

# SPATIAL ANALYSIS OPTIMISATION USING GENETIC ALGORITHM

**Mohamad M. Awad, Mohamad Khawlie**

**Abstract** Spatial analysis is used to help find solutions for geographical issues, it is considered the most important tool in GIS. Several programming techniques and methods are being used to optimise the solutions produced by spatial analysis tools in GIS. One example of these techniques is the Solver in Excel, which is an add-in tool. Solver adjusts the values in the cells one specifies to produce the result one wants from the formula of the objective function, which is subject to some constraints. Genetic Algorithm (GA) is one of the natural optimization methods, which produces many feasible optimal solutions they are better than the other linear programming tools, such as Solver in Excel, by selecting them produces only one solution which, in most cases is not an optimal solution and far in value from the closest available features. The paper tries to prove this idea by selecting several sights of wildlife habitats in Lebanon using ArcInfo and Arcview GIS software. A comparison between a Modified Genetic Algorithm (MGA) and Solver is carried out to find the Maximum area for a wildlife habitat with the lowest cost. This comparison is aimed to prove that (MGA) finds many solutions which are closer in area values to the available wild life habitat locations than the solution produced by Excel Solver. In addition, the cost of an (MGA) solution is always similar or less than the cost of Excel Solver solution.

**Keywords :** Geographic Information System (GIS), wildlife habitat, Genetic Algorithm, artificial intelligence, optimisation algorithm.

## 1 Introduction

In the early days of spatial analysis the analaog maps required rulers to measure distances, compasses to find directions and dot grids or planimeters to measure areas (Marble, 1990). Today, GIS spatial analysis is a set of tools that professionals in various disciplines (geography, urban planning, engineering, data processing, landscape architecture, environmental science, etc.) use to improve the way they work. These tools include automated mapping technology, data-base management, land records

information, topological data structures, and other spatial analysis capabilities (Huxhold, 1991). In other words, GIS analysis uses the power of the modern digital computer to measure, compare, and describe the contents of the databases. It allows ready access to the raw data, and allows aggregation and reclassification for further analysis. It can combine selected data sets in unique and useful ways far beyond a traditional map could provide on a single sheet (DeMers, 1991). Genetic Algorithms are based on the mechanics of natural evolution (Goldberg, 1989). They mimic natural populations reproduction and selection operations to achieve efficient and robust optimization. Through their artificial evolution, successive generations search for beneficial adaptations in order to solve a problem. Each generation consists of a population of chromosomes, also called individuals, and each chromosome represents a possible solution to the problem. The initial generation consists of randomly created individuals. Each individual acquires a fitness level, which is usually based on a cost function given by the problem under consideration.

The Darwinian principle of reproduction and survival of the fittest, and the genetic operations of recombination (crossover) and mutation are used to create a new offspring population from the current population. The reproduction operation involves selecting, in proportion to fitness, a chromosome from the current population of chromosomes, and allowing it to survive by copying it into the new population. Then, two mates are randomly selected from this population, and crossover and mutation are carried out to create two new offspring chromosomes. Crossover involves swapping two randomly located sub-chromosomes (within the same boundaries) of the two mating chromosomes. Mutation is applied to randomly selected gene, where the value associated with such a gene is randomly changed to another value within an allowed range. The offspring population replaces the parent population, and the process is repeated for many generations with the aim of maximizing the fitness of the individuals.

Genetic Algorithms are different from traditional methods in many ways, as they work with a coding of the parameters set and not with the parameters themselves. They search from a population of points not a single point. They have been adapted for solving a variety of engineering, science, and operational research problems. Some examples of such applications can be found (Fogel, 1998), and (Pham, 2000). An outline of Genetic Algorithm (GA) is given in Figure 1.

```
Random generation of initial population, size POP;  
Evaluate fitness of individuals;  
Repeat  
  Rank individuals and allocate reproduction trials  
  For(I=1 to POP step 2) do  
    Randomly select 2 parents from the list of reproduction trials  
    Apply crossover and mutation;  
  Endfor  
  Evaluate fitness of offsprings;  
Until(convergence criterion is satisfied)  
Solution = saved best feasible and the Fittest
```

Figure 1. Genetic Algorithm outline

In order to make Genetic Algorithm more efficient in comparison with Excel Solver, it is modified to save the best so far solution, this modification is called Modified Genetic Algorithm (MGA). In this paper GIS is used as a spatial analysis tool to locate and find how many areas are suitable for wildlife habitat in Lebanon. Esri's ArcInfo and Arcview GIS softwares are used to perform this task which begins by determining the suitable areas according to specific criteria, and the overlay of several layers.

