



Strathprints Institutional Repository

Catterson, Victoria M. and McArthur, Stephen D. J. and Moss, Graham (2010) Online conditional anomaly detection in multivariate data for transformer monitoring. IEEE Transactions on Power Delivery, 25 (4). 2556–2564. ISSN 0885-8977 , <http://dx.doi.org/10.1109/TPWRD.2010.2049754>

This version is available at <http://strathprints.strath.ac.uk/14806/>

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://strathprints.strath.ac.uk/>) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to Strathprints administrator: strathprints@strath.ac.uk

On-Line Conditional Anomaly Detection in Multivariate Data for Transformer Monitoring

V. M. Catterson, *Member, IEEE*, S. D. J. McArthur, *Senior Member, IEEE*, G. Moss

Abstract—Retrofitting condition monitoring systems to aging plant can be problematic, since the particular signature of normal behavior will vary from unit to unit. This paper describes a technique for anomaly detection within the context of the conditions experienced by an in-service transformer, such as loading, seasonal weather, and network configuration. The aim is to model the aged but normal behavior for a given transformer, while reducing the potential for anomalies to be erroneously detected. The paper describes how this technique has been applied to two transmission transformers in the UK. A case study of twelve months of data is given, with detailed analysis of anomalies detected during that time.

Index Terms—Power transformers, Monitoring, Fault diagnosis

I. INTRODUCTION

POWER transformers in the transmission network tend to be a focus for condition monitoring research and application, being expensive assets crucial to network operation. The anticipated return on condition monitoring comes from a reduction in maintenance costs and delayed replacement, but these benefits can only be realized by accurate, timely health information that truly reflects the current state of the plant.

To date, monitoring of transformers largely concerns single sensors, such as temperature [1], vibration [2], [3], or furan analysis [1]. In some cases, multiple sensors of the same type are used for corroboration and location of signals, including radio frequency (RF) sensors for partial discharge monitoring [4], [5]. Dissolved gas analysis (DGA) is a commonly employed technique for transformer health monitoring, which requires sampling of the transformer oil to determine constituent levels of key gases [6]. In all these cases, health analysis is based on the interpretation of a single type of data, using engineering expertise directly or through automated systems to find meaning in the data.

Multivariate data interpretation may increase the accuracy of health analysis, as it can offer corroboration across datasets. More diagnostic information can be gleaned from the co-occurrence or otherwise of events in, say, temperature and vibration data than in either set alone. However, extra data increases the complexity of the diagnostic task, as the interaction of different types of data may not be fully understood.

As an example application, two transmission transformers at a UK transmission substation are at the end of their design life, but it is hoped they can be kept in service with intensive condition monitoring. A combination of on-line sensors

including temperature, vibration, current, and dissolved gas sensors are installed on both units.

This multivariate monitoring scenario has certain features that mean diagnostic monitoring is not straight-forward. Sensors have been retrofit to both units, meaning that there is no historical body of data that can be used to help interpret new measurements. However, it is known that over the years of operation each unit has developed its own particular data signature that is different from a new unit. This includes a moderately high level of hydrogen in the DGA profile (which has remained static in recent years and is therefore not due to an ongoing problem). This means that both transformers have known, prior faults that have left residual low levels of fault indicators, while not requiring any immediate maintenance.

This situation is common. Condition monitoring is often applied towards the end of life when critical faults are expected; but by that stage years of operation have led to a particular ‘fingerprint’ representing normal behavior for one unit, which could be considered faulty behavior if it suddenly manifest in another. The key is to recognize changes in behavior, which can be more indicative of faults developing in older units than absolute values of measurements.

Anomaly detection techniques are a way of recognizing changes in plant behavior. Rather than simply matching patterns of expected faults, a model of behavior specific to each transformer under study can be trained to represent the normal operation of that particular asset. New measurements can be compared against the model to quantify how likely or anomalous they are. This allows for the natural differences between normal behavior in different transformers, and low level fault behavior can be trained into the model as normal for that unit. The combination of anomaly detection with fault classification can enhance the diagnosis of faults, by reserving fault classification for situations where an anomaly is detected, thus reducing the volume of data engineers must examine [7].

This paper describes the technique of Conditional Anomaly Detection, and details how it has been applied to the application transformers. This represents a new approach to condition monitoring, by considering normal and abnormal behavior in the context of plant operating conditions. The application is described in Section II, with an overview of anomaly detection techniques in Section III. Details of Conditional Anomaly Detection are given in Section IV, with some parameter requirements outlined in Section V. Section VI reports a case study of the technique applied to twelve months of transformer data, with the structure of the on-line anomaly detection software described in Section VII, and conclusions following.

Manuscript received August, 2009; revised February, 2010.

V. M. Catterson and S. D. J. McArthur are with the Institute for Energy and Environment, University of Strathclyde, Glasgow, UK (email: v.m.catterson@strath.ac.uk). G. Moss is with National Grid, UK.

TABLE I
SENSORS ON THE MORE-AGED T1 AND LESS-AGED T2

Subsystem	Sensors
T1 main tank	temperature (4), vibration (4), 3-phase current
T1 oil	top & bottom temperature and moisture, TRANSFIX
T1 pumps, fans	vibration (6) and load current
T1 tap changer	temperature, vibration, load current
T2 main tank	temperature (2), vibration (2), 3-phase current
T2 oil	top and bottom oil temperature
T2 pumps, fans	vibration (4)
T2 tap changer	temperature and vibration

II. APPLICATION DESCRIPTION

Two 275/132kV, 180MVA transformers have been in operation at a UK transmission substation for the span of their design life. One is showing more signs of aging than the other, and has been fitted with a Kelman TRANSFIX on-line dissolved gas analyzer¹ for over a year. A project was initiated to install a large number of sensors—primarily temperature, vibration, and current—at different locations on the main tanks and auxiliary systems, in order to investigate which techniques can assist with this particular health assessment need.

