High Voltage

Research Article

IET Journals
The Institution of Engineering and Technology

RF-based location of partial discharge sources using received signal features

ISSN 2397-7264
Received on 23rd April 2018
Revised 16th August 2018
Accepted on 20th September 2018
doi: 10.1049/hve.2018.5027
www.ietdl.org

Ephraim T. Iorkyase[1] ✉, Christos Tachtatzis[1], Ian A. Glover[2], Robert C. Atkinson[1]
[1]*Department of Electronic and Electrical Engineering, University of Strathclyde, Royal College Building, 204 George Street, Glasgow G1 1XW, UK*
[2]*Department of Engineering and Technology, University of Huddersfield, HD1 3DH Huddersfield, UK*
✉ *E-mail: ephraim.iorkyase@strath.ac.uk*

Abstract: Partial discharges (PDs) are symptomatic of some localised defects in the insulation system of electrical equipment. PD activity emits electrical pulses in the form of radio frequency (RF) signals which can be captured using appropriate sensors. The analysis of the measured RF signals facilitates localisation of PD. This study investigates the plausibility of using purely RF received signal features of PD pulses to locate PD at low cost. A localisation approach based on the analysis of these features has been developed, with the assumption that PDs generate unique RF spatial patterns due to the complexities and nonlinearities of RF propagation. In this approach, two distinct frequency bands which hold different PD information are exploited. PD location features are extracted from the main PD signal and the two sub-band signals. Correlation-based feature selection (CFS) is employed for feature selection and dimensionality reduction. Experimental results show that PD location can be inferred from the features of the PD pulses. The application of CFS to PD data reduces the memory/computational demand and improves localisation accuracy.

1 Introduction

Partial discharge (PD) is a result of electrical insulation breakdown irrespective of the causal mechanism. PD activity is generally accepted as a predominant cause for further insulation degradation which may eventually lead to catastrophic failure of the electrical equipment with severe social and economic consequences [1, 2]. It is therefore imperative that PD is detected and located at its early stages to permit repair/replacement prior to total failure.

PD localisation methods are based on measurement and analysis of the PD pulses. The radio frequency (RF)-based methods exploit the characteristics of electromagnetic waves [3] emanating from the discharge site to infer its location [4–7]. Much of the work done in this area focused on the use of time difference of arrival (TDoA) [8], direction of arrival (DoA) [9] and received signal strength (RSS) [10] gathered at multiple receivers to triangulate the PD location. RF PD source location has been implemented using TDoA measurement [6, 11–13]. A number of spatially separated antennas are used to pick up the RF signals. It is assumed the signals diminish with distance from the origin of the PD. The received signals are cross-correlated to yield the TDoA of the PD signal at each antenna and these TDoA are used to estimate the location of the PD source. TDoA requires accurate synchronisation, making it computationally complex and energy hungry, hence not possible to deploy for continuous monitoring and localisation of PD. DoA measurement has also been successfully used to locate PD sources [14, 15]. In this method, directional antenna arrays are used to estimate the DoA of the RF signals that propagate from the discharge generating defect. The location of the PD source is then determined by solving the two intersecting lines defined by the DoAs. As earlier mentioned, DoA requires an array of directional antennas and relies on line of sight path. DoA brings extra computational cost. Both TDoA and DoA are seen as uneconomic because they require additional hardware. RSS triangulation approach requires detailed models of RF propagation and does not account for existing variation. It is practically impossible to implement in situations where the transmit power is unknown which is the case under study.

An alternative approach is to use empirical measurements of received RF signals to infer PD location. By recording a database of PD ‘signatures’ along with their known locations, a PD location can be determined by acquiring its signature and comparing it with the known signatures in the database. The pattern matching scheme is considered as a low cost, low complexity technique compared to those based on distance/angle estimation [16]. Instead of exploiting signal timing or direction, it relies on signal structure characteristics. It turns the multipath phenomenon to surprisingly good use: by combining the multipath pattern with other signal characteristics, it creates a signature unique to a given location.

