

Kullback-Leibler divergence based wind turbine fault feature extraction

Yueqi Wu, Xiandong Ma

Engineering Department
Lancaster University
Lancaster, LA1 4YW

Email: y.wu31@lancaster.ac.uk, xiandong.ma@lancaster.ac.uk

Abstract— In this paper, a multivariate statistical technique combined with a machine learning algorithm is proposed to provide a fault classification and feature extraction approach for the wind turbines. As the probability density distributions (PDDs) of the monitoring variables can illustrate the inner correlations among variables, the dominant factors causing the failure are figured out, with the comparison of PDD of the variables under the healthy and unhealthy scenarios. Then the selected variables are used for fault feature extraction by using kernel support vector machine (KSVM). The presented algorithms are implemented and assessed based on the supervisory control and data acquisition (SCADA) data acquired from an operational wind farm. The results show the features relating specifically to the faults are extracted to be able to identify and analyse different faults for the wind turbines.

Keywords- Wind turbine, condition monitoring, kernel support vector machine, Kullback-Leibler divergence, Supervisory Control and Data Acquisition

I. INTRODUCTION

In the modern world, the increasing demand for electricity usage has influenced electricity production technology and fault diagnosis strategies. In particular, a great number of wind turbines have been installed in remote and offshore areas. The importance of condition monitoring (CM) and fault detection methods have increased dramatically in past decades [1], [2]. The global wind power operation and maintenance (O&M) cost is estimated to grow from 12 billion USD (2016) to 27 billion USD by 2026 [3]. As for the offshore wind farm, the O&M costs will be much higher due to the harsh environment, which can be between USD 0.027 and USD 0.048/kWh [4]. Hence, the CM method plays a vital role in the wind farm in order to monitor the condition of the wind turbines and thus reduce the O&M costs. Furthermore, the maintenance work can be scheduled in advance to reduce the shutdown time. There are various types of fault detection approaches, which can be classified into two categories: mechanistic methods and data-driven model-based approaches. As for the model-based approaches, it is not compulsory to consider the inner physical characteristic and process of the targeted system, compared with the mechanistic methods [5]. With regards to the model-based condition monitoring method especially in the detection of faults at their early stage, the multivariate statistical techniques have received a rising attention for condition monitoring of large-scale industrial systems in the past years [6].

The paper addresses the problem of fault feature extraction and selection of monitoring variables by employing Kullback-Leibler divergence (KLD) and kernel support vector machine (KSVM). In this paper, different faults are discussed and classified. The highly correlated variables with regards to the wind turbine failures are extracted. In Section II, the basic knowledge of KLD and SVM with radical basis function kernel are introduced. The SCADA data and the modelling procedures are introduced in Section III. In Section IV, the probability density distribution (PDD) of each variable under different operation scenarios is analysed. Then the PDD differences between the healthy turbine and faulty turbine for the same variables are analysed by KLD. Larger KLD values indicate the variable has a larger difference between the healthy and faulty conditions. By setting an appropriate threshold for the KLD value, the variables, which have larger chances to trigger a specific fault are selected. The fault features are also extracted based on the KLD selected variables.

II. METHODS

A. Kullback-Leibler Divergence

The KLD, also called Shannon entropy, is a probability measurement from information theory perspective. It is used to compare the one probability distribution against with another one [10]. To evaluate the discrimination between two PDDs, $p(x)$ and $g(x)$, of a random variable x , the Kullback-Leibler information $I(p \parallel g)$ is calculated:

$$I(p \parallel g) = \int p(x) \log \frac{p(x)}{g(x)} dx \quad (1)$$

The divergence of the two distributions is a symmetric version of the information:

$$KLD(p, g) = I(p \parallel g) + I(p \parallel f) \quad (2)$$

It is only produced when $\sum_x p = \sum_x g = 1$ and the two distributions share the same support set. When the two distributions are same, the KLD of the two distributions is 0. When the difference between the two distributions increases, the calculated value of KLD also increases. In other words, the bigger the KLD value, the larger the difference between two distributions that can be found.

B. Kernel Support Vector Machine

The support vector machine (SVM) is a machine learning algorithm that is developed from the statistical

learning with a better performance than many other methods, such as decision trees, discriminant analysis and nearest neighbour. It is a supervised learning method that has been widely applied in regression and classification [7], [8].

