A Complex Classification Approach of Partial Discharges from Covered Conductors in Real Environment (preprint)

S. Mišák, J. Fulneček, T. Vantuch, T. Buriánek and T. Ježowicz
Centre ENET, VSB- TU Ostrava
17. listopadu 15
708 33 Ostrava, Czech Republic

ABSTRACT

Partial discharges are known as indicators of degradation of insulation systems. In case of medium voltage overhead lines with covered conductors, the internal partial discharges indicate the degradation of insulation system, rupture or downfall of the line. The reliability and selectivity of methods to detect internal partial discharges in the covered conductors are mainly given by the level of background noise. The background noise distorts the pattern of partial discharges (PD-pattern) and decreases the capability of detection methods to recognize the features of PD-pattern corresponding to the degradation of an insulation system. This paper presents the results from the testing of a developed and optimized model for detection of covered conductor faults. The ability to identify covered conductor faults was verified in the real environment.

Index Terms — **Power transmission lines, Partial discharges, Pattern recognition, Pattern classification, Electromagnetic radiative interference, Aluminum power conductors, Wavelet transforms**

1 INTRODUCTION

COVERED conductors (CCs) are gradually replacing Aluminium Core Steel Reinforced (ACSR) medium voltage (MV) overhead lines in a forested and broken terrain. The main reason for the use of the MV overhead lines with the CCs is their high reliability [1, 2]. The high reliability is given by the XLPE insulation system [3]; therefore, in case of contact of phases there is no interphases short-circuit. Similarly, the interphase short-circuit does not arise in case a tree branch falls on the CCs phases [1].However, the disadvantage of the CCs operation is the problematic detection of faults, when the covered conductor (CC) is ruptured due to (i) atmospheric overvoltage or (ii) due to destructive degradation of the CCs insulation system. The degradation of the insulation system is the consequence of the inhomogeneity of the electric field. The inhomogeneity of the electric field can be caused by the CC contact with other dielectrics (tree branch in general) [1, 4]. The CC fault detection is problematic because the value of the current is very small (c. 10^{-6} Amps) and the starting element of the standard digital relay protection cannot detect this value [2, 4,5].

Presently, some methods can detect CC faults, where the faults can be caused by: (i) the downfall of a tree branch on the CCs and the subsequent multi-contact of the tree branch with the CC, (ii) the contact of a tree branch with CCs, (iii) the downfall of the CC to the ground. All the methods detect the CC faults on the basis of evaluating the impulse component of the signal (voltage, current) measured in the fault site. The impulse component of the signal is generated by internal partial discharges (PDs) activity in the fault site [6 - 8]; it is visible in the high-frequency band (mostly 10^4 to 10^7 Hz). The current signal is measured by a Rogowski coil [3, 5, 9 - 11] and the voltage signal can be measured by a sensor (inductor) placed on the insulation system [4, 12]. An example of the measured voltage signal for the CC fault is shown inFigure 1.

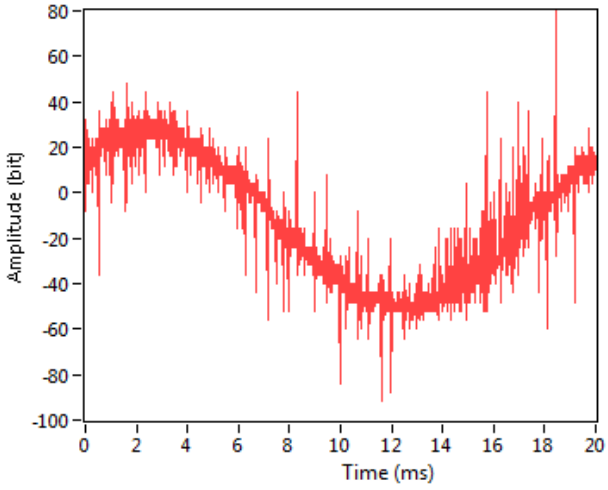

Figure 1: The measured signal during CC fault

The time course of the impulse component is generally named as a PD-pattern (partial discharge pattern) [5, 11]. The

methods of evaluating the PD-pattern from the voltage signal use a simple sensor (single layer inductor) to measure the PD-pattern [4, 12].

The issue of PD-pattern detection is a subject of many areas of study, artificial intelligence, signal processing, data analysis, statistic or applied maths. Fourier transformation is a frequently used method of feature extraction and analysis in the application of signal processing. Fourier transformation decomposes the signal's spectra across the frequency domain [13]. Another signal processing method is the wavelet transformation. The wavelet transformation [14] is able to perform a time-frequency analysis [15]. The models applied for final processing of the extracted features of a PD-pattern are mostly based on machine learning algorithms, which are reviewed in several studies [12, 13], where the mentioned experiments faced more or less one common problem. It is the presence of external background noise interference which degrades the quality of the observed data. This phenomenon directly implies a lower quality of the detection model [16].