Table I shows the number and types of sensors on each transformer component. The majority of the sensors are located on the more-aged transformer, with a subset on the less-aged. Most of these sensors are externally fitted, with the oil temperature and moisture measured internally. This meant most sensors could be deployed without needing an outage.

In addition to the transformer sensors, the substation environment is monitored by a weather station, which measures parameters such as ambient temperature, rainfall duration, and wind speed. Network parameters such as tap position are recorded separately, giving a complete view of the conditions under which the transformers are operating.

The dissolved gas analyzer takes nine separate measurements every hour. Management and storage of this data has been outsourced to a data warehousing company, which records the gas data for all of the utility's transformers. Operational data is recorded through the SCADA system, and controlled by the utility. All the other parameters are measured once every five minutes, and collected and stored by a second data warehousing company.

Since the first data warehouse and the SCADA data is highly commercially sensitive, access to this data is tightly controlled. Data relating to this site can be extracted and downloaded manually by the utility engineers as needed. The data stored in the second data warehouse relates only to this site, and is less sensitive. Automated off-site access to this data was available through a web service interface (described further in Section VII), allowing download of data by all partners.

This installation was intended for investigating how monitoring of transformers could keep them in service for longer as they reach the end of life. Known problems with the design family meant that hydrogen levels were relatively high, and fault diagnosis using Rogers' Ratio [6] indicated a low temperature thermal fault. However, gas levels had remained

static for a few years, and it was believed there was low risk in continued operation.

Any changes to the normal parameter values should be considered more significant than absolute levels, since it may signify the inception of a new problem. Techniques such as Rogers' Ratio would be of little use in recognizing new fault types, since the diagnosis would be skewed by the pre-existing moderate levels of hydrogen. Instead of diagnosis, anomaly detection was required. Further, the anomaly detection system should be on-line to ensure anomalous behavior is detected at the earliest opportunity. On-line analysis is key in evaluating the risk of keeping these aged assets in operation.

III. ANOMALY DETECTION

Anomaly detection, sometimes called novelty detection, is a general term for distinguishing outliers and unusual values from normal or expected data. Chandola et al [8] define three types of anomaly: point anomalies, contextual anomalies, and collective anomalies. The first is simply outlier detection: is a parameter at a point in the measurement space where it has not been seen before? This could be the case if data generally falls into one of two clusters, but a point is measured that falls outside of either cluster.

Contextual anomalies take into account the context of a measurement before deciding on its anomalousness. For example, the top oil temperature in a transformer may be significantly higher on one day of a month. This would appear to be a point anomaly, but if the weather conditions were particularly hot on that same day, the context suggests that the oil temperature is high because of external conditions, and is not an anomaly when viewed in context. Similarly, a top oil temperature which is normal for the height of summer may be considered anomalous when measured during winter.

The third type is a collective anomaly, where a collection of linked data points are out of place. The linkage may be in time (consecutive measurements in time series data) or space (neighboring measurements within a location). An example of time series data would be top oil measurements taken throughout one day; this could be expected to increase through the morning, reach a peak in the afternoon, then decrease through the evening and night, due to weather and load. Oil temperature would be more static throughout a day of an outage, but the ambient temperature would still cause the same pattern of behavior. Temperatures that remained fixed or even decreased during the day would be an anomalous collection of data, even if the specific temperature at any point in the day is normal, and may indicate a sensor failure.

Condition monitoring systems within the power domain tend to be point anomaly detection systems, allowing engineers to set thresholds and limits on expected values. This includes commercial systems such as the Kelman TRANSFIX dissolved gas analyzer. Work on diesel generators showed how thresholds could be extracted from a small set of training data when no historical data exists [9].

However, the most interesting of approaches for plant condition monitoring is contextual anomaly detection. There are many factors that could provide context for parameter

¹This unit measures oil levels of hydrogen, methane, ethane, ethylene, acetylene, oxygen, carbon dioxide, carbon monoxide, and water.

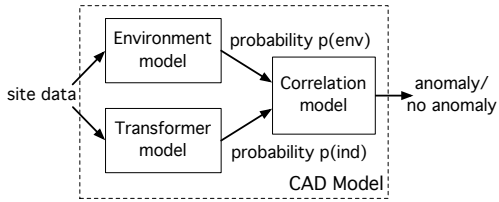


Fig. 1. CAD compares the likelihood of transformer indicator data against environmental conditions to detect anomalies.

changes, including load, aging of components, harmonics on the network, and seasonal and day-to-day weather variations. Considering the context of monitoring data should provide a means of limiting the number of false alarms generated by a monitoring system, while also being sensitive to values which are within normal thresholds, but unusual given other factors.

Considering the case study application outlined above and the need for on-line operation, the contextual data that can be accessed by an automated system is the weather and load data recorded in the second data warehouse. The ambient substation environment and the current load provide the context in which the transformers are operating, and so the specific technique for anomaly detection chosen for this application has to be capable of incorporating multiple weather and transformer data parameters. The selected technique is Conditional Anomaly Detection.

IV. CONDITIONAL ANOMALY DETECTION

Conditional Anomaly Detection (CAD) is a contextual anomaly detection technique based on the likelihood of *indicator* parameter values occurring at the same time as *environmental* parameter values [10]. Applied to the transformer monitoring application, indicator parameters are those measured from the plant under study, indicative of plant health, while environmental parameters are those measuring the weather in the substation and transformer load. The likelihood of indicator values being seen under given environmental conditions can be calculated from a CAD model, which comprises separate models for the statistical behavior of the plant and environment, and probabilistic links between the two (shown in Figure 1).

Since the statistical models of plant and environment behavior are separate, they could be used individually for point anomaly detection. The plant model can be used to calculate the probability of new transformer measurements; a low probability indicates anomalous values that are not likely to occur. However, the probabilistic link between the plant and the environment models provides the contextual reasoning ability about whether or not measurements are truly anomalous.

