Rubén Jaramillo-Vacio, Alberto Ochoa-Zezzatti, Armando Rios-Lira / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 3, Issue 4, Jul-Aug 2013, pp.974-984

Optimal Tuning Parameters of Self Organizing Maps for Classification of Partial Discharge Using Response Surface Methodology

Rubén Jaramillo-Vacio[1, 2], Alberto Ochoa-Zezzatti[3], Armando Rios-Lira[4]
[1]Comisión Federal de Electricidad-Laboratorio de Pruebas a Equipos y Materiales (LAPEM).
[2]Centro de Innovación Aplicada en Tecnologías Competitivas (CIATEC).
[3]Universidad Autónoma de Ciudad Juárez.
[4]Instituto Tecnológico de Celaya

Abstract

This paper presents an analysis for Self Organizing Map (SOM) using Response Surface Methodology (RSM) to find the optimal parameters to improve performance. This comparative explores the relationship between explanatory variables (numerical and categorical) such a competitive algorithm and learning rate and response variables as training time and quality metrics for SOM. Response surface plots were used to determine the interaction effects of main factors and optimum conditions to the performance in classification of partial discharge (PD).

I. Introduction

Cluster analysis, grouping or clustering have been studied to represent large amount data quite a long time. In principle, it is employed to generate some representative groups from numerous historical data or cases. Furthermore, it is used to achieve the aim of understanding the structure of original data distribution and even executing data analysis for inference. Competitive learning is an efficient tool for unsupervised neural networks like Self Organizing Maps.

In the field of data analysis two methodologies frequently encountered are supervised and unsupervised clustering methodologies. While supervised methods mostly deal with training classifiers for known class, unsupervised clustering provides exploratory techniques for pattern recognition in data. With the large amount data being generated from the different systems everyday, what makes a system intelligent is its ability to analyze the data for efficient decision-making based on known or new cluster discovery. The partial discharge is a common phenomenon which occurs in insulation of high voltage, this definition is given in IEC 60270 [1]. In general, the partial discharges are in consequence of local stress in the insulation or on the surface of the insulation. This phenomenon has a damaging effect on the equipments, for example transformers, power cables, switchgears, and others. Many investigators have studied the feasibility of selecting the different features to classify measured PD activities into underlying insulation defects or source that generate PD's. In particular for solid insulation like XLPE on power cables where a complete breakdown seriously damages the test object the partial discharge measurement is a tool for quality assessment [2].

Some research results in this field have been submitted in McGrail *et al* [3] where two variations of SOM are discussed using examples from different diagnostic measurements and tests. The use of Kohonen mapping applied to existing and new data is illustrated using examples of dissolved gas analysis, tap-changer monitoring and insulator testing. In [4] and [5] were used supervised neural networks for recognition between different sources formed of cylindrical cavities, the principal constraint was the recognition of different sources in the same sample. Kim et al [6] made the comparison between Back Propagation Neural Network and Fuzzy-Neural Networks, however, is necessary to improve performance in the multiple discharges and including defects and noises. Particle Swarm Optimization (PSO) was used for location of PD in power transformers, on site application should improve performance [7]. In [8] the SOM was used for PD pattern recognition and classification without quality measurement and optimization of the structure. Fadilah Ab Aziz *et al* [9] worked using the Support Vector Machine (SVM) for feature selection and PD classification, concluding that SVM is not reliable for small dataset. Hirose *et al* [10] showed that Decision tree applied for feature extraction and PD classification, that the allocation rules are sensitive to small perturbations in the dataset (Instability).

Venkatesh *et al* [11] proposed an exhaustive analysis is carried out to determine the role played by the free parameter (variance parameter) in distinguishing various classes of PD, number of iterations and its impact on computational cost during the training phase in NNs which utilize the clustering algorithms and the choice of the number of codebook vectors in classifying the patterns.

In [12], Agamolov shows the formalization of PD pulses analysis for forecasting a technical condition of the electrical machines and used information about change PD pulses in the difference transients. For it used a cluster algorithm, which is not demanding an aprioristic information about law distribution of PD pulses parameters for various levels of technical condition of an electrical machines. Also, in the paper it is shown the application example of offered PD pulses cluster analysis for data that was measured by means of electromagnetic sensors in turbo-generator. It aims to represent a multidimensional dataset in two or three dimensions such that the distance matrix in the original *k*-dimensional feature space is preserved as faithfully as possible in the projected space. The SOM, or Kohonen Map can also be used for nonlinear feature extraction, in [13] is analyzed the performance in classification of partial discharge on power cables using SOM.

Conventional methods of pattern recognition using neural networks and heuristic algorithms by maintaining other factors involved at unspecified constant levels does not depict the combined effect of all the factors. This method is time consuming and incapable. This calls for a research effort for developing, improving and optimizing the processing time and evaluate the significance of all the factors involved even in the presence of complex interactions between explanatory variables. It should be emphasized that the goal here is to find an optimal clustering for the data but to get good insight into the cluster structure of the data for data mining purposes. Therefore, the clustering algorithm using SOM must be fast, robust, and visually efficient after of this analysis.

II. Concepts used in Partial Discharge

They are generally divided into three different groups because of their different origins:

- **Corona Discharges** – Occurs in gases or liquids caused by concentrated electric fields at any sharp points on the electrodes.
- **Internal Discharges** – Occurs inside a cavity that is surrounded completely by insulation material; might be in the form of voids (e.g. dried out regions in oil impregnated paper-cables).
- **Surface Discharges** – Occurs on the surface of an electrical insulation where the tangential field is high e.g. end windings of stator windings.

The charge that a PD generates in a void (see Fig. 2) is called the physical charge and the portion of the void surface that the PD affects is called the discharge area. $E_{applied}$ is the applied electric field and $q_{physical}$ is the physical charge [14].

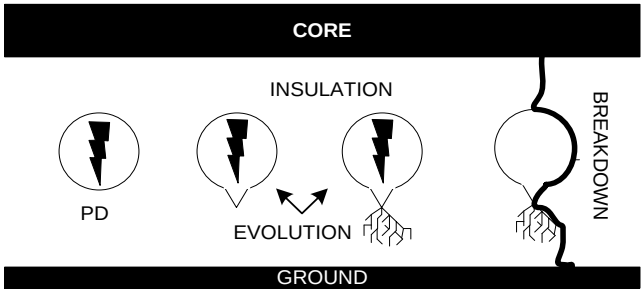

Fig. 1. Example of damage in a polymeric power cable from the PD in a cavity to breakdown.

