the available data set, since the determined target for faulty behavior of a Power Transformer is not reached.

Upon considering all the available data, faulty states only emerge a few times. These are presented in Figure 4.8. With these cases, showcasing the artificial labelling algorithm is worthwhile. However, only those examples that develop into a fault are of interest, and not those that are faulty throughout the whole time series. This is the case of Transformer M, displayed in Figure 4.8d, which detains a considered faulty condition during the nearly two years of its continuous operation.

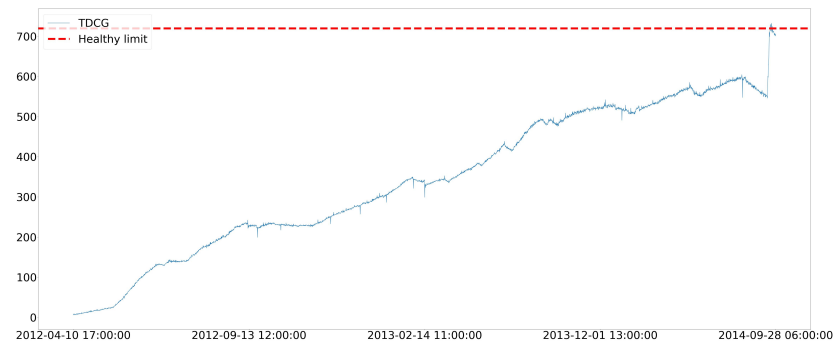
In Figure 4.9, the faulty states of the data are recognizable. These are concurrent with the earlier analysis, that is, fault type equals null when the TDCG is below the threshold value, represented by the red line in Figure 4.8, and fault type is greater than zero in the other cases.

The example from Transformer H, depicted in Figures 4.8a and 4.9a, illustrates well the algorithm for artificial labelling. After the TDCG reaches the limit, fault identification is executed, yielding a type 2 fault, or, according to Table 3.1, a low energy thermal fault. Throughout the development of the combustible gases, these fall below the condition for faulty behavior, but as quickly as they dip, they rise again. This is what causes the vertical strikes in the fault type plot, which represents the transition between thermal fault and non fault states of the transformer. However, this property could be rejected, since the interest is in predicting up to the faulty point, and not further.

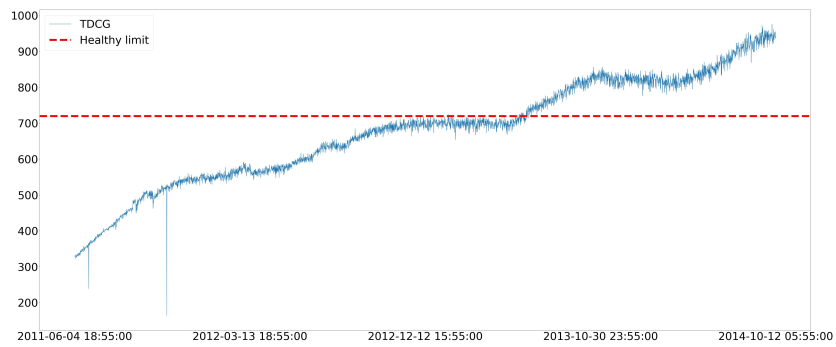
The case from Figure 4.9b is straightforward, seeing that it reaches a faulty state of low energy discharges at the end of the time series. Transformer L, represented by Figure 4.9c, reaches the faulty state at the end of 2012, but promptly falls back to a healthy state. It also reaches a more severe faulty thermal state around 2014. Although, these regards can be discarded, since the TDCG limit value is reached first before, substantiating the later evident faults. The time series can be



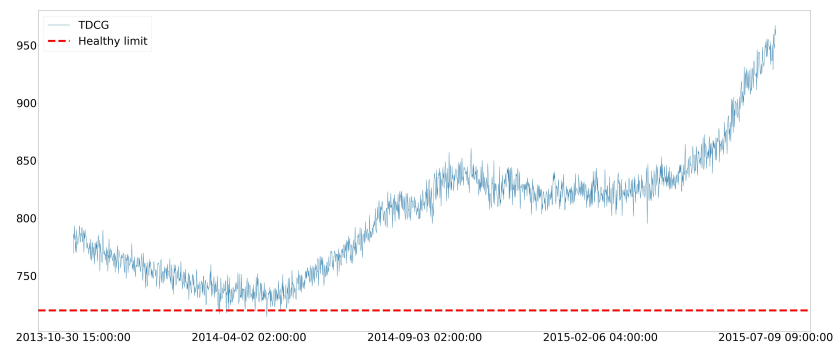
(a)



(b)

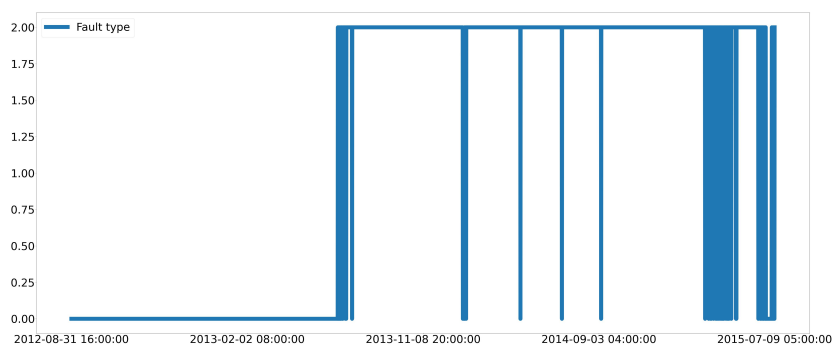


(c)

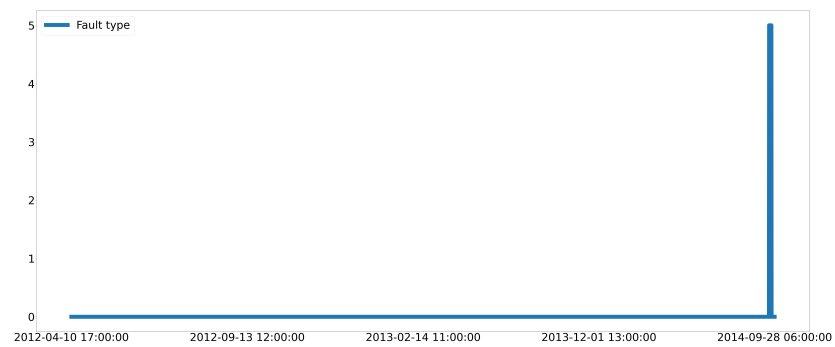


(d)

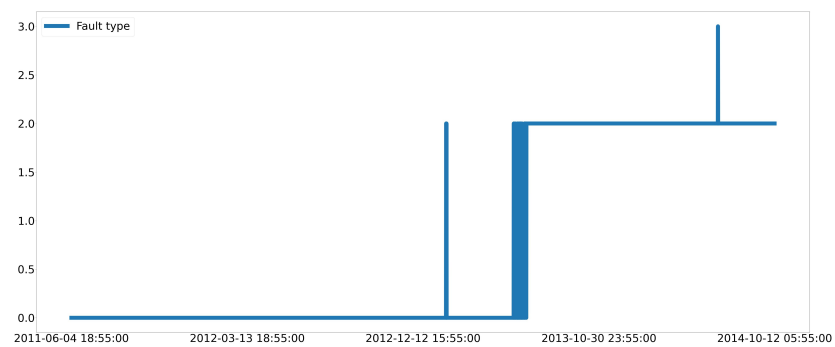
Figure 4.8: TDCG region plot of Transformers (a) H, (b) J, (c) L and (d) M.