We propose to deploy a network of low cost off-the-shelf commercial radio sensors in a matrix form. The low cost of the proposed solution allows a monitoring system to be permanently deployed and thus continuously monitor the substation in real time. In the proposed approach, sensor nodes emit an emulated PD signal in a round robin fashion. All other nodes monitor these emulated signals allowing the creation of a database with a map of spatial propagation characteristics across the substation. The database is interpolated and can be used to estimate the true location of PD. In this manner, our model of the radio environment is highly tuned to the individual characteristics of the substation site. Frequent retuning can be scheduled at regular intervals, such that any changes to substation topography can be incorporated into an updated model.

This work is aimed at investigating the plausibility of using purely RF received signals features of PD pulses to determine PD location. In our previous work [7], we obtained good results using ratios of RSS as PD features, however, by exploiting the frequency selective nature of multipath propagation, additional PD features can be extracted from RF signals to improve PD location error, and these additional features are included in this work. Furthermore, due to the high dimensionality of the features extracted in this work, correlation-based feature selection (CFS) algorithm [17, 18] is employed to select the best possible set of features that will be used in the localisation process to enhance accuracy while reducing computational cost. The obtained results are quite encouraging. With high probability, our system is able to estimate PD location of a single source to within a few metres of its actual location. This suggests that an autonomous and efficient substation-wide RF-based monitoring system can be built at low cost.

The remainder of this paper is structured as follows. Section 2 presents the experimental procedure and addresses the feature extraction procedure. In Section 3, a description of the extracted

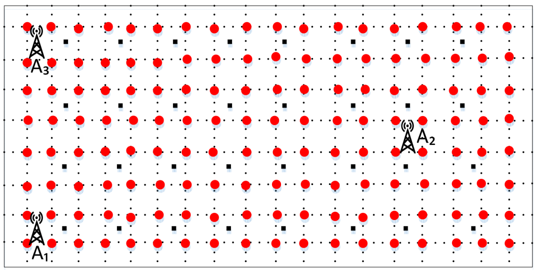

Fig. 1 *Grid for a measurement campaign*

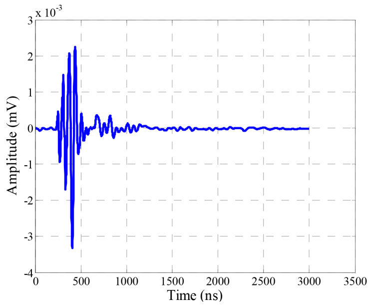

Fig. 2 *Recorded PD signal*

Table 1 PD time-domain features

S. No.	Feature parameter
1	impulse factor (IF)
2	variance (VAR)
3	root mean squared (RMS)
4	area under the waveform (A)
5	peak-to-peak value (PPV)

features as a function of location and localisation algorithm is presented. The experimental results are discussed in Section 4. Finally, the concluding remarks are stated in Section 5.

2 Experimental procedure

2.1 Data acquisition

The experiment was conducted in a 19.20 × 8.40 m^2 laboratory at the University of Strathclyde. The laboratory contained a great deal of clutter including metallic objects. Although we do not claim that the radio environment is representative of a typical substation (if there is such a thing), nonetheless the presence of clutter etc. gives rise to a complex, multipath-rich, propagation environment. A 1 m × 1 m grid was constructed in the floor of the laboratory (8 × 18 points), and measurements collected at each of the 144 grid intersection points which represent known training dataset. Another dataset was collected from 32 different points within the same laboratory referred to herein as test dataset and is used to test the performance of the localisation algorithm after training. The layout of the measurement space is as shown in Fig. 1. Pulse emulated PD sources were set up using a picosecond pulse generator and 20 PD pulses generated at each location in the grid. The RF measurements were made using three monopole antennas connected to a high-speed multichannel oscilloscope. The oscilloscope has a bandwidth of up to 9 GHz, however, for the purpose of this work, the received RF signals were sampled at 2 GSa/s. A sample waveform of the PD pulse is as shown in Fig. 2.

