Design and Implementation of Salinity Map for a Specific Area Using Statistical Parameters and Genetic Algorithm

Dr. Sattar B. Sadkhn 1*, Samaher Hussein Ali 2**, Ibtisam A. Rasheed 3***

*Computers Science Dept., Babylon University

drengsattar@yahoo.com

**Computers Science Dept., Babylon University

Samaher_hussein@yahoo.com

***Computers Science Dept., Babylon University

ibtisamalasady@googlemail.com

Abstract: This research presents a method to design a salinity map which describes the average of salinity in a specific area by using hybrid method represent by statistical Parameters and genetic algorithms. First, we input a samples of soil and analysis these samples in label then using statistical Parameters to extraction features from these samples (i.e., in this research, we extraction five types of features). These features are (moisture content, weight of pore or site water sample(water plus salt), weight of particulates in 1kg of sediment, weight of salt needed to bring dilution water for 1kg of sediment to the measured pore water salinity, weight of salt to bring column slurry to pore water salinity). After that, we using the genetic algorithm to design the salinity map of these equation's values and using the prepositional colors to represent the map where each area is characteristic by special color.

Key words: Salinity, Salinity Map, Statistical Parameters, Genetic Algorithm.

Introduction

The presence of salts in soil and water can be assessed by measuring salinity, the concentration of soluble salt in a soil, and sodicity, the relative concentration of Na+ compared to Ca+ and Mg+.

Salinity is most commonly measured with an electrical conductivity (EC) meter that estimates the concentration of soluble salts in a soil slurry or water solution by how well an electrical current passes through the medium. The ability of solution to conduct electricity increases with increasing salt content; therefore, a high EC value corresponds with high amounts of soluble salts, and Vice versa. EC values can be expressed in micromhos / (μ mhos/Cm), millim hos per centimeter (mmhos/Cm), or decisiemens per meter (ds/m). In seep discharge areas, soil samples should be taken from the 0 - 6 inch and 6 - 12 inch depths.

It has been known since at least the beginning of recorded history of crop cultivation that salts have an

affect on agriculture. Over five thousand year ago, the people of Mesopotamia farmed. The Fertile Crescent of the tigris and Euphrates Rivers (in modern day Turkey and Iraq), some of the richest farmland in the world at the time. As salt began accumulating in the soil, due to inadequate leaching and drainage of irrigation waters, the farming culture changed from growing wheat and barley to growing only salttolerant barley . Eventually, as salt took over, nothing grew in these once fertile valleys, and the land was abandoned [1]. Just before the time of Christ, Romans plowed the fields of conquered Carthage and applied salt to the fields, in an attempt to prevent the Carthaginians from reestablishing themselves. Efforts to recolonize the area 24 years later failed because the salted fields remaned unproductive [1] . Salts have been a known problem for thousands of years, particularly in arid and semiarid areas where there is insufficient rainfall to leach salts from the root zone [2]. During the last decade. In the Colorado River Basin alone, (eastern Utah, western Colorado, and Arizona) salt have cost agriculture more than \$100 million a year [3].

Irrigation water always includes some amount of dissolved substances, collectively called salts. These salts include dissolved solids derived from the weathering of rocks and soil by water and salts dissolved from soil the water previously passed through. Lime. Gypsum and other salt sources are dissolved over time, leading to varying degrees of salinity in irrigation water [2]. Whether or not water is suitable for irrigation, in terms of salinity, depends primarily on the kind and amounts of salts present, the soil type in question, specific plant species and growth stage, and the amount of water leached beneath the root zone, or the leaching fraction (LF) [4], [5].

1. Evolutionary Algorithms 1.1. Introduction:

Evolutionary algorithms are stochastic search methods that mimic the metaphor of natural biological evolution. Figure (1) shows how can you Problem Solution using Evolutionary Algorithm.

