Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Enabling Predictive Maintenance using Semi-supervised Learning with Reg-D Transformer Data ⋆

J du Toit *

* *Eskom Distribution, Western Cape Operating Unit, Eskom road, Brackenfell, South Africa (Tel: +27 (0)21 980 3915; e-mail: jacowp357@gmail.com).*

Abstract: Predictive maintenance has now become a possible avenue within the electricity distribution sector of Eskom. A recent roll-out of large-scale substation data acquisition projects have allowed this sector to use a predictive based maintenance scheduling plan instead of a previously used frequency based maintenance plan. This paper describes the design and implementation of a low-complexity anomaly detection algorithm, which is able to detect discrepancies, indicative of small electrical grid changes or substation equipment deterioration. The algorithm is based on projecting parametric multivariate Gaussian functions on to spatially distributed pre-selected substation data points. This method enables the utility to monitor critical variables, and their relationships, in an effort to foresee equipment or network distresses from high-value assets, particularly in the transmission and distribution sector. The results demonstrate an early positive detection of anomalous load behaviour from a live substation. The presented Semi-supervised learning methodology can form the underpinnings of an integrated approach, to aid with operational decision-making, and seems eminently suitable to reduce unscheduled asset downtime.

1. INTRODUCTION

Up until relatively recently the electricity distribution (ED) sector of Eskom, in the Western region, has relied on a frequency based maintenance scheduling plan on their high-value assets, including other substation equipment. After a recent roll-out of substation telemetry upgrades, Eskom is now in a position to capture and store electrical grid activity in the form of high resolution data. Some of these upgrades include: newer relays, data concentrators, dedicated servers, and fibre-optic communication infrastructure. This allows the utility to capture, transfer and store a wider range of variables at higher sampling intervals. Among other substation equipment, the utility is focused on utilising the new telemetry infrastructure, and data, to monitor their high-value assets — such as substation power transformers. The benefit of aggregating and utilising such data (information), to the point of adopting a prediction based maintenance plan, is paramount to the findings in this article.

In general, the most common causes of electrical equipment breakdown includes: mechanical equipment failure, environmental conditions, or work improperly performed. Predictive maintenance tools should be able to continuously, or periodically, monitor the condition of in-service equipment and detect equipment failure trending, well before complete failure occurs [1]. Having regular access to the current state of the equipment, in comparison to an operational state of the equipment, provides valuable information to determine when maintenance should be performed.

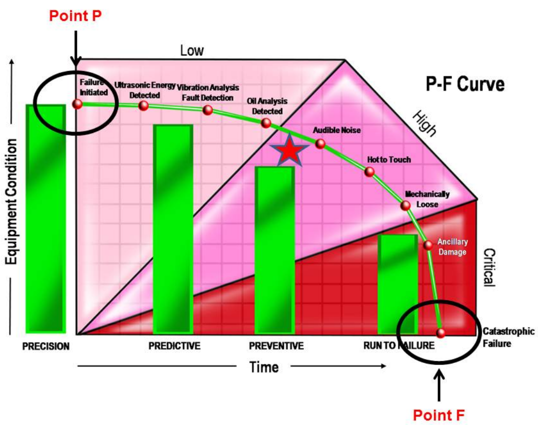

Fig. 1. The maintenance PF-curve.

A study in [2] highlighted that a reduction in maintenance costs can be expected between 25% to 30%, and the elimination of equipment breakdowns between 70% to 75%, when a proper predictive maintenance plan is implemented. A popular graph, illustrated in Fig. 1, indicates a general equipment failure curve and its related response category [1]. Here it is indicated how risk increases up until the point of critical consequences. The aim here is to detect anomalous equipment behaviour during the deterioration curve, where the risk is low and the necessary predictive- or preventive action can be applied before actual failure.

⋆ This work was supported by Eskom.

[1] Picture found at www.maintenancephoenix.com.

In this article, we describe the design and implementation of a low-complexity anomaly detection algorithm able to detect electrical grid discrepancies, in an effort to enable these predictive maintenance scheduling advantages for the ED sector.

The rest of this paper is organised as follows: in Section 2, we review the literature on predictive capable methodologies and their complexity. In Section 3, we present an exploratory data analysis, and the subsequent feature selection process. We formally explain the anomaly detection algorithm in Section 4, and show results using real substation data in Section 5. Finally, we state our conclusions, recommendations and future work in Section 6.