As an example, consider an environment model of ambient temperature and transformer load current, with an indicator model of top oil temperature. Viewed simplistically, the environment could be summarized as high load/hot day; high load/cold day; low load/hot day; or low load/cold day. Each of these four modes of operation will have associated likely ranges of top oil temperature. In order to decide if a given value of top oil temperature is anomalous or not, the current mode of operation is assessed based on the environment

model, and the likelihood of the top oil temperature given these operating conditions is calculated. With this model, anomalies could be flagged in cases such as:

- 1) The environment model gives a high probability of high load/hot day operation, and top oil temperature is higher than normal for this case; or,
- 2) The environment model gives a high probability of low load/cold day operation, and top oil temperature is higher than normal for this case (but may be within normal range for high load/hot day operation).

However, an anomaly would not be flagged if, say, a heat-wave caused the ambient temperature to be so unusually high that there is a low probability of being in any of the four modes of operation. Intuitively, in such a case we may expect top oil temperature to be unusually high as well, and so low probability indicator values will not be defined as anomalous.

Considering this case in terms of the probabilistic models, if the training data is very unlike the current environmental conditions, then the coverage of the models does not extend to the current situation, and we can make no firm judgement about how likely or not the indicator values are. It is better to label it as neither normal nor abnormal, than to flag something as anomalous when the coverage of the model does not provide information to support this. This approach will reduce the number of false anomalies reported to engineers.

The modeling technique suggested by Song et al in [10] for the environment and plant models is Gaussian mixture modeling. This is described below.

A. Gaussian Mixture Modeling

Some datasets can be represented by a set of Gaussian (normal) distributions, mixed together in different proportions. The set of Gaussian components and their mixing proportions constitute a probabilistic model of the dataset. The parameters of the model are $\Theta = \langle \theta_0, \theta_1, \dots, \theta_k \rangle$ for k components. Each θ comprises the Gaussian parameters $\langle \mu, \Sigma, \pi \rangle$ for mean, variance, and mix proportion.

An example follows. Figure 2 shows a histogram of wind direction from the first two weeks of September 2008, measured from the on-site environment. The x-axis denotes the values of maximum direction in degrees, while the y-axis is the number of times the corresponding value was seen, scaled to a probability of occurrence. The distributions overlaid on the histogram are the Gaussian components of a mixture model trained on the data. It can be seen that peaks in the data correspond with means of distributions, or peaks in the sum of overlapping components.

This mixture of Gaussian components is a model of the training data, and so may be expected to be generally representative of this parameter. New measurements of wind direction can be assessed for likeliness by comparing them with the probability density function of the above model. From the model graph, it can be seen that a measurement of 350° has probability 0.005 and is fairly likely to occur, while a measurement of 100° has probability 0.0013 (the sum of three components) and is less likely to occur. This can be calculated from the probability density function (PDF):

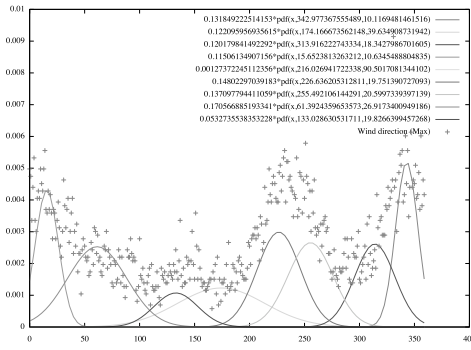


Fig. 2. Histogram of wind direction as probabilities with components of a GMM overlaid. The fifth component is not visible on this scale, due to its large σ_5 and small π_5 .

$$f_{GMM}(x) = \sum_{n=1}^k \pi_n P(x|\theta_n) \quad (1)$$

$$P(x|\theta_n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_n)^2}{2\sigma_n^2}\right) \quad (2)$$

where σ is the standard deviation, and variance is σ^2 .

This example describes the single parameter case, where Gaussian means are a single value. Gaussian mixture modeling may also be applied to the multi-dimensional case, where the means become vectors and variance translates to a covariance matrix. The probability density function in the multi-dimensional case is:

$$f_{GMM}(\mathbf{x}) = \sum_{n=1}^k \pi_n P(\mathbf{x}|\theta_n) \quad (3)$$

$$P(\mathbf{x}|\theta_n) = \frac{1}{((2\pi)^d \det|\Sigma|)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \quad (4)$$

where d is the dimension of the data.

The remaining question is how to learn the parameters of Θ that characterize the model. A common technique is to use the Expectation Maximization algorithm, detailed below.

B. Expectation Maximization

This is an iterative technique that converges towards a locally optimal set of values for unknown model parameters, that maximize the likelihood of the overall model [11]. It can be applied to a variety of learning problems, and is used three times in Conditional Anomaly Detection: once each for training the environment and indicator Gaussian mixture models, and once for learning the probabilistic mapping between the models. The fit of a model to the training data can be assessed by calculating the *log likelihood* of the model, which increases with a more accurate fit.

There are two stages to the algorithm: an Expectation (E) step, where values of the model parameters are fixed and the resulting likelihood of the model is calculated; and a

Maximization (M) step, where this likelihood is used to update the model parameter values. Each step is run alternately, and the new model likelihood compared at the end of every E-step with its previous value. The algorithm terminates when the change in likelihood is smaller than a given tolerance, or after a set number of iterations. The output model is the one with the highest log likelihood generated by the training process.

