

# Data Driven Transformer Thermal Model for Condition Monitoring

A. Doolgindachbaporn, G. Callender, *Member, IEEE*, P. L. Lewin, *Fellow, IEEE*, E. Simonson, and G. Wilson, *Member, IEEE*

**Abstract**—Condition monitoring of power transformers, which are key components of electrical power systems, is essential to identify incipient faults and avoid catastrophic failures. In this paper machine learning algorithms, i.e. nonlinear autoregressive neural networks and support vector machines, are proposed to model the transformer thermal behavior for the purpose of monitoring. The thermal models are developed based on the historical measurements from nine transformers comprised of two 180-MVA units, four 240-MVA units and three 1000-MVA units. The data consist of load profile, tap position, winding indicator temperature (WTI) measurement, ambient temperature, wind speed and solar radiation. The results are validated against field measurements, and it is clearly demonstrated that the alternative algorithms surpass the IEEE Annex G thermal model. An incipient thermal fault identification algorithm is then proposed and successfully used to identify an issue using measurements taken in the field. This algorithm could be used to alert the operator and plan intervention accordingly.

**Index Terms**— Condition monitoring, power transformer, transformer thermal model, winding temperature indicator.

## I. INTRODUCTION

The cost of replacing power transformers is high compared with other substation assets and has justified research focused on maximizing life expectancy while maintaining reliability and stability, in part through condition monitoring [1], [2]. Transformer hot-spot temperature is one of the most crucial parameters for monitoring as it captures the cooling performance of the transformer and is linked to the degradation rate of the paper insulation [3].

To achieve on-line thermal condition monitoring, an accurate dynamic transformer thermal model needs to be developed to predict the transformer temperature under different loading conditions. Divergence between the measurement and prediction could be then interpreted as changes in thermal behavior.

There are international transformer thermal models widely used for oil-immersed transformer as loading guides, i.e. IEEE

Std C57.91 (2011) Annex G [4] and IEEE 60076-7 [5]. A limitation is that the thermal models require specific thermal parameters for each transformer. There are a set of recommended values for those transformers that do not have their own parameters, but they may not be consistent with actual thermal performance. Refinements to these models have been proposed in the literature. Thermal influence of changes in winding losses and oil viscosity with temperature was discussed in [6]. The thermal impact of solar radiation and wind speed has been considered in [7], [8]. Alternatively, the transformer temperature distribution is modelled using thermal-hydraulic network model and computational fluid dynamics [9]. As the data associated with the transformer temperature are increasingly available, it is feasible to develop a transformer thermal model using machine learning techniques regardless of the thermal parameters.

Earlier work using machine learning for dynamic thermal modeling of power transformers showed promising results [10]. A power transformer thermal model that was developed using artificial neural network (ANN) was also more accurate compared with the traditional approach [11]. In addition, the ANN was used to forecast the hot-spot temperature for an instrument transformer [12]. Performance of the prediction made by binary regression tree, generalized linear model, gaussian process regression and support vector machine over three units lightly loaded was investigated [13]. However, long-term performance of the algorithms, especially ANN, over the course of the year, and across different units, was not considered due to lack of availability of field measurements. Furthermore, the thermal influence of environmental factors, i.e. wind speed and solar radiation were not considered.

The aim of this paper is to develop a tool for estimating hot-spot temperature by means of machine learning to aid thermal condition monitoring, for example in identifying: restricted oil and air flow, or faulty cooling pumps and fans [14]. The inputs comprise load, tap position, ambient temperature, solar radiation and wind speed. The model performance of different learning algorithms over testing datasets spanning 7 years for 9 transformers are provided. An incipient thermal fault identification algorithm is proposed based on the thermal model and used to identify a fault based on measurements taken in the field. The contribution of this work is the assessment of long-term performance of machine learning algorithms, especially ANN, across nine transformers. In addition, the impact of two weather factors, wind speed and solar radiation, is considered. Finally, a thermal anomaly identification algorithm has been

This work was supported by National Grid Electricity Transmission. A. Doolgindachbaporn, G. Callender and P. L. Lewin are with the Tony Davies High Voltage Laboratory, University of Southampton, Southampton, SO17 1BJ, United Kingdom (e-mail: ad2a15@soton.ac.uk; gmc1f17@soton.ac.uk and pll@ecs.soton.ac.uk) E. Simonson is with Southampton Dielectric Consultants Ltd, Southampton, SO17 1XS, United Kingdom (e-mail edward\_sdc@canburytech.net) G. Wilson is with National Grid Electricity Transmission, Warwick, CV34 6DA, United Kingdom (e-mail: gordon.wilson@nationalgrid.com)

developed and used to detect a period of raised transformer temperatures using field measurements.

## II. TIME SERIES FORECASTING

Transformer hot-spot temperature is dependent on many factors, most importantly: load profile, tap position, ambient temperature, wind speed and solar radiation. These factors all vary with time and are treated as time series hereafter. Traditionally, transformer thermal models calculate the hot-spot temperature through two governing equations: a top-oil thermal model and a hot-spot temperature rise above top-oil thermal model [4]. The top-oil thermal model has a relatively large time constant, often in the order of hours, and is dependent on environmental conditions and load. The hot-spot temperature has a much lower time constant, typically in the order of minutes, and the hot-spot temperature rise above top-oil is only dependent on load. In this work, WTI measurements, which are indicative of hot-spot temperature, are used with load and environmental data (ambient temperature, solar radiation and wind speed) directly to train a machine learning algorithm which removes the need for an independent top-oil thermal model. The machine learning algorithms introduced in this section are trained using these factors to estimate hot-spot temperature. The data are initially normalized to prevent factors with larger numeric values dominating.

### A. Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs) have been previously used in applications associated with electrical power systems, [10] and [15]. The idea of artificial neural networks (ANNs) is inspired by human learning processes [16]. In ANN architecture, there are three main layers: an input layer, one or more hidden layers and an output layer as illustrated in Fig. 1 [16]. The input layer is the first layer of the neural networks containing original inputs before passing to neurons. The hidden layers contain a series of artificial neurons. The artificial neurons are activation functions that map the sum of the weighted inputs plus bias into new values. The output layer then sums all the final information from the last hidden layer with bias. The activation function used in this work is a hyperbolic tangent.

