Classification of Field Leakage Current Waveforms using Genetic Algorithms and an Euclidian Classifier

D. Pylarinos, K. Theofilatos, K. Siderakis, E. Pyrgioti, T. Papazoglou, I. Vitellas, E. Thalassinakis

***Abstract*-- The shape of the leakage current waveform is strongly correlated to surface activity, and thus to surface condition, of high voltage insulators. Field monitoring is necessary to acquire an exact image of experienced activity and insulators' performance. However, in the field, peak and charge related values, and not the actual waveform, are commonly recorded and investigated, but a fully representative value of the waveforms' shape is yet to be determined. In this paper, a novel approach is proposed for the classification of field leakage current waveforms that portray discharges. Two classes are defined based on the duration of discharges. Twenty features are extracted, from both time and frequency domain and used as a pattern. The selection of features was based on the literature. Two simple classifiers are employed, one based on genetic algorithms and one based on the Euclidian distance and results are presented.**

***Index Terms*—insulator; discharges; classification; Euclidian; field; genetic algorithm; leakage current; waveform;**

I. INTRODUCTION

The performance of outdoor high voltage insulators is linked with several parameters related to the location, with the pollution phenomenon probably being the most significant problem regarding [1-3]. Leakage current (LC) monitoring is a technique commonly applied in order to monitor surface activity and thus surface condition and overall performance of insulators [4]. Field measurements portray an exact image of the performance under service conditions and such measurements are recorded either on insulators that are part of the network or on specially designed testing stations [5]. However, interpretation of such measurements is a major task [4]. In the field, several values, most commonly the peak value and charge, are extracted and recorded instead of the actual waveform [4]. However, further investigation of field waveforms has shown that field noise [6-7] and the complexity of field LC waveforms [8-9] may lead to misleading results when such values are employed. Therefore, the shape of field waveforms should be considered and classification techniques should be applied to field waveforms in order to maximize the efficiency of field monitoring. In this paper, field LC waveforms portraying discharges are classified in two different classes considering the duration of the discharges. Two simple classifiers are employed for the classification. The first incorporates a simple genetic algorithm approach and the second a simple Euclidian distance classification. Twenty different features, extracted from the time and the frequency domain are used as a pattern.

II. SET-UP AND MEASUREMENT SITES

The LC waveforms investigated in this paper have been recorded on insulators installed at two different 150kV Substations of the Transmission System of Crete, Greece. The Cretan Transmission Network is exposed to intense marine pollution and several techniques have been employed by the Greek Public Power Corporation to cope with the problem [10], with the construction of an open air test station being the most recent step [11, 12]. Eighteen different insulators (porcelain, RTV SIR coated and composite) have been monitored than more than six years. A collection ring was installed at the bottom side of each monitored insulator and the current was driven through a Hall current sensor in order to acquire the LC measurement, as shown in Figure 1. The acquired from the sensor data was then transmitted to a commercially available central Data Acquisition System (DAS), also shown in Figure 1. Sampling was performed continuously and simultaneously for all monitored insulators, at a rate of 2 kHz and resolution of 12bit.

The monitoring system incorporated the time-window technique [6-7] to record waveforms. Each waveform recorded has a duration of 480ms. The waveform portraying the highest peak value in the considered time window is recorded. The time-window is user defined (from 6 hours to 24 hours), with a 24 hours window being mostly employed due to hardware restrictions.

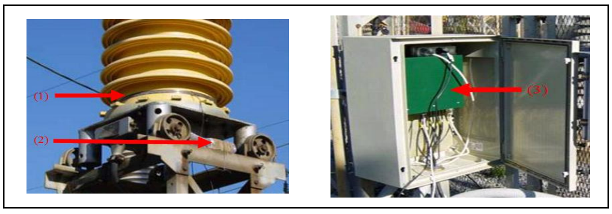

Figure 1. The measuring system: 1.collection ring 2.Hall sensor 3. the DAS

This work was supported in part by the Public Power Corporation (PPC SA), Greece.
D. Pylarinos is with the University of Patras, Greece, (e-mail:dpylarinos@yahoo.com).
K. Theofilatos is with the University of Patras, Greece
K. Siderakis is with the Technological Educational Institute of Crete
E. Pyrgioti is with the University of Patras, Greece
T. Papazoglou has been with Technological Educational Institute of Crete
I. Vitellas is with the Public Power Corporation, Greece
E. Thalassinakis is with the Public Power Corporation, Greece

III. Extracted Features and Waveform Data Set

A number of 387 discharge portraying waveforms are employed as a data set. Investigation of such waveforms has shown that values such as the peak value and the charge, is not representative of the waveforms' shape and therefore of the experienced surface activity [8-9, 12]. Therefore, added criteria should be employed for the classification. In this paper, the duration of portrayed discharges is used as a classification criterion. The waveforms are investigated and classified by hand in two difference classes. Class A includes waveforms that portray discharges that last four halfcycles or less, whereas class B includes waveforms that portray discharges that last five or more halfcycles. A set of 20 features are employed as a pattern, as shown in Table I. Features 1-10 derive from the time domain, and features 11-20 from the frequency domain, and have been chosen in order to evenly represent both domains and also considering the literature [4].