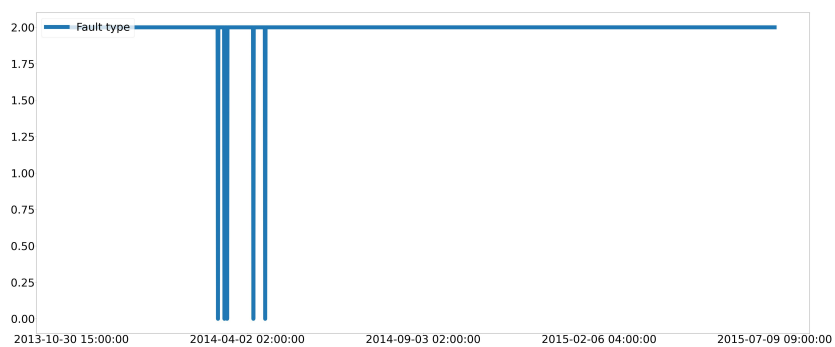
(a)



(b)



(c)



(d)

Figure 4.9: Faulty type plot of Transformers (a) H, (b) J, (c) L and (d) M.

cut from the point it first reaches a faulty behavior, considering that the prediction algorithm leads only to a faulty situation.

The example from Transformer M is distinct. Since it is constantly in fault, predicting algorithms cannot be applied. For this reason, examples with this characteristic are discarded from the data set and not used for fault prediction.

4.3 Data augmentation

As aforementioned in section 3.5, those time series that don't reach the determined target for the transformer to be considered faulty, an extension of the data set is performed. The process of augmenting the data sets is executed through a deep learning algorithm with the addition of biases. Timestamps are also extended to allow a time interval approach to the condition warning.

In Figure 4.7 a case for time series extension is showcased. The augmenting algorithm is feed a single variable from the data set, giving out a prediction for the next gas concentration. Then, the model is further trained with the new value to produce one more forecast. This is reproduced up to then thousand steps, circa 3 more years of data and to all gases in the data set. Displayed in Figure 4.10 is the result of this procedure applied to the recalled example. To help visualize the boundary between the real and augmented data, a red vertical line is drawn.

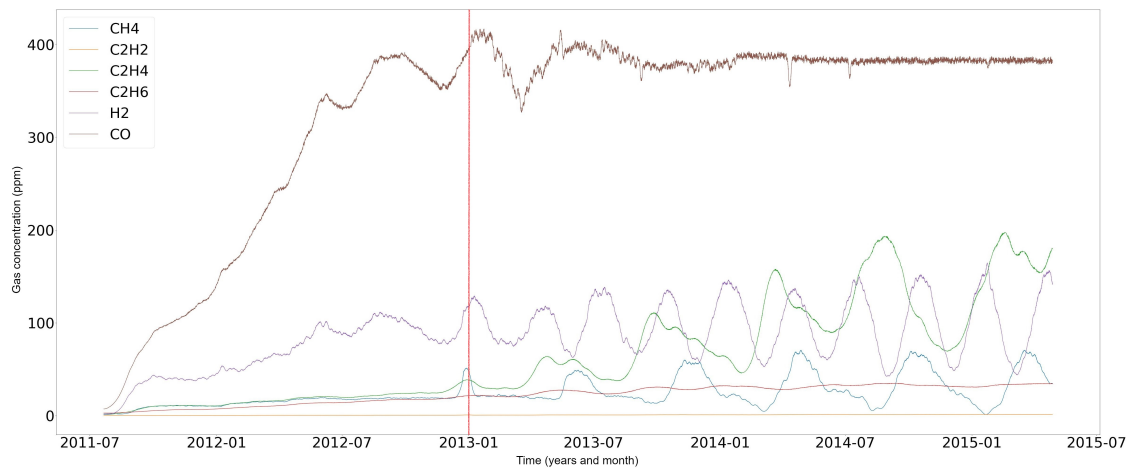


Figure 4.10: Augmented time series of region 2 of Transformer C.

Severe oscillations in gases' concentrations are noticeable throughout the augmented years. This behavior is expected, since the data extension algorithm is retrained each time with its own predictions, aggravating the hidden relations learned by the ANN. Nevertheless, because this issue worsens the further the data augmentation, it is acceptable as a background truth until the moment it reaches the defined limit for the first time.

In Figure 4.11, the total combustible gases reaches the healthy limit around May of 2014. Through examining Figure 4.7, the proposed fault is expected to occur very soon, yet, and keeping

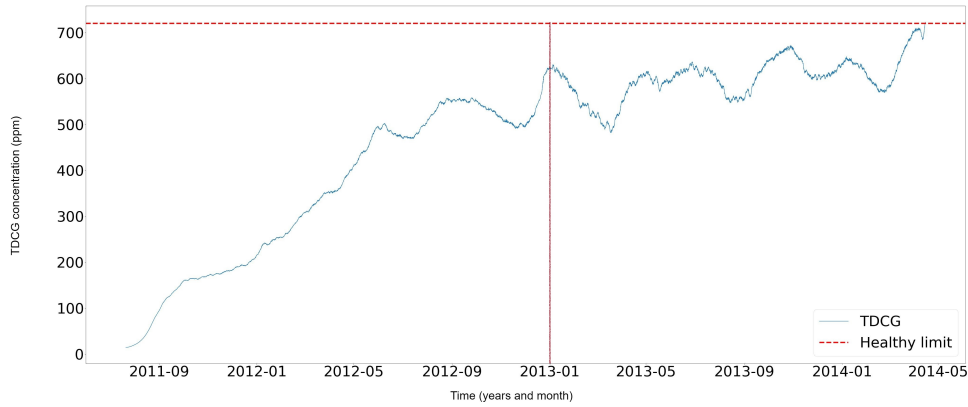


Figure 4.11: Total Dissolved Combustible Gases in region 2 of Transformer C extended data set.

in mind Figure 4.1, the TDCG healthy limit is reached just as the third region of Transformer C is about to end.

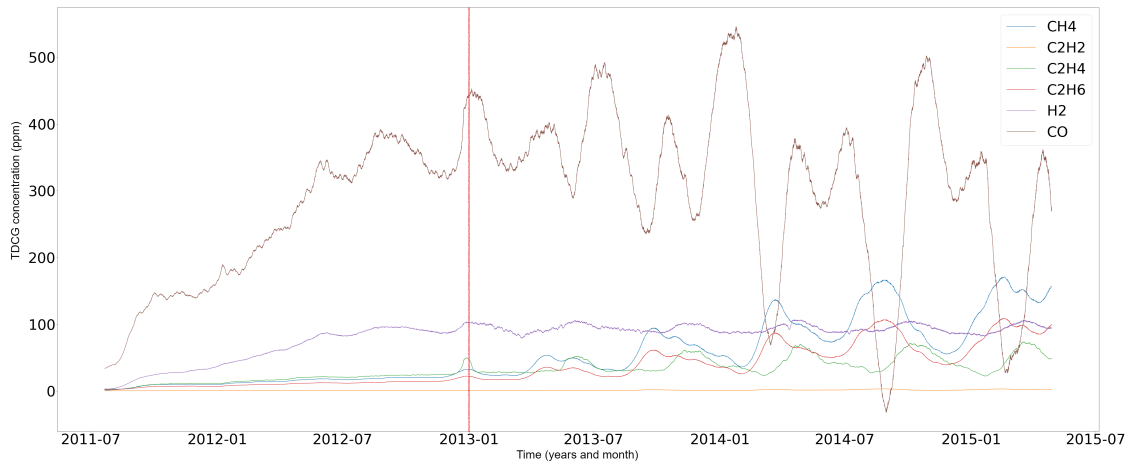


Figure 4.12: Augmented biased time series of region 2 of Transformer C.

Upon running the fault identification algorithm, a type T3 or high temperature thermal fault is diagnosed. Although the result is expected, for C_2H_2 concentration is inconsequential, leading to faults of the thermal type. This diagnosis could have been different, since CH_4 and C_2H_4 concentrations fluctuate in the expanded series. Still, it would remain on the edge of the Duval Triangle, as Figure 2.5 sustains.