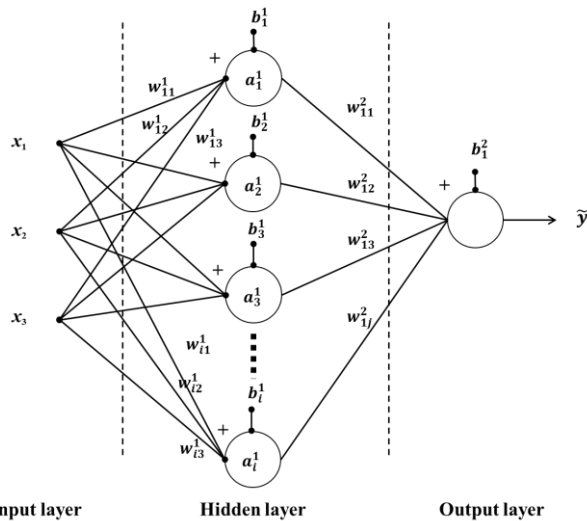


Fig. 1 Artificial neural network architecture. The figure is adapted from [16].  $x$  are inputs,  $w$  are weights,  $b$  are biases,  $a$  are the output from the activation function and  $\hat{y}$  are predicted outputs.

The following is the processes of a feedforward ANN. First each input is multiplied by its own weight and then they are added together with bias as a single input ( $n$ ) as follows:

$$n^{k+1}(i) = \begin{cases} \sum_{j=1}^{N^k} w_{ij}^{k+1} a_j^k + b_i^{k+1} & \text{if } k > 0 \\ \sum_{j=1}^{N^k} w_{ij}^{k+1} x_j + b_i^{k+1} & \text{if } k = 0 \end{cases} \quad (1)$$

where  $x$  is an input vector,  $a$  is the output of the activation function,  $w$  is weight,  $b$  is bias,  $k$  is the index of the layer itself,  $i$  is the index of the neurons,  $j$  is the index of the input features to the activation function and  $N$  is the total number of the input features for each layer. The outputs of each layer are passed to the next layer as inputs until the output layer. The sum of squared errors between the prediction and measurement ( $V$ ) is defined as follows:

$$V = \sum_{q=1}^Q (y_q - \hat{y}_q)^2 = \sum_{q=1}^Q E_q^2 \quad (2)$$

where  $y_q$  are the measured outputs,  $\hat{y}_q$  are the predicted outputs,  $E_q^2$  are the squared error between the measured and predicted outputs,  $q$  is the index of the points within dataset and  $Q$  is the total number of datapoints.

The weights and bias are adjusted to minimize the sum of square errors between the measurement and prediction of the training dataset. The Marquardt-Levenberg method is selected to solve the least square curve fitting because it requires a short time for training a model that has a few hundred weights or lower while still producing a reasonably accurate model [17].

A nonlinear autoregressive neural network with multiple inputs is used. The non-linearity between inputs and outputs comes from the activation function that is the hyperbolic tangent function, autoregressive means that the output at the previous time step is used as the input for prediction at the present time step.

The numbers of hidden layers and neurons are usually increased with a complexity of the system. Increasing the number of the hidden layers and neurons too much will result in the model overfitting [18]. The key idea is to optimize the number of hidden layers and neurons to produce a model with the highest accuracy when compared with the testing dataset.

### B. Support Vector Machine for Regression (SVM)

The concept of support vector machine (SVM) is to create a linear function between inputs and outputs that minimizes errors below a given margin ( $\epsilon$ ) [19]. However, most systems have a non-linear relationship between inputs and outputs within the original space. To create linear relationships between predictors and response, the data are first mapped into a higher dimensional space using a kernel function. The kernel function is a technique for measuring similarity between two sets of vectors [20]. The kernel method is selected for SVM as it has a fast computational time. The algorithm is referred to as ‘‘support vector machine’’ because some input vectors in the training dataset are used as reference, or support, vectors in the kernel

function to calculate the output prediction. The architecture of SVM regression is shown in Fig. 2. The predicted output based on SVM for regression is given as follows:

$$\hat{y} = \sum_{m=1}^S \lambda_m k(\mathbf{x}, \tilde{\mathbf{x}}_m) + b \quad (3)$$

where  $\tilde{\mathbf{x}}_m$  are the support vectors which are subsets of the input data in the training dataset,  $m$  is the index of the support vectors,  $S$  is the total number of the support vectors that are determined automatically,  $\lambda_m$  are weights for each support vector and  $k(\mathbf{x}, \tilde{\mathbf{x}}_m)$  is a kernel function. Similarities between an input vector ( $\mathbf{x}$ ) and support vectors  $\tilde{\mathbf{x}}_m$  are computed using a kernel function and then weighted and added with bias to form the predicted output.

There are many popular kernel functions such as linear, polynomial, Gaussian and hyperbolic tangent functions [19]. The Gaussian function is used in this work as it has been shown to outperform the others in various applications [21]. Explicitly the kernel function is:

$$k(\mathbf{x}, \tilde{\mathbf{x}}_m) = \exp\left(\frac{-\|\mathbf{x} - \tilde{\mathbf{x}}_m\|^2}{\gamma}\right) \quad (4)$$

where  $\gamma$  is the kernel scale and  $\|\ \ \|$  is Euclidean distance. The output of the kernel function is equal to one if an input vector is in the same position as the support vectors or decreases toward zero if an input vector is far away from the support vectors.

The algorithm determines a set of support vectors from the training dataset, weight ( $\lambda(m)$ ) and bias ( $b$ ) by minimizing the following nonlinear least square problem:

$$\begin{aligned} \text{minimise} \quad & \frac{V}{2} + C \sum_{q=1}^Q E_{\varepsilon}(y_q - \hat{y}_q) \\ \text{subject to } E_{\varepsilon}(z) = & \begin{cases} |z| - \varepsilon & \text{if } |z| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (5)$$

where  $C$  is the penalty factor and  $\varepsilon$  is the error margin.