The background noise is a form of noise pollution or interference of signals generated by another device than analysed, e.g. radio waves in case of CCs. When the analysed PD-pattern is distorted by the background noise disturbing signals, the impulse component consists of relevant and irrelevant (false hits) peaks.

Most of the studies attempt to simulate the real conditions by adding the background noise into the signal data and then by making use of compute-intensive approaches, they attempt to classify the PD-pattern properly. The goal of our contribution is significantly different. This experiment handles signal data acquired from the natural environment with high noise interference and low sampling rate, according to the placing of the sensor device that is also reflecting the necessity of the lowest possible price. The PD-patter detection needs to operate online with reasonably low computational power. Therefore, our motivation is to review and apply only the most relevant approaches for noise suppression, feature extraction and classification with possible soft-computing driven optimization. This contribution covers the design and evaluation of the entire PD-pattern detector as well as the relevancy of the most frequently known PD-pattern features that were applied. In case of success, the possible deployment as a PD-pattern detector can increase the reliability of overhead power lines with CC.

2 EXPERIMENT DESCRIPTION

This experiment can be described as a set of clearly defined steps, where each of them has its own purpose (see Figure 2). The starting point in the process is the platform for data acquisition (1), which collects raw signals measured in real environment (Section 3.1). These signals are later classified by an expert who adds the class labels to each signal (annotation). Most of the signals in the obtained database were labelled as failure-free signals because CC failures are extremely rare events. This requires us to solve the issue of an imbalanced dataset (2) (Section 3.3).

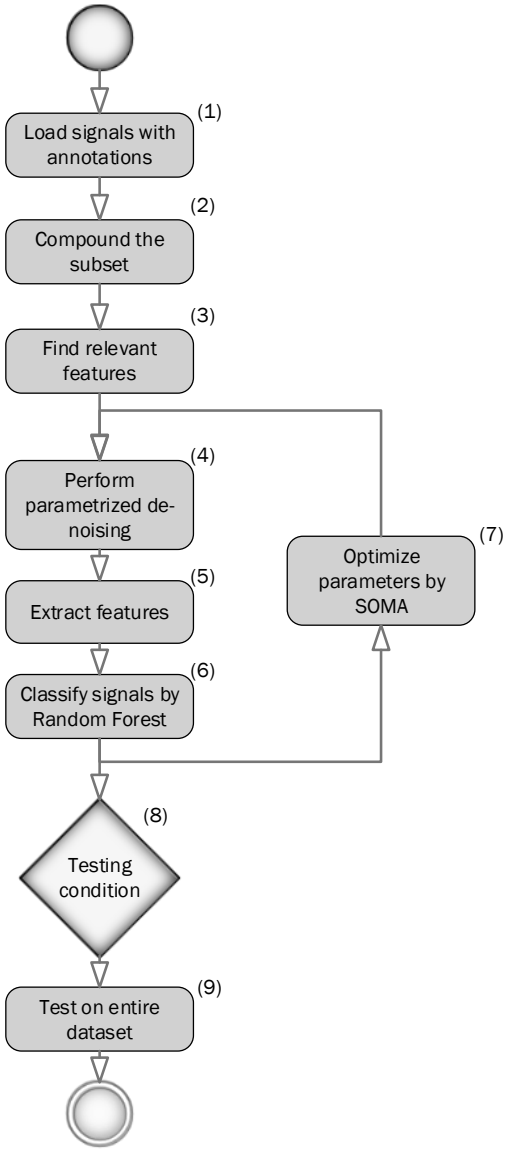

Figure 2: Diagram of experiment's workflow

The next step was to compute a set of features for best possible classification (3). The relevancy of the features was evaluated by their mutual information according to their class labels (Section 3.4). An increase in the relevancy of the applied features may lead to an increase in the model's total performance. This was the motivation behind the application of parameterized de-noising (4). The quality of a compounded feature matrix (5) depends on the previous step (4) and on the performance of the classification (6) (see Section 5.1). The final experiment in this paper aims to increase the total performance by optimizing the de-noising parameters by Self-Organizing Migrating Algorithm (7)

(Section 5.2). The last section contains a description of the settings, testing conditions and results of the experiment.