Regarding the time domain features, frequently used values such as the amplitude and charge, along with commonly used statistical values are employed. The charge is calculated from the waveform considering the digitization process and the applied sampling rate.

Regarding the feature domain features, it was considered that the content of odd harmonics is commonly correlated to the occurrence of discharges and the distortion of the waveforms' shape and therefore several commonly used ratios of odd harmonics [4] are used. It should be noted that the fundamental frequency is 50Hz and that the HD ratio is similar to the THD ratio, with the numerator being the sum of the odd harmonics' content.

Further, wavelet analysis and especially MRA is employed in order to acquire the STD_MRA VECTOR [13-14]. A ratio derived from such an analysis has previously been employed in order to export the S_R ratio to classify waveforms of low amplitude [7]. Similar ratios and also the distortion ratio [4] given by: $D_R = \frac{D_1 + D_2 + D_3 + D_4}{D_5}$, are considered in this paper. The frequency bands of the STD_MRA VECTOR's components are shown in Table II.

TABLE I – Feature Set

No.	Feature	No.	Feature
1	Amplitude	11	Third to First Harmonic Ratio
2	Mean	12	Fifth to First Harmonic Ratio
3	Median	13	Fifth to Third Harmonic Ratio
4	Variance	14	Total Harmonic Distortion Ratio (THD)
5	Standard Deviation	15	Harmonic Distortion Ratio (HD)
6	Median Absolute Deviation	16	STD_MRA VECTOR Ratio: D1/D5
7	Skewness	17	STD_MRA VECTOR Ratio: D2/D5
8	Kurtosis	18	STD_MRA VECTOR Ratio: D3/D5
9	Interquartile Range	19	STD_MRA VECTOR Ratio: D4/D5
10	Charge	20	Distortion Ratio: DR

TABLE II – Frequency Bands of MRA

Decomposition Level	(A) Approximation (Hz)	(D) Details (Hz)
1	0~500	500~1000
2	0~250	250~500
3	0~125	125~250
4	0~62.5	62.5~125
5	0~31.25	31.25~62.5
6	0~15.625	15.625~31.25

IV. A Simple Genetic Algorithms Approach

Genetic Algorithms (GAs), as proposed by Holland [15], are general search meta-heuristic algorithms based on evolution's principles of nature. Their initial setup proposed by Holland concerned the optimization of bit-strings which were called chromosomes. GAs have been proved useful and efficient in optimization problems where the search space is big and complicated or there is not any available mathematical analysis of the problem.

Their function is based on the initial creation of a population of candidate solutions, called chromosomes, and their iterative differentiation using the operators of evaluation, selection, crossover and mutation until some termination criteria are reached. Evaluation is responsible for evaluating the candidate solution using a problem specific fitness function. The various selection operators which have been proposed in the literature are used to enforce the search process to emphasize in the most promising areas of the search space. Crossover and mutation operators, are used to produce new probably better solutions using the population of the candidate solutions. The crossover operator is responsible for the recombination of existing solutions in the population whereas the mutation operator is responsible for the random differentiation of existing solutions. The probabilities, under which the mutation operator is applied, control the search behavior of the produced algorithm. For example, high mutation probability may degenerate the algorithm to a random search whereas a very small mutation rate may lead the algorithm in getting trapped in local optima.

The ability to define a different fitness function for each problem, and the flexibility of genetic algorithms to function deploying various representations of candidate solutions have enables their extensive use in a variety of problems. In the present paper, we applied GAs for classifying waveforms in two categories. Specifically, our approach attempts to classify the two classes of waveforms linearly using a set of 20 features. In order to achieve this linear classification, the two centers of the two classes should be determined. The positions of these two centers are what we assigned in the genetic algorithm to optimize. After the estimation of these centers is achieved, each waveform is classified in the class for which it has the minimum Euclidean distance from its centre. Our problem thus, is formulated as finding the optimal centers of each class with its one being composed by 20 numerical variables.