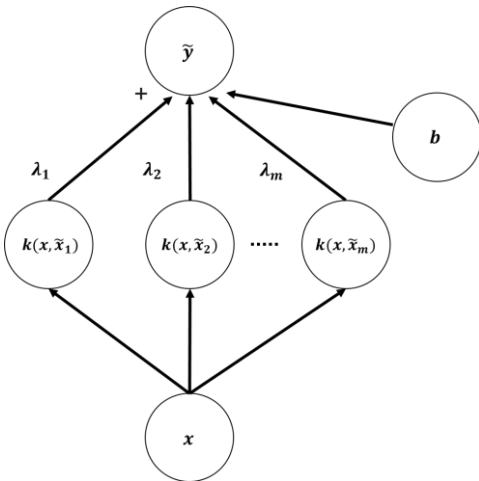


Fig. 2 Architecture of SVM regression. The figure is adapted from [21].  $x$  are inputs,  $\lambda$  are weights,  $b$  are biases,  $k(\mathbf{x}, \tilde{\mathbf{x}}_m)$  is a kernel function and  $\hat{y}$  are predicted outputs.

The kernel scale ( $\gamma$ ), penalty factor and the error margin ( $\varepsilon$ ) are all parameters that need to be specified by the user to optimize the accuracy of the model. Higher values of the penalty factor can cause overfitting problems because it emphasizes the model error of the training dataset and vice versa. Larger values of the kernel scale reduce ability to distinguish between 2 sets of data and this could therefore reduce the number of support vectors required to form a model and complexity of the model [22]. Larger values of the error margin reduce the contribution of the model errors to the least squares problem, and a lower number of support vectors are likely to be required to form a model which could lead to underfitting of the data [22].

### C. Hyperparameter Optimization

Hyperparameter optimization is a method used to tune learning parameters to optimize the error over a cross-validation dataset rather than the training dataset [23]. A grid search technique is used to determine the optimized number of the neurons and hidden layers because the number of the choices is small. The Bayesian optimization method is used to solve the hyperparameter optimization of the SVM algorithm because of the large number of the hyperparameter choices [24]. The Bayesian method is usually used to determine global optimization for black-box functions and functions that require high computational time to be evaluated. It is based on two processes; a Gaussian process model which is a probabilistic regression model formed by a combination of stochastic variables that have a joint normal distribution and an acquisition function which is a probability function used for selecting a next optimal combination [25]. Initially, several sets of the hyperparameters are evaluated to model the cost function using the Gaussian process model. Following this, the next candidate set of the hyperparameters is selected according to the acquisition function.

Training datasets, which in this work are half of the total dataset, are divided equally into five sections in this work. A model is trained with four sections which means 80% of the training dataset is used for training, or equivalently 40% of the total dataset. The remaining section of the training dataset is referred to as a cross-validation dataset, which is used for the purpose of the hyperparameter optimization. Each section of the training dataset is considered in turn. This means that there will be 5 trained models in total that use the same set of the hyperparameters. A set of optimized learning parameters, that provides a minimum of the sum square errors against the cross-validation data for each trained model, is finally selected. Table I provides restrictions of each hyperparameter of the two algorithms.

### D. Regression Order

The regression order represents how many hours the inputs extended are away from the output at the present time step; all data are in an hourly format. Instead of only using the input data from the previous time stamp to predict the hot-spot temperature at the present time step, the number of the inputs is extended to also include the input data beyond the previous time step. For example, regression order is equal to five means that the inputs for predicting hot-spot temperature at the present time step include the data from the last five hours. It should be

noted that a measured input of previous hot-spot temperature is not used to predict hot-spot temperature with regards to the testing data, i.e. the hot-spot temperature output from the algorithms is used as an input for the next time step.

### III. OPERATIONAL MEASUREMENTS

There are nine transformers considered in this analysis. They are comprised of two 180-MVA units (A and B), four 240-MVA (C, D, E and F) units and three 1000-MVA (G, H, and I) units. Their specifications are shown in the Appendix in Table VI. The inputs for the alternative thermal models are comprised of load profiles, tap position, winding temperature indicator (WTI) measurement, ambient temperature, wind speed and solar radiation, and the output is WTI temperature, see Fig. 3 for a snapshot of the measurements. It should be noted that the WTIs are calibrated to mimic the hot-spot temperature. Over the periods considered the cooling state of the transformers under consideration did not change, i.e. cooling systems were either continuously oil-forced-air-forced (OFAF) or oil-natural-air-natural (ONAN). That means the models in this work are trained to estimate the transformer temperature in either ONAN or OFAF mode only. When a sufficient amount of data is available for both ONAN and OFAF mode, two separate models should be created for each mode. As the data considered in this work do not contain periods of cooling transition between ONAN and OFAF modes and vice versa, the analysis was not included. It should be noted that the temperature changes rapidly in such periods and a higher sampling rate would be required to capture those changes.

The weather information is obtained from the closest UK Meteorological Office weather station [26]. The periods of the investigation are 2013 and 2019 inclusive. The data are in hourly format. Missing and unreliable values are deleted from the datasets. Half of the dataset is used to train the models and the other half is used to evaluate the models' accuracy. The cross-validation dataset discussed previously is a subset of the training dataset.

### IV. TRANSFORMER THERMAL MODEL PERFORMANCE

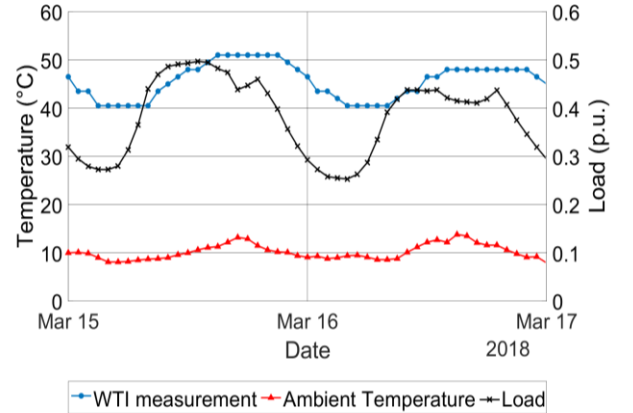
The goal is to predict the WTI temperature profile for a given loading condition as accurately as possible. The ANN and SVM algorithms are used to learn the thermal behavior of the WTI temperature from the historical measurements. The optimized hyperparameters are provided in Table II. The variation in the hyperparameters of the SVM model is large even in the same transformer family. The accuracy is not compromised with the variation. It is unlikely that there is a universal combination of the hyperparameters that is suitable for all transformers, and the hyperparameters should therefore be optimized for each transformer. It was found that the optimal hyperparameter combinations can be used for other transformers within the same family and still yield reasonable accuracy, typically errors increase by less than 0.1°C. Furthermore, the trained models of the ANN and SVM algorithms for Transformer G, H and I, which are in the same family, are used to predict the temperature for the other transformer in the same family, see Section VI for a further investigation. As the optimization algorithm usually provide local optimums, it is likely there might be different optimal combinations that provide similar performance.