2.1 PLATFORM DESCRIPTION

In this experiment, a single layer inductor is used as a voltage sensor. It is wound right on the surface of CC. This inductor is connected to a capacity divider. Raw signal from the divider is amplified and sampled. Detailed description of the platform can be found in [4].

The measurement is performed once per hour and it results in three data files, each for the respective phase. Each file contains 400,000 samples/bytes with 20MS/s sampling rate. The data are loaded by the given set of web services for further analysis and processing.

In laboratory conditions, PD measurements are normally carried out with narrow band detection instruments. In our experiment, wideband detection was used. Wideband detection provides better resolution of a pulse [17]. With 20 MS/s sampling rate, frequency band up to 10 MHz is examined.

2.2 SIGNAL CHARACTER

As it was mentioned previously, the raw signals from a real MV overhead line contain high amount of uncertain information. The reason is the presence of a unique background noise that is almost impossible to simulate in a laboratory. Besides the PD-pattern itself, every other signal in the impulse component of a raw signal is considered a background noise. According to [18], there are several sources of background noise:

1. Discrete spectral interference DSI (radio emissions).
2. Repetitive pulses interference (power electronics).
3. Random pulses interference RPI (lightning, switching operations, corona).
4. Ambient and amplifier noise.

The overhead line acts as a long wire antenna, so long-wave transmitters are the most significant permanent source of the DSI. Sometimes, more than 100 radio stations can be identified in the raw signal. Most of them are used for radio broadcasting, some for other purposes (time signals, communication). These sources of DSI can be easily recognizable with FFT based on their modulation [19]. Broadcasting of the radio waves is variable and depends on many factors [20]. That is why some of the radio stations are not present in the measured raw signal all day long or their signal is variable (note that not all of them transmit 24/7). Usually, DSI in the examined band is more significant during the night because of the ionosphere condition. A high level of DSI can cover the PD-pattern in a raw signal.

Another significant source of interference on MV overhead lines with CC is RPI, which is most often represented by a corona discharge. Generally, RPI creates false hit peaks in a time domain of a raw signal, that may be mistaken as a PD-pattern, seeFigure 3.

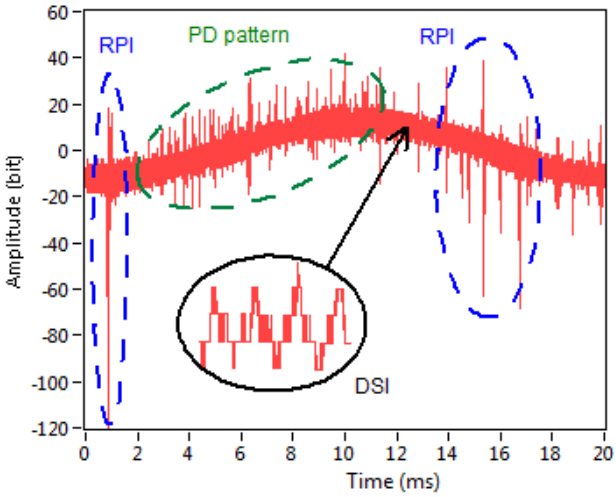

Figure 3: PD-pattern with noise during CC fault

False hit peaks can be identified in a raw signal according to their position, shape, amplitude and periodicity. The biggest permanent source of DSI on the measured site was radio transmitter “Solec Kujawski”. Its carrier wave (225 kHz) is clearly visible in almost all of acquired signals (see DSI detail in Figure 3). This transmitter with output power of 1 MW is situated 350 km from measured site

2.3 SUBSET SELECTION

The entire dataset contains more than 10,000 classified raw signals, which came from a single location. Because PD activity on MV overhead lines was not appearing on a daily basis, the number of raw signals containing PD-pattern presenting degradation of CC insulation system was significantly smaller than the rest of the signals (there were only 96 PD-pattern containing raw signals). This condition (imbalanced dataset) leads the machine learning algorithm to converge into a solution, where all the signals are marked by the dominant class label and other classes are ignored. The imbalanced dataset problem was addressed in some previous studies [21, 22] and it can be solved by using a proper design of a representative subset (over-sampling or under-sampling method [21]), which is later applied as the training dataset during the experiment.

In this paper, the simple under-sampling method was applied for subset construction. The subset should contain all the kinds of class labels equally distributed because there are various signals in both classes with various amounts of background noise. The background noise can cover the PD-pattern or form some false hit peaks, which should also be recognized by final processing. This phenomenon has to be reasonably represented in the chosen subset.