2. LITERATURE REVIEW

Most utilities, including Eskom, have relied on periodic on-site inspections, and manual evaluations of electrical equipment in the past. Furthermore, some equipment manuals still recommend technicians to write down readings from meters or gauges that is evaluated in comparison to readings obtained from normal operating conditions [3]. This kind of performance monitoring methods, and their subsequent maintenance scheduling plans, has disadvantages that include: on-site safety risks, human-error, traveling expenditure, etc.

In other studies, modelling techniques are used, such as heat circuit based thermal modelling, to perform on-site condition monitoring [9]. These implementations have their advantages, however require additional on-site components. Some methods comprise of utilising historic data and modelling of the operational behaviour of the equipment by using Neural networks or Fuzzy logic algorithms [7]. This requires non-linear modelling techniques, which introduces additional complexity (extra hidden layers to capture nonlinearities) in terms of large matrix multiplication [8]. Such complexities will introduce constraints on the utility's communication channel and the computational capabilities of the back-end services. In general, many efforts are made to incorporate expensive sensors and other on-site measuring devices to gain access to important variables (data) that is used in combination with complex algorithms.

The focus of this paper is the utilisation of historic data, as a non-intrusive approach, to identify and develop a statistical model of the operational workings of a substation by using low-complexity algorithms. This approach places fewer constraints on the available communication bandwidth, and the computational requirements of the back-end services. The following section describes the exploratory data analysis process, with focus on the data pre-processing requirements.

3. EXPLORATORY DATA ANALYSIS

The data was collected from a Reg-D relay in a live substation, over a period of one month, which has two functioning transformers operating in parallel with a master-slave configuration. Data from both transformers on the 11kV switch board side (load side) were inspected for this study. The data was normalised after all the missing (dead zones) values was removed. This is indicated in Fig. 2. The data was extracted from an *OSIsoft PI Server* using a *Microsoft Excel plug-in*. The measurement interval was set at 30 seconds. An interpolation [2] function in the *PI-Excel plug-in* was used to align all asymmetric data measurements.

All outliers were kept for supervised labelling purposes. Most of the correlations between the variables were investigated to determine their implications on the model. The pre-selected data consist of the following variables, captured from both power transformers [3]:

- Power factor,
- Reactive power,
- Apparent power,
- Frequency,
- Real power,
- Circulating current,
- Voltage,
- Power reserve,
- Current,
- Current phase degree.

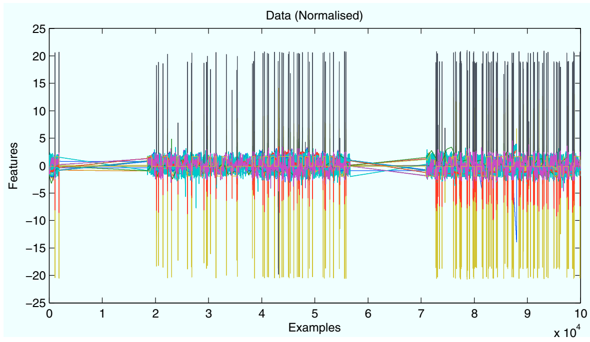

Fig. 2. All the features normalised.

A visual representation of the correlations between the different variables is shown in Fig. 3, where the diagonal represents a histogram of some of the different features. Following the diagonal, the variables are labeled as follows:

(a) Real power (T1),
(b) Circulating current (T1),
(c) Output voltage (T1),
(d) Output current (T1),
(e) Output current (T2),
(f) Circulating current (T2),
(g) Real power (T2),
(h) Output voltage (T2),
(i) Tap position (T2).

In Fig. 3 the normal operational regions of the data points are indicated. From this representation (variable relationships) the variable selections process, with regard to the modelling technique, can be validated.

The outliers where inspected, including their corresponding row entries, where each event was flagged (these events represent the anomalies). Some outliers are a result of *transition effects* between the two transformers just before, or after, a transformer's tap position changed. This can

[2] Although this is not a favoured approach, the data was misaligned and needed an alignment approximation technique. The influence of this will be explained in Section 6.

[3] T1 and T2 represents the two transformers.