The repetition rate *n* is given by the number of PD pulses recorded in a selected time interval and the duration of this time interval. The recorded pulses should be above a certain limit, depending on the measuring system as well as on the noise level during the measurement.

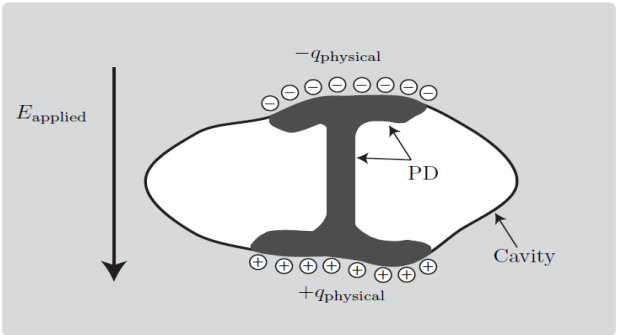

Fig.2. Schematic picture of a partial discharge in a cavity

The pulse repetition frequency *N* is the number of partial discharge pulses per second in the case of equidistant pulses. Furthermore, the phase angle ϕ and the time of occurrence t_i are information on the partial discharge pulse in relation to the phase angle or time of the applied voltage with period *T*:

$$\phi i = 360(ti / T) \quad (1)$$

In the measurement equipment is recorded the number of the detected PD with a certain combination of phase and charge (see Fig. 3). This graph shows the behavior of partial discharge in a void under high voltage rising. For PD diagnosis test, is very important to classify measured PD activities, since PD is a stochastic process, namely, the occurrence of PD depends on many factors, such as temperature, pressure, applied voltage and test duration; moreover PD signals contain noise and interference [15].

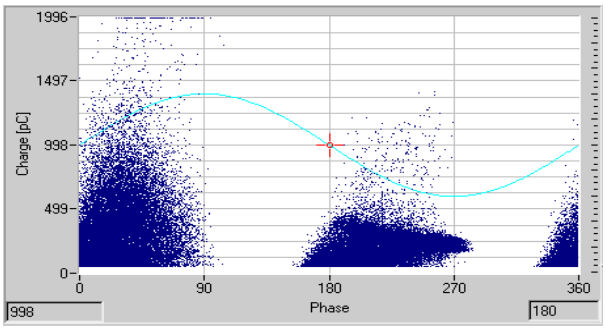

Fig. 3. Example of a partial discharge pattern

Therefore, the test engineer is responsible for choosing proper methods to diagnosis for the given problem. In order to choose the features, it is important to know the different source of PD, an alternative is though pattern recognition. This task can be challenging, nevertheless, features selection has been widely used in other field, such as data mining [16] and pattern recognition using neural networks [4]-[6]. This research only shows test on laboratory without environment noise source, and it is a condition that does not represent the conditions on site, Markalous [17] presented the noise levels on site based on previous experiences.

The phase resolved analysis investigates the PD pattern in relation to the variable frequency AC cycle. The voltage phase angle is divided into small equal windows.

The analysis aims to calculate the integrated parameters for each phase window and to plot them against the phase position (ϕ).

- (n −ϕ) : the total number of PD pulses detected in each phase window plotted against the phase position.
- (q_a −ϕ) : the average discharge magnitude in each phase window plotted against the phase position ϕ, where q_a is average discharge magnitude.
- (q_m −ϕ) : the peak discharge magnitude for each phase window plotted against ϕ, where q_m is peak discharge magnitude.

III. Self Organizing Map (SOM)

3.1 Winner takes all

The Self Organizing Map developed by Kohonen, is the most popular neural network models. The SOM algorithm [12] is based on unsupervised competitive learning called winner – takes – all, which means that the training in entirely data-driven and that the neurons of the map compete with each other.

Supervised algorithms [4, 5], like multi-layered perceptron, required that the target values for each data vector are known, but the SOM does not have this limitation. The SOM is a two-layer neural network that consists of an input layer in a line and an output layer constructed of neurons in a two-dimensional grid as is shown in Figure 4. The neighborhood relation of neuron i, an n-dimensional weight vector w is associated; n is the dimension of input vector. At each training step, an input vector **x** is randomly selected and the Euclidean distances between **x** and **w** are computed. The image of the input vector and the SOM grid is thus defined as the nearest unit w_{ik} and best-matching unit (BMU) whose weight vector is closest to the **x** [18]:

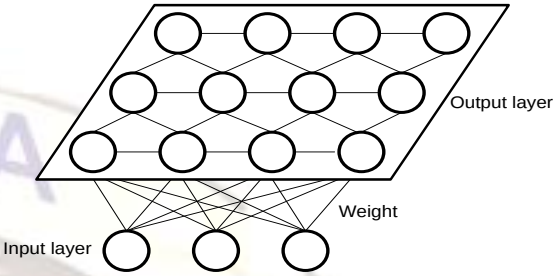

Fig. 4. Basic structure of a SOM.

$$D(x, w_i) = \sqrt{\sum_k (w_{ik} - x_k)^2} \quad (2)$$

The weight vectors in the best-matching unit and its neighbors on the grid are moved towards the input vector according the following rule (WTA):

$$\Delta w_{ij} = \delta(c,i)\alpha(x_j - w_{ij})$$
$$\Delta w_{ij} = \alpha(x_j - w_{ij}) \quad to \quad i = c$$
$$\Delta w_{ij} \quad to \quad i \neq c \quad (3)$$

where **c** denote the neighborhood kernel around the best-matching unit and α is the learning rated and δ is the neighborhood function (see Fig. 5).

The number of panels in the SOM is according the A x B neurons, the U-matrix representation is a matrix U ((2A-1) x (2B-1)) dimensional [19]. The selection of the distance criterion depends on application. In this paper, Euclidean distance is used because it is widely worn with SOM [20].