2.2 Feature generation and selection

In previous work [7], we have examined the potential of relative RSS (since PD will be of unknown magnitude) to infer location. In this work, we additionally exploit the knowledge that multipath effects will influence the delay spread (time duration) of a signal, and this, in turn, can provide information on the location of the source. Furthermore, multipath is highly frequency selective and so

Fig. 3 *Frequency response of the antenna*

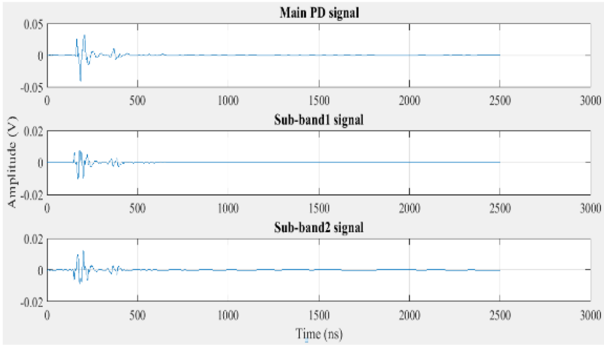

Fig. 4 *Bandpass filtered PD signals*

the ratio of specific frequency components will also yield useful information. The use of relative frequency measures will account for the different frequency components inherent in different types of PD. Each of these signal properties represents a feature. The features considered in this paper for PD location are as shown in Table 1. By way of explanation, the RMS feature represents the ratio of RMS values across receiving antenna pairs rather than absolute RMS values. These features are computed from PD pulse waveform. Intuitively, the derivation and use of these features for PD location are based on the fact that any discharge event produces RF signal that will propagate as travelling wave through the environment. As a result, the radiated signal amplitude and shape are modified by the propagation environment between the PD source and the receiving antennas due to path loss attenuation, signal shadowing and multipath effect. These effects may vary quite markedly within an electricity substation leading to the uniqueness of the signal to the location of measurement. The nature of the received PD pulse suggests that manipulating the amplitude and peaks of the signals can provide a rich database for inferring PD location. The received PD signal is not only a function of the source but also the measuring device. With the response of the measuring device known, more characteristics of the signals at defined intervals may be explored and used for localisation.

The PD data are collected using three identical monopole antennas. The frequency response of the antenna produces two peaks at 151 and 200.7 MHz, respectively, as shown in Fig. 3.

These peaks indicate the regions where most of the information about the discharge are captured. Based on the information from the frequency response, two bandpass filters are designed to filter the PD signals in these regions for more information. The centre frequencies of the filters corresponding to the peaks observed from the frequency response. The upper and the lower cut-off frequencies are 145.5 and 155.5 MHz for band 1 and 195.35 and 205.35 MHz for band 2; these correspond with the 3 dB point of each peak. This provides us with two more distinct PD sub-signals as shown in Fig. 4 for analysis.

Each of the observed radio characteristics for each antenna produces 15 features making it a total of 45 features for the three antennas. There is a trade-off between the number of features used as inputs to a machine learning algorithm and performance. A delicate balance must be achieved between too few features (insufficient information leading to under fitting) and too many features (the curse of dimensionality). Moreover, the various