For the Gaussian mixture model case, the E-step of the algorithm involves computing γ_{ij} , the probability that datapoint i was generated by (comes from distribution) component j , for all i up to N , and j up to k . The equation for this is:

$$\gamma_{ij} = \frac{\det|\Sigma_j|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_j)^T \Sigma_j^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_j)\right)}{\sum_{l=1}^k \det|\Sigma_l|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_l)^T \Sigma_l^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_l)\right)} \quad (5)$$

From this set of probabilities, the log likelihood of the model can be computed:

$$\log(\mathcal{L}(\Theta|\mathbf{X})) = \sum_{i=1}^N \log \sum_{j=1}^k \pi_j \gamma_{ij} \quad (6)$$

Finally, the M-step of the algorithm computes updated values for all $\langle \boldsymbol{\mu}_j, \Sigma_j, \pi_j \rangle$:

$$\pi'_j = \frac{1}{N} \sum_{i=1}^N \gamma_{ij} \quad (7)$$

$$\boldsymbol{\mu}'_j = \frac{\sum_{i=1}^N \mathbf{x}_i \gamma_{ij}}{\sum_{i=1}^N \gamma_{ij}} \quad (8)$$

$$\Sigma'_j = \frac{\sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}'_j)(\mathbf{x}_i - \boldsymbol{\mu}'_j)^T \gamma_{ij}}{\sum_{i=1}^N \gamma_{ij}} \quad (9)$$

C. Learning the CAD Model

The process of learning the Conditional Anomaly Detection model has three steps [10]. The first is to learn a Gaussian mixture model U of the environmental parameters. The second is to learn a Gaussian mixture model V of the indicator parameters, in this case the transformer parameters. The third is to learn the mapping function between U and V , which is formulated as an Expectation Maximization problem.

The E-step is to calculate all b_{kij} , the probability that environment datapoint \mathbf{x}_k was generated by component U_i while indicator datapoint \mathbf{y}_k was generated by component V_j , given that U_i was the generating component in the environment. The equation for this is:

$$b_{kij} = \frac{f_{GMM}(\mathbf{x}_k|U_i)\pi_i f_{GMM}(\mathbf{y}_k|V_j)p(V_j|U_i)}{\sum_{t=1}^{n_U} \sum_{h=1}^{n_V} f_{GMM}(\mathbf{x}_k|U_t)\pi_t f_{GMM}(\mathbf{y}_k|V_h)p(V_h|U_t)} \quad (10)$$

where n_U and n_V are the numbers of components in U and V respectively.

The M-step maximizes the $p(V_j|U_i)$ terms:

$$p(V_j|U_i)' = \frac{\sum_{k=1}^N b_{kij}}{\sum_{k=1}^N \sum_{h=1}^{n_V} b_{kih}} \quad (11)$$

As with training Gaussian mixture models, the log likelihood is used to assess the fit of the model. This is calculated by:

$$\log(\mathcal{L}(\Theta|\mathbf{X}, \mathbf{Y})) = \sum_{k=1}^N \log f_{CAD}(\mathbf{y}_k|\Theta, \mathbf{x}_k) \quad (12)$$

where the f_{CAD} function is:

$$f_{CAD}(\mathbf{y}_k|\Theta, \mathbf{x}_k) = \frac{\sum_{i=1}^{n_U} f_{GMM}(\mathbf{x}_k|U_i)\pi_i}{\sum_{t=1}^{n_U} f_{GMM}(\mathbf{x}_k|U_t)\pi_t} \cdot \sum_{j=1}^{n_V} f_{GMM}(\mathbf{y}_k|V_j)p(V_j|U_i) \quad (13)$$

D. Detecting Anomalies

After the Gaussian mixture models and mapping probabilities have been learned, they constitute the CAD model which can be used for anomaly detection. This section describes the method for anomaly detection proposed in the original Song et al paper, explains why it is not directly applicable to the case of transformer monitoring, and proposes a more appropriate technique for this application.

Detection of anomalies involves calculating f_{CAD} for new measurements (eqn. 13). This function can be summarized in English as the probability of indicator parameters having their measured values, given the likelihood of seeing the current environment parameter values. f_{CAD} values decrease in the presence of an anomalous indicator and a normal environment.

Song et al expected Conditional Anomaly Detection to be applied to datasets which contain predominantly “normal” behavior, that is, the data had been gathered during situations where no major faults or deviations were observed, and therefore it can be assumed to be representative of normality. However, the reason for applying unsupervised learning to the problem is that the contents of the dataset are unknown, and it may therefore contain low level, non-critical anomalies such as aging trends and slow fault inception. In order to allow outliers in the training data to be recognized as anomalies when encountered in the future, Song et al suggest picking a threshold value of f_{CAD} based on the percentage of points in the training dataset that are expected to be anomalous.

For example, if it is thought that 1% of training measurements are anomalous, the f_{CAD} values of each training datapoint should be sorted, and the lowest 1% examined. The highest f_{CAD} of this subset becomes the threshold for detecting anomalies: any future measurement with an f_{CAD} value less than this threshold is deemed anomalous.

This threshold requires careful selection: if it is set too high, the technique will be too sensitive to anomalies and engineers may be overwhelmed with spurious alarms; but if set too low, anomalies may be missed. In the case of the transformer monitoring project, labeling 1% of data as anomalous would result in 3 measurements per day being raised as anomalous, seriously overestimating the true likelihood of an anomaly that would be of interest to engineers. Visual analysis of site data from the month of September revealed nothing anomalous,

indicating that the threshold for anomalous data in the training dataset may be 0%. Therefore, another technique for setting the anomaly threshold was required.

One way of approximating an appropriate threshold is to select a reasonable value based on some test data, and refine the estimate with more operational experience. Results of this method applied to a case study are described in Section VI. Additionally, through experimentation with site data it was found that additional information can be extracted from the CAD model which aids with the detection of “true” anomalies.

The most anomalous situations are indicated when $f_{CAD} = 0$. It can be seen from eqn. 13 that this could be caused when the probability density function of either U or V renders a zero probability of the measurement, that is, when either $\forall i f_{GMM}(\mathbf{x}_k|U_i) = 0$ or $\forall j f_{GMM}(\mathbf{y}_k|V_j) = 0$. This would mean that either the environmental conditions or the indicator conditions differ so greatly from the training dataset that there is no chance of one of the learned components covering those values.

However, the meaning of these two cases is very different. In the first case, the environment is so unusual compared with the training examples that its probability is zero. This situation ought to be considered non-anomalous, since an anomaly in the environment means it cannot be predicted what the indicator values should normally be, as the situation has not been encountered in the training data. If this is the case, it means a result of $f_{CAD} = 0$ should not automatically be labelled as an anomaly.