Later, an optimal wild life habitat is to be found which will maximize the area and minimize the cost by using one of the important methods of linear programming, which is Excel Solver, and one of the important natural evolution algorithms which is Modified Genetic Algorithm (MGA). We are going to prove that (MGA) finds more accurate and feasible solutions than the other conventional linear programming methods.

This paper is organized as follows. Section 2 describes locating wild life habitat in Lebanon. Section 3 Linear programming methods to locate an optimal area. Section 4 The Modified Genetic Algorithm method to locate an optimal area. Section 5 contains the conclusion.

## 2 Locating wildlife habitat in Lebanon

Lebanon's geographic location, physiography and its richness in natural resources, especially forests and water, have made its terrain a refuge for many different migrating animals and birds. However, the lack of information in the government sectors and lack of public awerness about the importance of saving these species resulted in destruction of habitat, and several sepecies have disapeared since many decades. This study is an attempt to find the best place as a wildlife habitat for protecting mountain animals.

The first step is to create the forest layer and other relevant natural resources layers from the land cover map (17 classes). In addition, the size of the forest to be protected depends on the number of mountain animals (here the number of those animals is assumed to be 200). It is assumed that each animal needs at least 500 m<sup>2</sup> of forests, and 10 m wide and 5 to 15 meter in length on the field which has common border with the forest. These edge lands are needed by the animals to escape from predators, from forest fires, or other dangers. However, the total area should not exceed 200000 m<sup>2</sup>. The second step, a 10 meter wide buffer is made in the forest layer, and a 1 meter wide buffer is made in the fields layer. The last step, the two buffers are intersected to create edged fields 11 meter in width, and the edged fields is erased by the original forest layer to create the final edged fields with forest index. Figure 2 illustrates this process and the process of creating forest and fields layers. Another constraint for managing this wildlife habitat is the financial budget , which should not exceed \$75000 per year (\$ 0.25 / m<sup>2</sup> for forests and \$1 / m<sup>2</sup> for fields). The objective is to maximize the size of the reserve while keeping the project economically feasible.

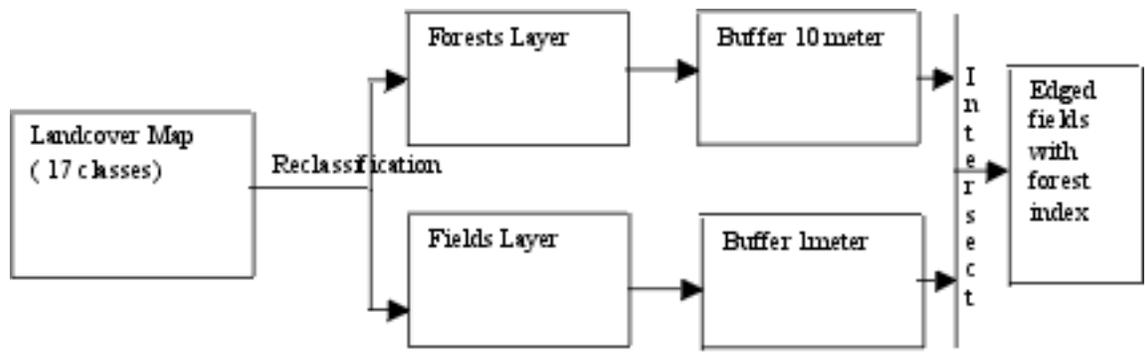


Figure 2. Selection and interaction of different layers for locating habitats

## 3 Linear and nonlinear programming methods to locate an optimal wildlife habitat

Linear and nonlinear programming and GIS tools available today afford better implementation and integration of spatial data analysis so that optimally suitable sites can be identified. Or at least sites can