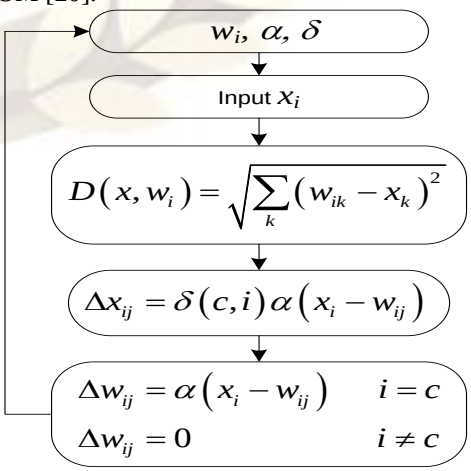

Fig. 5. Flow chart of SOM.

It is complicated to measure the quality of an SOM [21]. Resolution and topology preservation are generally used to measure SOM quality. SOMs with growing topology have been proposed for generating topology-preserving mappings and for data visualization. In addition, Kantardzic [12] noted that typically the two performance evaluation criteria that can be used to measure criteria are topology preservation and resolution. There are many ways to measure these two properties. For simplicity, this research chooses the quantization error **(*qe*)** and the topographic error **(*te*)** .

3.2 The frequency sensitive competitive algorithm

To solve the "dead units" problem it has been introduced the so called "frequency sensitive competitive learning" algorithm (FSCL) [22] or competitive algorithm "with conscience". The FSCL algorithm is an extension of *k*-means algorithm, obtained by modifying relation (2) according to the following one:

$$j = \arg\min \gamma_i \left\| x(n) - c_i(n) \right\| \quad i = 1, \ldots, N \qquad (4)$$

where n is the inputs, N represents the centres numbers, the relative winning frequency γ_i of the centre c_i defined as:

$$\gamma_i = \frac{s_i}{\sum_{i=1}^{n} s_i} \qquad (7)$$

where s_i is the number of times when the centre c_i was declared winner in the past. So the centers that have won the competition during the past have a reduced chance to win again, proportional with their frequency term γ. After selecting out the winner, the FSCL algorithm updates the winner with next equation:

$$c_i(n+1) = c_i(n) - \eta\left[x(n) - c_i(n) \right] \qquad (8)$$

η is the learning rate, in the same way as the *k*-means algorithm, and meanwhile adjusting the corresponding s_i with the following relation:

$$s_i(n+1) = s_i(n) + 1 \qquad (9)$$

3.3 The rival penalized competitive learning algorithm

The rival penalized competitive algorithm (RPCL) [22] performs appropriate clustering without knowing the clusters number. It determines not only the winning centre *j* but also the second winning center *r*, named rival

$$r = \arg\min \gamma_i \left\| x(n) - c_i(n) \right\|, \quad i = 1, \ldots, N \;\; i \neq j \qquad (10)$$

The second winning centre will move away its centre from the input with a ratio □, called the de-learning rate. All the other centres vectors will not change. So the learning law can be synthesized in the following relation:

$$c_i(n+1) = \begin{cases} c_i(n) + \eta\left[x(n) - c_i(n) \right] & if \;\; i = j \\ c_i(n) - \beta\left[x(n) - c_i(n) \right] & if \;\; i = j \\ c_i(n) \;\; if \;\; i \neq j \;\; and \;\; i \neq r \end{cases} \qquad (11)$$

If the learning speed η is chosen much greater than β, with at least one order of magnitude, the number of the output data classes will be automatically found. In other words, suppose that the number of classes is unknown and the centres number *N* is greater than the clusters number, than the centres vectors will converge towards the centroids of the input data classes. The RPCL will move away the rival, in each iteration, converging much faster than the *k*-means and the FSCL algorithms.

IV. Design and Analysis of Experiments

Design of Experiment (DOE) techniques are well known in industrial areas, where production processes, product developments and *cost* and time reductions are often optimized using the relevant methods and models.

Typically, an experiment may be run for one or more of the following reasons [23]:
(a) to determine the principal causes of variation in a measured response, (b) to find the conditions that give rise to a maximum or minimum response, (c) to compare the responses achieved at different settings of controllable variables, (d) to obtain a mathematical model in order to predict future responses. Response Surface Methodology (RSM) is a collection of mathematical and statistical techniques useful for analyzing the effects of several independent variables on the response. RSM has an application in the process design and optimization as well as the improvement of existing design [24]. This methodology is more practical compared to theoretical models as it arises from experimental methodology which includes interactive effects of the variables and, eventually, it depicts the overall effects of the parameters, as in this case, in the SOM algorithm.

In multivariable systems, the classical approach of changing one variable at a time to study the effects on other variables for a particular response is time consuming. Consequently, an alternative strategy involving statistical approaches, e.g., RSM, was applied to solve for multiple variables in this complex system. This method of parametric optimization can also be combined to monitor the impact of factors affecting the training time and to quality of SOM.

The advantage of the RSM method is the minimization of the number of experiments and time needed. In the optimization procedure we studied the response of the statistically designed combinations, estimated the coefficients by fitting the experimental data to the response functions, predicted the response of the fitted model and checked the adequacy of the model. We used the D-Optimal under response surface methodology (RSM) to investigate competitive effects in the ***qe, te*** and **training time**.

A D-optimal design is a computer aided design which contains the best subset of all possible experiments. Depending on a selected criterion and a given number of design runs, the *best* design is created by a selection process. Levels of factors are shown in Table 1 and Table 2.

TABLE 1 EXPERIMENTAL RANGE AND LEVELS OF NUMERICAL VARIABLES

Factor	Coded levels	
	-1	1
η (eta)	1E-5	1
β (Beta)	0.01	1

TABLE 2 EXPERIMENTAL RANGE AND LEVELS OF CATEGORICAL VARIABLES

Factor	Levels		
Algorithm	WTA	FSCL	RPCL
Discharge	Internal	External	Surface

V. Analysis of Data

Were considered the pattern characteristic of univariate phase-resolved distributions as inputs, the magnitude of PD is the most important input as it shows the level of danger, for this reason the input in the SOM the raw data is the peak discharge magnitude for each phase window plotted against (qm −ϕ). Figure 6 shows the conceptual diagram training. In the cases analyzed, the original dataset is 1 million of items, was used a neurons array of 100×100 cells to extract features.