The computational time of the models, on an ordinary desktop computer, was approximately 10 minutes for the ANN algorithm and approximately an hour for the SVM algorithm. It should be noted that the computational time is also dependent on size of training dataset, which is approximately 30,000 data-points, equivalent to 3.5 years, in this work.

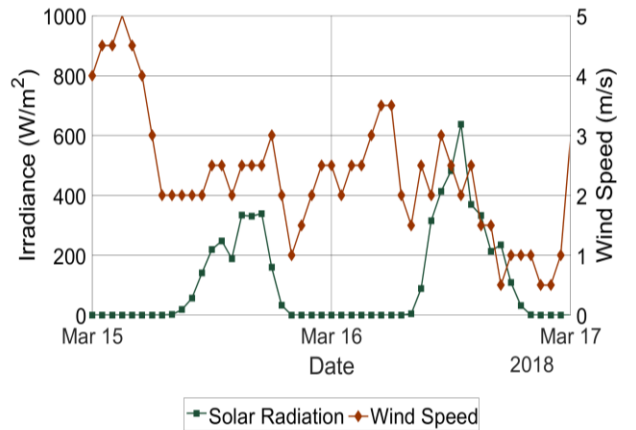
TABLE I  
HYPERPARAMETER RANGES

Algorithm	Parameter	Lower	Upper
ANN	The number of neurons	1	10
	The number of hidden layers	1	3
SVM	Epsilon ( $\epsilon$ )	$10^{-3} \sigma(y_q)$	$10^3 \sigma(y_q)$
	Kernel scale ( $\gamma$ )	$10^{-3}$	$10^3$
	Box constraint ( $C$ )	$10^{-3}$	$10^3$

$\sigma(y_q)$  = Standard deviation of  $y_q$



(a) Load, WTI and ambient temperature measurements



(b) Solar radiation and wind speed

Fig. 3 Snapshots of load, WTI, ambient temperature, solar radiation and wind speed measurements.

The accuracy of the estimations made by the ANN and SVM algorithms is evaluated using the testing datasets. It should be noted that a measured input of previous WTI temperature is not provided to predict WTI temperature with regards to the testing data, i.e. the WTI temperature output from the algorithms is used as an input for the next time step. The accuracy of the prediction made by the IEEE Annex G algorithm is also investigated for the sake of comparison. It should be noted that the top-oil temperature rise at rated load and oil time constant, which are key parameters in the IEEE Annex G model, are tuned to the operational measurement to increase the accuracy of the prediction. Adjusting the other parameters does not improve the accuracy significantly. A snapshot of the measurement and predictions made by the algorithms is provided in Fig. 4.

The root mean squared errors (RMSEs) and mean absolute errors (MAE) between the measurement and predictions made by the IEEE, ANN and SVM algorithms over the testing datasets are provided in Table III. Both machine learning algorithms surpass the traditional algorithm for predicting the thermal profiles by having the overall RMSEs of 1.5°C and 1.6°C for the ANN and SVM algorithms, respectively. Whereas the overall RMSE of the IEEE algorithm is 2.6°C. In addition, the error duration curves of the three algorithms for all the transformers are shown in Fig. 5. Clearly, the ANN algorithm surpasses the other algorithms and the highest absolute error is within 5°C while the maximum error produced by the traditional algorithm could go up to 8.5°C.

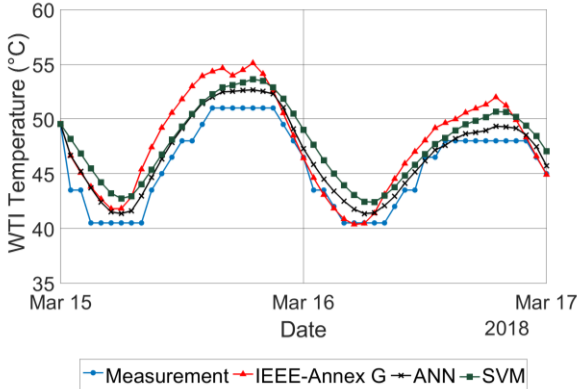


Fig. 4 Snapshot of the WTI measurement and predictions made by the three algorithms for Transformer A.

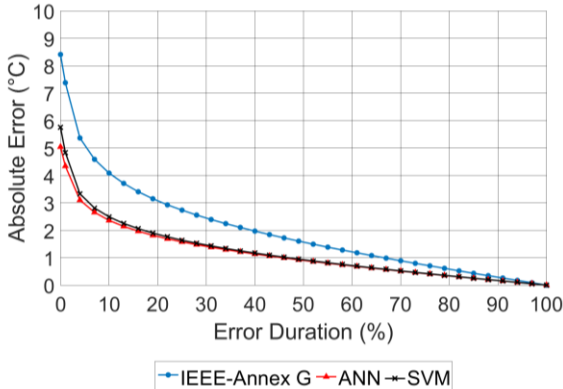


Fig. 5 Error duration curves of the IEEE, ANN and SVM algorithms for all the transformers.

TABLE II  
OPTIMIZED HYPERPARAMETERS

Transformer	ANN	SVM		
	Neurons	$C$	$\gamma$	$\epsilon$
A	8	207.5	17.18	0.8008
B	8	339.6	5.722	0.1264
C	10	16.9	1.599	0.0087
D	7	717.9	84.415	0.1182
E	8	26.5	0.731	0.1218
F	10	966.4	185.203	0.0159
G	8	891.3	1.91	0.0167
H	7	138.5	1.063	0.13
I	7	324.6	4.673	0.0237

The number of the hidden layers is equal to 1 for all the models.