be rated as to their closeness (Guerra, et al., 2002). Guerra's method is used to calculate the optimal solution for the wildlife habitat by using an add-in tool Microsoft Excel Solver. The Microsoft Excel Solver combines the functions of a graphical user interface (GUI), an algebraic modeling language like GAMS (Brooke, et al., 1992) or AMPL (Fourer, et al., 1993), and optimizers for linear, nonlinear, and integer programs. Optimization in Microsoft Excel begins with an ordinary spreadsheet model. The spreadsheet's formula language functions as the algebraic language used to define the model. Through the Solver's GUI, the user specifies an objective and constraints by pointing and clicking with a mouse, and filling in dialog boxes. The Solver then analyzes the complete optimization model and produces the matrix form required by the optimizers in much the same way that GAMS and AMPL do. The optimizers employ the simplex, generalized reduced gradient, and branch and bound methods to find an optimal solution and sensitivity information.

To use Solver in order to obtain an optimal solution, the following parameters, objective, and constraints must be defined.

Objective function = Max (E + F), where E is the edge field and F is the forest.

Criteria:

- 1- Field edge  $E \geq 10 \times L \times 200 \text{ m}^2$  ( here L is a variable [5..15])
- 2- Forest  $F \geq 200 \times 500 \text{ m}^2$
- 3- Wildlife habitat size  $\leq 200000 \text{ m}^2$
- 4- Cost  $\leq \$75000$

By taking  $L = 15$  and by using Excel Solver, the solution is  $E = 33333.3$ ,  $F = 166666.7$ , and the cost is \$ 75000.

From the above results, each forest polygon in the forests layer is given a value equal to the difference between its actual area and the computed area above (F), then Arcview is used to link both edged fields layer and forests layer. The new computed values in the forests layer are used to select the closest values to the above computed areas (E and F). One of these best selections is shown in Figure 3, which is not a feasible solution since it exceeds the maximum cost of \$75000. It also exceeds the maximum needed area for both layers which is  $200000 \text{ m}^2$ .