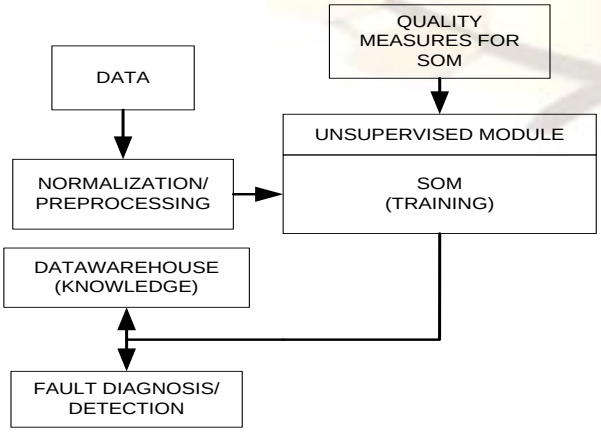

Fig. 6. The component interaction between SOM

5.1 Regression Analysis

Regression analysis was executed to determine the surface response as function of second order polynomial equation:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i \le j}^{k} \beta_{ij} x_i x_j + \varepsilon \quad (12)$$

Where Y is predicted response, β_i , β_{ii} , β_{ij} represent linear, quadratic and interaction effects. β_0 is the intercept term and x_i , x_j ,..., x_k are the input which affect the output value.

The scheme of experiments carried out in this study was presented in Table 3.

In figure 7, 8 and 9 is showed an example of the performance and convergence of the competitive learning algorithms to different PD source, also in figure 10, 11 and 12 is presented the dataset of each PD source and prototype vector resulting of the training of SOM.

TABLE 3 D-OPTIMAL DESIGN MATRIX

Inputs				Outputs		
Eta	Beta	Alg.	Discharge	*qe*	*te*	Time
A	B	C	D			
1	0.5	WTA	External	0.86	0.99	176.5
1E-5	0.01	WTA	Surface	10.22	0.97	157.9
1	1	FSCL	External	0.96	0.98	228.9
1	0.5	FSCL	Surface	9.61	0.96	215.9
0.25	0.25	WTA	External	0.76	1	171.6
1E-5	0.01	RPCL	External	1.12E6	1	243.6
0.5	1	WTA	Surface	11.84	0.99	159.7
0.5	1	FSCL	Internal	7.84	0.98	43.9
1	0.01	FSCL	Internal	5.52	0.97	44.0
1	0.01	RPCL	Internal	7.32	0.98	47.2
0.25	0.5	FSCL	External	0.85	0.92	234.5
0.5	1	WTA	Surface	10.81	0.96	163.0
1E-5	1	RPCL	Internal	1.15E6	1	47.5
1E-5	0.5	WTA	Internal	6.39	0.97	32.9
1E-5	1	FSCL	Surface	11.41	0.96	219.1
1E-5	0.01	RPCL	External	5.82E6	1	248.4
1	1	RPCL	Surface	118.23	1	230.2
1E-5	0.01	FSCL	Internal	6.21	0.97	46.1
0.5	0.01	WTA	Internal	8.71	0.99	159.0
1E-5	0.5	RPCL	Surface	2.50E6	1	47.2
0.5	0.01	FSCL	Surface	8.68	0.98	215.0
0.5	0.5	RPCL	Internal	169.61	1	46.9
1	1	RPCL	External	50.14	1	244.1
1E-5	0.01	FSCL	Internal	6.13	0.98	44.0

1E-5	1	WTA	External	2.06	0.99	170.5
1	1	WTA	Internal	7.80	0.98	32.8
1	1	WTA	Internal	7.47	0.99	32.5
1	1	RPCL	External	57.55	1	246.5
1	0.01	RPCL	External	1.19	0.97	251.7
0.75	0.25	RPCL	Surface	119.92	1	230.0
0.75	0.25	FSCL	External	0.76	0.96	229.5
1	0.01	WTA	Surface	8.50	0.96	164.2

5.1 Analysis of Variance

The statistical significance of the model equation and the goodness of fit were evaluated by R^2 and by the *F*-test analysis of variance (ANOVA), which is a statistical technique that subdivides the total variation in a set of data into components associated with specific sources of variation to test hypotheses on the parameters of the model. A large *F*-value indicates that most of the variation can be explained by a regression equation whereas a low *p*-value (<0.05) indicates that the model is considered to be statistically significant (Table 4).

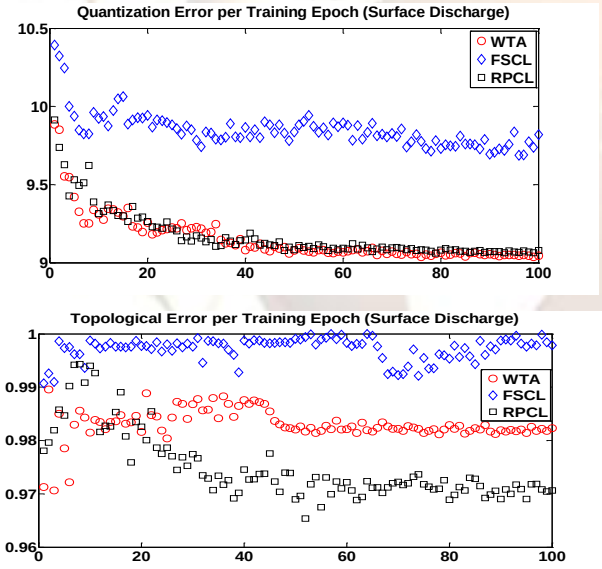

Fig. 7. Quantization and Topological error per Training Epoch (Surface Discharge)

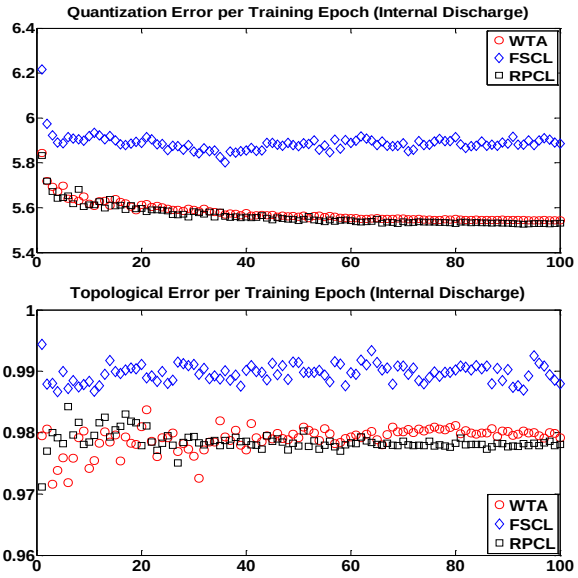

Fig. 8. Quantization and Topological error per Training Epoch (Internal Discharge)