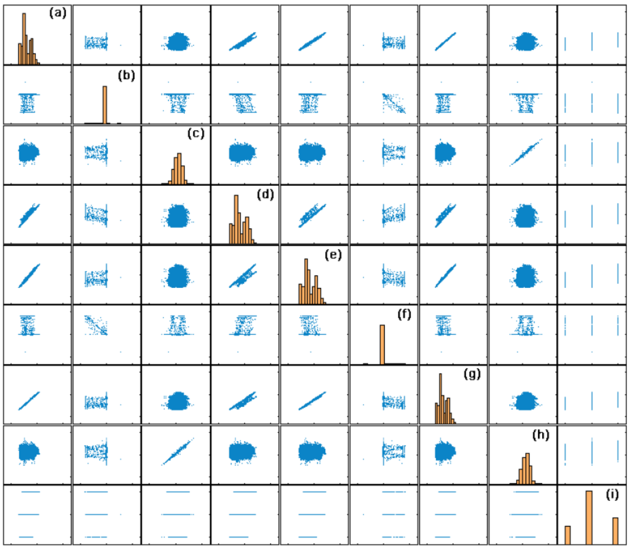

Fig. 3. Some feature correlations and histograms.

be seen in Fig. 4. The root cause of why the data points are not complying to the $y = -x$ linearity, is still being investigated. Moreover, it is suspected that it could be the result of linear estimations introduced by the interpolation technique.

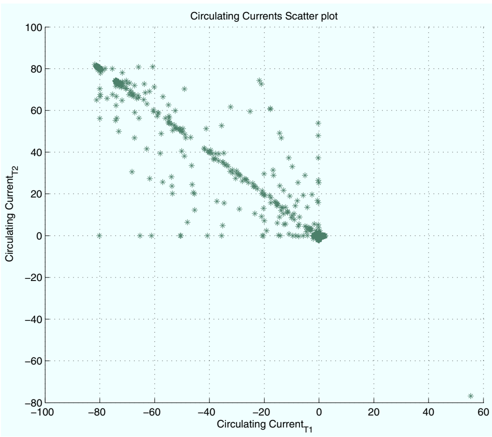

Fig. 4. Anomalous circulating currents from the two transformers.

During or just before a tap change the circulating currents, output currents, and current phases exhibit erratic responses. However, in limited cases this transition effect occurs without the presence of a tap change. The latter cases were captured and labelled as possible anomalies. One such case is demonstrated in Fig. 5. Both circulating current measurements (from both transformers), indicate unusually low and high values (negatively correlated). The output currents, from both transformers, also increase and decrease significantly during this event. These "abnormalities" occurred without the presence of a tap change.

For the rest of this paper, the data being utilised is normalised and subdivided into a training set x_{train}, cross-validation set x_{cv}, and test set x_{test}. The anomalous labels used for the supervised part in the cross-validation set and test set is represented by y_{cv} and y_{test} respectively. It is common practice to consider the anomaly detection algorithm described in this paper, due to the sparsity of

2.7633e+06	0.0360	1.1558e+04	148.5188	150.0300	-0.0935	2.8008e+06	1.1579e+04	7
2.6236e+06	0.0370	1.1649e+04	141.0236	145.8590	-0.0947	2.7105e+06	1.1670e+04	7
2.5940e+06	0.0380	1.1659e+04	139.0868	142.6894	-0.0959	2.6865e+06	1.1681e+04	7
2.6894e+06	0.0390	1.1737e+04	146.1268	148.1651	-0.0971	2780241	1.1769e+04	7
2.6121e+06	0.0400	1.1767e+04	140.6885	144.9678	-0.0982	2709769	1.1789e+04	7
2716749	-16.7579	1.1751e+04	169.3870	139.3160	31.4188	2.7009e+06	1.1813e+04	7
2.6885e+06	-0.0101	1.1654e+04	144.6871	145.4651	0.0541	2.7625e+06	1.1674e+04	7
2693881	-0.1285	11674	144.3691	148.0321	0.0540	2.7286e+06	1.1690e+04	7
2.7957e+06	-0.2469	1.1660e+04	148.9518	153.7467	0.0539	2.8694e+06	1.1679e+04	7
2.7185e+06	-0.3653	1.1653e+04	146.9308	147.7047	0.0538	2758098	1.1677e+04	7
2.6878e+06	-0.4837	1.1659e+04	144.1144	147.9759	0.0537	2.7423e+06	1.1678e+04	7
2.7098e+06	-0.6021	1.1663e+04	149.0956	151.8519	0.0537	2.8413e+06	1.1683e+04	7

Fig. 5. Example of an outlier.

anomalous training examples. If ample anomalous examples were obtainable, it is recommended to explore supervised classification methods instead [4, 5].