It is assumed that there are two classes for separation, which are denoted as $\{-1, 1\}$. For a given training data set $T(x, y)$:

$$T(x, y), (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), \\ x_n \in R^n, y_n \in \{-1, 1\}. \quad (3)$$

where x are the training vectors and y are the indicator vectors indicating the class of x , respectively. In this case, $\{-1, 1\}$ indicates the turbine under faulty and healthy condition, respectively. The hyperplane created by SVM is used to separate the two different classes that is specified by its weight w and the bias b [8]. The hyperplane can be described as:

$$(w, x) + b = 0 \quad w \in R^n, b \in R \quad (4)$$

It yields a corresponding decision function:

$$f(x) = \text{sgn}((w, x) + b) \quad (5)$$

The sign of $f(x)$ depends on the side of the hyperplane where the sample lies. An optimal hyperplane is the one that maximises the distance between the hyperplane and the nearest points to the hyperplane of both classes. The points of having the minimal distances to the hyperplane are called support vectors. The distance between the support vectors is called margin. In order to obtain the hyperplane, the weight and bias of support vectors should satisfy the equation $|(w, x) + b| = 1$. The optimal hyperplane with maximal margin $\tau(w)$ can thus be formulated as follows:

$$\min_{w \in R^n, b \in R} \tau(w) = \frac{1}{2} \|w\|^2 \quad (6)$$

$$\text{subject to} \quad y_i((w, x) + b) \geq 1 \quad \forall i = 1, \dots, n \quad (7)$$

The decision function $f(x_i)$ yields $+1$ for $y_i \in \{1\}$ while -1 for $y_i \in \{-1\}$ when the constraint (7) is satisfied. Furthermore, the Lagrange multipliers α_i are introduced to adapt the Karush Kuhn Tucker conditions in order to optimise the w and b [9].

$$\max_{\alpha \in R^n} L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i, x_j) \quad (8)$$

$$\text{subject to} \quad \alpha_i \geq 0 \quad \forall i = 1, \dots, n \quad (9)$$

$$\text{and} \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (10)$$

The optimal parameters w^* and b^* can be written as:

$$w^* = \sum_{i=1}^n \alpha_i y_i x_i \quad (11)$$

$$b^* = -\frac{1}{2} (w^*, x_a + x_b) \quad (12)$$

where $x_a \in \{1\}$ and $x_b \in \{-1\}$ indicate the support vectors from two classes. Then the decision function from (5) can be written as:

$$f(x) = \text{sgn}((w^*, x) + b^*) \quad (13)$$

By adapting the optimal parameters w^* and b^* from (11-12), the (13) in the feature space can be updated as

$$f(x) = \text{sgn}((w \cdot \phi(x)) + b) \\ = \sum_{i=1}^n y_i \alpha_i \phi(x) \phi(x_i) + b^* \quad (14)$$

The above equations are only suitable for linear separation problem. In general, most of the cases are non-linearly separable. In order to solve the non-linearly separable problem, the kernel function is introduced for improving the computation efficiency. In this specific condition, the Radical Basis Function (RBF) or Gaussian kernel is used.

The dot product (x, x_i) can be substituted by the kernel

$$k(x, x_i) = (\phi(x) \cdot \phi(x_i)) \quad (15)$$

where $\phi(x)$ and $\phi(x_i)$ are the mapping of x, x_i in hyper dimension. By adapting a RBF kernel, the kernel function can be written as:

$$K(x, x_i) = e^{\left(\frac{-\|x-x_i\|^2}{2\sigma^2}\right)} = e^{(-\gamma\|x-x_i\|^2)} \quad (16)$$

where σ is the width of the kernel, which controls the smoothness of the decision boundary in the feature space and $\gamma = -\frac{1}{2\sigma^2}$.