2.4 FEATURE’S EXTRACTION

In case of discrete variables, MI is defined as the joint probability ($p(x,y)$) and marginal probabilities ($p(x),p(y)$) of X and Y.

$$I(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)} \quad (1)$$

In case of continuous variables, when the probability distribution function (PDF) is unknown, the MI estimation is not an easy task and it can be performed in different ways.In this paper, we applied the Kraskov estimation approach [26].

Table 1: Features' relevancies

Feature name	Relevancy (MI)
Mean value	0.3787
Standard deviation	0.0516
Skewness	0.0483
Kurtosis	0.0311
Entropy of signal	0.0708
Energy of decomposition	0.0708
Entropy of detail coeffs.	0.0721
Fractal dimension	0.2820
Number of peaks	**1.7676**
Mean width of peaks	**1.8311**
Mean height of peaks	**0.7818**
Max. width of peaks	**1.9384**
Max. height of peaks	**0.9907**
Min. width of peaks	**1.8822**
Min. height of peaks	**0.9474**

As it can be seen in Table 1, the most relevant feature for classification of our dataset is the number of peaks and the features derived from the detected peaks. These features will be described, implemented, parameterized and optimized in the following sections. The rest of the features were not applied in later experiments because of their lower MI value.

3 NOISE AND FALSE-PEAK SUPPRESSION (NFS)

According to the presence of previously mentioned background noise interference, our proposed model is meant to highlight most of the relevant known features and cancel the highest possible amount of false hit peaks (noise interference) to bring the best possible performance. Its workflow is shown in Figure 4. It should be modular to be maintained and extended as easily as possible and it should also be adjustable by as many reasonable parameters as possible. These parameters will be discussed in the following sections. Those parameters are available to be expertly adjusted and there is of course the possibility of optimization based on soft computing approach as well.

3.1 UNIVARIATE WAVELET DE-NOSISNG AND PEAKS EXTRACTION

The raw signal from the input is processed to suppress the major part of the noise. The basic univariate wavelet de-noising is the operation applied for this purpose as it was applied in many previous studies [27, 28].The applied mother wavelet and the number of the decomposition levels are in Table 2.

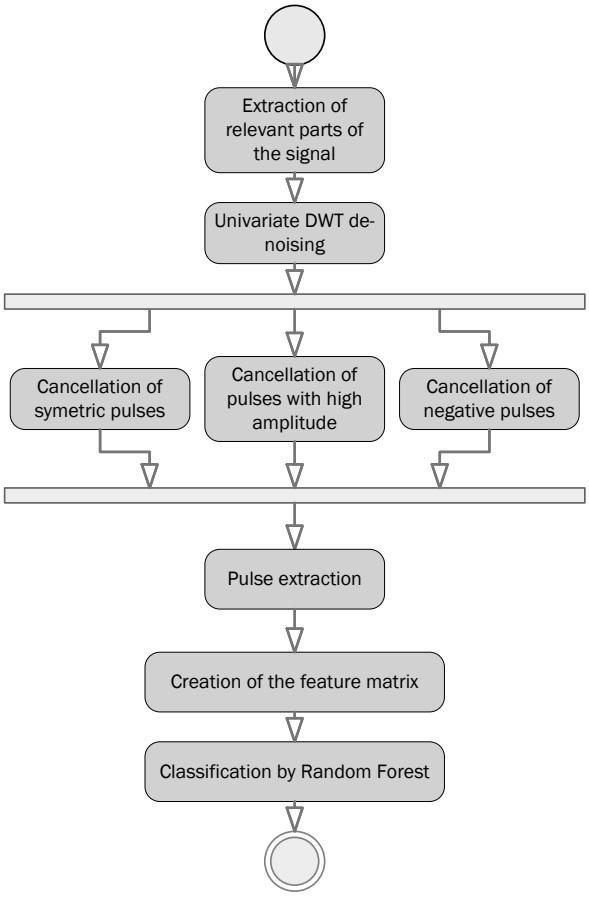

Figure 4: Diagram of the NFS algorithm workflow

Applying hard thresholding, the major amount of irrelevant small peaks is suppressed and the most significant peaks remain for further analyses.

Next, the preserved peaks are examined. Each peak is described with its starting index, its amplitude and width. Width of the peak is defined as a number of observations (samples) from the beginning of the peak to its end. Amplitude of the peak is a data point between the start and end of the peak, where its absolute value is maximal.

3.2 CANCELLATION OF FALSE HIT PEAKS

During data processing, it became clear that a statistically significant part of false hit peaks reaches much higher amplitude than the PD-pattern. A false hit peak is also very often followed by another one with the opposite polarity, creating a symmetric pair. We used this knowledge to cancel the false hit peaks. A great source of false hit peaks in the analysed raw signals was corona discharges. These corona discharges create typical “pulse trains” which can be easily recognizable in the time domain of a signal.