A simple GA is employed to solve this optimization problem [16]. The data set gets normalized and two thousands generation are employed with a crossover probability of 90% and a mutation probability of 10% and different initial populations as shown in Table III. For our problem, we selected as fitness function the classification accuracy which is derived when the centers of each candidate solutions are used. Then, for the initial population showing the best results, different crossover and mutation probabilities are employed as shown in Table IV. Fifty present of the data was used as a training set and fifty present as a test set. The mean value of the results of five runs is presented in each table.

TABLE III–GA RESULTS (CROSSOVER: 90%, MUTATION: 10%)

Initial Population	Accuracy (%)
20	0.558162
40	0.561224
60	0.560203

TABLE IV–GA RESULTS (INITIAL POPULATION: 40)

Crossover (%)	Mutation (%)	Accuracy (%)
0.9	0.2	0.557142
0.9	0.3	0.560203
0.7	0.1	0.561224
0.7	0.2	0.560203
0.7	0.3	0.560203

V. A MINIMUM EUCLIDIAN DISTANCE CLASSIFIER

The employed Euclidian classifier is a rather simple classifier based on the Euclidian distance given by $d(x,y)=\sqrt{\sum_{i=1}^{n}(y_i-x_i)}$, where y_1, y_2, ..., y_n the components of array y and x_1, x_2,...,x_n the components of array x.

At first a mean values array per class is created, containing the mean values for each feature for each class. Then, the Euclidian distance between each data array and the mean values array is calculated. The array is classified to the class that is less distant from it. An overall accuracy of 77.2% is accomplished using this classifier.

VI. DISCUSSION & FUTURE WORK

Results show a low accuracy percentage (~56%) for the GA employed, regardless the different settings used. This shows that the simple GA approach followed, using constant values for the parameters of crossover probability, mutation probability and population size, was unable to adapt the search procedure in order to verge on the optimal solution. Thus, as it turns out it gets trapped in local optimal solutions. The Euclidian classifier shows significantly better results (~77%). Despite its improved classification results in comparison to genetic algorithms, Euclidean classifier's performance is not very satisfactory.

The justification for these results, is that the examined classification problem is with high probability a non linear one. Thus, the performance of the linear classification methods which were developed in the present work could be improved by using modern non-linear classification methods. At present, we are using different non-linear classification and feature selection algorithms in order to acquire more accurate results and select the most informative feature subset from the initial set of the 20 features.

VII. CONCLUSION

The shape of the LC waveform is correlated to surface activity and condition of high voltage insulators. Field monitoring is essential to acquire an actual image of insulators' performance and experienced surface activity under service conditions. However, extracting a representative value from the waveform is a rather complex task and a fully representative value is yet to be presented. In this paper, a new approach for the classification of activity portraying field LC waveforms is proposed. Based on the literature, twenty different features, from both time and frequency domain, are calculated and used as a pattern. Then two simple classifiers, one employing GAs and one employing the Euclidian distance, are applied in order to classify the waveforms in two different classes based on the duration of discharges. Results show low accuracy for the GA classifier and significantly better for the Euclidian classifier. Future work will focus on applying different classification and feature selection algorithms in order to acquire higher accuracy percentages.

VIII. REFERENCES

[1] CIGRE WG 33-04, TF 01, *A review of current knowledge: polluted insulators,* CIGRE, 1998

[2] IEC/TS 60815, *Selection and dimensioning of high-voltage insulators intended for use in polluted conditions* (IEC, 2008

[3] CIGRE WG 33-04, *"The measurement of site pollution severity and its application to insulator dimensioning for a.c. systems",* Electra Vol. 64, p. 101-116, Cigre, 1979

[4] D. Pylarinos, K. Siderakis, E. Pyrgioti, *"Measuring and analyzing leakage current for outdoor insulators and specimens",* Reviews on Advanced Materials Science, Vol. 29, No. 1, pp. 31-53, 2011

[5] CIGRE WG B2.03, *Guide for the establishment of naturally polluted insulator testing stations*, CIGRE, 2007

[6] D. Pylarinos, K. Siderakis, E. Thalassinakis, E. Pyrgioti, I. Vitellas, S. L. David, "Online applicable techniques to evaluate field leakage current waveforms", Electric Power Systems Research, Vol. 84, No. 1, pp. 65-71, 2012

[7] D. Pylarinos, K. Siderakis, E. Pyrgioti, E. Thalassinakis, I. Vitellas, "Impact of noise related waveforms on long term field leakage current measurements", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 18, No. 1, pp. 122-129, 2011