Thus, the decision function $f(x)$ for an unknown input x in the feature space can be extended to:

$$f(x) = \text{sgn}((w \cdot \phi(x)) + b) \\ = \sum_{i=1}^n y_i \alpha_i e^{(-\gamma\|x-x_i\|^2)} + b^* \quad (17)$$

In the regression cases, the SVM classification function from (13) can be expressed as follows to predict the value $h(x)$

$$h(x) = (w^*, x) + b^* \quad (18)$$

By adapting RBF kernel to the regression function (18) in feature space, the prediction value can be extended to:

$$f(x) = (w^*, x) + b^* \\ = \sum_{i=1}^n \alpha_i e^{(-\gamma\|x-x_i\|^2)} + b^* \quad (19)$$

III. SCADA DATA

The SCADA systems were initially developed independent of the control systems and then have evolved to the generation incorporated with the industrial control systems (IDS). Recently, SCADA systems have been developed and embedded into the distributed control system (DCS). It is a data control and acquisition system, which has been widely used for high-level supervisory management through controller such as programmable logic controller to interact with machineries or plants [11], [12]. The data used in this paper are from an operational wind farm in one-year duration consisting of 26 turbines with each turbine having 128 monitoring variables, including temperatures, active power and various speeds. The SCADA system applied in this particular wind farm sampled signals at 10 minutes intervals to cover a range of physical and electrical variables and digital control signals. However, the digital signals cannot reveal the working condition of the wind turbine; it is thus necessary to eliminate those variables before further processing [13].

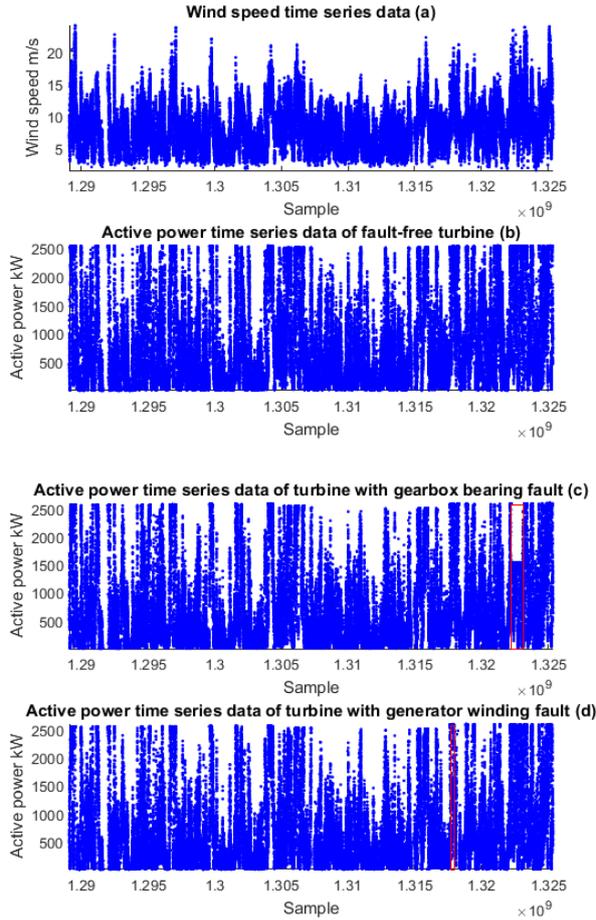


Figure 1. Time series data of (a) wind speed, (b) active power of a fault-free turbine, (c) active power of a turbine with gearbox bearing fault and (d) active power of a turbine with generator winding fault, from top to bottom

Fig. 1 shows wind speed and the output active power of the turbines working in different conditions over one year. Fig. 1(a) shows the wind speed data measured on both healthy and faulty turbine. Fig. 1 (b-d) show the active power outputs over one year for the fault-free turbine and turbines with gearbox bearing fault and generator winding fault, respectively. The occurrence of the faults are labelled in red boxes. Compared with the turbine having a gearbox fault, the turbine with a generator fault clearly has a shorter faulty time duration. In order to prevent overheating to the critical components and further damage the whole turbine, the turbines were operated with power output being reduced to half.

Fig. 2 illustrates the procedures of variable selection and fault feature extraction. In order to obtain the PDD ideally, the data samples should be selected as much as possible. In the fault classification and feature extraction procedures, 50000 samples measured over one year are selected. Two turbines are selected for modelling: one has a gearbox fault while another has a generator winding fault. The faulty data can be selected based on time sequential information. First, the PDDs of all the monitoring variables are calculated with the data under healthy and faulty conditions. Then the two types of PDD are compared by calculating the KLD values between them. By sorting the KLD values of the monitoring variables in descending

order, the monitoring variables related specifically to the fault can be discovered. Finally, the features associated with the gearbox bearing fault and generator winding fault are extracted by the KSVM.