Taking a look at the biasing technique, gases' known relationships are able to be enforced in the new extended data in three forms. These are all done through a linear regression algorithm, that computes relations of gases from the entire data set, from unique transformers and from each region.

Despite carefully choosing the bias weights, oscillations tend to magnify and gases' concentrations even attain negative values, as Figure 4.12 testifies. However, as evidenced by Figure 4.13,

by adding to this case a small universal bias and a 55% and 40% bias to transformer and region weights respectively, the condition of fault is possible to fulfill earlier. This approach could be considered pessimistic, comparing with the unbiased extension of Figure 4.11, since the healthy limit is reached months before. Albeit, upon fault diagnosis, the identification algorithm labels the fault as a T2 type thermal fault. This inquiry proves appropriate, considering that the earlier identified fault is of the same type, but of lower energy.

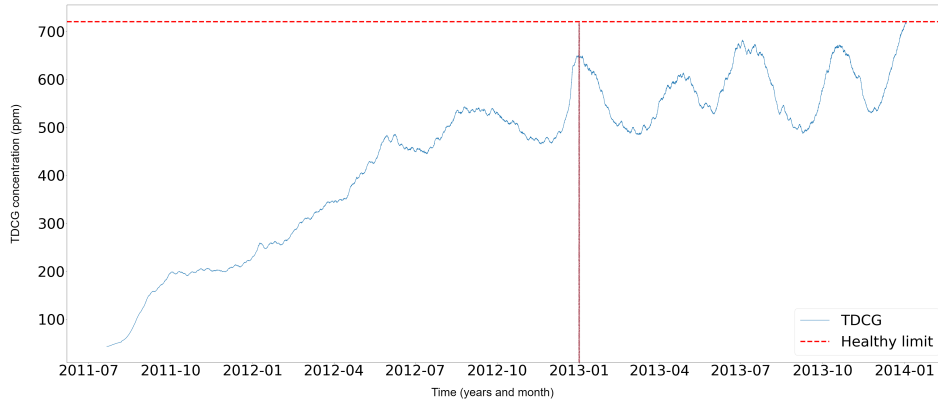


Figure 4.13: Total Dissolved Combustible Gases in region 2 of Transformer C extended and biased data set.

Hence, each unique region was extended and added optimal individual biases for a pessimistic approach. In Figure 4.14 are present those time series that achieved the considered condition for fault. To note that, from 36 transformer DGA regions only 8 turn out to be faulty, being 3 of those time series already faulty in the original non augmented data. These 8 time series present diversified behaviors. Therefore, this set is adequately representative for the evaluation of the algorithms. Thus, the 8 DGA time series set is the established background truth for the next section.

4.4 Fault Prediction

In this section, the models applied to problem of predicting faults in Power Transformers and their results are explored. Simpler models, such as regressions, are used as a base comparison for other more complex models, as is the LSTM and ARIMA.

4.4.1 Regression models

The regression algorithm is applied to all cases of DGA data that eventually leads to a fault. The best model, among multi linear regressions, such as exponential and polynomial of various degrees, is chosen by means of the coefficient of determination or R^2 , calculated during the validation process while training.

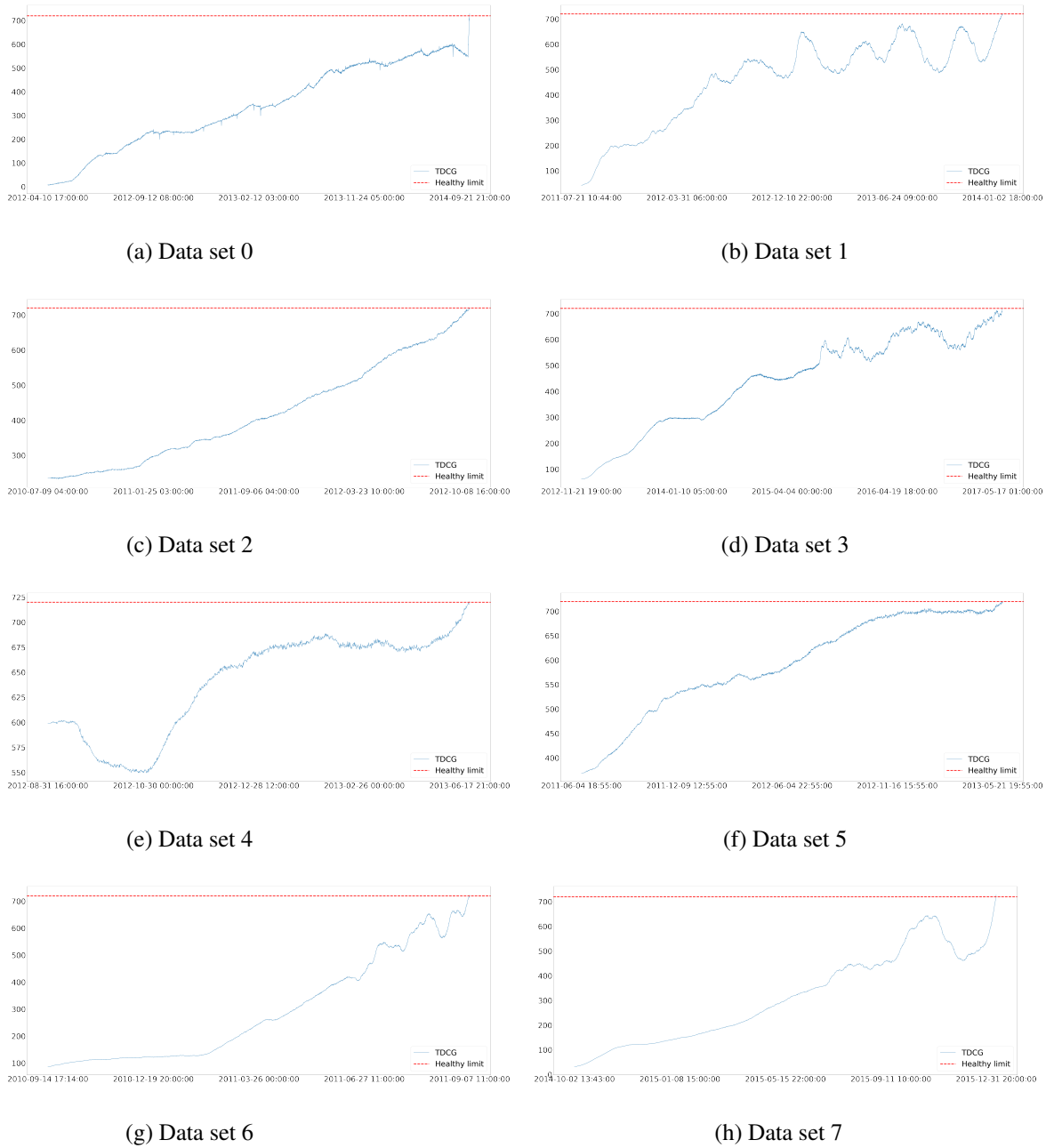


Figure 4.14: Transformer (a) J, (b) C region 2, (c) F region 1, (d) F region 2, (e) H, (f) L, (g) C region 1, (h) C region 4.

For a first approach, predicting a single step ahead, regression models behave remarkably, as Figure 4.15 shows. The presented section of the time series is the testing set with size equal to twenty percent of the total original length. For all time series, the polynomial regressor presents as the most appropriate model.