4. METHODOLOGY

In this section, a Gaussian distribution over the retrieved data points will be assumed, and a model is derived based on the mean and covariances of the dataset, limited to the degrees of freedom encapsulated in that sample set. Among other variables, the voltage and power factor measurements were found to be uncorrelated [4] and was used to demonstrate the effectiveness of the algorithm. Projections are made in two dimensional space for illustrative purposes, however, the model can be scaled to a multidimensional space if required.

4.1 Gaussian distribution

If $x \in \mathbb{R}$ and a Gaussian, with mean μ and variance σ^2, the distribution is a normal bell shaped curve centered at μ and σ. The standard deviation is indicative of the width of the bell curve. This can be expressed as the distribution: $x \sim N(\mu, \sigma^2)$. The equation for computing the mean and variance is shown in Equation 1 and Equation 2. The probability distribution calculation is shown in Equation 3.

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} \tag{1}$$

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)} - \mu)^2 \tag{2}$$

$$p(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\ \sigma} exp\left(-\frac{(x-\sigma)^2}{2\sigma^2}\right) \tag{3}$$

As the mean and variance parameters are altered, the shape of the distribution function changes, which describes a Gaussian property that can be used to project the distribution over the preferred data points.

4.2 Spatial anomaly detection

When projecting the Gaussian distribution on to the selected data the assumption being made is that the provided dataset, x_{train}, is non-anomalous and represents normal operation of the equipment, or state of the electrical network. We need to build a model to predict the probability of the power factor and voltage being appropriate, and if $p(x_{test}) < \epsilon$ then we need to flag a possible anomaly.

Assuming that each feature is distributed as per the Gaussian probability distribution then, $x_n \sim N(\mu_n, \sigma_n^2)$,

[4] Other studies have shown that correlated data can provide useful boundaries for outlier detection. We refer the reader to [6].

where n represents the number of features. The computed probability can be expressed as the *density estimation*, $p(x) = \prod_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2)$ [4]. The density estimation follows the random variable independence assumption. However, in practice it works quite well even if the features are not independent. The features, x_i chosen for this exercise, indicated anomalous examples, which were unusually high or unusually low values not corresponding to any obvious transition effects in the system.

The majority of the training set consisted of a non-anomalous subset of the data. Ideally this data should not contain any anomalous data points. Both the cross validation set $(x_{cv}^{(1)}, y_{cv}^{(1)})$, ..., $(x_{cv}^{(m_{cv})}, y_{cv}^{(m_{cv})})$ and the test set $(x_{test}^{(1)}, y_{test}^{(1)})$, ..., $(x_{test}^{(m_{test})}, y_{test}^{(m_{test})})$ contained some abnormal entries. The training and testing proceeds as follows:

- Use the training set to compute μ_n and σ_n^2.
- Fit the model $p(x)$ by computing the appropriate density estimation.
- On the cross validation set, apply different values of ϵ, and choose the value that maximises the $F1$ score.
- On the test set, calculate the prediction of $y = p(x)$, $(p(x) < \epsilon$ (anomaly), $p(x) > \epsilon$ (normal)) and determine the $F1$ score.

The $F1$ score, depicted in Fig. 4, is used as the test's accuracy metric ("cost function"). This method incorporates both the *precision* (number of correct results divided by the number of all returned results) and *recall* (the number of correct results divided by the number of results that should have been returned) of the test.

$$F1 = 2\left(\frac{precision \cdot recall}{precision + recall}\right) \tag{4}$$

Parameter ϵ can be altered by changing the anomaly detection "sensitivity" threshold, until a satisfactory result is obtained. The next section will illustrate a useful way to train a more variable and multi-dimensional Gaussian fitting function.

4.3 Multivariate Gaussian distribution

The algorithm described in the previous section largely assumes variables to be independent. This poses difficulties when variables are strongly correlated and might introduce inaccurate results (skewed decision boundary etc.). This section describes a technique, which can be used to alter the shape of the Gaussian, which effectively helps to fit the contour profiles of the Gaussian — to fit elliptically orientated data points. This is done by introducing additional parameters. By changing the values on the diagonal of the covariance matrix, in the multivariate Gaussian distribution the contours can be made either broader or narrower.