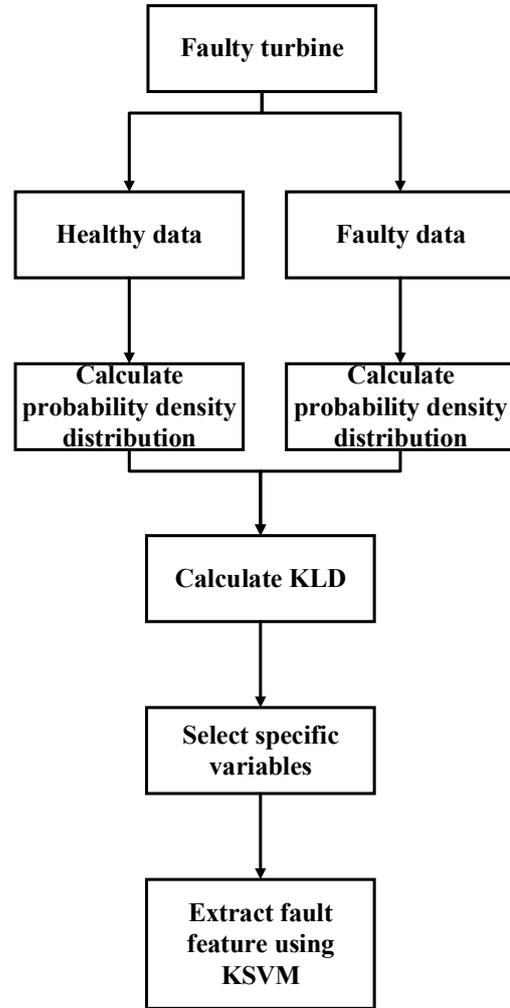


Figure 2. Fault classification and feature extraction

IV. RESULT AND DISCUSSION

A. Variable selection using KLD

As mentioned above, the digital constants and control signals are removed in the data pre-processing, totally 78 variables are then retained for further analysis. The KLD values are calculated based on same variable under healthy and faulty conditions. By sorting the KLD values in descending order and eliminate the variables that are infinite or cannot be computed, the KLD values of 22 variables related to the gearbox fault are given in Table I.

TABLE I. KLD VALUE OF GEARBOX BEARING FAULT VARIABLES

Variable	Nacelle temperature	Wind direction	Nacelle position	Pitch converter temperature 2
KLD	15.89	14.95	14.51	7.47
Variable	Hydraulic fluid temperature	Pitch angle 2	Pitch Converter Temperature 1	Pitch Converter Temperature 3

KLD	6.49	5.52	4.87	4.68
Variable	Power Switchboard Temperature	Gearbox Oil Pressure Behind Pump	Temperature of generator cooling water return	Temperature of gearbox oil heat exchanger output
KLD	4.68	2.44	1.69	1.14
Variable	Hygrometer 2 temperature	Generator speed	Generator speed (to-default)	Wind speed right-hand
KLD	1.11	0.86	0.7	0.51
Variable	Wind speed	Wind speed left-hand	Minimum pitch angle	Oscillation signal Z
KLD	0.51	0.5	0.39	0.19
Variable	Maximum pitch speed	Temperature of Cooling water return		
KLD	0.16	0.16		

All 22 KLD values show the differences between healthy condition and faulty condition. The larger value indicates larger difference. It can be observed from the Table I that some variables are related to the environmental condition, such as wind speed and direction. The environmental related variables can be ignored in the variable selection because they do not have a direct impact on the fault. By eliminating these variables, pitch system, gearbox and generator related variables are retained.