Different main schools of evolutionary algorithms evolved during the last 30 years: genetic algorithms mainly developed in the USA by J. H. Holland, evolutionary strategies, developed in Germany by I. Rechenberg & H.P. Schwefel, and evolutionary programming each of these constitutes a different approach; however, they are inspired in the same principles of natural evolution.

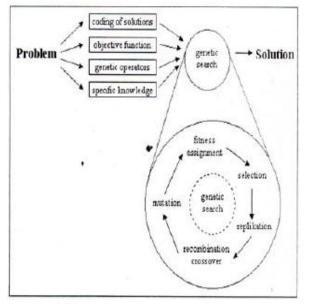


Fig (1): Problem Solution using Evolutionary Algorithm

1-Selection of the fittest individuals, yielding the so-called gene pool.

- 2- Recombination / crossover of the previously selected individuals forming the gene pool, giving rise to an offspring of new individuals.
- 3- Mutation of the newly created individuals.

By iterating this three – step mechanism. it is hoped that increasingly better individuals will be found. This reasoning is based on the following ideas

• The selection of the fittest individuals ensures that only the best ones will be allowed to have offspring. Driving the search towards good solutions.

• By recombining the genetic material of these selected individuals. The possibility of obtaining an offspring where at least one child is better than any of its parents is high.

4- Replacement criterion is the last operation that basically says which elements, among those in the current gene-pool and their newly generated offspring are to be given a chance of survival on to the next generated. And there are two basic strategies of replacement [6].

• The plus strategy:-denoted by $(\mu + \lambda)$ where the letter μ denotes the population size and the letter $\lambda > \mu$ the number of offspring to be generated out of the μ elements. In this strategy both the parents and their offspring will be taken into account to form a new generating of agin μ elements.

• The comma strategy: - denoted by (μ, λ) the parents after generating offspring die off and are not taken into account to form the next generation.

1.2. Overview

Evolutionary algorithms operate on a population of potential solutions applying the principle of survival of the fittest to produce better and better approximations to a solution. At each generation, a new set of approximations is created by the process of selecting individuals according to their level of fitness in the problem domain and breeding them together using operators borrowed from natural genetics. This process leads to the evolution of populations of individuals that are better suited to their environment than the individuals that they were created from, just as in natural adaptation.

Evolutionary algorithms model natural processes, such as selection, recombination, mutation, migration, locality and neighborhood. Figure (2) shows the structure of a simple genetic algorithm. Evolutionary algorithms work on populations of individuals instead of single solutions. In this way the search is performed in a parallel manner.

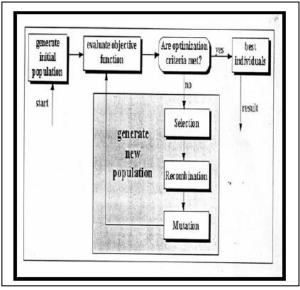


Fig (2): Flowchart of a Single Population Evolutionary Algorithm

At the beginning of the computation a number of individuals (the population) are randomly initialized. The objective function is them evaluated for these individuals. The first / initial generation is produced. If the optimization criteria are not met the creation of a new generation starts. Individuals are selected according to their fitness for the production of offspring. Parents are recombined to produce offspring. All offspring will be mutated with a certain probability. The fitness of the offspring is them computed. The offspring are inserted into the population replacing the parents, producing a new generation. This cycle is performed until the optimization criteria are reached. Such a single population evolutionary algorithm is powerful and performs well on a broad class of problems. However, better results can be obtained by introducing many populations, called subpopulations. Every subpopulation evolves for a few generations isolated (like the single population evolutionary algorithm) before one or more individuals are exchanged between

the subpopulations. The Multi population evolutionary

algorithm models the evolution of a species in a way more similar to natural that the single population evolutionary algorithm. From the about discussion, it can be seen that evolutionary algorithms differ substantially from more traditional search and optimization methods. The most significant differences are[7]:

• Evolutionary algorithms search a population of points in parallel, not a single point.