[8] D. Pylarinos, K. Siderakis, E. Thalassinakis, E. Pyrgioti, I. Vitellas, "Investigation of leakage current waveforms recorded in a coastal high voltage substation", Engineering, Technology & Applied Science Research, Vol. 1, No. 3, pp. 63-69, 2011

[9] D. Pylarinos, K. Siderakis, E. Pyrgioti, I. Vitellas, E. Thalassinakis, *"Monitoring leakage current waveforms in the field"*, DEMSEE 5th international conference and technical exhibit on Deregulated Electricity Market issues in South-Eastern Europe, Sitia, Greece, September 23-24, 2010

[10] K. Siderakis, D. Pylarinos, E. Thalassinakis, E. Pyrgioti, I. Vitellas, *"Pollution maintenance techniques in coastal high voltage installations"*, Engineering, Technology & Applied Science Research, Vol. 1, No. 1, pp. 1-7, 2011

[11] TALOS High Voltage Test Station, www.talos-ts.com

[12] D. Pylarinos, K. Siderakis, E. Thalassinakis, I. Vitellas, E. Pyrgioti, "Recording and managing field leakage current waveforms in Crete. Installation, measurement, software development and signal processing", ISAP 16th International Conference on Intelligent System Applications to Power Systems, Hersonissos, Crete, Greece, September 25-28, 2011

[13] S. G. Mallat, *A Wavelet Tour Of Signal Processing*, Academic Press, 1999.

[14] S. G. Mallat, *"A Theory for Multiresolution Signal Decomposition: The Wavelet Representation"*, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 11, pp. 674-693, 1989.

[15] J. Holland, *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*, Cambridge: Mass ;MIT Press, 1995.

[16] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, 3rd edition, 1996.

IX. Biographies

Dionisios Pylarinos was born in Athens, Greece in 1981. He received a Diploma degree in Electrical and Computer Engineering from the University of Patras, Greece in 2007 and a Ph.D. degree from the University of Patras in 2012. He is a member of the Technical Chamber of Greece. His research interests include outdoor insulation, electrical discharges, signal processing and pattern recognition

Konstantinos Theofilatos was born in Patras in 1983. He graduated in 2006 from the Department of Computer Engineering and Informatics of the University of Patras, Greece. In 2009, he received a Master's degree from the same department. Since 2009, he has been a PhD degree candidate in the same department. He is a member of the Pattern Recognition Laboratory (prlab.ceid.upatras.gr) since 2006. His research interests include computational intelligence, machine learning, data mining, bioinformatics, web technologies and time series forecasting and signal processing.

Kiriakos Siderakis was born in Iraklion in 1976**.** He received a Diploma degree in electrical and computer engineering in 2000 and the Ph.D. degree in 2006 from the University of Patras. Presently, he is a Lecturer at the Department of Electrical Engineering, at the Technological Educational Institute of Crete. His research interests include outdoor insulation, electrical discharges, high voltage measurements and high voltage equipment diagnostics and reliability. He is a member of the Greek CIGRE and of the Technical Chamber of Greece.

Eleftheria Pyrgioti was born in 1958 in Greece**.** She received her Diploma degree in electrical engineering from Patras University in 1981 and the Ph.D. degree from the same University in 1991. She is an assistant professor at the department of Electrical and Computer Engineering at the University of Patras. Her research activity is directed to the high voltage, lightning protection, insulation coordination, distributed generation. She participates in the permanent committee TE 63 of ELOT (Hellenic Organization for Standardization) since 1989 and is a member of the Greek CIGRE and the Tehcnical Chamber of Greece.

Thales Papazoglou was born in 1945 in Greece. He is an electrical engineer, has a M.Sc. from Caltech and a Ph.D. from the University of Arizona. He has work experience at NASA's JPL as an engineer, at IVO Finland as a Visiting Professor and at TEIC as Professor and Head of Electrical Engineering Dept. He has actively participated in 12 Working Groups and Task Forces in CIGRE and he is at present the Convener of WG C2.13 on Voltage and VAr Support in System Operation. He is the founder of the annual DEMSEE and he is the Convener of the Steering Committee of DEMSEE.

Isidoros Vitellas was born in 1954 in Greece. He has a diploma in electrical engineering and the Ph.D. degree in the same field. He is currently Director of the Islands Network Operations Department in PPC (Public Power Corporation) Athens, Greece.

Emmanuel Thalassinakis received the Diploma in electrical and mechanical engineering and also the Ph.D. degree from the National Technical University of Athens. After working for the Ministry of the Environment, in 1991 he joined the Public Power Corporation (PPC) where he is now Assistant Director of the Islands Network Operations Department.