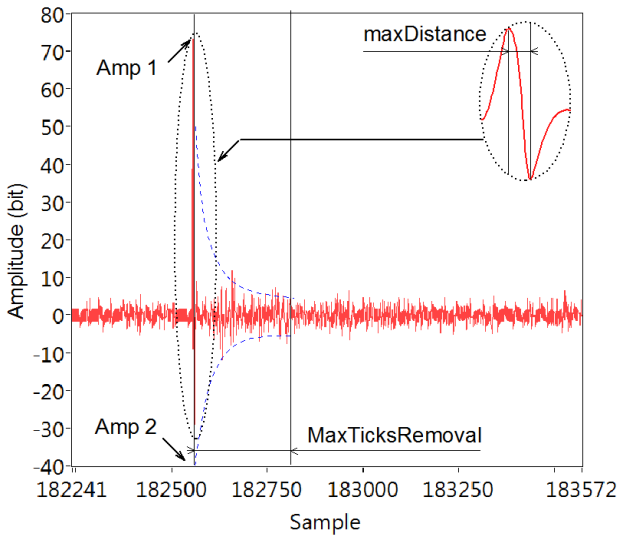

Figure 5: Corona discharge peak detail

The collection of peaks extracted from the previous step is analysed to find and remove the symmetric peak pairs. In this process, each peak is compared to the next peak. If their distance in the signal is under a defined limit *maxDistance*, then a further analysis is performed. When the signs of two consecutive peaks are opposite and the ratio of their amplitudes is higher than the defined limit maxHeightRatio, the peak pair is considered symmetric. A symmetric pair of peaks is very often followed by dumped oscillations, as it can be seen in Figure 5in case of the corona discharge. When the first symmetric pair is removed from the signal, the following dumped oscillations can be misdetected as a PD-pattern. The length of this oscillation cannot be predicted because of the variable dump factor. Therefore, all the peaks in a defined distance (*maxTicksRemoval*) behind the symmetric pair are also cancelled. In another step, all the peaks with higher amplitude than the defined limit *maxHeight* are removed from the extracted set of peaks.

3.3 SELECTION OF RELEVANT AREAS IN RAW SIGNALS

According to many studies of PDs detection [29, 30, 31], the appearance of peaks is not random. These peaks are mostly clustered in specific subparts of the sinusoidal signal. The correct placing of the peaks is variable and it depends on several conditions [31].

We decided to cover this effect by dividing the processed signal into four parts of the same length (100000 samples). The first part of the signal is ignored. The second and the last part are considered as relevant. The third part (irrelevant) is also taken into account, because hypothetically a high difference on peak-based features between the relevant and irrelevant parts can imply the occurrence of a PD pattern and on the other hand, a high number of peaks equally distributed on all the parts imply peaks originated from noise.

All the previously mentioned features were computed on all of the three parts of the processed signal and their values form a classification matrix containing 28 feature columns and one class column for each of the signal. All the features were derived from the extracted peaks (number of positive peaks, number of negative peaks, max/min width of peaks, max/min amplitude of peaks, mean value of width/amplitude of peaks) and four (one for each sinusoidal part and one for all of them) added columns containing standard deviation of histograms of peaks positions, widths, and amplitudes. These last histogram-based features should represent the character of peak's distributions on the signal, which experimentally appeared as valuable information.

4 CLASSIFICATION

The matrix of computed features (section 3.3) is applied as an input dataset for the step which follows. The motivation behind this step is to use the machine learning (ML) algorithm to create a mapping strategy for the recognition of signals containing PD-patterns from all the measured signals. The final performance of the applied ML algorithm will be able to confirm or reject the suitability of NFS application. For this purpose, Random Forest (RF) ML algorithm was applied.

Random Forest is an ML based algorithm with a wide range of applications, fast learning ability comparing to neural networks and minimal presence of over-fitting.

The performance of the ML part depends on the quality of the obtained dataset. We used this dependency by application of expert settings on all of the NFS parameters and to move the performance on the higher level, we applied the optimization algorithm based on artificial intelligence technique. This algorithm is called Self-organized migrating optimization (SOMA) and through various applications it is a suitable choice for such task.

Table 2: NFS parameters with their expert's adjustment and possible ranges of SOMA optimization

Name of the parameter	Experts setting	SOMA's range
maxDistance (ticks)	10	<4,10>
maxHeightRatio (%)	0.25	<0.05,0.5>
maxHeight (%)	100	<80, 140>
maxTicksRemoval	500	<50,500>
Threshold coef.	1	(0,5>
Mother wavelet	db4	all members of the wavelet families
Level of decomposition	1	{1,…6}

Table 2 summarizes all the NFS parameters (mentioned in previous section) with their expert's adjustment and defined ranges for SOMA's optimizations.The RF algorithm

and the SOMA optimization are described in following sections.