• Evolutionary algorithms do not require derivative information or other auxiliary knowledge; only the objective function and corresponding fitness levels influence the directions of search.

• Evolutionary algorithms use probabilistic transition rules, not deterministic ones.

• Evolutionary algorithms are generally more straightforward to apply.

• Evolutionary algorithms can provide a number of potential solutions to a given problem. The final choice is left to the user. (Thus, in cases where the particular problem does not have one individual solution, for example a family of pareto – optimal solutions, as in the case of multiobjective optimization and scheduling problems, then the evolutionary algorithm is potentially useful for identifying these alternative solutions simultaneously.

1.3. Implementation Details

A population of individual is maintained within search space for a GA. Each representing a possible solution to a given problem. Each individual is coded as a finite length vector of characters. A fitness value is assigned to each solution representing the ability of an individual to "compete". The goal is to produce an individual with the fitness value close to the optimal.

By combining information from chromosome, selective "breeding" of individual is utilized to produce "offspring" better than the parents. Continuous improvement of average fitness value from generation to generation is achieved by using the

genetic operators. The basic genetic operators are[8][9]:-

- Selection used to achieve the survival of the fitness.
- Crossover used for mating between individual.
- Mutation used to introduce random modifications.

The genetic operators are used in the genetic algorithm optimization procedure according to the flowchart given in figure (3). It is not necessary to employ all of these operators in a genetic algorithm because each operates independently of the other. The

choice or design of operators depends on the problem and the representation scheme employed. For instance, operators designed for binary strings cannot be directly used on strings coded with integers or real number.

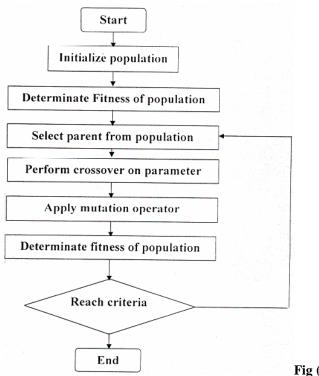


Fig (3): Flowchart of a Simple Genetic Algorithm

2. Salinity Determination

Salinity is most commonly measured with an electrical conductivity (EC) meter that estimates the concentration of soluble salts in a soil slurry or water solution by how well an electrical current passes through the medium.

To Determine salinity using a bucket of sediment that has separated into solid and liquid phases, determine the salinity or the Total Dissolved Solids (TDS) concentration of the pore water or dredging site

water using a salinity or TDS meter or gravimetrically, If using a meter[10]:

A. if needed, clean the probe with a solution of nitric acid.

B. Following the instructions for the meter, calibrate the meter using water with a known salinity. Be sure to check the batteries and determine the temperature of the liquid.

C. Determine the salinity /TDS of the pore water or dredging site water in ppt .

3. The Proposed Method

The objective of this paper is present a method to design salinity map using hybrid technique represented by statistical Parameters and genetic algorithms, this search concerns itself with design salinity map of specific area in our country Iraq and we can say its the first method using evaluation algorithm represent by genetic algorithm in design salinity map. Then we can consider it the seed in this field and we hope to develop this method in the future. The main steps of the proposed method are explained in Figure 4 and these steps will discussed in details:

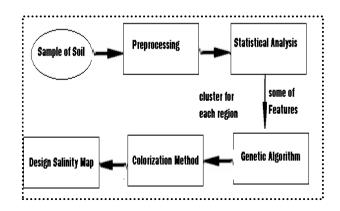


Fig (4): A Block Diagram of Proposed Method

3.1. Preprocessing

According to this method, one needs input a samples of soil and analysis these samples in label to get the initial values of salt for each sample. For this purpose, the samples were take into different regions. For now, many point have been defined for the specific area (i.e., each point represent sample). These points are necessary for describing a salinity map. The salts of these points are given in Table 1.