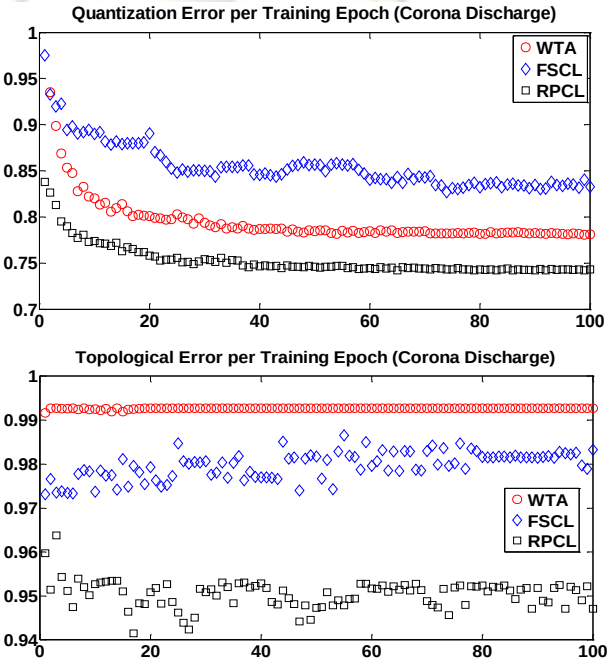

Fig. 9. Quantization and Topological error per Training Epoch (Corona Discharge)

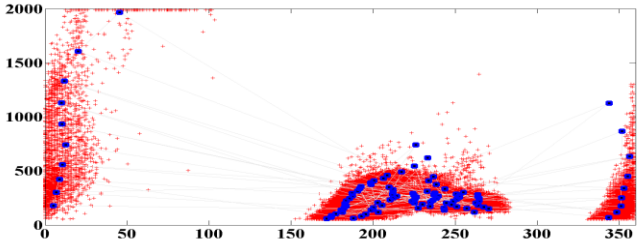

Fig. 10. Dataset and Prototype Vector (Surface Discharge)

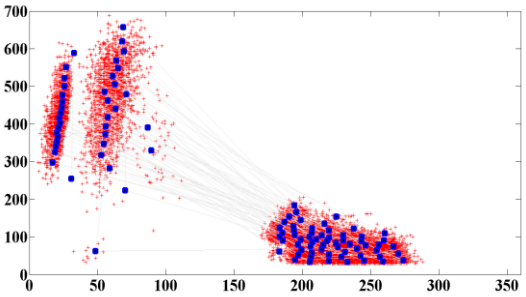

Fig. 11. Dataset and Prototype Vector (Internal Discharge)

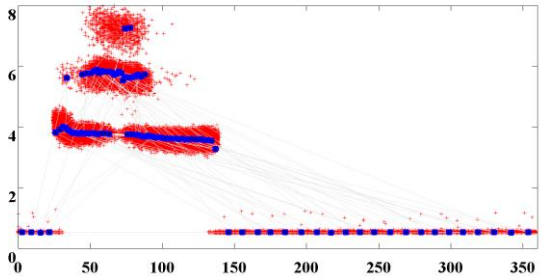

Fig. 12. Dataset and Prototype Vector (Corona Discharge)

A relationship between the response and input variables expressed by the following response surface quadratic model equations:

qe = 3.38-1.67 *A+0.46 * B-1.97 * C[1]-2.02*C[2]-4.223E-003*D[1]+0.81* D[2]+0.22*AB+1.67*AC[1]+ 1.57*AC[2]+0.36*AD[1]+0.21*AD[2]-0.41*BC[1]-0.13*BC[2]- 0.42*BD[1]+0.082*BD[2]+0.50*C[1]D[1] +0.46*C[2]D[1]+0.11*C[1]D[2]+0.20*C[2]D[2] (13)

te=-0.21+1.575*A+7.729*B+1.551C[1]-0.17*C[2] +3.032*D[1]-1.489* D[2]+8.981*AB+8.344*AC[1] +9.763*AC[2]-4.087*AD[1]- 2.446*AD[2]+1.602*BC[1] +7.662*BC[2]-1.931*BD[1]- 1.990*BD[2]6.174*C[1]D[1] +6.882*C[2]D[1]-0.16*C[1]D[2]+6.520*C[2]D[2]-0.12*A^2+0.17*B^2 (14)

Time=143.71+10.73*A-8.59*B-14.10*C[1]+10.77*C[2]- 99.70*D[1]+27.39*D[2]+1.30*AB-4.85*AC[1]-4.85*AC[2]-15.15*AD[1]+25.15*AD[2]-18.28*BC[1] +12.57*BC[2]-10.13*BD[1]+13.49*BD[2]+38.46* C[1]D[1]-32.99*C[2]D[1]-20.53*C[1]D[2] +29.06*C[2]D[2]-30.94*A^2+40.27*B^2 (15)

The statistical significance of the model equation and the goodness of fit were evaluated by R^2 and by the *F*-test analysis of variance (ANOVA), which is a statistical technique that subdivides the total variation in a set of data into components associated with specific sources of variation to test hypotheses on the parameters of the model.

A large *F*-value indicates that most of the variation can be explained by a regression equation whereas a low *p*-value (<0.05) indicates that the model is considered to be statistically significant (Table 4).