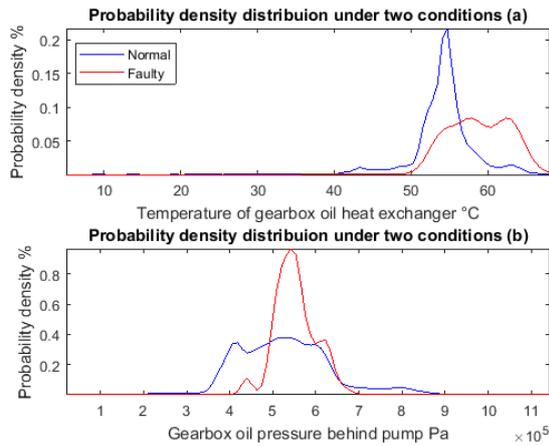


Figure 3. PDDs between healthy and faulty conditions of (a) gearbox oil heat exchanger temperature and (b) gearbox oil pressure behind pump

Fig. 3 shows examples of the PDDs of the gearbox oil exchanger temperature and gearbox oil pressure behind pump under healthy and faulty conditions. It can be observed from the top plot of Fig.3 that when the turbine is under the healthy condition, the PDD of gearbox oil exchanger is similar to Gaussian distribution with an average temperature around 55 °C. In addition, the gearbox oil pressure behind pump also indicates the abnormal behaviour of this turbine. Consequently, the pressure tended to be higher when fault occurred, compared with the healthy condition.

TABLE II. KLD VALUES OF GENERATOR WINDING FAULT VARIABLES

Variable	Temperature of cooling water return	Wind speed left-hand	Wind speed	Pitch converter
-----------------	-------------------------------------	----------------------	------------	-----------------

KLD	37.89	11.73	11.7	11.62
Variable	Wind speed right-hand	Hygrometer 2 temperature	Gearbox oil pressure Behind Pump	Pitch converter temperature 1
KLD	11.25	10.26	8.26	5.5
Variable	Temperature of generator cooling water return	Oscillation signal Z	Hydraulic fluid temperature	Nacelle temperature
KLD	3.7	2.92	1.81	1.43
Variable	Temperature of gearbox oil heat exchanger output	Pitch angle 1	Minimum pitch angle	Pitch angle 3
KLD	1.06	0.45	0.45	0.37
Variable	Pitch angle average 1-3	Generator speed	Generator speed (to-default)	Pitch angle 2
KLD	0.33	0.28	0.28	0.26
Variable	Generator temperature			
KLD	0.2			

Table II shows KLD values of the selected variables related to generator winding fault. There are 21 variables are selected for this specific fault. It can be observed from the table that the pitch system, gearbox and generator related variables are selected.

Fig. 4 shows two variables' PDDs of the turbine with generator winding fault. For this specific fault, both the temperatures of generator cooling water and generator winding tended to shift towards higher temperature zone compared with the normal condition. This clearly indicates the abnormal behaviour of generator when fault appears.

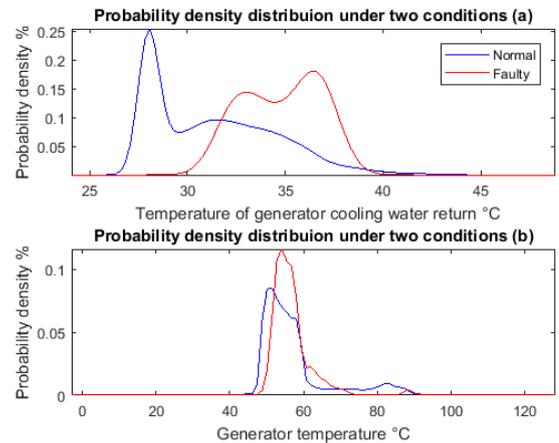


Figure 4. PDDs between healthy and faulty conditions of (a) generator cooling water return temperature and (b) generator temperature

B. Feature extraction using KSVM

It can be concluded from the both Table I and II that the pitch system, generator system and gearbox system tended to be the three most abnormal systems of the turbines when faults occur. The variables are chosen from gearbox,

generator and pitch systems to separate the fault and normal conditions.

The relationship between generator speed and wind speed can be hard to discover, which is shown at top plot of Fig. 5. In order to extract the real behaviour of the generator speed against the wind speed, the KSVM is applied as a regression model to find the relationship between generator speed and wind speed. A threshold is also added to help the data selection for this regression model. The bottom plot of Fig. 5 shows the comparison between original generator speed curve and filtered generator speed curve, where the yellow curve indicates the filtered generator speed curve. The relationship between generator speed and wind speed can be revealed clearly now by adapting KSVM regression.