3.2 Statistical Analysis

After we got the salt expected in each sample we can extracted some of the coefficients(features). In this research, we extraction five types of features from these samples. These features are (moisture content, weight of pore or site water sample (i.e., water plus salt), weight of particulates in 1kg of sediment, weight of salt needed to bring dilution water for 1kg of sediment to the measured pore water salinity, weight of salt to bring column slurry to pore water salinity). Then use these 5 values as feature vectors of genetic algorithm.

In this work, we use the following equations to determine the salinity before design salinity map of any specific area:-

A. Calculate the weight of pore or site water sample (Water + salt)

$$W_s = \frac{W_w S}{1000 - S} \tag{1}$$

Where: Ws= weight of salt in sample, g

$$Wpw = \frac{(1000 * Ws)}{S}$$
(2)

Where S = salinity of pore water, ppt Wpw = weight of pore or site water sample (water + salt), g.

B. Compute average moisture content. Average the three particulates concentrations to obtain the average particulates concentration of the sediment. Compute the average moisture content of the sediment sample based on the average particulates concentration using the following equation[11]:

The points	The Salt
491	4.0
584	53.3
555	15.2
492	20.6
511	5.3
588	17.6
456	6.2
504	16.0
460	15.2
459	6.7
509	1.4
458	20.1
464	30.6
452	20.3
503	11.5
476	7.3
550	4.7
515	5.4
471	14.4
470	8.1
465	31.5
472	32.5
559	1.5
519	5.0
554	13.8
453	2.9
479	4.2
527	5.0
512	12.5
496	8.8

Table (1) : The Salt of each Sample

1. Compute the weight of particulates in each sample using the following equation

$$W_p = W_d - W_s \tag{3}$$

Where Wp = weight of dry particulates, g.

Wd = weight of dried sample (particulates and salt), g

2. Determine the particulate concentration of each sample using the following equation

$$C_{p} = \frac{1000 W_{p}}{(W_{p} / G_{s}) + W_{w}}$$
(4)

Where Cp = particulates concentration of sediment/L Gs = specific gravity of particulates

Ww= weight of water in the sample, g.

$$w = \left(\frac{1000 - \left(C_p / G_s\right)}{C_p}\right) \ 100\% \tag{5}$$

3. Determine the moisture content using the following equation

Where w =moisture content, %

Cp = particulates concentration of sediment, g/L

Gs = specific gravity of particulates.

C. compute unit weight of sediment. Calculate the weight of pore water, salt, and particulates in 1 kg of sediment as follows[12]:

1. Calculate the weight of the pore water in 1 kg of sediment:

$$W_{wkg} = \frac{1000 \, w}{100 + w} \tag{6}$$

Where Wwkg = weight of pore water in 1 kg of sediment, g/kg 2. Calculate the weight of salt in 1 kg of sediment:

$$W_{skg} = \frac{W_{wkg} S}{1000 - S} \tag{7}$$

Where Wskg = weight of salt in 1 kg of sediment, g/kg

3. Calculate the weight of the particulates in 1 kg of sediment:

$$W_{pkg} = 1000 - W_{wkg} - W_{skg} \tag{8}$$

Where Wpkg = weight of particulates in 1 kg of sediment, g/L.

D. Compute unit requirements for dilution of sediment. Determine the amounts needed to dilute 1 kg of sediment to the desired slurry concentration as follows:

1. Calculate the total volume after dilution of 1 kg of sediment:

$$V_{kg} = \frac{W_{pkg}}{C_{sl}} \tag{9}$$

Where Vkg = total volume of slurry created from 1 kg of sediment, L/kg 2. Calculate the volume of pore water in 1 kg of sediment:

$$V_{pwkg} = \frac{W_{wkg}}{1000}$$
(10)

L/kg.