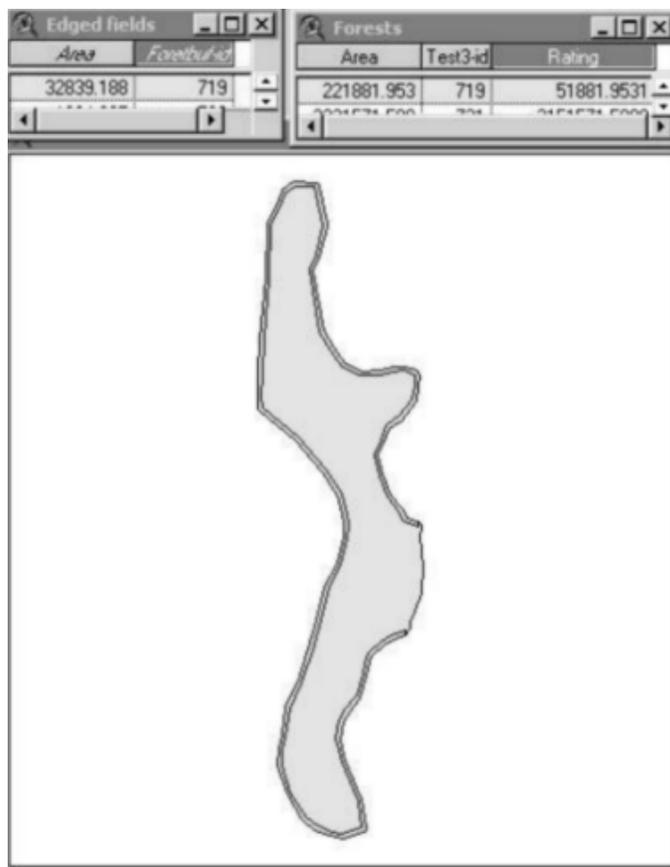


Figure 3. Solver result found by Arcview spatial analyst

If the problem is changed by taking  $L = 10$ , and Excel Solver is run, the result is again  $E = 33333.3$ ,  $F = 166666.7$ , and the cost is \$ 75000. Again the same results are obtained if Excel Solver is run for  $L = 5$ . The cost of the closest available polygons (in distance) to the computed value for  $L = 10$  is \$ 64892. Table 1 shows the solutions after Excel Solver is run for different values of  $L$ , and the polygons' identification (ID), which are selected from the forest and fields layers, and which have area values close or similar to the Excel Solver solutions

Table 1. Area values of Solver solutions and the selected polygons

Polygon ID	Area of Forest Solver	Area of Field Solver	Area of Forest (polygons)	Area of Field (polygons)
141	166666.7	33333.3	169395.56	22544.625
11	166666.7	33333.3	132624.609	20160.242
605	166666.7	33333.3	134833.563	35182.582

# 4 The Modified Genetic Algorithm method to locate an optimal habitat

Genetic Algorithm has been modified in order to increase and improve the number of solutions obtained. This is done by using a procedure which saves best so far of the population. In addition, another procedure is used which checks for duplicates before saving best so far solution, Figure 4. This is necessary because the same solution remains unchanged until after several iterations.

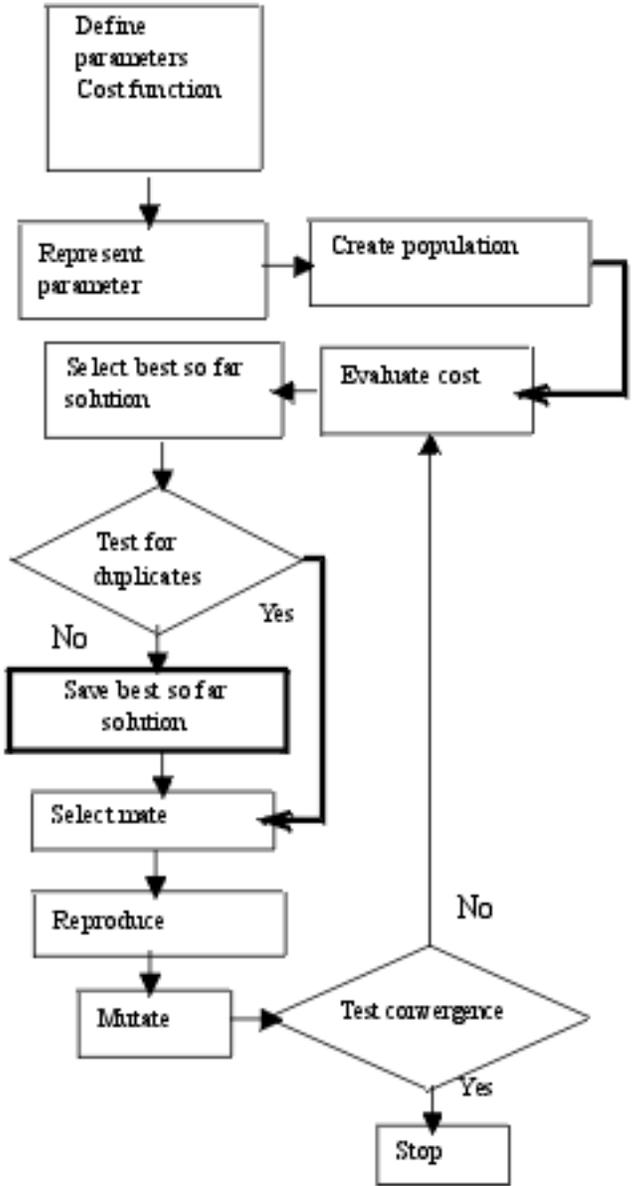


Figure 4. Modified Genetic Algorithm Outline

In the Modified Genetic Algorithm (MGA), the population number is 40 and the number of generations (iterations) is 100, and one point cross over is used. The constraints on the edge field area were varied (the same values for L = 15, 10, and 5 are used), to get several solutions by (MGA). In order to select the best one (the closest to the best available polygons) the selection index  $\psi$  formula (1) is used where  $0 < \psi \leq 1$ , and if  $\psi = 1$  this is the perfect match.