4.1 RANDOM FOREST

Random Decision Forest was firstly published by Ho in 1995 [24]. This model is also widely applied in many ML studies [25, 32]. The learning ability of the RF is supported by application of a Decision Tree (DT) model which is able to learn every specific aspect from the training set, which leads to decreasing the bias, but increasing the variance.

The core idea of the RF, which makes it so useful, is the application of boosting and bagging technique. Boosting is the application of an ensemble of learned classifiers (DTs), where each of them is trained on different parts of the dataset. This is supposed to provide the learned algorithms to be uncorrelated as much as possible, because the final decision of the RF comes from voting of this ensemble. The second aspect of the training ability is the bagging principle (boosted aggregating), which simply multiplies samples of the dataset which are hard to learn.

Those techniques help to increase the total performance of the model and decrease the over-fitting in case of application on untrained samples.

4.2 OPTIMIZATION BY SELF-ORGANIZING MIGRATING ALGORITHM (SOMA)

This optimization algorithm is inspired by social interactions inside a group of individuals and it was proposed by Zelinka [24] in 2004. Each of the individuals (candidates) is defined by a vector of coordinates. These coordinates represent constants of the model that has to be optimized. During the initialization of the algorithm, all the candidates are randomly distributed over the entire search space (all of the constants are chosen randomly). The algorithm goes through iterations and during each of them, a position of all the candidates is evaluated by the given fitness function. The leader of the population is selected by the highest obtained performance (fitness value).

During the next iteration, all non-leader candidates modify their position according to their leader. This is repeated in a given number of loops or until the stopping criteria is reached (threshold of error, etc.).

This is the main idea how the algorithm effectively achieves the position close to the global extreme of the problem. More details of the algorithm are described in [22, 23]. Concrete settings and results will be described in the following section.

5 SETTINGS AND RESULTS

The algorithms were developed, optimized and tested in the Matlab programming language. The setting of the RF and SOMA can be seen in Table 3.

Table 3: Random Forest and SOMA setting

Random Forest	
Learning algorithm	Decision Tree
Number of algorithms	200
Applied variables per tree	All of them
Applied samples per tree	All of them
Splitting criteria	Information Gain
SOMA	
Number of particles	15
Number of migrations	100
Length of path	0.09
Step size	0.02
Pertubation	0.25

5.1 TESTING METHOD

The previously mentioned modules of NFS were applied in different combinations, which lead to several variations of the classification model. Those variations were tested by Cross-validation (CV) [33]. The results of the detection problem are represented by four statistical criteria -accuracy, precision, recall and f-score [34].

5.2 RESULTS

Table 4 contains the testing parameters measured on a chosen subset (green) and parameters measured on the entire dataset (blue).

Table 4: NFS results (RA: Relevant areas feature, SP: Symmetric peaks cancellation, HA: High amplitude peaks cancellation, ND: No de-noising used)

Results (%)	**Expert settings**				**SOMA**	**ND**
	RA	**RA SP**	**RA HA**	**RA SP HA**	**RA SP HA**	
Accuracy	85.8	89.7	86.1	89.3	93.2	84.8
Precision	71.9	84.3	72.7	84.1	95.2	65.7
Recall	43.6	57.7	46.1	55.6	68.4	43.4
F-score	53.2	67.8	55.3	66.1	79.1	51.4
Accuracy	99.9	99.9	99.9	99.9	99.8	98.8
Precision	70.7	81.6	72.7	80	83.6	57.8
Recall	43.7	58.1	46.1	58.1	66.7	37.4
F-score	52.7	66.7	55.2	65.7	72.8	44.2

This experiment revealed that PD detection performed significantly better when the method was focused on sine's relevant parts, which could potentially contain a PD-pattern. On the other hand, the performance of the detection dropped in case of processing the irrelevant parts or even the entire period of the raw signal.

With experts setting, the best results were received when the "relevant areas" (RA) feature was combined with "symmetric peaks" (SP) cancellation. When 'high amplitude' (HA) peaks cancellation was used the results were slightly worse. This was caused by imperfection of the expert setting and result was later significantly improved by SOMA algorithm.

All of the combinations of applied modules for

parameterized de-noising obtain better performance compared to the model without de-noising (ND) which proves the requirement of proper handling of noise interference. The SOMA was able to provide the model with best performance by iterative optimization.