Despite the fact these regression models perform well on a single step ahead prediction, in actuality they are only forecasting up to 4 hours in the future (in average, each DGA point is 4

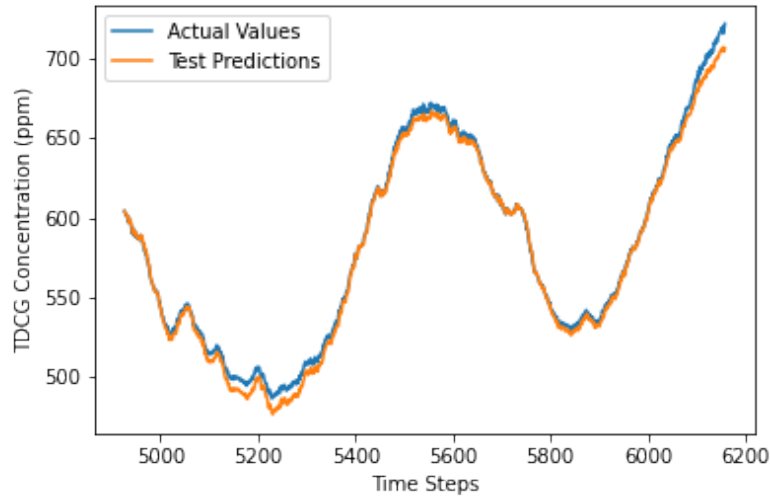


Figure 4.15: Regression and one time step prediction of Transformer C region 2.

hours apart from the others). Therefore, a 5 day, or 30 time steps, prediction is executed.

To achieve a multi step ahead forecast with these simple algorithms, indeed thirty different models are trained and combined for each time series prediction. Then, this process is done several times, in order to obtain an average of predictions. The chosen prediction is actually an average of several assembled models of the same regressor type.

In Figure 4.16 are present three case results, representing each possible situation, that is to say, a favorable prediction, an inferior one and a failed forecast. Plotted in the figures are the real and forecasted values, but also the curve fitting. The horizontal blue line defines the defined threshold, and the vertical ones represent the portion of the data on which the models were trained.

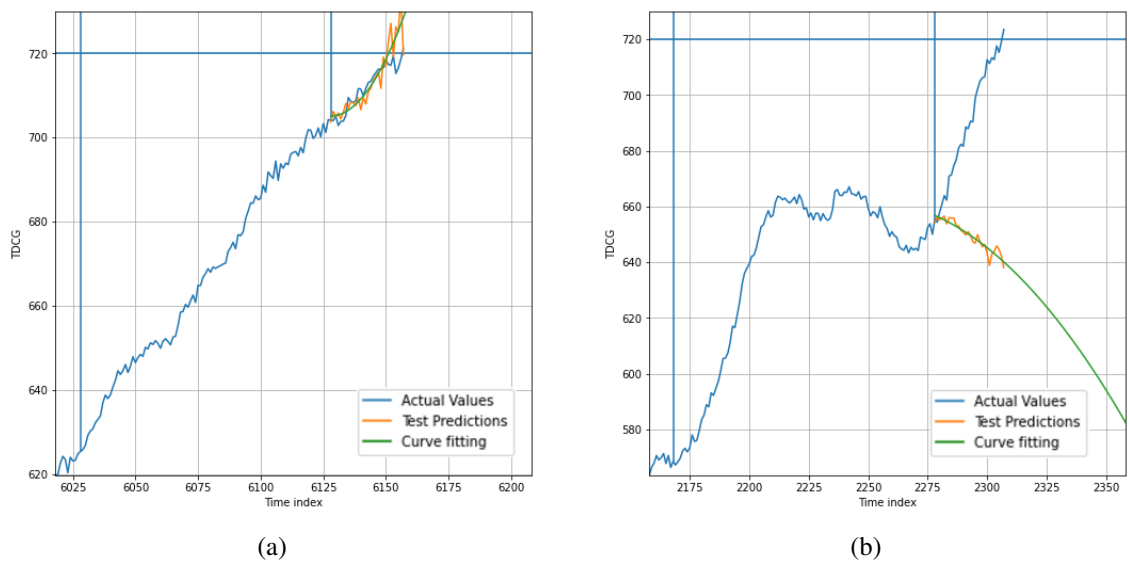
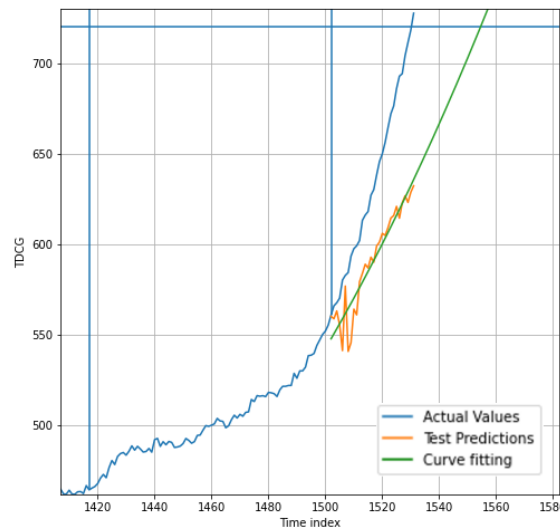


Figure 4.16: Regression model 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7.

In Figure 4.16a is present a positive prediction, that reaches the agreed limit before the real time series. Hence, from a pessimistic point of view, this forecast is very well suited. Per contra, Figure 4.16b offers an opposite prediction, failing to reproduce the trend. This could be due to chosen point or instant of prediction, that sits on the turning point of the data. Figure 4.16c displays a prediction example that required fitting a curve for it to reach the limit. In green, a curve is drawn on the predictions to further extend the forecast and obtain the time to fault.



(c)

Figure 4.16: Regression model 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7.

In Table 4.5 are summarized the results of the predictions using regression algorithms, by determining the estimated time for fault against the actual time to fault, differences and errors.

Data Set	Time for fault	Predicted time	Difference	Error
0	120 hours	None	-	-
1	120 hours	96 hours	-24 hours	20%
2	120 hours	108 hours	-12 hours	10%
3	120 hours	164 hours	44 hours	36.667%
4	120 hours	92 hours	-28 hours	23.333%
5	120 hours	None	-	-
6	120 hours	None	-	-
7	120 hours	212 hours	92 hours	76.667%

Table 4.5: 30 step ahead regression prediction performance.

Although some examples don't fulfill the established limit in the predicted time steps, a fitted curve allows the extension of the forecast, and, therefore, the estimation of the time to warning. Yet, this approach is not favorable, considering that large errors seem to be associated with it. Besides, failing to predict a fault in useful time is not valuable, as sometimes is better to predict a fault sooner than later. Still, as is the case with data set 0, 5 and 6, the models fail to predict the positive trends.

Accomplished the benchmark for comparison, next, the ARIMA model, an advanced ML algorithm for time series analysis, is considered.

4.4.2 Auto-regressive Integrated Moving Average

Parameters for an ARIMA model are optimized through the auto-ARIMA functions of the pyramid library. The best model, according to the employed Akaike Information Criteria is used for each time series. Similarly, 30 time steps are predicted from the fixed instant.

Illustrating the functioning of this model, in Figure 4.17 are present the above examined examples. To note that, in these examples the supplementary curve fits near perfectly on the ARIMA predictions.