TABLE III  
STATISTICAL QUANTITIES OF THE ERRORS.

Transformer	IEEE		ANN		SVM	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
A	2.2	1.7	1.5	1.1	1.6	1.2
B	2.1	1.7	1.2	0.9	1.2	0.9
C	2.8	2.1	1.7	1.3	1.9	1.5
D	3.5	2.6	2.2	1.5	2.5	1.7
E	2.3	1.8	1.5	1.2	1.5	1.1
F	2.4	1.8	1.4	1.1	1.4	1.1
G	2.5	1.9	1.3	1	1.4	1
H	2.2	1.7	1.2	0.8	1.2	0.9
I	2.8	2.2	1.4	1	1.5	1.1
All Transformers	2.6	2.0	1.5	1.1	1.6	1.2

The benefit of the proposed model over a traditional approach is that it can easily include weather factors: wind speed and solar radiation. Furthermore, transformer thermal parameters which are derived from a heat run test such as the top-oil and hot-spot temperature rise at rated load are not required. Based on the authors' experience, the computational time for training the ANN models, which are trained with the data size of about 30,000 datapoints, was approximately 10 minutes for each model using an ordinary desktop computer. Many standard programming languages have built in machine learning toolboxes, which removes the need for a user who has expertise in machine learning.

It is of interest to know to what extent the present WTI temperature is related to the past information. The regression order is varied from one to ten to investigate whether they could improve the model accuracy. As the ANN algorithm outperforms

the SVM algorithm, the following investigation is undertaken using the ANN algorithm. A plot of the overall RMSE against the regression order is shown in Fig. 6. The accuracy decreases with increasing the regression order. This means that the information of the last one hour before the present time is sufficient to predict the temperature. It is noteworthy that increasing the regression order decreased model accuracy due to overfitting issues.

Further investigations are made regarding the impacts of weather factors on the accuracy of the prediction. Models using the ANN algorithm are trained with different sets of input features: 1) load, ambient temperature and WTI, 2) load, ambient temperature, WTI and wind speed, 3) load, ambient temperature, WTI and solar radiation and 4) load, ambient temperature, WTI, wind speed and solar radiation. The RMSEs for all nine transformers based on the input features are provided in Table IV. It is found that considering the load, ambient temperature and WTI temperature plus wind speed could improve the accuracy by decreasing the error by  $0.5^{\circ}\text{C}$  in general while taking into account solar radiation improves the accuracy in some units only. In fact, in some units considering the solar radiation makes the prediction less accurate. This is probably because of the overfitting issue where the algorithm tries to fit the output that do not actually relate to that input features. The accuracy of the alternative models that do not consider wind speed and solar radiation is slightly improved compared to the IEEE model, however, the main advantage of the alternative algorithm is that the wind speed and solar radiation could be easily considered, unlike a traditional model where a new governing equation would have to be introduced.

The impact of increasing the number of hidden layers and neurons on model performance was also investigated. ANN models with five hidden layers containing 50 neurons for each layer have been trained and tested. No significant difference was observed compared against a single hidden layer that has the number of neurons between seven and ten.

## V. INCIPIENT THERMAL FAULT IDENTIFICATION

An algorithm for identifying incipient thermal faults, such as restricted oil flow, is proposed using the thermal model, see Fig. 7. for a flow chart of the algorithm. Firstly, a transformer thermal model is developed using historical operational measurements. It is strongly recommended that the model should be trained with at least 1-year data so that the model can be exposed to yearly load and weather cycles. Secondly, a fault identification threshold (FIT) is calculated based on the training data as follows:

$$\text{FIT} = \overline{|E|} + 4\sigma(|E|) \quad (6)$$

With regards to the error probabilities of the nine transformers, typically 99.7% of the error lies within the threshold FIT. Following that, the load profile, ambient temperature, solar radiation, wind speed and the hot-spot temperature at the previous time step and the hot-spot temperature at the present time step are inputted and then the thermal model predicts the present hot-spot temperature which is compared with the measured hot-spot

temperature. Subsequently, the difference between the measurement and prediction in the testing dataset ( $E_t$ ) is compared to the FIT to identify whether there is an incipient fault. There are 5 scenarios: 1) the difference between the measurement and prediction is less than FIT (No indication of an incipient fault), 2) the difference is greater than FIT but less than FIT + STD (Level 1), 3) the difference is greater than FIT + STD but less than FIT + 2STD (Level 2), 4) the difference is greater than FIT + 2STD but less than FIT + 3STD (Level 3) and 5) the difference is greater than FIT + 3STD (Level 4). The higher level means it is more likely that there is an incipient thermal fault.

To ensure that the reliability of the algorithm, a counting technique is introduced. A warning will be raised only if the errors have exceeded the thresholds for a certain period continuously. The periods of the exceeding time vary with the magnitude of the divergence. The higher difference between the measurement and prediction is, the shorter period for that the excessive errors must have persisted is. For Level 1, to Level 4, the periods of time that the excessive errors must have persisted to raise a warning are 2 days (48 datapoints), 1 day (24 datapoints), 12 hours (12 datapoints) and 6 hours (6 datapoints), respectively. To reset all the counters, the errors must have been less than FIT for 1 day (24 datapoints) consecutively.

TABLE IV  
RMSES USING DIFFERENT SET OF INPUTS

Inputs	RMSE ( $^{\circ}\text{C}$ )
Load, WTI temperature, tap position (if available) and ambient temperature	2.2
Load, WTI temperature, tap position (if available) ambient temperature and wind speed	1.6
Load, WTI temperature, tap position (if available) ambient temperature and solar radiation	2.1
Load, WTI temperature, tap position (if available) ambient temperature, wind speed and solar radiation	1.6

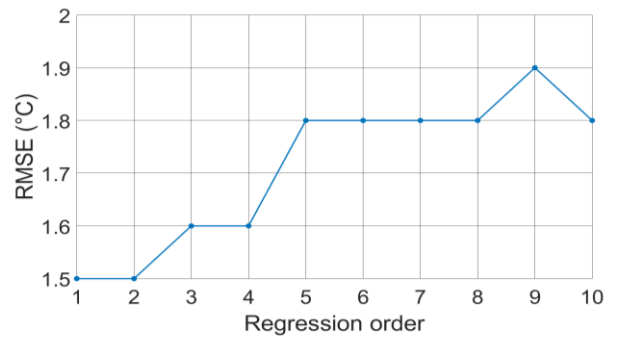


Fig. 6 Overall RMSE against the regression orders for all the transformers.