$$\psi = \frac{((F_t + E_t) - D_{f,e})}{(F_t + E_t)} (1)$$

$D_{f,e}$  is the distance between the selected polygons of the forest and edge layers which satisfy the criteria of minimum and maximum area, and the (MGA) solutions Figure 5.  $F_t$  is the area of each polygon of the selected forests, and  $E_t$  is the corresponding edge field ( according to the index).

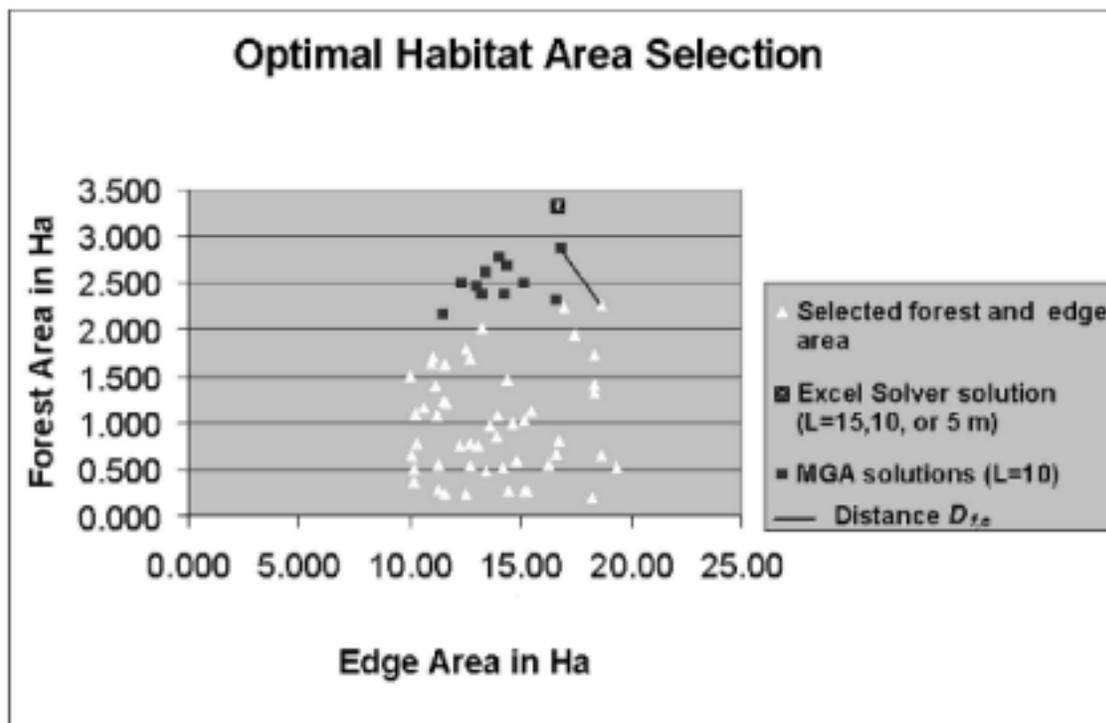


Figure 5, The distance  $D_{f,e}$  between forest (Y axis) and Field Area ( X axis) and (MGA) solutions

A program is written in Avenue language to read each of the results of (MGA) and to find the closest match to each of these solutions. These values are presented to the decision maker to choose one of them. Table 2 shows some of the values of the selected polygons and their corresponding solutions of (MGA).