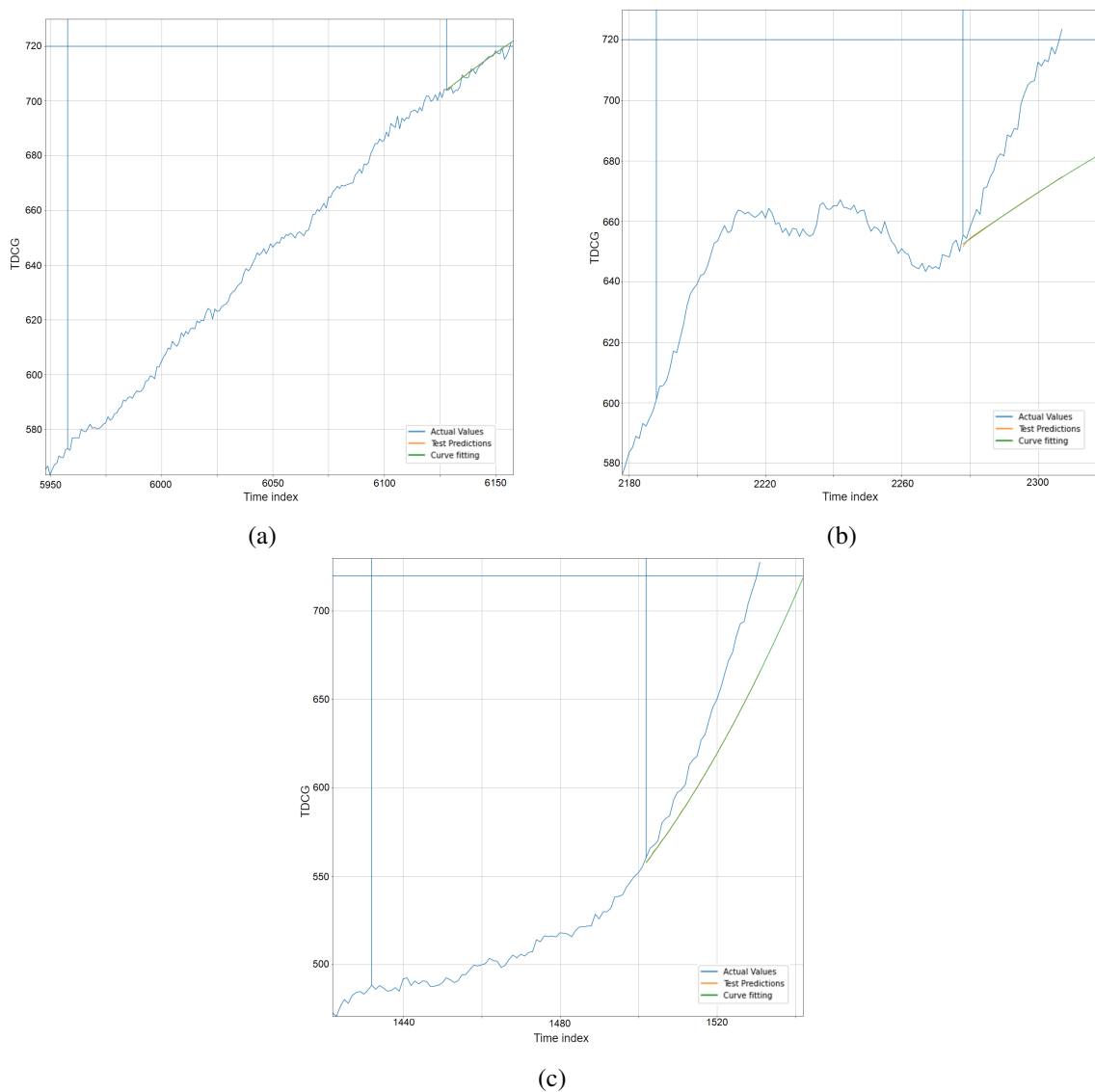


Figure 4.17: ARIMA model 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7.

Figure 4.17a, as expected, presents an exceptional example. The prediction fails its target by only one time step, or 4 hours. In Figures 4.17b and 4.17c the estimations are not quite there, although they reproduce positively the trend, they still require the aid of the curve fitting. Particularly in Figure 4.17b, since the instant of estimation is on a tipping point of the curve, the predicted slope is lower than the true tendency.

Compiling the results, Table 4.6 is assembled. The ARIMA model presents the best results so far, in terms of number of successful predictions and overall accuracy. This model fails to predict only data set 0, as it was foreseen. Once more, those predictions that do not reach the limit are extended by fitting a curve. Consequently, predictions of this type carry large errors. Nonetheless, the ARIMA model predicts with only one time step of inaccuracy 4 time series.

Data Set	Time for fault	Predicted time	Difference	Error
0	120 hours	None	-	-
1	120 hours	116 hours	-4 hours	3.333%
2	120 hours	116 hours	-4 hours	3.333%
3	120 hours	124 hours	4 hours	3.333%
4	120 hours	256 hours	136 hours	113.333%
5	120 hours	124 hours	4 hours	3.333%
6	120 hours	456 hours	336 hours	280.000%
7	120 hours	164 hours	44 hours	36.667%

Table 4.6: 30 step ahead ARIMA prediction performance.

With promising results, the following algorithm to be tested is the Long Short Term Memory ANN, which presents more flexibility and generalization capacity.

4.4.3 Long Short-Term Memory

A Long Short Term Memory ANN is build and trained with optimal values for each time series. In fact, developing one LSTM model per data set, for each training size, input window and dropout ratios.

As foreseen, the LSTM behaves exceptionally when predicting a single time step ahead. In Figure 4.18, the near perfect following of the trend is well captured by the LSTM. Although, when comparing this result to the simpler regression method present in Figure 4.15, the LSTM prediction falls short.

Taking a look at the 30 time steps prediction, leading results are present in Table 4.7. These are the best performing LSTM models obtained through the parameter search algorithm. Considering all the outcomes, the ones that perform better, according to the specified metrics, are from data set 1 and 5.

The noticeable difference between varying windows and training sizes, as well as dropout ratios suggests that the optimal parameterization and architecture is unique for each time series and each instant of prediction.

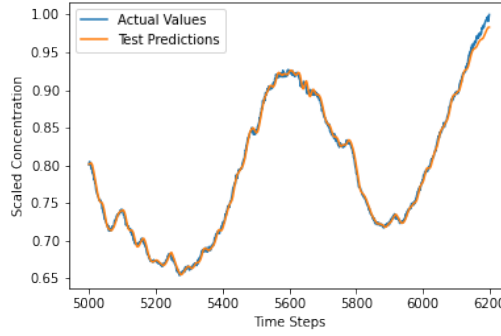


Figure 4.18: LSTM one time step ahead prediction of Transformer C region 2.

Data Set	MAPE	Training size	Window size	Dropout ratio
0	7.38%	0.5	400	0.2
1	0.839%	0.5	500	0
2	1.265%	0.25	100	0
3	1.13%	0.3	500	0
4	8.179%	1	400	0
5	0.964%	0.5	200	0
6	4.558%	0.25	400	0.2
7	7.597%	0.25	100	0.8

Table 4.7: Best 30 step ahead LSTM prediction performance parameters.

In Figure 4.19, the examples used so far are presented. Although these cases perform well, according to MAPE metric, they don't reach the established threshold. More, they completely fail to capture the final increasing trend. Altogether, it appears that only in Figure 4.19b a trend was somewhat captured and reproduced.

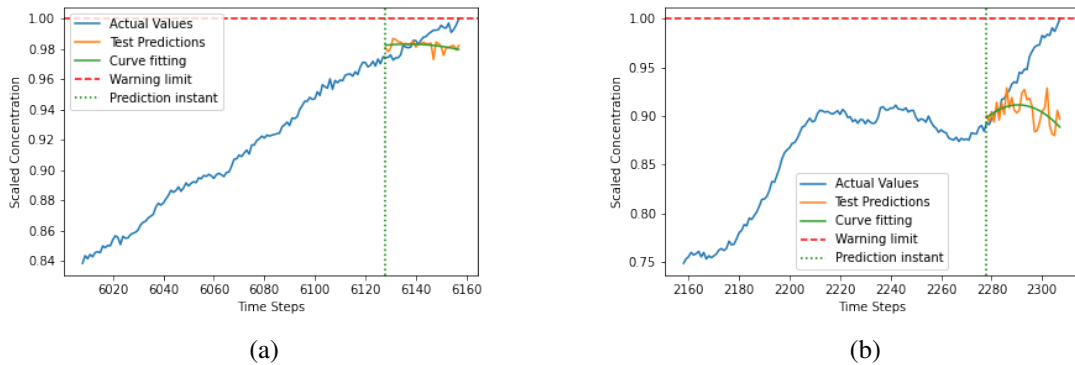
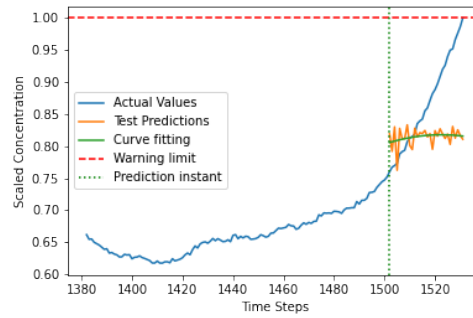


Figure 4.19: LSTM model 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7.