This algorithm has been applied to identify incipient faults from the measured data. It is found that there was one transformer where an incipient fault was detected by the algorithm, which is Transformer E. The predictions and the errors between the measurement and predictions between 1<sup>st</sup> January 2013 and 31<sup>st</sup> December 2019 were then calculated by the algorithm using the load, thermal and weather measurements, see Fig. 8 and Fig. 9.

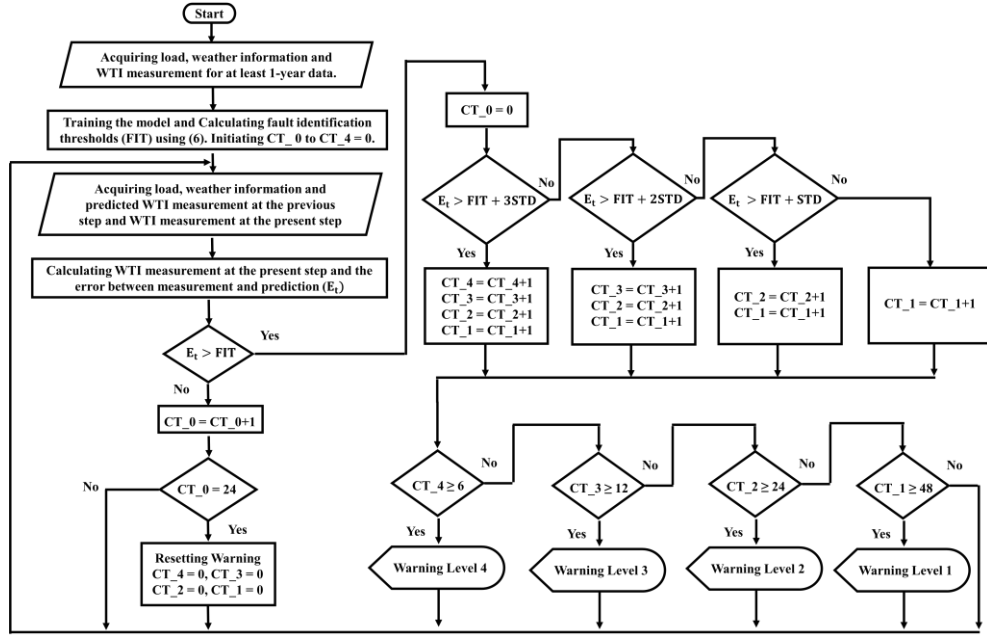


Fig. 7 Flow chart of the algorithm of the incipient thermal fault identification.

There had not been a significant difference between the measurement and prediction until 3<sup>rd</sup> January 2018. A warning level 4 was raised because there were continuous excessive divergences between the measurement and prediction greater than the level 4 threshold. The divergence had persisted until mid April 2019 and the measurement has approached the prediction again since then. Although the WTI temperature is not significantly higher than the WTI temperature in the adjacent year, the average load in the suspicious periods, 0.3 p.u., is relatively smaller than the average load in the adjacent year which is 0.4 p.u. A list of maintenance activities for this transformer was reviewed but there were not any activities, which were undertaken regarding this incident. It is believed that there may have been a partial blockage of oil sludge in an oil pipe. The blockage in the oil pipe results in a reduction in the oil flow, which carries the heat away from a transformer tank to the environment via a cooler bank, causing a lower heat transfer rate.

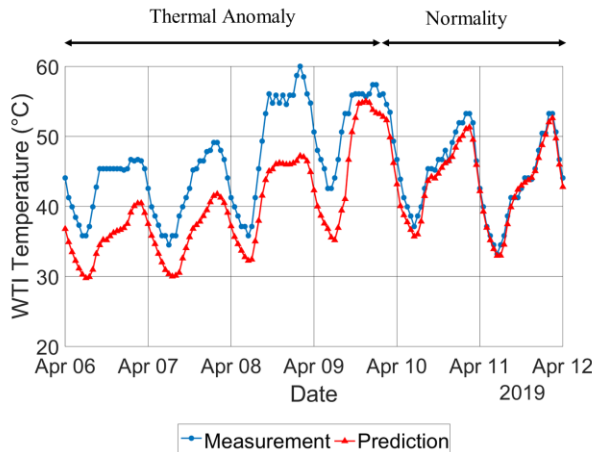


Fig. 8 Time series of WTI measurement and prediction over thermal anomaly and normal periods for Transformer E.

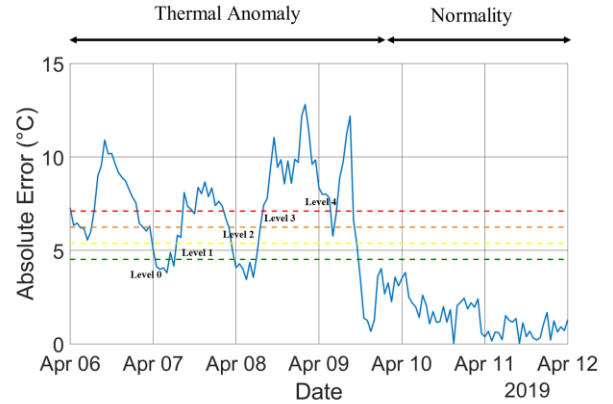


Fig. 9 Time series of the absolute errors between the measurement and prediction from Fig. 8 showing periods that thermal anomaly was detected.

Although the anomalous thermal fault was remedied without intervention, it took approximately 1 year before the problem disappeared. Using the algorithm, the issue could have been identified within one day and subsequently a proper maintenance could be planned to fix the fault.

It should be noted that the proposed approach will detect anomalous performance provided that the transformer is behaving “normally” in the training data. It would therefore be recommended that these algorithms are trained by the data from the commencement of service.