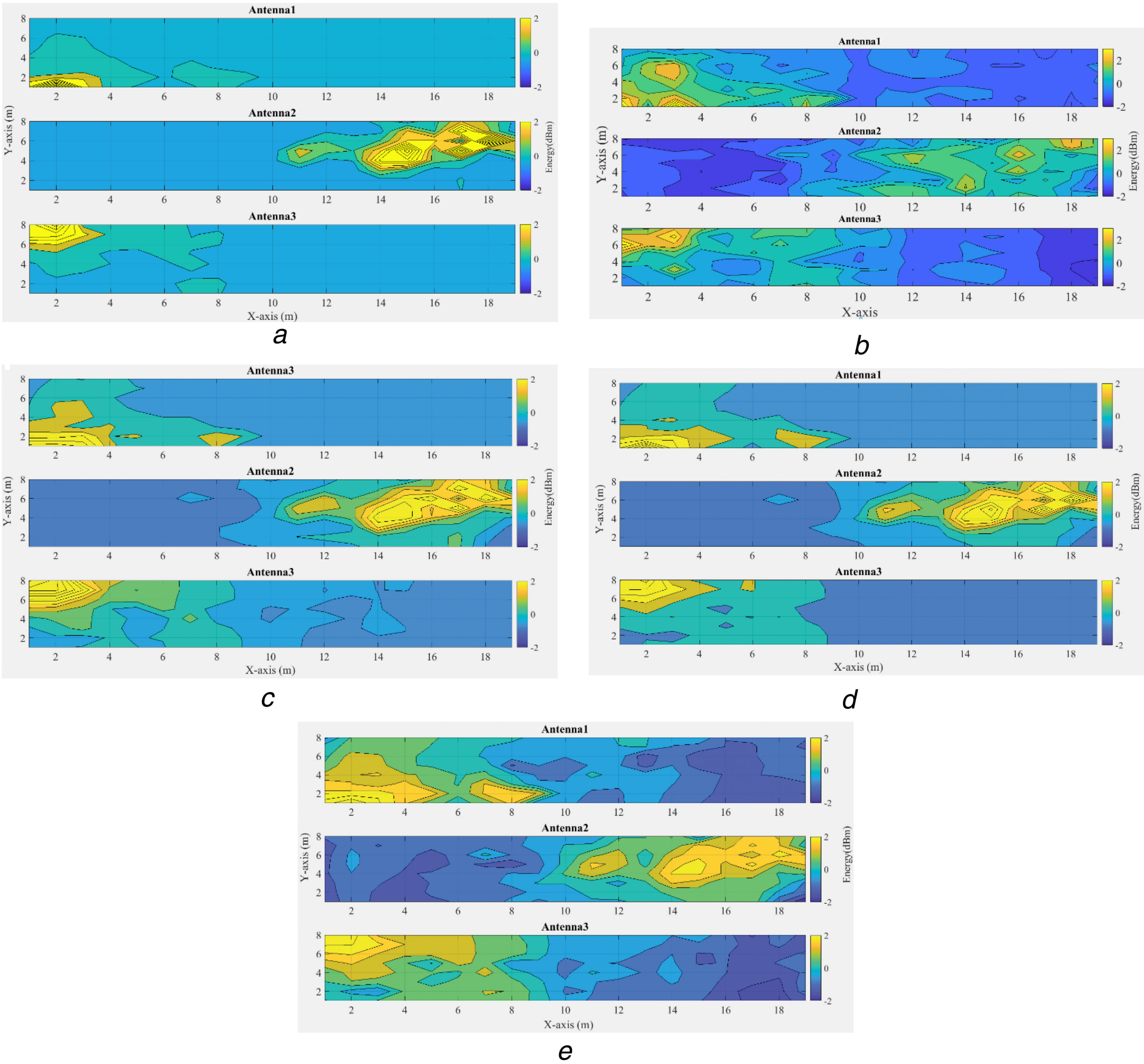

Fig. 5 *Spatial pattern for*
(a) Variance, ***(b)*** Impulse factor, ***(c)*** Peak-to-peak value, ***(d)*** RMS, ***(e)*** Area of received PD pulse

features are often correlated: two features may be highly correlated with the output, but also with each other, hence providing little additional information. Accordingly, we seek to transform our original features set to a potentially smaller set of decorrelated features using CFS algorithm. CFS is based on correlation which is a similarity measure between two random variables. Two variables are said to be linearly dependent when their correlation coefficient is ± 1, and uncorrelated when the correlation coefficient is 0. Generally, in CFS, the mutual correlations of all feature pairs are evaluated and the feature with the highest average absolute mutual correlation is removed at each iteration step of the algorithm [18]. However, in this paper, we develop a new feature selection criterion for the CFS algorithm. Instead of removing features with the highest average correlation, at each iteration step of the selection algorithm, we first pick features with higher correlation and then among them discard the feature with the highest average mutual correlation. For example, if two features have a correlation of 1 between them, we discard the feature with the highest average correlation and keep the other. When a feature is removed from the feature set, it is also discarded from the remaining average correlation. This continues until average absolute mutual correlation of all remaining feature is less than 0.4. The resulting feature matrix from the CFS algorithm is the optimal lower dimensional subset from the original feature set. This method is applied to the PD feature data for feature selection. The result brings the total number of features from the three antennas used for localisation from 45 to 9. The nine best features selected by the algorithm are the area under the curve of sub-band 1, sub-band 2 and the broadband signal on each of three antennas (i.e. each antenna produces three features in the two sub-bands and the main signal for a particular signal).