Then the KSVM is applied to divide the normal and fault data into two groups. Fig.6 shows the extracted feature of the gearbox bearing fault. The blue dots indicate the normal feature and the red crosses indicate the gearbox bearing fault, which is separated by the yellow plane. It can be observed from the figure in order to prevent dramatic damages to the turbine, the power output is reduced to half. This is due to the fact that the changing of pitch angle controls the output power. Any point within the yellow plane are considered as the abnormal date related to the fault. The plane created by KSVM is used to extract the fault feature can be described as $(x - 1136.12)^2 + (y - 2319.04)^2 + (z + 1916.93)^2 - 9300391.69 < 0$. When the equation is above zero, it can be considered as healthy condition. When the gearbox bearing fault occurs, the inequality is satisfied.

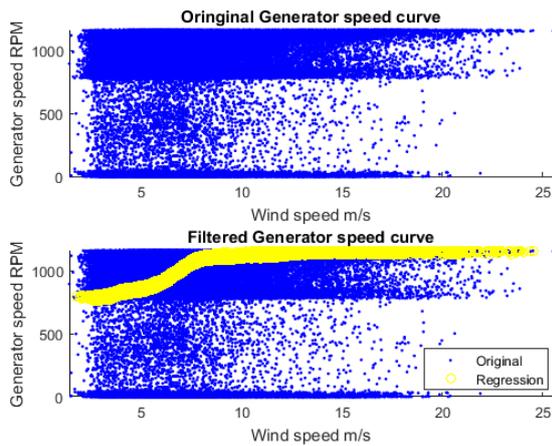


Figure 5. Original generator speed curve vs. filtered generator speed curve

Fig. 7 shows the generator speed curve against wind speed for the turbine with generator winding fault. In order to extract the key information, same method has been applied. As the bottom plot of Fig. 7 shows, the feature is revealed by KSVM.