## VI. RELATIVE THERMAL PERFORMANCE

Using the proposed thermal model, relative thermal performance amongst a transformer family could also be identified. The thermal model, which is developed using the measurements of a unit in the fleet, should generally be able to be used for predicting the temperature of the other transformers in the same family without significant errors if those transformers are exposed to similar installation. Building on this hypothesis, a thermally underperforming transformer could be distinguished from the family by creating an error matrix. The error matrix is

a matrix of mean errors between the measurement and predictions made by the thermal model that is developed using its own historical data and the data of the other transformers in the same family. Typically, the mean errors between the measurement and prediction made by the thermal model using its own data should be approximately zero, however, the mean errors that are produced by the thermal model of the other transformer might be positive or negative, if the thermal behavior between them is significantly different. If the mean error ( $\bar{E}$ ) is positive, it means the transformer whose measurement data are used for testing is operating at a hotter temperature than the transformer used to train the model.

In the datasets considered, there are three transformers that are in the same family (G, H and I). An error matrix is provided in Table V where Transformer ID in the first column indicates the thermal model used for prediction and Transformer ID in the row headers indicates the measured data compared with the model prediction. It can be seen that the diagonal elements are approximately zero because the measurement and thermal model are derived from the same transformer while the elements outside the diagonal could be zero, positive or negative depending on the differences between their thermal behavior. In the transformer family under consideration, Transformer I has a higher operating temperature, indicating poorer cooling performance compared to the other units in its family, see Fig. 10 and Table V.

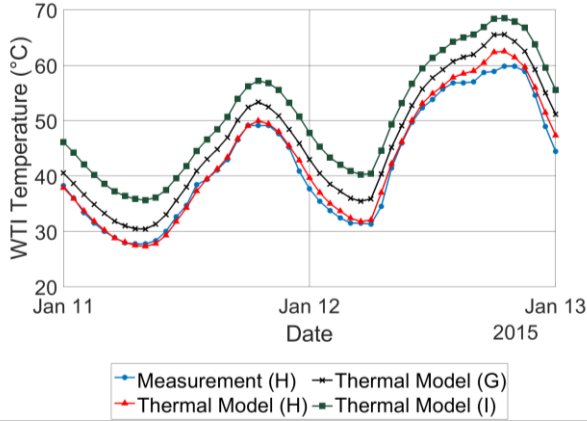


Fig. 10 Snapshot of measurement and prediction made by Transformer H's model for Transformer G, H and I.

## VII. CONCLUSION

The machine learning algorithms, i.e. ANN and SVM algorithms are used to predict the WTI temperature under various loading conditions. The thermal models consider load profile, tap position, ambient temperature, wind speed and solar radiation. Both alternative algorithms generally surpass the traditional algorithms exhibiting closer agreement with the measured data. Increasing the regression order results in overfitting issues, which decreases the model accuracy. Considering wind speed and solar radiation improved model accuracy for some units. A method for incipient thermal fault identification is introduced using the thermal model. The algorithm notifies the operator if the differences between measurement and prediction have been significantly greater than a specified threshold over a given period. The greater the divergence between the meas-

urement and prediction, the shorter time duration that the divergence must have persisted to display a warning. The proposed method identified thermally anomalous behavior in field measurements. Relative thermal performance among a transformer family is also investigated using machine learning thermal models.

TABLE V  
ERROR MATRIX

Model	Measurement		
	G	H	I
G	-0.05°C	-3.67°C	3.95°C
H	3.66°C	-0.01°C	7.91°C
I	-3.28°C	-7.2°C	0.37°C

## VIII. APPENDIX

TABLE VI  
TRANSFORMER SPECIFICATIONS

Quantities	Transformers						
	A	B	C	D	E	F	G to I
Voltages(kV)	275/66		400/132				400/275
Nameplate Capacity (MVA)	180	180	240				1000
Typical cooling mode	ONAN	OFAF	ONAN	ONAN	ONAN	ONAN	ONAN
Load losses (kW)	960	992	748	596	792	848	1772
No-load losses (kW)	128	129	55	68	51	63	75
$\Delta\theta_{OR}$ (K)	40.6	39.4	38.5	46.3	33.8	34.2	45.0
$Hg_R$ (K)	7.3	22.0	8.7	5.7	14.5	5.2	15.4

## IX. ACKNOWLEDGMENT

The authors would like to acknowledge National Grid Electricity Transmission for supporting the project and agreeing to publish the results. Project Condition and Climatic Environment for Power Transformer (NIA\_NGET0213) was made possible through the Network Innovation Allowance.

## X. REFERENCES

- [1] M. Wang, A. J. Vandermaar and K. D. Srivastava, "Review of condition assessment of power transformers in service," *IEEE Electr. Insul. Mag.*, vol. 18, no. 6, pp. 12-25, 2002.
- [2] S. Li and J. Li, "Condition monitoring and diagnosis of power equipment: review and prospective," *High Voltage*, vol. 2, no. 2, pp. 82-91, 2017.
- [3] *Power transformer- Part 2: Temperature rise for liquid-immersed transformers*, IEC 60076-2, 2011.
- [4] *IEEE Guide for Loading Mineral-Oil-Immersed Transformers and Step-Voltage Regulators*, IEEE std C57.91, 2011.