3 PD location based on extracted features

This work is based on the premise that the CFS selected features provide meaningful information for inferring PD location. To demonstrate that this is a reasonable premise, Figs. 5*a*–*e* show the maps of how the PD features (impulse factor, variance, root mean square, peak-to-peak value, and area under waveform) measured at each of the three receiving sensors vary with location.

Unsurprisingly, the values measured closer to a PD source are stronger than the ones received from a distance. The effect of multipath and signal distortions add to the unique signatures created at different locations. The spatial map shows a clear correlation between the features and distinctive PD locations. This unique spatial pattern is an indication that the selected features are informative and can be used to infer PD location by matching the patterns. Fig. 6 shows the architecture of the pattern matching technique for PD location. It should be noted that in this study the

PD events were generated with a pulse generator to ensure that the emulated PD traces are identical for each location and repetitive. In practical reality where different sources of PD generate different shapes and amplitudes, the absolute values of the features differ significantly across PD types. Robust PD location features can be reconstructed as ratios of the features on each antenna location.

3.1 K-nearest neighbour (KNN) regression

KNN algorithm is one of the best-known machine learning algorithms [19]. For PD location, KNN is considered as a regression problem which consists in mapping PD signal input features onto dual output representing PD location coordinates. The idea of KNN is based on the assumption of locality in feature space.

KNN training phase constitutes the creation of a database (radio map) with reference PD location patterns (features, location). In the estimation phase, the distance between the unknown PD and each stored neighbour is calculated in feature subspace using Euclidean distance metric and k nearest neighbours (those with the shortest distance) are taken into account. The unknown PD location coordinate is modelled as the average of the coordinates of the k nearest neighbours, thus

$$(\hat{x}, \hat{y}) = \frac{1}{k}\sum_{i=1}^{k}(x_i, y_i) \tag{1}$$

4 Experimental results and discussions

In order to verify the use of time-domain statistical features to infer PD location and the effectiveness of CFS in reducing dimensionality/enhancing accuracy, all the RF signal samples collected for each of the $144 \times 3 = 432$ combinations of PD location and sensors as described in Section 2 are used to train the pattern-matching algorithm. During the training phase, PD data from reference locations (known locations) are used and the algorithm is fed two inputs; the PD features and their corresponding location. Each (feature, location) pair constitute training data point. In the case of CFS-based technique, instead of the generated features, the selected features are used in the input. The training algorithm used to evaluate the performance of using time-domain features and the effect of CFS technique in inferring PD location is KNN algorithm described in Section 3.1. Any other pattern matching algorithm could be used, however, KNN was chosen for its direct applicability. The predicted location is returned as the average of the locations of the K nearest neighbours. The choice of k has a significant impact on the localisation performance of KNN and must be tuned appropriately. In this paper, the heuristic optimal number of nearest neighbours is obtained as $k = 6$ using cross-validation [20].

4.1 Enhancing location accuracy

The experimental results processed at 32 test locations before and after the application of CFS to the PD features are presented. The performance is evaluated in terms of distance error defined as the Euclidean distance between the estimated location and the true PD location. Fig. 7 shows the cumulative error distribution of the KNN model before and after the application of CFS. In comparison, the figure clearly shows that KNN after CFS (KNN-CFS) outperforms KNN without CFS. In terms of the mean distance estimation error, the PD location accuracy after applying CFS is 1.65 m compared to 2.60 m without CFS. More specifically, the application of CFS improves the location accuracy by 36.54, 37.71 and 47.17% in terms of mean error, 50th and 75th percentile of the estimated location error, respectively. The summary of results is given in Table 2. The results indicate that using CFS to reduce the dimension of the data by selecting uncorrelated PD features before learning the model for PD location is better than using the generated features directly.