The additional parameters used to realise such orientations are: $M \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$. The overall position of the contour profile can be moved by changing M. The probability density is computed as follow:

$$p(x; M, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(x-M)^T\Sigma^{-1}(x-M)), \tag{5}$$

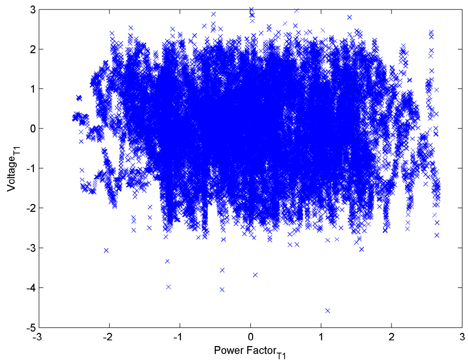

Fig. 6. Voltage and power factor correlation of T_1.

where $|\Sigma|$ is the determinant of Σ [4].

The multivariate Gaussian distribution is often used to describe, or at least approximate, any set of (possibly) correlated real-valued random variables each of which clusters around a mean value. Each correlation between the variables in Fig. 3 was inspected. The most preferred relation is indicated in Fig. 6. The voltage and power factor of T_1 (transformer one) had clear outliers, approximated a Gaussian distribution, and seemed to follow the independence assumption ($R^2 = 0.0039$ correlation coefficient [5]).

The multivariate Gaussian distribution can model certain arrangements of data points better. Therefore, by using a modified probability function we can predict anomalies more accurately in the case where features inhabit correlation. Any of such efforts will rely on the effective variability of the decision boundary.

5. RESULTS

A summary of the main results are indicated in Table 1. The best boundary was found using the cross-validation set with $\epsilon = 6.48\times10^{-4}$. The best $F1$ score for this set was 0.73. After training the detected outliers are flagged and plotted, as shown in Fig. 7.

Table 1. Summary of results.

ϵ	6.48×10^{-4}
$F1$	0.73
Anomalies	16

The flagged anomalies compared to the anomalous cross-validation set data points is also indicated. This illustrates the distance from the mean and position of the possible outliers in the data set compared to the detected set.

The trained model was implemented and tested on a completely new dataset, which was known to contain an occurring peculiarity. The new captured anomalies are illustrated in Fig. 8. This result reflects a planned *feeder shift-over* event, which was commissioned at the substation at the time the data was captured. A feeder shift-over is usually part of a contingency plan in the case of scheduled maintenance, and entails the reconfiguration

[5] A mutual information measure (from Shannon's theory on information) was also considered [10], however not regarded as necessary due to the noticeable co-variances between the variables.

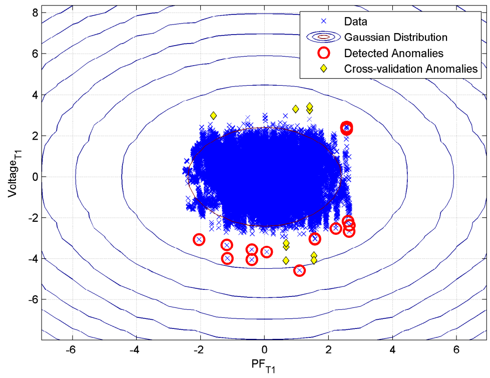

Fig. 7. Detected anomalies and trained examples.

of electrical power to a different load. The implemented anomaly detection scheme was capable of detecting this event, since the load characteristics predominantly affected the power factor variable's relationships in the data. This validates the sensitivity of the developed anomaly detection algorithm in terms of grid conditions. It is believed that such methods can add tremendous value to the future *Smart grid* movement, as it can be integrated into a large-scale grid anomaly detection systems.

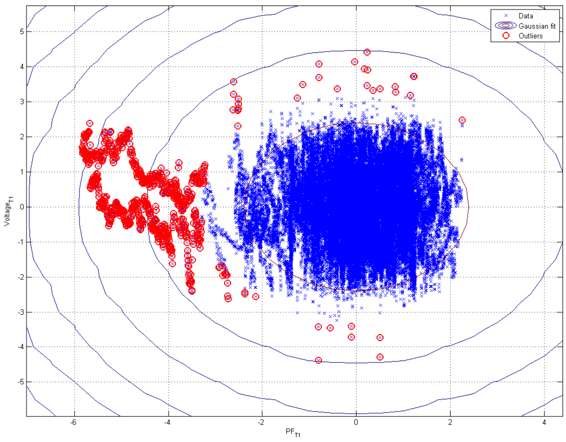

Fig. 8. Detected anomalies after transformer changes.