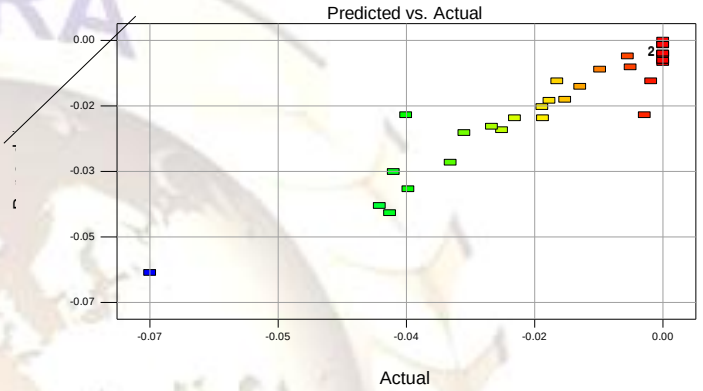

Fig. 13. The actual and predicted values of response of ***te***

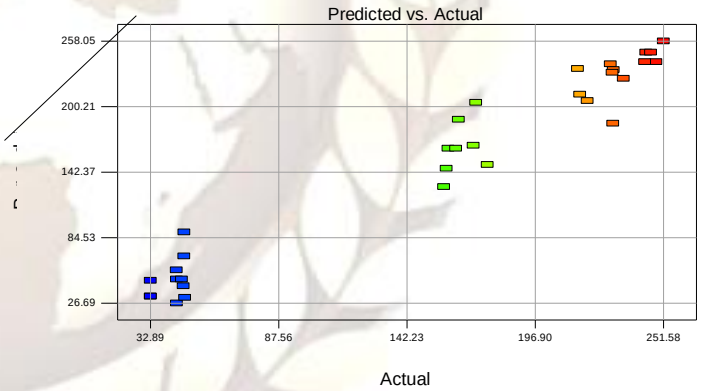

Fig. 14. The actual and predicted values of response of ***te***

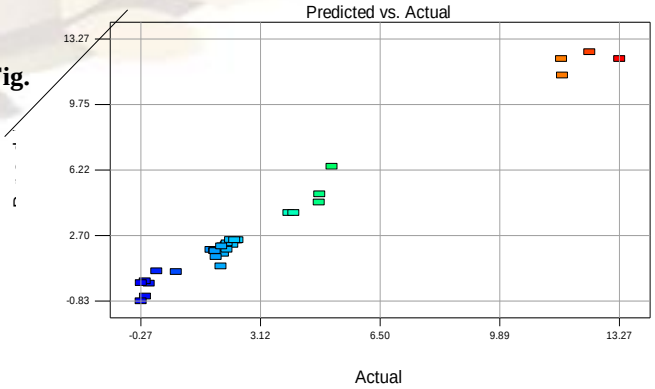

Fig.

15. The actual and predicted values of response of ***time***

The ANOVA analysis for the responses ***te, qe*** and training time was shown in table 4. P-value for the models is less than 0.05 indicates that the model terms are statistically significant. The actual and predicted values of responses were shown in figures 13, 14 and 15.

TABLE 4 ANOVA RESULTS

	Sum of Square	**DF**	**Mean Square**	**F-Value**	**P-value**
For ***qe***					
Model	425.37	19	22.39	45.82	<0.0001
Residual	5.86	12	0.49		
Lack of fit	0.2	1	0.2	0.72	0.484
Pure error	1.37	5	0.27		
Cor total	428.24	31			
R^2 = 0.9864					
R^2_{adj} = 0.9649					
For ***te***					
Model	843.7	21	40.17	143.48	<0.0001
Residual	7.86	10	0.65		
Lack of fit	0.25	1	0.25	0.91	1.245
Pure error	1.4	5	0.28		
Cor total	850.2	31			
R^2 = 0.9764					
R^2_{adj} = 0.9539					
For ***time***					
Model	2134	21	101.61	23.82	<0.0001
Residual	946.2	10	78.85		
Lack of fit	9.5	11	0.86	0.2024	0.742
Pure error	21.33	5	4.26		
Cor total	2182.53	31			
R^2 = 0.9575					
R^2_{adj} = 0.9286					

Actual values are the measured values for a particular experiment, whereas predicted values are generated by using the approximating functions. The values of R^2 and adjusted R^2 have advocated a high correlation between actual and predicted values.

The *qe* model *F*-value of 45.82 and value of p <0.0001 indicate statistical significance of a quadratic model. The *te* model *F*-value of 143.48 and value of p <0.0001 indicate statistical significance of a quadratic model. The *time* model *F*-value of 23.82 and value of p <0.0001 indicate statistical significance of a quadratic model. On the basis of this investigation, the relationship between the independent variables and the response can be explained according to the regression model. The goodness of the each model can be confirmed by the coefficient of determination R^2 and the adjusted R^2. Both values are closed to 1, which are very high and indicate a high correlation between the observed and the predicted values.