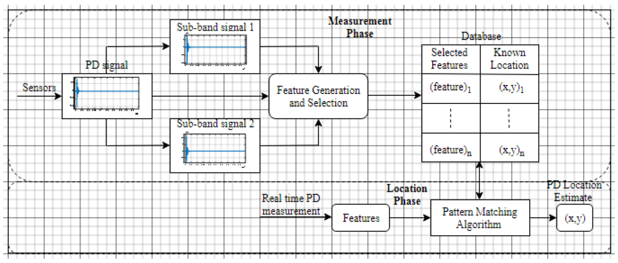

Fig. 6 *Architecture for PD pattern matching technique*

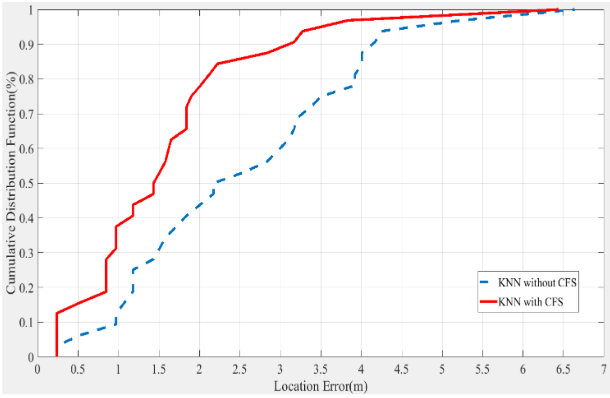

Fig. 7 *CDF of localisation error based on PD original features and CFS selected features*

Table 2 Model location accuracy before and after applying CFS

Method	Error in location estimate, m		
	Mean	50th	75th
KNN-original features	2.60	2.36	3.71
after applying CFS	1.65	1.47	1.96
CFS improvement, %	36.54	37.71	47.17

4.2 Computational complexity

The PD location system considered in this paper is centralised, which means each computation is done at the so-called central station. KNN algorithm used for location estimation stores all training samples in the radio map and search for nearest neighbours by comparing the distances between the test data and all training data. Suppose there are n training samples with dimension d, $O(d)$ is the complexity to compute the distance to one training sample, $O(nd)$ to compute distances to all training samples. In addition, there is a complexity of $O(nk)$ to find k closest neighbours. In total, KNN generates a complexity of $O(nk + nd)$. This can be costly for our application with limited resources. However, with the integration of simple, low complexity CFS algorithm, it can be seen from the results that the size of the data is greatly reduced while improving accuracy. This can be explained by the ability of CFS to intelligently select uncorrelated features into fewer dimensions for model learning. The initial dimension of the PD features space is reduced by 80% when CFS is applied. It can be concluded that applying CFS reduces the computational requirement of PD location determination in general.

5 Conclusions

A method for locating PD sources based on the application of CFS to received radio signal features has been described. By examining the frequency response of the antennas, two distinct frequency bands which hold different PD information (based on the knowledge that multipath propagation is frequency selective) were exploited for more PD features and integrated for fingerprinting. This provides a rich database for PD localisation via pattern matching algorithm. The proposed technique finds an effective way to select the most informative and decorrelated features of the generated PD data based on correlation theory such that the dimensionality of the feature space is reduced with improved accuracy/performance. KNN algorithm has been used to demonstrate the effectiveness of the proposed technique. Two separate models were developed and evaluated based on the original PD features and the selected features. The proposed technique based on CFS shows a considerable reduction in computation cost while offering a better accuracy performance. Experimental results indicate that the proposed method provides a significant improvement with the mean error reduced by 36.54% compared to using original features. It also offers an 80% reduction in computational load. This method can be extended to multiple PD type scenarios without any limitation by taking the ratios of the features on each antenna location. A possible future extension is to investigate some nonlinear approaches that can better extract more information from PD pulses for PD localisation.

6 References

[1] Boya, C., Rojas-Moreno, M.V., Ruiz-Llata, M., *et al.*: 'Location of partial discharges sources by means of blind source separation of UHF signals', *IEEE Trans. Dielectr. Electr. Insul.*, 2015, **22**, (4), pp. 2302–2310

[2] Portugues, I.E., Moore, P.J., Glover, I.A., *et al.*: 'RF-based partial discharge early warning system for air-insulated substations', *IEEE Trans. Power Deliv.*, 2009, **24**, (1), pp. 20–29

[3] Hiroki, T., Daiki, T., Yoshihiro, B., *et al.*: 'Finite-difference time-domain simulation of partial discharges in a gas insulated switchgear', *High Volt.*, 2016, **1**, (1), pp. 52–56