6 CONCLUSION

The goal of our work is to develop a simple and cheap tool for online PD detection. Such a tool should help to increase the reliability of overhead power lines with CC. For this purpose, a special measuring platform was designed. To get the price of the platform as low as possible, data acquisition system with sampling rate only 20 MS/s was used.
A dataset of raw signals which had been collected from the MV overhead line in real environment was created. For proper PD-pattern detection, it is necessary to de-noise this raw signal before its classification. We were able to extend the standard de-noising method with help of fundamental knowledge of the noise.

De-noised signals were classified by Random Forest algorithm. Random Forest as a representative of machine learning algorithms with boosting application was a valuable part of the model because of its great performance, stability and minimal presence of over-fitting. The optimization of SOMA significantly increases the total performance, but on the other side, the SOMA optimization can cause a small occurrence of over-fitting. This problem will be addressed in our future work.

Finally, the process for background noise elimination in raw signals was verified within real experiments in the MV overhead lines with CC and the total performance of the PD-activity detection was increased. The described experiment is unique in two ways:

1. Data from real environment were used; these data were acquired at relatively low sampling rate.
2. An extended de-noising method was used, developed specially for this purpose.

When this de-noising method is used, result of the experiment shows significant increase in total performance of detection.This total performance was defined by selected parameters, such as *accuracy*, *precision* and *recall.* Precision is the parameter with the highest relevancy. This significance of the precision module is given by request of the MV overhead lines operator on the reduction of the number of false CC faults detection. With proposed de-noising method, the detection of PD activity with a described low-cost platform is possible.

Our future work aims into computation of a higher amount of relevant signals' features to improve the detection performance and decrease over-fitting.

ACKNOWLEDGMENT

This research was conducted within the framework of the project LO1404: Sustainable development of ENET Centre, Students Grant Competition project reg. no. SP2015/170, SP2015/178, project LE13011 Creation of a PROGRES 3 Consortium Office to Support Cross-Border Cooperation (CZ.1.07/2.3.00/20.0075), project TACR: TH01020426. Next, we would like to thank for language corrections to Ms. Alena Kasparkova, Ph.D.

REFERENCES

[1] P. Pakonen, "Detection of Incipient Tree Faults on High Voltage Covered Conductor Lines", Ph.D. dissertation, Dept. El. Eng., University of Tampere. Tampere, 2007.

[2] G. Hasmi, M. Lehtonen and M. Nordman, "Modeling and experimental verification of on-line PD detection in MV covered conductor overhead networks", IEEE Transaction on Dielectrics and Electrical Insulation, vol. 17, no. 1, pp. 167-180, 2010.

[3] S. Dabbak, H. Illias, A. B. Chin and M. A. Tunio, "Surface Discharge Characteristics on HDPE, LDPE and PP", Applied Mechanics and Materials, Vol. 785, pp. 383-387, 2015.

[4] S. Misak and V. Pokorny, "Testing of a Covered Conductor's Fault Detectors", IEEE Transaction on Power delivery, vol. 30, no. 3, pp. 1096-1103, 2015.

[5] R. Bartnikas, "Partial discharges. Their Mechanism, detection and measurement", IEEE Transaction on Dielectrics and Electrical Insulation, vol. 9, no. 5, pp. 763-808, 2002.

[6] H. A. Illias, M. A. Tunio, A. H. A. Bakar, H. Mokhlis and G. Chen, "Partial discharge phenomena within an artificial void in cable insulation geometry: experimental validation and simulation", IEEE Transaction on Dielectrics and Electrical Insulation, vol. 23, no. 1, pp. 451-459, 2016.

[7] F. Alvarez, J. Ortego, F. Garnacho and M. A. Sanchez-Uran, "A clustering technique for partial discharge and noise sources identification in power cables by means of waveform parameters", IEEE Transaction on Dielectrics and Electrical Insulation, vol. 23, no. 1, pp. 469-481, 2016.

[8] M. S. Abd Rahman, P. L. Levin, and P. Rapisarda, "Autonomous localization of partial discharge sourceswithin large transformer windings IEEE Transaction on Dielectrics and Electrical Insulation, vol. 23, no. 2, pp. 1088-1098, 2016.

[9] E. Hemmati and S. Shahrtash, "Evaluation of unshielded Rogowski coil for measuring partial discharges signal", in Proc. 2012 International Conference Environment and Electrical Engineering (IEEE), pp. 434-439.

[10] M. Safiq, G. A. Hussain, L. Kutt and M. Lehtonen, "Effects of geometrical parameters on high frequency performance of Rogowski coil for partial discharge measurements", Measurement, vol. 49, no. 1, pp. 126-137, 2013.