The second case, where $\forall j f_{GMM}(\mathbf{y}_k|V_j) = 0$, is a true anomaly. This result means that the indicator parameters are so far outside expected values that the probability of them being generated by the model is zero. As long as the environment values have a non-zero probability of being generated, this is a truly anomalous situation.

As a result, a novel means of using the CAD model to detect anomalies is proposed:

- 1) Select an approximate threshold of f_{CAD} value for recognizing anomalies, τ , either by Song et al’s percentage method, or through experimentation;
- 2) Calculate f_{CAD} for a new measurement;
- 3) If $f_{CAD} > \tau$, it is not anomalous;
- 4) If $0 < f_{CAD} \leq \tau$, it is anomalous;
- 5) If $f_{CAD} = 0$, inspect the independent probabilities of the environment and indicators:

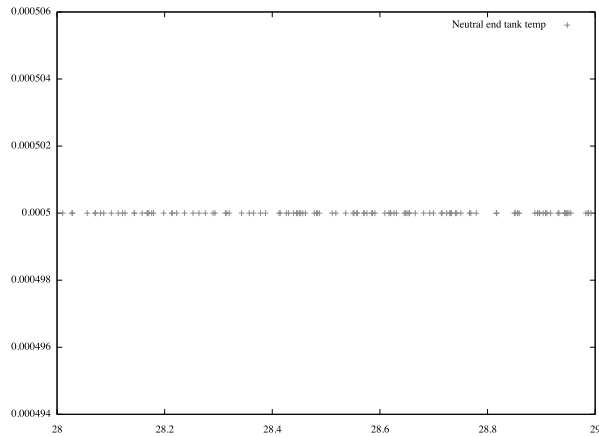
$$p(\text{env}) = \sum_{i=1}^{n_U} f_{GMM}(\mathbf{x}_k|U_i)\pi_i \quad (14)$$

$$p(\text{ind}) = \sum_{j=1}^{n_V} f_{GMM}(\mathbf{y}_k|V_j)\pi_j \quad (15)$$

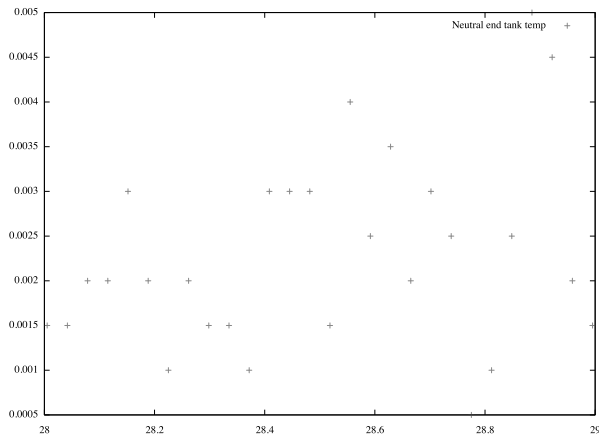
- 6) If $p(\text{env}) = 0$, it is not anomalous;
- 7) If $p(\text{ind}) = 0$, it is anomalous.

V. MODEL PARAMETER SELECTION

Careful selection of the input parameters to CAD is required to gain best results. Since the environmental and indicator parameters are to be modeled as Gaussian mixtures, an accurate



(a) Before binning, showing equal probabilities



(b) After binning, showing varying probabilities

Fig. 3. Section of a histogram before and after binning.

model requires that they truly look like a mixture of Gaussian distributions. Further, if there is a linear dependency between two model inputs, i.e. they are highly correlated, the Gaussian mixture model collapses and cannot be trained. This is because linear dependency results in a singular covariance matrix, yet training requires the inverted covariance matrix in eqn. 5.

These requirements immediately discount the use of some of the site parameters in the CAD model. For example, the values of maximum, minimum, and average wind speed are highly linearly dependent, meaning a model of environmental parameters could not contain all three. Similarly, the three-phase load currents are highly correlated, so only one phase could be used in a model.

Many of the transformer parameters that could form the indicator model look highly non-Gaussian. Histograms of parameters such as the temperature of the tank at the neutral end (“Neutral end tank temp”) show many measurement values with single occurrences, which makes them unsuitable for Gaussian mixture modeling. However, these occurrences often have approximately the same values that differ only in the third or fourth significant figure (see Figure 3(a)).

Rather than discounting these parameters from modeling, the data was preprocessed to make it more suited to Gaussian

components. This was done by binning each measurement: sorting datapoints into 400 equal-width bins per parameter. The result is that the measurements differing by low significant figures were counted as the same value for the purposes of histogram calculation, which in turn suits Gaussian components better (Figure 3(b)). For consistency, this process was applied to all parameters selected for modeling.

VI. CASE STUDY RESULTS

This section presents a case study of Conditional Anomaly Detection applied to the site under study. Nine months of operation of the more-aged transformer is considered, with detail of the model training, test results, and analysis of the anomalies observed.

An engineer’s visual inspection of site data from September 2008 concluded that the transformers were operating normally during that time. No significant deviations in transformer behavior were seen during October, although the ambient temperature and other weather parameters were rather different. As a result, data from September and October 2008 were used as training data representing normal behavior and conditions.

The environment model was trained on ambient temperature, the load current of the yellow phase, solar radiation, wind speed, and wind direction. Gaussian Mixture Models ranging in size from five to 17 components were trained, with the log likelihood and visual inspection guiding selection of the best fit model. The 16-component model was chosen.

For the transformer behavior indicator model, parameters relating to the transformer oil were chosen. This subsystem is presented here for detailed study; however, it is only one of a set of indicator models trained for different transformer subsystems. For the application as a whole the parameters were split into subsystems in order to support the engineer during on-line operation, and this is discussed further in the following section. This section describes one particular model in detail.