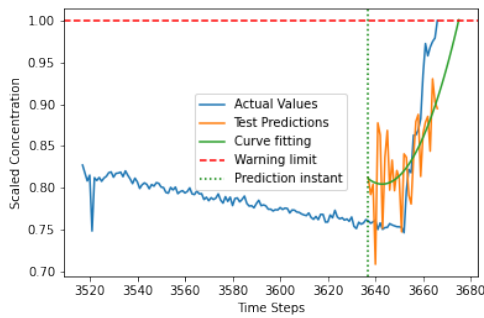
In fact, the LSTM merely generated one case that fulfils the warning condition independently, and another one that, after curve fitting, reaches the limit. In Figure 4.20 these cases are present. Peculiarly, the time series correspondent to data set 0 is only successfully predicted when employing



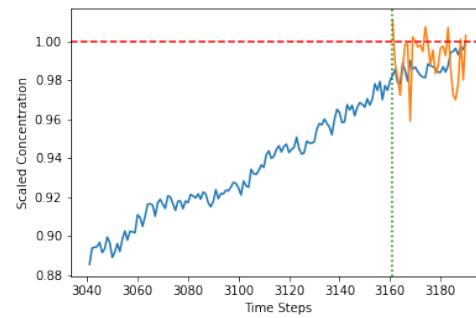
(c)

Figure 4.19: LSTM model 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7.

the LSTM, as Figure 4.20a depicts. Since this time series presents a such great increase, it was expected that it wouldn't be predictable. Thus, and as it is apparent, this good fortune is product of over fitting and is not an accurate prediction. As previously, results are compiled in the time Table 4.8.



(a)



(b)

Figure 4.20: LSTM model 30 time steps prediction of data sets (a) 0 and (b) 2.

Data Set	Time for fault	Predicted time	Difference	Error
0	120 hours	152 hours	32 hours	26.667%
1	120 hours	None	-	-
2	120 hours	0 hours	-120 hours	100%
3	120 hours	None	-	-
4	120 hours	None	-	-
5	120 hours	None	-	-
6	120 hours	None	-	-
7	120 hours	None	-	-

Table 4.8: 30 step ahead LSTM prediction performance.

Lastly, the curve fitting algorithm is tested on its own, and also an average of all model results is executed.

4.4.4 Curve fitting

To assess the curve fit prediction capabilities in an isolated way, the algorithm is applied in a small window before the point of forecast. In Figure 4.21 the studied cases are presented. To note that the employed fitting is polynomial of degree 2 in the case of Figures 4.21a and 4.21c, and of degree 3 in the case of Figure 4.21b.

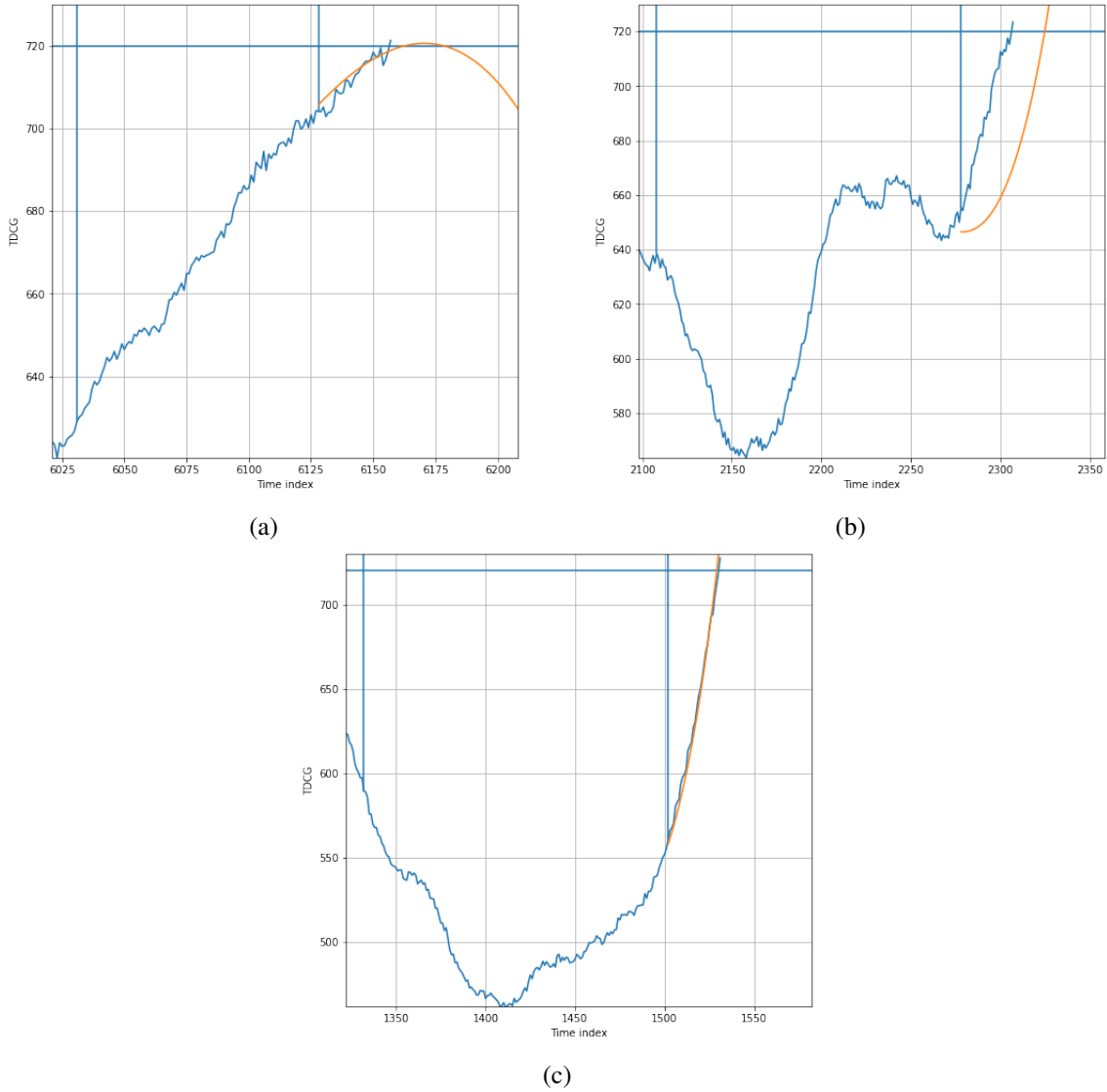


Figure 4.21: Curve fitting 30 time steps prediction of data sets (a) 1, (b) 6 and (c) 7.

Consistently, Table 4.9 presents the algorithms' results. Although the outcome is not admirable, in a couple of cases, the curve fitting is more appropriate than the regression models. Expressly in data sets 1 and 4. Though, in data set 1 the fitted curve estimates the intersection with the limit after the fault. In addition, this method produced predictions for every data set, with the exception of data set 0, whereas the linear regressions failed to detect two of them.

In light of a multi paradigm approach, Table 4.10 contains an average of all model predictions.