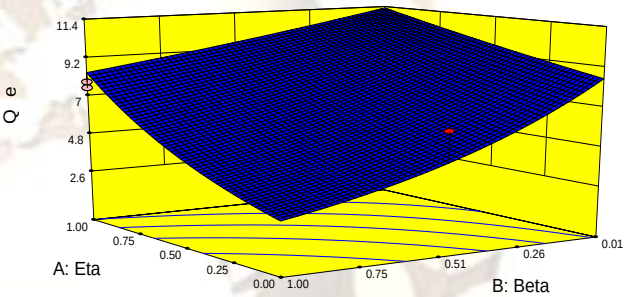

Fig. 16. Surface plot of the combined effects of the ***qe.***

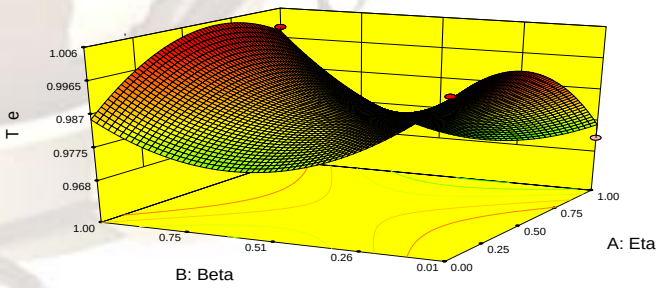

Fig. 17. Surface plot of the combined effects of the ***te.***

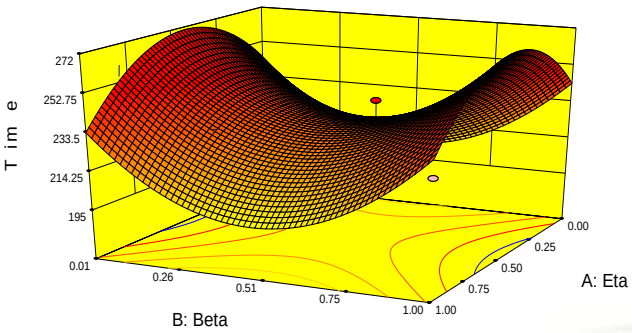

Fig. 18. Surface plot of the combined effects of the ***time.***

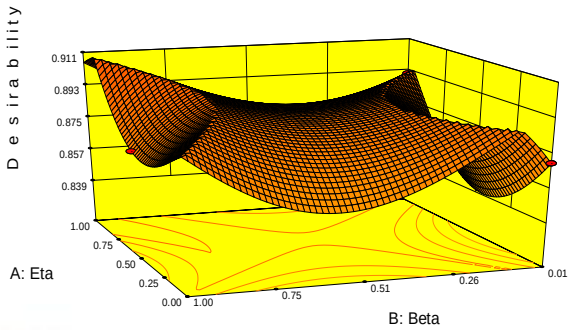

Fig. 19. Desirability plot to optimize multiple responses**.**

For example, value of adjusted R^2 of 0.9864 seggest that the total variation of 98.64% in qe is atributted to the independient variable and only about 1.36% of total variation cannot be represented by the model. On the basis of the evaluation of ANOVA outputs, the statistical significance of a quadratic model for the response was confirmed and it can be concluded that the model can be used for further analysis of effect of process variables.

5.1 Optimization by Response Surface Methodology

RMS was used to estimate the effect of four factors at different levels, in this analysis; 3D surface plots were drawn by using RMS to demonstrate that the interactions of these factors also have a significant effect on the response. For example in figure 16 shows the treatments to minimize ***qe*** and is clearly identify where a minimum in the figure. However for responses ***te*** and ***time*** (figures 17 and 18) is observed very interesting behavior, as it a saddle point, and in this context are a combined minimization/maximization problem of convex-concave functional. In this case it is important to establish the most suitable area of operation of the algorithm in order to obtain the most accurate response. The optimum conditions of all factors were found for the simultaneous eta, beta, competitive learning algorithms and discharge kind. The response surface plot at optimum condition is shown in figure 19. The basic idea of the desirability function approach is to transform a multiple response problem into a single response problem by means of mathematical transformations.

In the equation 16 is presented a desirability function (***dj***) with a range of value between 0 and 1, where 1 is the desired value (optimum).

$$dj = \begin{cases} \left(\dfrac{Y_j - Y_{\min j}}{T_j - Y_{\min j}}\right)^s & if\ Y_{\min j} \le Y_j(x) \le T_j \\ \left(\dfrac{Y_j - Y_{\max j}}{T_j - Y_{\max j}}\right)^s & if\ T_j \le Y_j(x) \le Y_{\max j} \\ 0 & otherwise \end{cases} \quad (16)$$

The best solution based in the different levels of the SOM's parameters is presented in Table 5.

TABLE 5 SOLUTIONS FOR DIFFERENT COMBINATIONS OF NUMERICAL AND CATEGORICAL FACTOR LEVELS

INPUTS				RESPONSE			
Eta	**Beta**	**Alg.**	**PD**	**Qe**	**Te**	**Time**	**Desirability**
0.9	1.0	FSCL	Internal	8.0	1.0	32.9	**0.9112**
1.0	1.0	WTA	Internal	8.4	1.0	32.9	**0.9047**
1.0	1.0	WTA	Internal	8.4	1.0	32.2	**0.9039**
1.0	0.3	RPCL	Internal	4.2	1.0	32.9	**0.9026**
1.0	0.3	RPCL	Internal	5.2	1.0	32.9	**0.8971**
0.0	0.9	WTA	Internal	2.9	1.0	42.9	**0.8951**
0.0	0.9	WTA	Internal	2.9	1.0	42.0	**0.8951**
0.9	0.1	FSCL	Internal	7.0	1.0	32.9	**0.8945**
0.9	0.1	FSCL	Internal	7.0	1.0	32.9	**0.8945**
0.9	0.1	FSCL	Internal	6.9	1.0	32.9	**0.8945**
0.9	0.1	FSCL	Internal	7.0	1.0	32.9	**0.8945**
0.9	0.1	FSCL	Internal	7.0	1.0	32.9	**0.8945**
0.8	0.1	FSCL	Internal	6.9	1.0	32.9	**0.8945**
1.0	0.9	RPCL	Internal	12.2	1.0	32.9	**0.8856**
1.0	0.7	RPCL	Internal	8.4	1.0	16.8	**0.8818**
0.0	0.6	WTA	Surface	8.2	1.0	73.3	**0.7547**
0.0	0.7	WTA	Surface	8.0	1.0	74.7	**0.7528**
1.0	0.7	WTA	External	0.7	1.0	143.8	**0.7251**
1.0	0.7	WTA	External	0.7	1.0	143.8	**0.7251**
1.0	0.6	WTA	External	0.6	1.0	144.4	**0.7232**
0.0	0.7	WTA	External	1.9	1.0	151.1	**0.6635**
0.0	0.7	WTA	External	1.9	1.0	151.1	**0.6635**
1.0	0.7	WTA	Surface	14.1	1.0	136.5	**0.6152**
0.0	0.3	FSCL	Surface	8.7	1.0	149.8	**0.5911**
0.0	0.3	FSCL	Surface	8.7	1.0	150.1	**0.5911**
0.0	0.2	FSCL	Surface	8.4	1.0	155.0	**0.5874**
0.0	0.6	FSCL	Surface	10.0	0.9	155.1	**0.5536**
1.0	0.4	RPCL	Surface	16.7	1.0	180.3	**0.5173**
1.0	0.4	RPCL	Surface	17.3	1.0	180.1	**0.5173**
1.0	0.3	RPCL	Surface	15.3	1.0	181.2	**0.5167**
1.0	0.5	FSCL	External	0.5	1.0	195.9	**0.4915**
1.0	0.5	FSCL	External	0.5	1.0	195.9	**0.4915**
1.0	0.5	FSCL	External	0.5	1.0	196.2	**0.4914**
1.0	0.6	FSCL	External	0.5	1.0	196.2	**0.4913**
1.0	0.6	FSCL	External	0.5	1.0	196.5	**0.4910**
1.0	0.4	FSCL	Surface	10.0	1.0	208.9	**0.4064**
1.0	0.4	FSCL	Surface	9.9	1.0	208.9	**0.4064**
1.0	0.5	RPCL	External	10.8	1.0	212.8	**0.3935**
1.0	0.5	RPCL	External	11.0	1.0	212.7	**0.3935**
1.0	0.5	RPCL	External	11.2	1.0	212.7	**0.3934**