Bearing in mind the requirement of low linear dependency between parameters in the model, only one of top and bottom oil temperature, and top and bottom oil moisture could be selected, so the model parameters were chosen to be top oil temperature, bottom oil moisture, and hydrogen levels. Gaussian Mixture Models ranging from five to 17 components were trained on these parameters, and the 14-component model was found to be the best fit.

After selecting the best environment and indicator models, the CAD model was trained. The same training set of two months of data was used, to find the correlations between environment and indicator conditions.

The test set of data comprises the twelve months of operation following the training data, 1st November 2008 to 31st October 2009. Based on the results of this testing, the threshold for anomaly detection was set at 1×10^{-20} . The full test set results are summarized in Table II.

The results from the month of March are most informative. Figure 4 shows the values of f_{CAD} , $p(env)$, and $p(ind)$ for the measurements taken every five minutes throughout March 2009.

The most obvious outlier in this graph is measurement 8378, corresponding to 19:25 on March 30th, where the

TABLE II
ANOMALIES DETECTED BY CAD MODEL IN THE 12 MONTH TEST SET

Month	f_{CAD} value		Cause of $f_{CAD} = 0$		Anomalies
	$0 < f < 10^{-20}$	$f = 0$	$p(env)$	$p(ind)$	
Nov	0	2	2	0	0
Dec	3	11	0	11	14
Jan	0	0	0	0	0
Feb	1	3	3	0	1
Mar	4	39	38	1	5
Apr	0	4	4	0	0
May	0	30	30	0	0
Jun	0	0	0	0	0
Jul	0	1	0	1	1
Aug	0	0	0	0	0
Sep	0	0	0	0	0
Oct	0	0	0	0	0

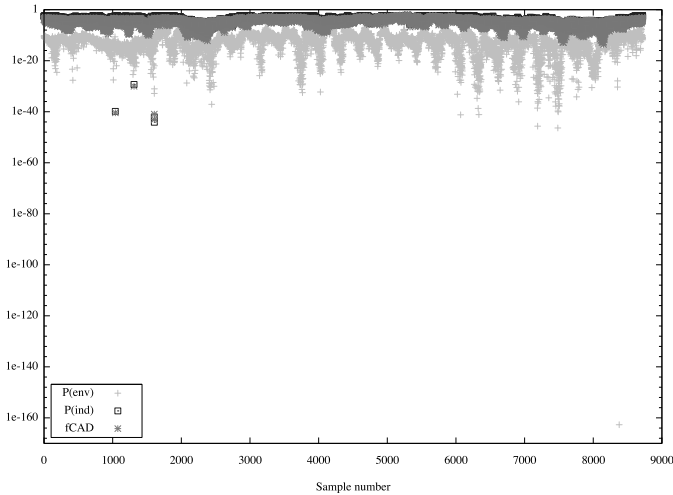


Fig. 4. CAD model output from March 2009, showing $p(ind)$ has a higher likelihood than $p(env)$ for most of the month, with f_{CAD} in between.

environmental conditions have a very low probability. Ambient temperature, solar radiation, and load current are all within normal ranges, but wind speed and direction are both abnormally high at 1012.3° and 21m/s respectively. This suggests temporary sensor failure as the cause. Since the environment is very abnormal while the indicators are relatively normal, no anomaly is detected.

Four outliers in f_{CAD} can be seen, corresponding to unlikely transformer conditions while the environment is relatively normal (detailed in Figure 5). These are diagnosed as anomalous, since f_{CAD} falls below the anomaly threshold. The exact times of these points are 14:35 on 4th, 13:15 on 5th, and 13:50 and 13:55 on 6th March. In all four cases, the anomaly is due to very low bottom oil moisture levels between 0.1 and 10.1ppm, while normal levels are 55–95ppm. Top oil temperature and hydrogen levels are within normal ranges.

Considering that these moisture measurements occur between normal readings, it is likely that these are due to sensor problems. For example, at five minute intervals between 13:00 and 13:30 on 5th March, moisture levels are reported as 66.8, 64.8, 62.8, 10.1, 61.8, 62.8, and 64.8, which strongly suggests the value at 13:15 is erroneous. Furthermore, an anomalously low level of moisture is less of a concern than anomalously high levels, especially since the top oil temperature is be-

tween 15 and 20°C at these anomalous points. As a result, these datapoints do not represent a serious problem with the transformer.

Not shown in Figure 4 due to the scale of the graph is the datapoint where the transformer indicator conditions have a zero probability. This occurs at 14:30 on 4th March. All three transformer parameters have highly anomalous values: top oil temperature is recorded as -95°C , moisture is -30.0ppm , and hydrogen is -125.0ppm . As with the previous cases, this is a problem with the sensors or data logger rather than a true problem with the transformer.

Finally, the 38 cases of the environmental conditions having zero probability are also sensor faults. In particular, the ambient temperature is measured as various unlikely high temperatures such as 255 and 19140°C . No anomaly is diagnosed at these points since the environmental conditions are so unlikely.

This study shows that the CAD model can identify anomalous values in transformer measurements, while ignoring any datapoints with anomalous environmental conditions. During the period under study, 21 anomalies were found. In all cases, these were found to be due to sensor or logging problems, and not indicative of true transformer behavior. However, in some of these cases the sensors returned anomalous but plausible values, such as oil moisture of 10.1ppm, showing that CAD accurately models the behavior of the given transformer, and detects anomalies specific to this unit, which a simple thresholding technique may not find.

VII. ON-LINE OPERATION

On-line anomaly detection was desirable for the site under study, in order to gain early warning of potential deviations in transformer behavior. This section describes the models used for on-line data analysis, how the on-line system operates, and what information is presented to engineers about anomalous behavior.

Since Conditional Anomaly Detection is based on statistical correlations within datasets, rather than knowledge of the links between transformer components, it simply returns a value for f_{CAD} without any explanation of why measurements seem anomalous. Within an on-line monitoring system, the reason behind any anomalous behavior is of prime importance, as any maintenance or operational decisions taken by the engineer must be based on some rationale. With this in mind, the on-line operation of Conditional Anomaly Detection aimed to give as much information to the engineer as possible about potential reasons for anomalies.