- [5] *Power transformer- Part 7: Loading guide for mineral-oil-immersed power transformers*, IEC 60076-7, 2018.
- [6] D. Susa and M. Lehtonen, "Dynamic thermal modeling of power transformers: further Development-part II," *IEEE Trans. Power Del.*, vol. 21, no. 4, pp. 1971-1980, 2006.
- [7] A. A. Taheri, A. Abdali and A. Rabiee, "A Novel Model for Thermal Behavior Prediction of Oil-Immersed Distribution Transformers With Consideration of Solar Radiation," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1634-1646, 2019.
- [8] A. Doolindachbaporn, G. Callender, P. Lewin, E. A. Simonson and G. Wilson, "A Top-Oil Thermal Model for Power Transformers that Considers Weather Factors," *IEEE Trans. Power Del.*, doi: 10.1109/TPWRD.2021.3105459.
- [9] CIGRÉ Working Group A2.38, *Transformer Thermal Modelling*, CIGRÉ, Paris, France, Technical Brochure 659, 2016.
- [10] Q. He, J. Si, and D. J. Tylavsky, "Prediction of Top-Oil Temperature for Transformers Using Neural Networks," *IEEE Trans. Power Del.*, vol. 15, no. 4, pp. 1205-1211, 2000.
- [11] W. Bengang, W. Xinye, Y. Zhoufei and H. Hua, "A method of optimized neural network by L-M algorithm to transformer winding hot spot temperature forecasting," *2017 IEEE Electr. Insul. Conf. (EIC)*, Baltimore, MD, USA, 2017, pp. 87-91.
- [12] E. A. Juarez-Balderas, J. Medina-Marin, J. C. Olivares-Galvan, N. Hernandez-Romero, J. C. Seck-Tuoh-Mora and A. Rodriguez-Aguilar, "Hot-Spot Temperature Forecasting of an Instrument Transformer Using an Artificial Neural Network," *IEEE Access*, vol. 8, pp. 164392-164406, 2020.
- [13] M. Kunicki, S. Borucki, A. Cichoń and J. Frymus, "Modeling of the Winding Hot-Spot Temperature in Power Transformers: Case Study of the Low-Loaded Fleet," *Energies*, vol. 12, no. 3561, 2019.
- [14] CIGRÉ Working Group A2.24, *Thermal Performance of Power Transformers*, CIGRÉ, Paris, France, Technical Brochure 393, 2009.
- [15] P. A. Gonzalez and J. M. Zamarreno, "Prediction of hourly energy consumption in buildings based on a feedback artificial neural network," *Energy Build.*, vol. 37, no. 6, pp. 595-601, 2005.
- [16] P. A. Jansson, "Neural Networks: An Overview," *Anal. Chem.*, vol. 63, no. 6, pp. 357-362, 1991.
- [17] M. T. Hagan and M. B. Menhaj, "Training Feedforward Networks with the Marquardt Algorithm," *IEEE Trans. Neural Networks*, vol. 5, no. 6, pp. 989-993, 1994.
- [18] R. Aggarwal and Y. Song, "Artificial neural networks in power systems. II. Types of artificial neural networks," *Power Eng. J.*, vol. 12, no. 1, pp. 41-47, 1998.
- [19] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 20, no. 1, pp. 199-222, 2004.
- [20] J. Shawe-Taylor and N. Cristianini, "Basic kernels and kernel types," in *Kernel Methods for Pattern Analysis*, Cambridge, Cambridge University Press, 2004, pp. 291-326.
- [21] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, Cambridge: MIT Press, 2002.
- [22] S. Christou, "Thermal Prognostic Condition Monitoring for MV Cable," Ph.D. Thesis, Dept. Elect. Comput. Sci., Univ. Southampton, Southampton, UK, 2016.
- [23] K. Duan, S. S. Keerthi and A. N. Poo, "Evaluation of simple performance measures for tuning SVM hyperparameters," *Neurocomputing*, vol. 51, pp. 41-59, 2003.
- [24] J. Moćkus, "On Bayesian methods for seeking the extremum," *Optimization techniques IFIP technical conference 1975*, Springer, Berlin, Heidelberg, pp(400-404).
- [25] J. Snoek, H. Larochelle and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, vol. 2, p. 2951-2959, 2012.
- [26] Met Office (2006): MIDAS: UK Hourly Weather Observation Data. NCAS British Atmospheric Data Centre, Leeds, UK. [Online]. Available: <https://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c>. [Accessed 1 January 2019].



**Atip Doolindachbaporn** was born in Bangkok, Thailand in 1992. He received M.Sc. degree in electrical engineering from the University of Southampton, UK in 2017. He is currently a Ph.D. degree student at the University of Southampton. His research interests include power transformer rating and condition monitoring of power transformers.



**George Callender** was born in Basildon, UK in 1991. He received M.Sci (Hons) in Natural Sciences (Maths and Physics) from the University of Durham, UK in 2013. He received a Ph.D. degree in electrical engineering from the University of Southampton, UK in 2018. He is currently a Research Fellow in High Voltage Numerical Modelling at the University of Southampton. His research interests include partial discharge phenomena and the thermal modelling of high voltage plant.



**Prof. Paul L. Lewin** (M'05-SM'08-F'13) was born in Ilford, Essex in 1964. He received the B.Sc. (Hons) and Ph.D. degrees in electrical engineering from the University of Southampton, UK in 1986 and 1994, respectively. He joined the academic staff of the University in 1989 and is Head of Electronics and Computer Science, where he is also Director of the Tony Davies High Voltage Laboratory. His research interests are within the generic areas of applied signal processing and control. Within high voltage engineering this includes condition monitoring of HV cables and plant, surface charge measurement, HV insulation/dielectric materials and applied signal processing. Since 1996 he has received funding and grants in excess of £30M, supervised 50 graduate students to successful completion of their doctoral theses and published over 500 refereed conference and journal papers in these research areas. He is a Chartered Engineer, a Fellow of the IET, and former president of the IEEE Dielectrics and Electrical Insulation Society.



**Edward Simonson** was born near Gloucester, UK in 1965. He received B.Sc. (Hons) in Physics from the University of Durham, UK, in 1986. He received a D.Phil. Degree in Astronomy from the University of Sussex, UK in 1990. He joined the then new National Grid Company in 1990 to work on detailed thermal-hydraulic models of transformer windings and was involved in the measurement and analysis of the first fiber-optic temperature measurements of transformers in the UK. In 2003 he joined Southampton Dielectric Consultants and continued to work on transformer thermal ratings. He has been a member of the maintenance team for the IEC 60076-7 loading guide and a member of the CIGRE working group A2.38 on transformer thermal modelling.



**Gordon Wilson** (M '08) was born in Weymouth, UK in 1972. He is an innovation lead at National Grid managing projects on HV substation assets. Gordon joined National Grid in 1999 after completing a National Grid-sponsored Ph.D. in chemistry at the University of Surrey. He has primarily worked on transformers and insulating materials over the last 20 years. Throughout his career, Gordon has been involved in research and development and has sponsored PhD students at Cardiff University, Cranfield University, University of Manchester and University of Southampton, where he is a visiting fellow. Gordon is a member of BSi National Committee GEL/10 for electrotechnical fluids, CIGRE Advisory Group D1.01 for liquid and liquid impregnated insulation systems and he has represented the UK in several IEC and CIGRE groups related to oil and transformer ratings.