3. Calculate the volume of particulates in 1 kg of sediment:

$$V_{pkg} = \frac{W_{pkg}}{1000 G_s} \tag{11}$$

Where Vpkg = volume of particulates in 1 kg of sediment, L/kg. 4. Calculate the volume of dilution water per 1 kg of sediment:

$$V_{dwkg} = V_{kg} - V_{pwkg} - V_{pkg}$$
(12)
Where

Where Vdwkg = volume of tap water needed to dilute 1 kg of

Where Vpwkg = volume of pore water in 1 kg of sediment,

sediment to desired slurry concentration, L/kg.5. Calculate the weight of salt required to slurry 1 kg of sediment:

$$W_{dskg} = \frac{1000 \, S \, V_{dukg}}{\left(1000 - S\right)} \tag{13}$$

(1000 - 0) Where Wdskg = weight of salt needed to bring dilution water for 1 kg of sediment to the measured pore water salinity, g/kg.

E. compute slurry requirements. Determine the amounts of sediment, tap water, and salt to mix the required volume of slurry to fill the settling column to the desired height[13].

1. Calculate the required weight of sediment: $W_t = V_t / V_{kg}$ (14)

Where Wt = required total weight of sediment sample for settling test, kg. Vt = volume of slurry required for settling test, L (typically, 90 to 120L)

(15)

2. Calculated the volume of tap water needed to create slurry for settling test

$$V_{wt} = W_t V_{dwkg}$$

Where Vwt = volume of tap water, L.

3. Calculate the weight of salt needed to create slurry for settling test: $W_{st} = W_t W_{dskg}$ (16)

Where Wst = weight of salt to bring column slurry to pore water salinity, g.

3.3. Genetic Algorithm for Design Salinity Map

Genetic algorithms (GAs) belong to a class of search techniques that mimic the principles of natural selection to develop solutions of large optimization problems. GAs operates by maintaining and manipulating population of potential solutions called chromosomes. Each chromosome has an associated fitness value which is a qualitative measure of the goodness of the solution encoded in it. This fitness value is used to guide the stochastic selection of chromosomes which are then used to generate new candidate solutions through crossover and mutation. Crossover generates new chromosomes by combining sections of two or more selected parents. Mutation acts by randomly selecting genes which are then altered; thereby preventing suboptimal solutions from persisting and increases diversity in the population. The process of selection, crossover and mutation continues for a fixed number of generations or until a termination condition is satisfied [14].

3.3.1. Representation of Solution

The chromosomes are make up of real values (representing the values of the statistical analysis drawing from salinity equations). The length of a chromosome equal N genes while the length of gene is taken to be five numbers (i.e., five numbers represents a values of moisture content, weight of pore or site water sample (water plus salt), weight of particulates in 1kg of sediment, weight of salt needed to bring dilution water for 1kg of sediment to the measured pore water salinity, weight of salt to bring column slurry to pore water salinity) where each individual gene position divide into five locations each one represents one of these values. The representation of gene is shown in Figure(5).

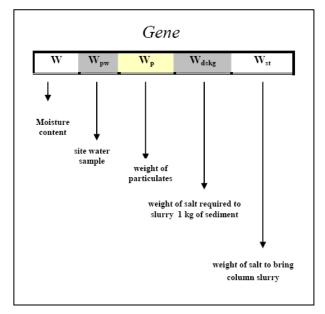


Fig (5): The Representation of Gene

3.3.2. Fitness Function

The fitness of a chromosome is computed using the average Euclidean distance of the vectors, is computed as

$$d(X, Z^{(i)}) = \sum_{k=1}^{K} w_k \left| \frac{(X_k - Z_k^{(i)})}{r_k} \right|$$
(17)

Where, k=1,2,...,K, x is feature vector, Z(i) is center of gen i, k number of features in feature vector (in this work k=5), wk is weight of feature k (0<wk<1), rk is the range of feature k. The objective is to minimize the d(x,z(i)) for achieving proper clustering. The fitness function for chromosome j is defined as 1/ dj(x,z(i)), where dj is the Euclidean distance computed for this chromosome, where the maximization of the fitness function will ensure minimization of the Euclidean distance. **3.3.3. Selection** The selection used to a achieve the survival of the fitness. It's mechanism favors the individual with high fitness value. It allows these individual better chance for reproduction into next generation. There are several selection methods (e.g., Roulette wheel, Ranking-based selection...ect). in this paper, we used "Roulette wheel" to select two pair of chromosome.