One way of enhancing information is to train and deploy a suite of CAD models looking at different sub-components within the transformers under study. The previous section detailed the case study of a CAD model based on the oil parameters of the more-aged transformer. By training multiple, targeted indicator models, such as the oil model, a model of tank temperatures, and a model of tap changer-related parameters, the system provides inherent correlation of anomalies across subsystems. The engineer can be informed that, for example, anomalies occurred at the same time in the oil and tank models, or that an anomaly was isolated to the tank

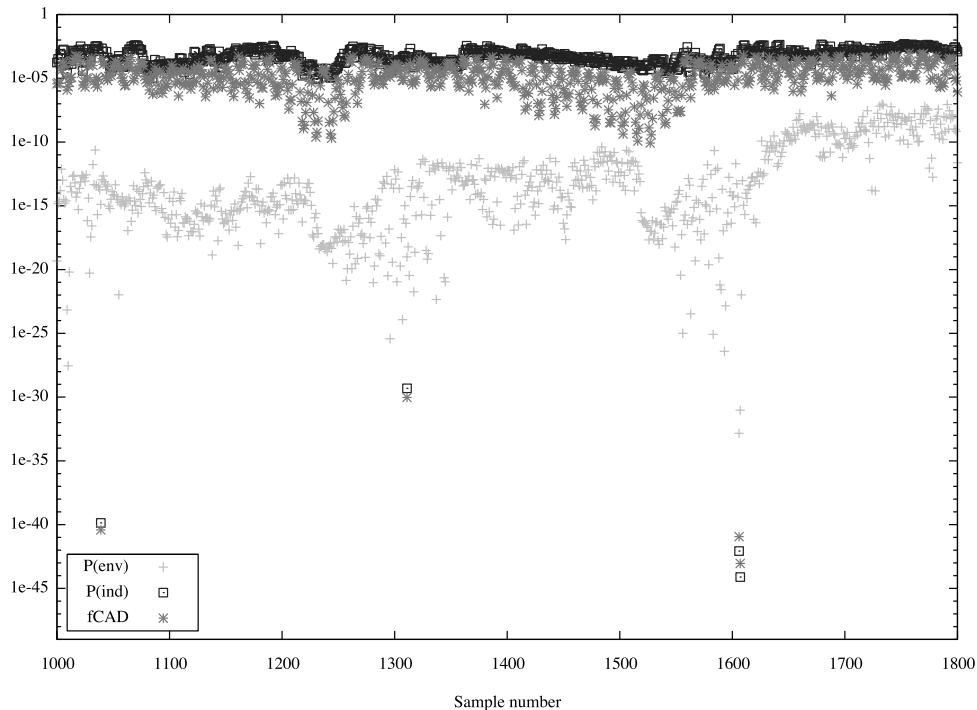


Fig. 5. CAD model output shows significant anomalies successfully detected in samples 1039, 1311, 1606, and 1607 from March 2009.

temperatures. This is more informative than anomaly detection in a single, large CAD model covering all parameters.

One environment model can be reused across all the indicator models of one transformer. The weather and loading conditions always provide the environment in which the transformer is operating, and so training a new CAD model for the suite involves training the indicator model, and learning the correlations between it and the existing environment model. However, loading may differ between units on the same site, so the environment model should be tailored to the specific transformer.

After the required suite of CAD models have been trained, they are run in parallel for on-line anomaly detection. Multi-agent systems technology [12] provides the software framework for this on-line system, allowing data processing modules to be flexibly added and upgraded as the system is running. The system architecture is shown in Figure 6.

The first stage of on-line operation is to collect new measurements as they are recorded. As mentioned in Section II, data is archived to a data warehouse for long term storage, which can be accessed on-line through a web service interface [13]. Every five minutes, the Data Provider Agent connects to the web service and requests the most recent measurements. The Data Director Agent filters the site data to find groups of measurements needed by the CAD Model Agents (that is, the appropriate weather, load, and transformer parameters that are inputs to the CAD models). Finally, any identified anomalies that result from the CAD models are recorded by the CAD Report Agent, which keeps a record of the number and frequency of anomalies in the last seven days, and the last 30 days. The output of all analysis is logged with an agent for archiving.

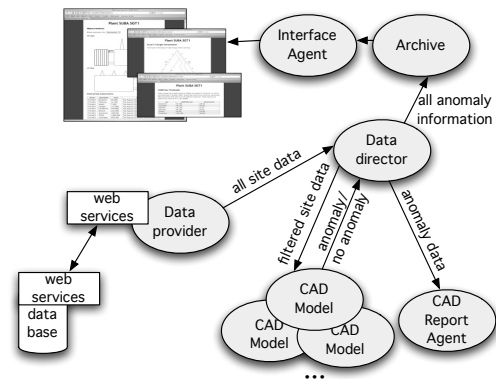


Fig. 6. On-line system architecture, showing data processing tasks performed by separate agents.

The CAD Report Agent performs simple comparisons of anomalies across multiple CAD models, and also across multiple transformers in the same substation. The knowledge of which transformer subsystems recorded anomalies simultaneously can help the engineer assess the cause and risk; and a comparison of the two sister units can be similarly informative. If anomalies are detected on both transformers at the same time, the cause may be network-related or substation-wide, whereas problems confined to one transformer suggest a change in that unit.

In addition to detecting and recording anomalies, the information must be presented to engineers. This is achieved through the interface agent, which populates a web-based engineer's interface with site information. At the first level, engineers are presented with a table of the current CAD models, and whether any anomalies have been detected recently

by one or more of them. By selecting one of the models, the engineer can see histograms of anomalous behavior during the last seven and 30 days, giving an indication of the frequency of anomalous behavior. Finally, the most detailed level of information allows investigation of the CAD model output, showing each parameter of the environment and indicator models in graphical form, and the position and probability of a given measurement. These three levels of information allow the engineer to find correlations between anomalies in sub-components, investigate whether anomalous behavior is increasing in frequency, and visualize exactly which parameters are behaving anomalously.