[11] G. Stone, "Partial discharge diagnostics and electrical equipment insulation condition assessment", IEEE Transaction on Dielectrics and Electrical Insulation, vol. 12, no. 5, pp. 891-904, 2005.

[12] N. Sahoo, M. Salama and R. Bartnikas, "Trends in partial discharge pattern classification: a survey", IEEE Transaction on Dielectrics and Electrical Insulation, vol. 12, no. 2, pp. 248-264, 2015.

[13] W. J. K. Raymond, H.A. Illias, A. H. A. Bakar and H. Mokhlis, "Partial discharge classifications: Review of recent progress", Measurement, vol. 68, pp. 164-181,. 2015.

[14] T. Hucker and H. G. Krantz, "Requirements of automated PD diagnosis systems for fault identification in noisy conditions", IEEE Transaction on Dielectrics and Electrical Insulation, vol. 2, no. 4, pp. 544-556, 1995.

[15] C. K. Chui, *An introduction to wavelets*, San Diego: Academic Press, , 1992.

[16] J. Jun, "Noise reduction and source recognition of partial discharge signals in gas-insulated substation" Ph.D. dissertation, Dept. El. Eng, National University of Singapore, Singapore 2005.

[17] M. Pompili, C. Mazzeti and R. Batnikas, " Simultaneous ultrawide and narrowband detection of PD pulses in dielectric liquids", IEEE Transaction on Dielectrics and Electrical Insulation,, vol. 5, no.3, pp 402-407, 1998

[18] H. Zhang, T. R. Blacjbuirn, B. T. Phung and D. Sen, "A novel wavelet transforms technique for on-line partial discharge measurement", IEEE Transaction on Dielectrics and Electrical Insulation, vol. 14, no.1, pp. 3-14,2007.

[19] European Communication Committee, *The European table of frequency allocations and applications in the frequency range 8.3 kHz to 3000 GHz (ECA table)*, ECO 2015. Available at: http://www.erodocdb.dk/docs/doc98/official/pdf/ERCRep025.pdf (visited 15.1.2016)

[20] L. W. Barclay, "*Propagation of Radio Waves (2nd edition)*", London: IET, 2003.

[21] H. Julio, J. A. Carrasco-Ochoa and J. F. Martinez-Trinidad, "An empirical study of oversampling and undersampling for instance selection methods on imbalance datasets", in Proc. 2013 Iberoamerican Congress, pp 262-269.

[22] V. Ganganwar, "An overview of classification algorithms for imbalanced datasets", *International* Journal of Emerging Technology and Advanced Engineering, vol. 2, no. 4, pp. 42-47, 2012.

[23] I. Zelinka, "SOMA – self-organizing migrating algorithm", *New Optimization Techniques in Engineering*, pp. 167-217. Berlin: Springer Heidelberg, 2004

[24] T. K. Ho, "Random decision forest" in Proc.. 1995 Document Analysis and Recognition International Conference. pp 278-283.

[25] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R.P. Sheridan and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and QSAR modeling", Journal of Chemical Information and Computer Sciences, vol. 43, no. 6, pp 1947-1958, 2003.

[26] A. Kraskov, H. Stogbauer and P. Grassberger, "Estimating mutual information", Physical Review E, vol. 69, no. 6, 2004.

[27] D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*, Cambridge: University press, 2006

[28] B. Vidakovic, *Statistical modeling by wavelets*, New York: John Wiley & Sons, 2009.

[29] M. G. Niasar, "Partial Discharge Signature of Defects in Insulation Systems Consisting of Oil and Oil-impregnated paper", Licentiate thesis, University of Stockholm 2012.

[30] H. Illias, T. S. Yuan, A. H. A. Bakar, H. Mokhlis, G. Chen and P. L. Lewin, "Partial Discharge Patterns in High Voltage Insulation", in Proc 2012 International Conference on Power and Energy, pp. 750-755.

[31] R. V. Brunt, "Stochastic properties of Partial-discharge Phenomena", IEEE Transactions on Electrical Insulation, vol. 26, no. 5, pp. 902-948, 1991.

[32] A. Prinzie and D. V. Poel, "Random forests for multiclass classification and QSAR modeling", Expert Systems with Applications, vol. 34, no. 3, pp. 1721-1732, 2008.

[33] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection", in Proc. 1995 International Joint Conference on Artificial Intelligence, vol. 14, pp. 1137–1145.

[34] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation", Journal of Machine Learning Technologies, Vol. 2, No. 1, pp. 37-63, 2011.