6. CONCLUSIONS AND RECOMMENDATIONS

This paper explored two statistical techniques that can be used to develop a learning algorithm capable of recognising unexpected transformer or grid operation. These techniques were coupled with recently obtained Reg-D substation data, which validates the utilisation thereof. The outliers used in this study were assumed to indicate anomalous events, since it could not be argued differently. The presented results reflect a system that is capable of detecting electrical grid infrastructure changes, network reconfiguration and possible transformer degradation.

The benefits of using the presented anomaly detection algorithm includes: emphasis on the operational regions of the transformer/grid data, continuous and autonomous monitoring of the present operating condition of the transformer/grid and non-intrusive low-complex detection of abnormal conditions. Furthermore, the methods explained are well suited to accommodate the current substation- and communication infrastructure.

Most features, captured from substations, will exhibit correlations, since the dynamics of the electrical grid is governed by the balancing of power supply-and-demand operations throughout generation, transmission, distribution, and reticulation. Therefore, the spatial multivariate solution seems eminently suitable for fitting this specific data. For large scale commissioning of the suggested algorithm it should be noted that each substation will require a uniquely trained model. This is due to the diversity of electrical network layouts, which will exhibit different grid characteristics. Inevitably, this will produce unique relational data points. In the event where electrical loads are reconfigured from one load to another, due to planned or unplanned maintenance, the substation measurements can sometimes vary significantly, as shown in Fig. 8. This affects the model accuracy over time, since the mean and variance of the data, reflecting the operational region, will change. This particular occurrence might hold between some variables only, and should be investigated in future work. Bayesian methods in [5] coupled with Gaussian mixture models can be incorporated, to actively adapt the model over time.

For future work it is recommended that more features (transformer oil quality, temperature, etc.) be added to explore possible accuracy improvements. The availability and expert recognition of transformer data, constituting unacceptable conditions, are anticipated to reinforce the model as explained in [3]. Other issues concerning the activation and availability of transformer diagnostic measurements can also improve the quality of predictive analytics on such equipment.

The biggest concern regarding this exercise was the misaligned substation data, due to dissimilar measurement intervals from different relays, or delayed data concentrator time-stamping. Although not many of these events were encountered, it is not ideal to train models with such data, with the risk of inaccurate results. In addition, the delta window technique, which is used to track and log larger deviations above or below a certain threshold in the measurements, is contributing to random sampling interval times. This poses a risk to the reliability and use of meta-data. Introducing such randomness is proportional to uncertainty within the model. A case in point is Fig 4.

REFERENCES

[1] S. Frank Waterer, *Enhance Power Equipment Reliability with Predictive Maintenance Technologies*, Schneider Electric USA Inc., (2012)

[2] G.P. Sullivan, J.D. Dean, D. R. Dixon, *Top Operations and Maintenance (O&M) Efficiency Opportunities at DoD/Army Sites*, Energy Engineering Analysis Program (EEAP), (2007)

[3] *User's Manual Operation and Maintenance for Power Transformers*, 1ZCL000002EG-EN rev. 1, (2007)

[4] Andrew Ng, *Machine Learning Class Notes*, Stanford Coursera.org, (2012)

[5] David Barber, *Bayesian Reasoning and Machine Learning*, Cambridge University Press, (2012)

[6] Parminder Chhabra, Clayton Scott, Eric D. Kolaczyk, and Mark Crovella, *Distributed Spatial Anomaly Detection*, (2008)

[7] K. Tomsovic and A. Amar, *On Refining Equipment Condition Monitoring using Fuzzy Sets and Artificial Neural Nets*, (1994)
[8] Rahmat Shoureshi, Tim Norick and Ryan Swartzendruber, *Intelligent Transformer Monitoring System Utilizing Neuro-Fuzzy Technique Approach* (2004)
[9] J.Q. Feng, P. Sun, W.H. Tang, D.P. Buse, Q.H. Wu, Z. Richardson, and J. Fitch, *Implementation of a power transformer temperature monitoring system* (2002)
[10] Cover, Thomas M., and Joy A. Thomas. *Entropy, relative entropy and mutual information*, Elements of Information Theory (1991): 12-49.