VIII. CONCLUSIONS

This paper describes the technique of Conditional Anomaly Detection, and how it can be applied to transformer monitoring. A particular application is given, where two in-service transformers are being monitored by a range of temperature, vibration, and other sensors. The results of applying CAD to a twelve month set of site data showed that 21 anomalies were detected. Analysis of these anomalies found that they were all attributable to sensor faults giving anomalous measurements. The utility's analysis concurs with this assessment, as no transformer faults were found during this time. This indicates that the CAD technique can be applied to on-line monitoring of transformers, and can detect unusual behavior. Finally, the system for on-line anomaly detection is described, with detail of the system components needed to provide a solution that supports engineers in decision making.

While previous work presented a method of automatically learning thresholds for point anomaly detection [9], this paper shows how to provide contextual anomaly detection, with the potential to reduce "false alarm" anomalies due to unusual operating and environmental conditions. Further work will investigate two avenues. The suite of CAD models can be increased to cover new types of sensor as they are added to plant, and in particular, a model of the transformer's insulation would be beneficial for end-of-life monitoring. Secondly, the potential of collective anomaly detection techniques for transformer monitoring will be studied. Such techniques learn patterns of plant behavior over time, and can detect when a change in parameter value is anomalous, compared to contextual anomalies which occur when a value is anomalous given other parameters. A type of collective anomaly detection has recently been applied to a single parameter: partial discharge data [7], and a multivariate version of this technique teamed with CAD could enhance the information about anomalies presented to engineers.

ACKNOWLEDGMENT

This work is funded through the EPSRC Supergen V, UK Energy Infrastructure (AMPerES) grant in collaboration with UK electricity network operators working under Ofgem's Innovation Funding Incentive scheme—full details on <http://www.supergen-amperes.org>.

REFERENCES

- [1] M. K. Pradhan, "Assessment of the Status of Insulation during Thermal Stress Accelerated Experiments on Transformer Prototypes," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 13, no. 1, pp. 227–237, Feb. 2006.
- [2] B. Garcia, J. C. Burgos, and A. M. Alonso, "Transformer tank vibration modeling as a method of detecting winding deformations—Part I: Theoretical Foundation," *IEEE Trans. Power Del.*, vol. 21, no. 1, pp. 157–163, Jan. 2006.
- [3] E. Rivas, J. C. Burgos, and J. C. Garcia-Prada, "Condition Assessment of Power OLTC by Vibration Analysis using Wavelet Transform," *IEEE Trans. Power Del.*, vol. 24, no. 2, pp. 687–694, Apr. 2009.
- [4] M. D. Judd, L. Yang, and I. B. B. Hunter, "Partial Discharge Monitoring for Power Transformers using UHF Sensors Part I: Sensors and Signal Interpretation," *IEEE Electr. Insul. Mag.*, vol. 21, no. 2, pp. 5–14, Mar./Apr. 2005.
- [5] I. E. Portugues, P. J. Moore, I. A. Glover, C. Johnstone, R. H. McKosky, M. B. Goff, and L. van der Zel, "RF-Based Partial Discharge Early Warning System for Air-Insulated Substations," *IEEE Trans. Power Del.*, vol. 24, no. 1, pp. 20–29, Jan. 2009.
- [6] T. K. Saha, "Review of Modern Diagnostic Techniques for Assessing Insulation Condition in Aged Transformers," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 10, no. 5, pp. 903–917, Oct. 2003.
- [7] A. J. Brown, V. M. Catterson, M. Fox, D. Long, and S. D. J. McArthur, "Learning Models of Plant Behavior for Anomaly Detection and Condition Monitoring," *Engineering Intelligent Systems*, vol. 15, Jun. 2007.
- [8] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, Jul. 2009.
- [9] S. D. J. McArthur, C. D. Booth, J. R. McDonald, and I. T. McFadyen, "An agent-based anomaly detection architecture for condition monitoring," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1675–1682, Nov. 2005.
- [10] X. Song, M. Wu, C. Jermaine, and S. Ranka, "Conditional Anomaly Detection," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 5, pp. 631–645, May 2007.
- [11] Z. Ghahramani and M. I. Jordan, "Learning from incomplete data," MIT Center for Biological and Computational Learning, Technical Report 108, available from <http://hdl.handle.net/1721.1/7202>.
- [12] S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziaargyriou, F. Ponci, and T. Funabashi, "Multi-Agent Systems for Power Engineering Applications—Part I: Concepts, Approaches and Technical Challenges," *IEEE Trans. Power Systems*, vol. 22, no. 4, pp. 1743–1752, Nov. 2007.
- [13] Z. Jun, "Web services provide the power to integrate," *IEEE Power Energy Mag.*, vol. 1, no. 6, pp. 40–49, Nov./Dec. 2003.



Victoria M. Catterson (M'03) received the B.Eng. (Hons) and Ph.D. degrees from the University of Strathclyde, Glasgow, U.K., in 2003 and 2007 respectively. She is a Research Fellow within the Institute for Energy and Environment at the University of Strathclyde. Her research interests include agent architectures, condition monitoring, and diagnostics.



Stephen D. J. McArthur (M'93–SM'07) received the B.Eng. (Hons.) and Ph.D. degrees from the University of Strathclyde, Glasgow, U.K., in 1992 and 1996 respectively. He is a Professor in the Institute for Energy and Environment, Director of the British Energy Advanced Diagnostics Centre, and Co-director of the Strathclyde Centre for Applied Intelligent Systems. He currently chairs the IEEE PES Multi-Agent System Working Group. His research interests include intelligent system applications in power, condition monitoring, and agent systems.

Graham Moss is a Condition and Capability Engineer with National Grid, Warwick, UK, in Asset Stewardship.