Table 2. Area values of (MGA) solutions and the selected polygons

Polygon ID	Area of Forest (MGA)	Area of Field (MGA)	Area of Forest (polygons)	Area of Field (polygons)
141	166585	20539	169395.56	22544.625
11	123432	21463	132624.609	20160.242
605	123432	21463.2	134833.563	35182.582

In order to compare (MGA) Solutions and Solver Solution, the distance  $D_{f,e}$  and  $\Psi$  are computed for each selected polygon and the results are compared, Table 3 shows this comparison between (MGA) and Solver. It can be seen clearly that (MGA) solutions are closer to the available selected polygons in the layers of forest and fields than Solver's solution. This is due to the fact that MGA creates feasible population at the beginning. Other possibility is to build the population from the attribute tables of both layers. In addition, the cross over and mutation creates diversity and changes an infeasible solution to a feasible one. In Branch and Bound algorithm, which is used by Excel solver, considers the problem as a directed acyclic graph, where at each node a bound on the possible value of any solution that might lie further on in the graph is calculated. If the bound shows that any such solution must be worse than the best solution found so far, then one needs not go on exploring this part of the graph.

Table 3. Index  $\Psi$  and  $D_{f,e}$  values for Solver and (MGA)

Polygon ID	Index $\Psi$ Solver	Distance $D_{f,e}$ Solver	Index $\Psi$ (MGA)	Distance $D_{f,e}$ (MGA)
141	0.9420	11128.441	0.982	3452.795
11	0.7611	36501.964	0.9392	9284.491
605	0.8124	31886.807	0.8951	17838.64

## 5 Conclusion

The research does not aim to show how to manage wildlife habitats, but it aims to show how to optimize this management using new algorithmics techniques. In general, Spatial Analysis is used to select wildlife habitats, this selection is performed by using spatial themes overaly or according to their adjacency and proximity, but what if nonspatial criteria are used such as cost. This paper presents a new technique in which optimisation methods are combined with spatial analysis to obtain an optimal solution where an objective is given subject to constraints. These optimisation methods work with non spatial data, and the solution is located in the spatial themes using scripts that is written in Avenue language. This example is not a complex one, the objective however, can be changed and many constraints can be added to increase the complexity of wildlife habitats selection such as road network, elevation, NDVI for forest cover density, hydrographic network, aspect...etc. In addition, optimal sites are selected according to certain objective and constraints using Excel Solver and Modified Genetic Algorithm. However, these results show that (MGA) proved to be more accurate in predicting optimal solutions. This is due to the rich diversity of genetic information in (MGA) population which performs very well when applied to any complex problem.

## References

BROOKE, A., KENDICK, D., MEERAUS, A. (1992), *GAMS, A User's Guide*, Boyd and Fraser, Danvers, Massachusetts.

DEMERS, MICHAEL N. (1991): Classification and Purpose in Automated Vegetation Maps, *Journal of*

*Geographical Review* 81(3);267-280.

FOGEL, D. (ED.) (1998): *Proc. IEEE Int. Conference on Evolutionary Computation*, Piscataway, NJ : IEEE Press.

FOURER, R., GAY, D.M., KERNIGHAN, B.W. (1993), *AMPL: A Modeling Language for Mathematical Programming*, Duxbury Press, Pacific Grove, California.

GOLDBERG, D. E. (1989): *Genetic Algorithms in Search, Optimisation and Machine Learning*. Boston, Addison-Wesley, Reading.

GUERRA, G., LEWIS, J. (2002): Spatial Optimisation and GIS, *ArcUser* 5(2); 32-34.

HUXHOLD, W. E. (1991): *An Introduction to Urban Geographic Information Systems*. New York, Oxford University Press.

MARBLE, D. F. (1990): The Potential Methodological Impact of Geographic Information Systems on the Social Sciences. In *Interpreting Space GIS and Archaeology*, Allen, Zubrow, and Green, (eds). National Academy Press.

PHAM, D., KARABOGA, D. (2000): *Intelligent Optimisation Techniques*, London: Springer.

## **Author information**

### **Mohamad Awad**

National Council for Scientific Research

National Center for Remote Sensing

Lebanon, Beirut

[mawad@cnrs.edu.lb](mailto:mawad@cnrs.edu.lb)

### **Mohamad Khawlie**

National Council for Scientific Research

National Center for Remote Sensing

Lebanon, Beirut

[mkhawlie@cnrs.edu.lb](mailto:mkhawlie@cnrs.edu.lb)