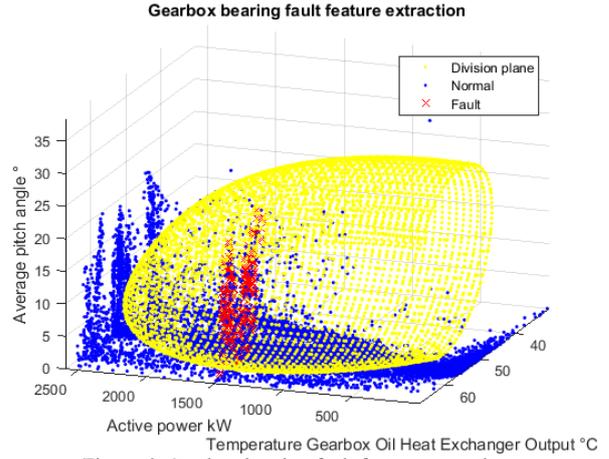


Figure 6. Gearbox bearing fault feature extraction

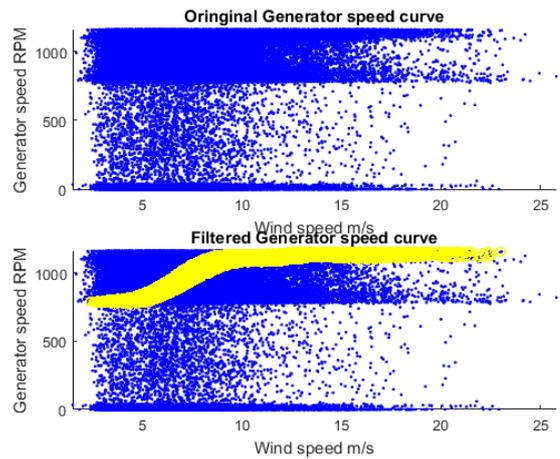


Figure 7. Original generator speed curve vs. filtered generator speed curve

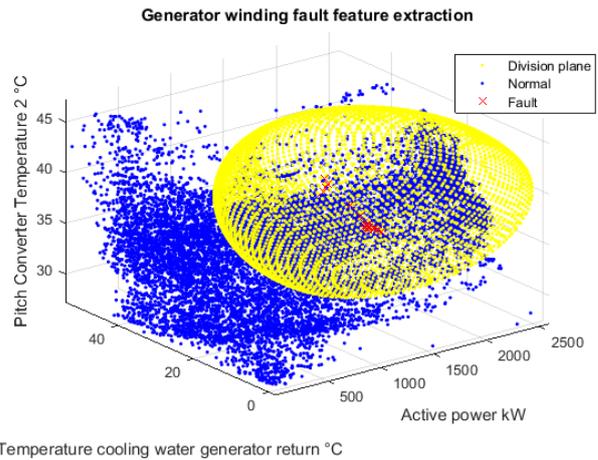


Figure 8. Generator winding fault feature extraction

Based on the active power, generator cooling water temperature and pitch converter temperature, the generator winding fault feature can be extracted as shown in Fig. 8. The plane which is used to extract the fault feature is also created by KSVM. It is described as the inequality $(x - 1251.33)^2 + (y + 5895.003)^2 + (z - 6673.873)^2 - 79617582.07 < 0$. When the inequality is established, the generator winding fault can be found. It can be observed from the figure that when the fault appears, the pitch

converter temperature tends to be higher while the generator temperature is still in a low temperature zone.

V. CONCLUSIONS

This paper presents a novel fault diagnosis and detection method based on a multivariate statistical technique combined with the machine learning algorithm based on KSVM.

The turbines with a gearbox bearing fault and a generator winding fault are tested by using the proposed model. The results are drawn as follows.

- The fault can be localised via calculating the KLD values between healthy and faulty conditions.
- The amplitudes of KLD values indicates the contributions brought by the variables to the fault.
- By selecting the most significant variables for classification model training, it reduces computation load and improve efficiency.
- The fault feature separation plane can extract the gearbox bearing fault the generator winding fault accurately.

Further work will focus on applying the fault feature model to the turbines with other faults based on deep learning algorithms. Other multivariate statistical techniques such as Wasserstein metric will be considered and incorporated for fault diagnosis.

ACKNOWLEDGMENT

Yueqi Wu gratefully acknowledges the support provided by Engineering Department of Lancaster University in terms of the Ph.D. studentship. The Wind Prospect Ltd is also gratefully acknowledged in terms of the permission of SCADA data use.

REFERENCES

- [1] Y. Wang, X. Ma and P. Qian, "Wind turbine fault detection and identification through PCA-based optimal variable selection", *IEEE Transactions on Sustainable Energy*, 2018, DOI: [10.1109/TSTE.2018.2801625](https://doi.org/10.1109/TSTE.2018.2801625).
- [2] F. Ferracuti, A. Giantomassi, S. Iarlori, G. Ippoliti & S. Longhi, "Induction motor fault detection and diagnosis using kde and kullback-leibler divergence", *Industrial Electronics Society, IECON 2013-39th Annual Conference of IEEE*, pp.2923-2928
- [3] IRENA, "Renewable Power Generation Costs in 2017", 2017.
- [4] J. Harmouche, C. Delpha and D. Diallo, "Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis: Part II", *Signal Processing*, vol. 109, pp. 334-344, 2015.
- [5] Zhang, W., & Ma, X. Simultaneous fault detection and sensor selection for condition monitoring of wind turbines. *Energies*, 9(4). 2016.
- [6] J. Shi, W. Lee, Y. Liu, Y. Yang and P. Wang, "Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines", *IEEE Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064-1069, 2012.
- [7] V. Vapnik, *The nature of statistical learning theory*. New York: Springer, 2010.
- [8] B. Scholkopf and A. Smola, *Learning with Kernels*. Cambridge: MIT Press.
- [9] J. Harmouche, C. Delpha and D. Diallo, "Incipient fault detection and diagnosis based on Kullback–Leibler divergence using Principal Component Analysis: Part I", *Signal Processing*, vol. 94, pp. 278-287, 2014.
- [10] T. Erven and P. Harremoës, "Rényi Divergence and Kullback–Leibler Divergence", *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797 - 3820, 2014.
- [11] S. Boyer. SCADA. Research Triangle Park, NC: International Society of Automation, 2010.
- [12] P. Tavner, *Offshore Wind Turbines*. Stevenage: The Institution of Engineering and Technology, 2012.
- [13] P. Qian, X. Ma and P. Cross, "Integrated data-driven model-based approach to condition monitoring of the wind turbine gearbox", *IET Renewable Power Generation*, vol. 11, no. 9, pp. 1177-1185, 2017.
- [14] M. Lydia, S. Kumar, A. Selvakumar and G. Prem Kumar, "A comprehensive review on wind turbine power curve modeling techniques", *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 452-460, 2014.