Data Set	Time for fault	Predicted time	Difference	Error
0	120 hours	None	-	-
1	120 hours	136 hours	16 hours	13.333%
2	120 hours	100 hours	-20 hours	16.667%
3	120 hours	60 hours	-60 hours	50.000%
4	120 hours	96 hours	-24 hours	20.000%
5	120 hours	124 hours	4 hours	3.333%
6	120 hours	188 hours	68 hours	56.667%
7	120 hours	112 hours	-8 hours	6.667%

Table 4.9: 30 step ahead Curve Fitting prediction performance.

As proposed, combinations of assorted models and techniques provides an improved analysis on the problem. The blending of unique model characteristics, in most cases, presents results as good as the best performing model.

Data Set	Time for fault	Predicted time	Difference	Error
0	120 hours	152 hours	32 hours	26.667%
1	120 hours	116 hours	-4 hours	3.333%
2	120 hours	81 hours	-39 hours	32.5%
3	120 hours	116 hours	-4 hours	3.333%
4	120 hours	148 hours	28 hours	23.333%
5	120 hours	124 hours	4 hours	3.333%
6	120 hours	322 hours	202 hours	168.333%
7	120 hours	162.667 hours	42.667 hours	35.556%

Table 4.10: Final 30 step ahead prediction performance.

4.5 Result Discussion

The main results of this work were the data augmentation and the fault prediction. Both were achieved, yet, with room for improvement.

The restricted availability of DGA records has a direct consequence on diversity and usefulness of data. When augmenting the data sets, the results preserved only already observed gases' relationships. Especially in the case of C_2H_2 , which is not relevant in normal functioning of Power Transformers, but is a key gas when faults occur.

When it comes to fault prediction, the proposed models performed acceptable, in consideration of the available data. By all means, no actual faults were predicted, since none were present in acquired data sets. However, the proposed threshold for prediction can serve as a warning. Particularly, allowing for gas concentration surveillance to alert any off limit behavior.

Possibly the least expected result was that of the LSTM. Such an adaptable ML algorithm should provide better results than those obtained. Albeit, the ANN results suggest that different configurations and composition of networks could demonstrate better outcomes.

4.6 Chapter Conclusion

In this chapter the results of the proposed methodology were presented. Data sets were described, statistically analyzed and augmented. ML models were built and explored for fault prediction. A table with the final predictions against the real events is bestowed.

Chapter 5

Conclusion

In this section, general and final conclusions are presented. This dissertation's contributions and limitations are described, as well as the overall fulfillment of the planned objectives.

5.1 General conclusion

This thesis presented a multi-paradigm approach to the fault prediction problem in Power Transformers, combining, not only traditional DGA methods for fault identification, but also advanced forecasting ML algorithms.

Inasmuch as DGA hasn't a unique established method, the purposeful combination of different algorithms and procedures demonstrates the improvement capability of this analysis. The blending of different paradigms allows a more flexible approach to problems, since it incorporates the best features of each one. Thus, the central question of this dissertation has not a single answer, but rather, a broad solution.

Besides the main focus of the work, a method for overcoming the problem of DGA data scarcity was proposed. The designed procedure for data augmentation proved to be an innovative solution, allowing a consistent application of the predicting methodology. It further proved, to some extent, that is possible to extend individual time series and relate it numerically to other gases.

Although prediction results are not satisfactory, they establish facts. That is, they demonstrate that predicting faults in Power Transformers using in oil concentration of gases is possible. The proposed method proved to be insufficient in obtaining meaningful results, nevertheless evidence regarding tendencies for simpler and linearizable local models was reported. Accordingly, the applied ARIMA model proved to be the better solution in this context.

In spite of being state of the art, the LSTM presents difficult configuration, suggesting that a more wary approach is necessary. Still, the fact that neural networks' potential is continually expanding, contributes to ever improving and advanced applications.

5.2 Contributions

Contributions of this thesis range from managing limited data, such as unlabelled and restricted time series, to an overview of the state of the art procedures for fault prediction in Power Transformers using DGA. Furthermore, the work consolidated the indispensable use of combinations of established and advanced ML models in this field.

Artificial labelling of time series is straightforward by the use of a semi-supervised approach. Again, combinations of different methods for health diagnosis in Power Transformers confirmed to be the more valuable solution. Likewise, augmenting gases' concentrations is achievable, though imposing statistical relationships demands a scrupulous application. Since Power Transformers can operate individually and in parallel, and each unique configuration has its own relations, when extending the data sets, all associations must be taken into account.

Perhaps the most relevant contribution of this work is that the fault prediction problem has not an isolated solution, considering DGA time series are erratic, albeit capable of being described locally by linear models. Hence, the method for solving the proposed question must rely on combination of elementary regressions and cutting edge models.

5.3 Limitations

During the development of this work, numerous and unaccounted hurdles arose. Particularly, the restricted availability of DGA time series data, the absence of labels in data sets and the difficulty of properly assemble and adjusting neural networks. More, the proposed methodology was likely much simplified for the problem.

The limited available DGA records delivered many uneventful and analogous cases, which consequently, required the process of augmentation. But enlarging data sets this way, perpetuated only already known gases' relations, which does not necessarily represent real phenomena.

The data extension, though acceptable, could be enhanced, had there been labeling of the data, such as load factor, temperature and other operation conditions. Nevertheless, a process of augmenting simultaneously multi variable time series would spare the biasing application.

As to computational limitations, the logarithmic and high degree polynomial transformations of multi linear regressions proved to be an impediment.

Neural networks were not exploited to their fullest, mainly because of an oversimplification of the problem. That is, solely using the TDCG content for fault prediction, gave the ANN fewer information to work with, preventing it from detecting hidden relations between gases. These intrinsic relationships could be of extreme importance for accurately predicting DGA time series.

5.4 Future work

The work developed in this dissertation, even if yielded not desired results, enables a more knowledgeable next approach. For better accomplishments, along with improving the methodology,

broader and more representative, as well as longer data sets would be valuable.

Concerning the applied method, the use of not only limits for total dissolved gases but for each unique gas would provide a wider supervision on DGA. More, positive gradient thresholds could be followed, providing an extra layer for faulty behavior estimation. A stability, trend seasonality and decomposition analysis could also present more insight into gas in oil evolution. These deeper examinations of the data, could facilitate the understanding and prediction of gas concentrations turning points.

In regards to modelization, convenient and more suitable algorithms can be used. Expressly, a combination of linearizable algorithms, for modeling basic tendencies, and adequate advanced models. ANNs should be trained with as much information as possible. Ideally, training neural networks on every gas for forecast and health index estimation. This indicator should be a combination artificial labelling and time series, such as a labeled TDCG.

Finally, this work could be adapted to forecast longer time periods, since the further the prediction, the better the usefulness.

Taking all things into consideration, the proposed question was undoubtedly tackled but not answered in its entirety, nevertheless giving definite directions for the future.