VI. Conclusions

The present investigation was carried out to study combined effect of the performance of SOM using different competitive learning algorithms to find the optimum conditions to classify measured PD activities into underlying insulation defects or source that generate PD's. Two types of factors were successfully tested: numerical (η and β) and categorical (Competitive learning algorithm and PD kind). D-optimal design based on four inputs and three responses was used to estimate the effects in training time and SOM quality measures. By conducting the validation experiments at optimum treatments, it was concluded that the developed models could precisely fit to the models developed with acceptable values of percentage errors (R^2 and adjusted R^2). Optimization was carried out by RSM and the major findings are: Undoubtedly RSM is a good technique to provide optimum treatments of a process by studying the effect of main factors and their interactions on response with minimum number of experiments in contrast to classical method of changing of one variable at a time. An alternative modeling approach to optimize the multi-response system based on desirability functions has been presented. The approach proposed aims to identify the settings of factors to maximize the overall minimal level of satisfaction with respect to all the responses. The optimal treatments algorithms are shown in Table 5.

References

[1] IEC 60270 Ed. 2. *High-voltage test techniques - Partial discharge measurements*. (2000). 15 – 16.

[2] William A, *Electrical Power Cable Engineering*. Ed. Marcel Dekker, Inc. (1999). 219 – 221.

[3] A. J. McGrail, E. G. *Data mining techniques to assess the condition of high voltage electrical plant*. CIGRÉ. (2002). 01 – 12.

[4] Mazroua, A. PD pattern recognition with neural netwoks using the multilayer perception technique. *IEEE Transactions on Electrical Insulation*, *28*, (1993). 1082-1089.

[5] Krivda, A. Automated Recognition of Partial Discharge. *IEEE Transactions on Dielectrics and Electrical Insulation*, 28, (1995). 796-821.

[6] Kim, J. Choi, W. Oh, S. Park, K. Grzybowski, S. Partial Discharge Pattern Recognition Using Fuzzy-Neural Networks (FNNs) Algorithm. I*EEE International Power Modulators and High Voltage Conference*. (2008). 272-275.

[7] Ri-Cheng, L. Kai, B. Chun, D. Shao-Yu, L. Gou-Zheng, X. Study on Partial Discharge Localization by Ultrasonic Measuring in Power Transformer Based on Particle Swarm Optimization. *International Conference on High Voltage Engineering and Application*. (2008). 600-603.

[8] Chang, W. Yang, H. Application of Self Organizing Map Approach to Partial Discharge Pattern Recognition of Cast-Resin Current Transformers. *WSEAS Transaction on Computer Research. Issue 3. Volume 3*. (2008). 142-151.

[9] Fadilah Ab Aziz, N. Hao, L. Lewin, P. Analysis of Partial Discharge Measurement Data Using a Support Vector Machine. *5th Student Conference on Research and Development*. (2007). 1-6.

[10] Hirose, H. Hikita, M. Ohtsuka, S. Tsuru, S. Ichimaru, J. Diagnosis of Electric Power Apparatus using the Decision Tree Method. *IEEE Transactions on Dielectrics and Electrical Insulation*, 15, (2008). 1252-1260.

[11] Venkatesh S., Gopal S., Jayalalitha S. Role of Partitioning based Clustering Algorithms in Classifying Multi-Source Partial Discharge Patterns Using Probabilistic Neural Networks and Its Adaptive Version – A Review. *Journal of Theorical and Applied Information Technology*. Vol. 36. No.1. (2012)

[12] Agamalov O. Cluster Analysis of Partial Discharge in Electrical Machines. CIGRE. A1-202. (2012) 10 – 19.

[13] Jaramillo-Vacio, R.; Ochoa-Zezzatti, A.; Rios-Lira, A.; Cordero, D., "A comparative study of partial discharge by classification's kind," *12th International Conference on Hybrid Intelligent Systems (HIS)*, vol., no., pp.34,40, 4-7 Dec. 2012

[14] Forssén, C. *Modelling of cavity partial discharges at variable applied frequency*. Sweden: Doctoral Thesis in Electrical Systems. KTH Electrical Engineering. (2008).

[15] Edin, H. *Partial discharge studies with variable frequency of the applied voltage*. Sweden: Doctoral Thesis in Electrical Systems. KTH Electrical Engineering. (2001).

[16] K. Lai, B. P. Descriptive Data Mining of Partial Discharge using Decision Tree with genetic algorithms. *AUPEC*. (2008).

[17] Markalous, S. *Detection and location of Partial Discharges in Power Transformers using acoustic and electromagnetic signals*. Stuttgart University: PhD Thesis. (2006).

[18] Kohonen T. Engineering Applications of Self Organizing Map. *Proceedings of the IEEE*. (1996).

[19] Rubio-Sánchez, M. *Nuevos Métodos para Análisis Visual de Mapas Auto-organizativos*. PhD Thesis. Madrid Politechnic University. (2004).

[20] Vesanto J., Alhoniemi E., Clustering of the Self Organizing Map. *IEEE Transactions on Neural Networks* , Vol. 11, Nr. 3, pages 1082-1089. (2000).

[21] Pölzlbauer, G. Survey and Comparison of Quality Measures for Delf-Organizing Maps. *Proceedings of the Fifth Workshop on Data Analysis* (2004). 67-82.

[22] C. Ahalt, A. K. Krishnamurthy, P. Chen, and D. E. Melton "*Competitive learning algorithms for vector quantization*", Neural Networks, vol. 3, no. 3, pp. 277-290. (1990).

[23] Dean A., Voss D. Design and Analysis of Experiments. Springer. (1999)

[24] Fristak V. Remenarova L. Lesny. Response surface methodology as optimization tool in study of competitive effect of ca2+ and mg2+ ions in sorption process of Co2+ by dried activated sludge. Journal of Microbiology Biotechnology and Food Science. Vol. 1. Nr. 5. (2012) 1235-1249.