[4] Mitchell, S.D., Siegel, M., Beltle, M., *et al.*: 'Discrimination of partial discharge sources in the UHF domain', *IEEE Trans. Dielectr. Electr. Insul.*, 2016, **23**, (2), pp. 1068–1075

[5] Homaei, M., Moosavian, S.M., Illias, H.A.: 'Partial discharge localization in power transformers using neuro-fuzzy technique', *IEEE Trans. Power Deliv.*, 2014, **29**, (5), pp. 2066–2076

[6] Moore, P.J., Portugues, I.E., Glover, I.A.: 'Radiometric location of partial discharge sources on energized high-voltage plant', *IEEE Trans. Power Deliv.*, 2005, **20**, (3), pp. 2264–2272

[7] Iorkyase, E.T., Tachtatzis, C., Lazaridis, P., *et al.*: 'Radio location of partial discharge sources: a support vector regression approach', *IET Sci. Meas. Technol.*, 2018, **12**, (2), pp. 230–236

[8] Zeng, F., Tang, J., Huang, L., *et al.*: 'A semi-definite relaxation approach for partial discharge source location in transformers', *IEEE Trans. Dielectr. Electr. Insul.*, 2015, **22**, (2), pp. 1097–1103

[9] Hou, H., Sheng, G., Jiang, X.: 'Localization algorithm for the PD source in substation based on L-shaped antenna array signal processing', *IEEE Trans. Power Deliv.*, 2015, **30**, (1), pp. 472–479

[10] Iorkyase, E.T., Tachtatzis, C., Atkinson, R.C., *et al.*: 'Localisation of partial discharge sources using radio fingerprinting technique'. Loughborough Antennas Propagation Conf. (LAPC), Loughborough, UK, November 2015, pp. 1–5

[11] Sinaga, H.H., Phung, B.T., Blackburn, T.R.: 'Partial discharge localization in transformers using UHF detection method', *IEEE Trans. Dielectr. Electr. Insul.*, 2012, **19**, (6), pp. 1891–1900

[12] Robles, G., Fresno, J.M., Sánchez-Fernández, M., *et al.*: 'Antenna deployment for the localization of partial discharges in open-air substations', *Sensors*, 2016, **16**, (4), pp. 541–551

[13] Yang, Y., Tang, L., Hu, Y., *et al.*: 'A study on the mobile partial discharge locating system', *Energy Power Eng.*, 2013, **5**, (4), pp. 589–592

[14] Polak, F., Sikorski, W., Siodla, K.: 'Location of partial discharges sources using sensor arrays'. Int. Conf. on High Voltage Engineering and Application (ICHVE), Poznan, Poland, September 2014, pp. 1–4

[15] Ishimaru, H., Kawada, M.: 'Locating multiple partial discharge sources using MAP estimation and ray tracing', *IEEJ Trans. Electr. Electron. Eng.*, 2014, **9**, (1), pp. 1–7

[16] Qiu, L., Kennedy, R.A.: 'Radio location using pattern matching techniques in fixed wireless communication networks'. Int. Symp. on Communications and Information Technologies, Sydney, Australia, October 2007, pp. 1054–1059

[17] Haindl, M., Somol, P., Ververidis, D.: 'Feature selection based on mutual correlation'. Iberoamerican Congress on Pattern Recognition, Cancun, Mexico, November 2006, pp. 569–577

[18] Cui, Y., Jin, J.S., Zhang, S., *et al.*: 'Correlation-based feature selection and regression'. Pacific-Rim Conf. on Multimedia, Shanghai, China, September 2010, pp. 25–35

[19] Genming, D., Zhang, J., Zhang, L., *et al.*: 'Overview of received signal strength based fingerprinting localization in indoor wireless LAN environments'. IEEE 5th Int. Symp. on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE), Chengdu, China, October 2013, pp. 160–164

[20] Kohavi, R.: 'A study of cross-validation and bootstrap for accuracy estimation and model selection'. Int. Joint Conf. on Artificial Intelligence (IJCAI), Montreal, Quebec, Canada, 1995, vol. 14, no. 2, pp. 1137–1145