References

- [1] IEC. Iec standard 60599 - guide to the interpretation of dissolved and free gas analysis, 1999.
- [2] Issouf Fofana and Yazid Hadjadj. Electrical-based diagnostic techniques for assessing insulation condition in aged transformers. *Energies*, 9(9):679, 2016.
- [3] Janvier N’Cho, Issouf Fofana, Yazid Hadjadj, and Abderrahmane Beroual. Review of physicochemical-based diagnostic techniques for assessing insulation condition in aged transformers. *Energies*, 9(5):367, 2016.
- [4] Reza Khalili Senobari, Javad Sadeh, and Hossein Borsi. Frequency response analysis (fra) of transformers as a tool for fault detection and location: A review. *Electric Power Systems Research*, 155:172–183, 2018.
- [5] Venera Nurmanova, Yerbol Akhmetov, Mehdi Bagheri, Amin Zollanvari, Gevork B. Gharehpetian, and Toan Phung. A new transformer fra test setup for advanced interpretation and winding short-circuit prediction. IEEE.
- [6] Venera Nurmanova, Mehdi Bagheri, Amin Zollanvari, Kamilla Aliakhmet, Yerbol Akhmetov, and Gevork B. Gharehpetian. A new transformer fra measurement technique to reach smart interpretation for inter-disk faults. *IEEE Transactions on Power Delivery*, 34(4):1508–1519, 2019.
- [7] A. Akbari, A. Setayeshmehr, H. Borsi, E. Gockenbach, and I. Fofana. Intelligent agent-based system using dissolved gas analysis to detect incipient faults in power transformers. *IEEE Electrical Insulation Magazine*, 26(6):27–40, 2010.
- [8] IEEE. Ieee standard c57.104 - guide for the interpretation of gases generated in oil-immersed transformers, 2019.
- [9] M. Duval. A review of faults detectable by gas-in-oil analysis in transformers. *IEEE Electrical Insulation Magazine*, 18(3):8–17, 2002.
- [10] Jeff Golarz. Understanding dissolved gas analysis (dga) techniques and interpretations. IEEE.
- [11] N. Moodley and C. T. Gaunt. Developing a power transformer low energy degradation assessment triangle. IEEE.
- [12] N. Moodley and C. T. Gaunt. Low energy degradation triangle for power transformer health assessment. *IEEE Transactions on Dielectrics and Electrical Insulation*, 24(1):639–646, 2017.

- [13] N. Pattanadech and W. Wattakapaiboon. Application of duval pentagon compared with other dga interpretation techniques: Case studies for actual transformer inspections including experience from power plants in thailand. *IEEE*.
- [14] Jose Ignacio Aizpurua, Victoria M. Catterson, Brian G. Stewart, Stephen D. J. McArthur, Brandon Lambert, Bismark Ampofo, Gavin Pereira, and James G. Cross. Improving the accuracy of transformer dga diagnosis in the presence of conflicting evidence. *IEEE*.
- [15] Seifeddine Souahlia, Khmais Bacha, and Abdelkader Chaari. Svm-based decision for power transformers fault diagnosis using rogers and doernenburg ratios dga. *IEEE*.
- [16] A. G. C. Menezes, O. M. Almeida, and F. R. Barbosa. Use of decision tree algorithms to diagnose incipient faults in power transformers. *IEEE*.
- [17] Youcef Benmahamed, Omar Kherif, Madjid Tegar, Ahmed Boubakeur, and Sherif S. M. Ghoneim. Accuracy improvement of transformer faults diagnostic based on dga data using svm-ba classifier. *Energies*, 14(10):2970, 2021.
- [18] Alex R. Soto Enriquez, Shigeaki L. Lima, and Osvaldo R. Saavedra. K-nn and mean-shift algorithm applied in fault diagnosis in power transformers by dga. *IEEE*.
- [19] Jose Ignacio Aizpurua, Victoria M. Catterson, Brian G. Stewart, Stephen D. J. McArthur, Brandon Lambert, Bismark Ampofo, Gavin Pereira, and James G. Cross. Power transformer dissolved gas analysis through bayesian networks and hypothesis testing. *IEEE Transactions on Dielectrics and Electrical Insulation*, 25(2):494–506, 2018.
- [20] Jun Jiang, Ruyi Chen, Min Chen, Wenhao Wang, and Chaohai Zhang. Dynamic fault prediction of power transformers based on hidden markov model of dissolved gases analysis. *IEEE Transactions on Power Delivery*, 34(4):1393–1400, 2019.
- [21] M. Duval and A. Depabla. Interpretation of gas-in-oil analysis using new iec publication 60599 and iec tc 10 databases. *IEEE Electrical Insulation Magazine*, 17(2):31–41, 2001.
- [22] Adriana Rosa Garcez Castro. *Knowledge extraction from artificial neural networks: Application to transformer incipient fault diagnosis*. Thesis, 2004.
- [23] Md Umar Farooque, Shufali A. Wani, and Shakeb A. Khan. Artificial neural network (ann) based implementation of duval pentagon. *IEEE*.
- [24] Hazlee Azil Illias, Xin Rui Chai, Ab Halim Abu Bakar, and Hazlie Mokhlis. Transformer incipient fault prediction using combined artificial neural network and various particle swarm optimisation techniques. *PLOS ONE*, 10(6):e0129363, 2015.
- [25] Anyi Li, Xiaohui Yang, Huanyu Dong, Zihao Xie, and Chunsheng Yang. Machine learning-based sensor data modeling methods for power transformer phm. *Sensors*, 18(12):4430, 2018.
- [26] Shigeaki Leite de Lima. *Diagnóstico de falhas incipientes e tomada de decisão em transformadores de potência*. Thesis, 2013.
- [27] V. Miranda and A. R. G. Castro. Improving the iec table for transformer failure diagnosis with knowledge extraction from neural networks. *IEEE Transactions on Power Delivery*, 20(4):2509–2516, 2005.

- [28] D. Saravanan, Ahmad Hasan, Ajit Singh, Hannan Mansoor, and Rabindra Nath Shaw. Fault prediction of transformer using machine learning and dga. *IEEE*.
- [29] Xiaohui Yang, Wenkai Chen, Anyi Li, and Chunsheng Yang. A hybrid machine-learning method for oil-immersed power transformer fault diagnosis. *IEEE Transactions on Electrical and Electronic Engineering*, 15(4):501–507, 2020.
- [30] Hazlee Azil Illias, Xin Rui Chai, and Ab Halim Abu Bakar. Hybrid modified evolutionary particle swarm optimisation-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis. *Measurement*, 90:94–102, 2016.
- [31] Huang Yann-Chang and Sun Huo-Ching. Dissolved gas analysis of mineral oil for power transformer fault diagnosis using fuzzy logic. *IEEE Transactions on Dielectrics and Electrical Insulation*, 20(3):974–981, 2013.
- [32] Luis Dias, Miguel Ribeiro, Armando Leitão, Luis Guimarães, Leonel Carvalho, Manuel A. Matos, and Ricardo J. Bessa. An unsupervised approach for fault diagnosis of power transformers. *Quality and Reliability Engineering International*, 37(6):2834–2852, 2021.
- [33] Sunuwe Kim, Soo-Ho Jo, Wongon Kim, Jongmin Park, Jingyo Jeong, Yeongmin Han, Daeil Kim, and Byeng Dong Youn. A semi-supervised autoencoder with an auxiliary task (saat) for power transformer fault diagnosis using dissolved gas analysis. *IEEE Access*, 8:178295–178310, 2020.
- [34] Nad Moodley. Low energy degradation triangle for power transformer health assessment. 2018.
- [35] Abdul Wahid Mahrukh, Gao Xue Lian, and Sohail Sajeel Bin. Prediction of power transformer oil chromatography based on lstm and rf model. *IEEE*.
- [36] G. Nandagopan, K. Sabeena Beevi, A. S. Kunju Lekshmi, K. J. Vishnupriya, Deepa S. Kumar, S. Abhijith, and K. K. Rishika. Online prediction of dga results for intelligent condition monitoring of power transformers. *IEEE*.
- [37] Moustafa Alzantot, Supriyo Chakraborty, and Mani Srivastava. Sensegen: A deep learning architecture for synthetic sensor data generation. *IEEE*.
- [38] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [39] Cheolhwan Oh, Seungmin Han, and Jongpil Jeong. Time-series data augmentation based on interpolation. *Procedia Computer Science*, 175:64–71, 2020.
- [40] Shixiong Wang, Chongshou Li, and Andrew Lim. Why are the arima and sarima not sufficient. *arXiv preprint arXiv:1904.07632*, 2019.
- [41] Zoë M. Lewis, James A. Wild, and Matthew Allcock. Uk power station transformer dissolved gas analysis data (2010-2015), 2021.
- [42] Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, 2019.