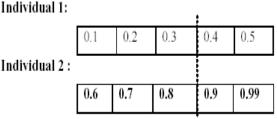
3.3.4. Crossover

The crossover operator divides a population into pairs of the individual and performs recombination of their genes with a certain probability. Crossover is performed immediately after selection and reproduction. And it is used to combine the pairs of selected strings (parents) to create new strings that potentially have a higher fitness than either of their parents. It is usually performed with a probability called crossover probability (pc) usually is chosen to be near one to preserve some of the good strings found previously. This operation dose not change the value of bits. There are several ways of doing the crossover. Some common crossover operations are single – point crossover, multi point– crossover and uniform crossover. In this search, we used a **Single – point crossover (1x):**

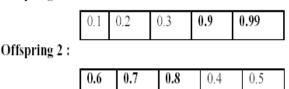
In single – point crossover one crossover position Q [1. 2. N-1], Q: number of variables of an individual, is selected uniformly at random and the variables exchanged between the individual about this point, then two new offspring are produced.

Example:

Consider the following two individuals each individual have five genes :



Then, let the crossover position equal 3 (x=3) after crossover the new individuals are created: Offspring 1:



3.3.5. Mutation

The mutation operator is applied to every string resulting from the crossover process. When using mutation operator a portion of the new individuals will have some of their bits flipped with a predefined probability (mutation probability Pm) usually quit low, e.g. 0.001. The propose of mutation is to maintain diversity within the population and prevent premature convergence. The usage of this operator allows the search of some reigns of the search space, which would be otherwise unreachable. Eventually, it helps the genetic algorithm avoid premature convergence, getting trapped at local optima and find the global optimal solution.

3.3.6. Elitism

When creating new population by genetic algorithm processes, we might lose the best chromosome since the selection of chromosomes (or candidate solutions) is more or less done at random. Elitism is the name of method, which first copies the best chromosome (or a few best chromosomes) to new population for further evolution.

Elitism can very rapidly increase performance of GA because it prevents losing the best found solution. We have implemented elitism at each generation by preserving the best string seen up to that generation in a location outside the population. Thus on termination, this location contains the best values of the final result of salinity equations.

4. Result

In our research, we attempted to design salinity map of different types of samples using hybrid method represent by statistical equations and genetic algorithms. We used statistical equation values to represent the different genes in chromosomes.

How these method works? To answer about this query first we need to know the role of each tool in this method and we thank a paragraph number 3 illustrations these tool in detail.

Genetic algorithm applied to find the best values of the final results of salinity equations. Before this, we need to determine some of parameters relate to GA such as (population size= 100 individuals, the probability of

crossover=90%, the probability of mutation = 10%, number of generation =75). And by experiment we find fitness of best solution=0.92119.

After we get the best values of salinity equations by genetic algorithm. Colorization each gene in best chromosome (i.e., the chromosome have the maximization value of fitness function in the last generation) with specific color.

In this work, an salinity map is colorization depend on the salinity ratio where red and green color represent the filed has higher salinity while the other color represent the filed has middle or lower salinity figure(6) explains result of the proposed method.

In this research, the length of chromosome equal 14 genes and a gene itself divide into five location each location represents one of the salinity features. Therefore the salinity map of the study area is divide into 14 region each region represents one gene in chromosome.

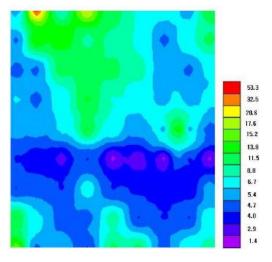


Fig (6): Salinity Map of Study Area

5. Conclusion

The relationship between salinity and genetic algorithm is complex and unknown. But from above steps, we can say the basic principles of compute salinity, depend on salts, and the basic principles of design (determined) salinity map, depend on genetic algorithm and colorization method, are grasped, the relationship between salinity and algorithm becomes apparent.

By experiments, we see the following points :

- 1. The features of salinity may be compute manually with out using computer techniques in lab but the manual method takes more time and also may be lead to error in computation values. Therefore the using of computer less from the time required to compute these features and the probability of down in error (i.e., the results givens more accuracy with short time).
- 2. The salinity features that used in this research succeeds in description all sample in study area where the genetic algorithm enables from determine all salinity regions. And we can consider this features is enough to determination any salinity map.
- 3. The GA is given already the optimal solution of any problem therefore by merge with mathematical equations, we obtain optimal salinity map and also we enable from determined salinity map more accuracy and generality.

Future work, We can used another types of Evolution Algorithms such as Breeder Genetic Algorithm or Genetic Programming or evaluation strategies. Also we can use another types of fitness function such as Minkonasky Distance. Finally, we can extraction another type of mathematical equations related to salinity computation.

REFERENCES

[1] Hadas, Moses, 1965. Imperial Rome. Time, New York.

pp. 38-39.

[2] Miller, R.W. and R.L. Donahue, 1995. Soils in Our Environment, Seventh Edition. Prudence Hall, Englewood, Cliffs, NJ. p. 323.

[3] Anonymous, 1989. California's Problem Passing Water, Agrichemical Age, 33, no. 11.

[4] Bauder, J.W, 2001. Interpretation of chemical analysis of irrigation water and water considered for land spreading. Personal communication. Montana State University, Bozeman, Montana.

[5] USDA, Natural Resources Conservation Service, 2002. Soil Conservationists. Salinity Management Guide – Salt Management. Available: http://www.launionsweb.org/sali nity.htm. 2002.

[6] Samaher H. Ali, 2005. Designing a Software for Classifying Objects for Air Photos & Satellite Images using Soft Computing ,M.Sc. Thesis. Babylon University.

[7] Quagliarella.D, Periaux.J, Poloni.C, Winter.G, 1998. Genetic Algorithms and Evoluation Strategy in Engineering and Computer Science, John Wiley and Sons, New Yourk.

[8] Nizar.G,Michel.Cand Nazha.B, 2004. Unsupervised and Semi-Supervised Clustering :aBrief Survey, France.

[9] Sanghamitra.B and Ujjwal.M, 2002 .Genetic Clustering
 For Automatic Evolution OF Clusters and Application to Image Classification, the Journal of the Pattern Recognition Society, Vol.35, PP. 1197-1208. Site:http://www.elsevier.com/locate/patcog

[10] Nikos J. Warrence, Dr. James W. Bauder, and Krista E. Pearson, 2002. Basics of Salinity and Sodicity Effects on Soil Physical Properties. Department of Land Resources and Environmental Sciences, Montanta State University- Bozeman.

[11] Jeffrey C. Silvertooth, 2001. Soil Management and Soil Testing for Irrigated Cotton Production. The University of Arizona, college of Agriculture and life sciences.

[12] Jeffrey G. Paine, 2003. Determining salinization extent,

identifying salinity sources, and estimating chloride mass using surface, borehole, and airborne electromagnetic induction methods. Water Resources Research, v. 39, no. 3, p. 3-1 -- 3-10.

[13] John S. Adams, 2005. Nueces Bay Salinity Monitoring Project. Texas A&M University-Corpus Christi.

[14] Tawfiq A, Samaher H. Ali and Israa H, 2007. Object Oriented Classification of Forest Images Using Soft Computing Approach, IEEE, 4rth International Conference: Sciences of Electronic, Technologies of Information and Telecommunications March 25-29, 2007 – TUNISIA