GENETICLAND: MODELING LAND USE CHANGE USING EVOLUTIONARY ALGORITHMS

45th Congress of the European Regional Science Association, Land Use and Water Management in a Sustainable Network Society, Vrije Universiteit Amsterdam, 23-27 August 2005

Júlia SEIXAS[1*], João P. NUNES[1] Pedro LOURENÇO[1] Fernando LOBO[2], and Paulo CONDADO[2]

[1]Departmento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Monte de Caparica
2829-516 Caparica, Portugal; *mjs@fct.unl.pt

[2]Departamento de Engenharia Electrónica e Informática
Faculdade de Ciências e Tecnologia, Universidade do Algarve, Campus de Gambelas, 8000-117 Faro, Portugal

Abstract:

Land use planning concepts and methods have evolved new approaches due to the perception of the very long term impact of global change, particularly climate change. Future land use configurations provide valuable knowledge for policy makers and economic agents, especially under expected environmental changes such as decreasing rainfall or increasing temperatures. Considering the time frame requirements from climate change issues, usually 100 years, this paper proposes an optimization approach to study future land uses. Modelling land use change is designed as an optimization problem in which landscapes (land uses) are generated through the use of evolutionary algorithms (EA). GeneticLand is an evolutionary algorithm, designed for a multi-objective function, minimization of soil erosion, and maximization of carbon sequestration, and a set of local restrictions (e.g. physical constraints and landscape spatial structure). GeneticLand has been applied for a Mediterranean landscape, located in Southern Portugal. This paper presents the GeneticLand algorithm design and results obtained show the feasibility of the generated landscapes, whose main characteristic is an increase in spatial heterogeneity.

Keywords: land use, climate change, optimization, evolutionary computing, Mediterranean

This work was funded by the Portuguese Foundation for Science and Technology (FCT) under project GeneticLand - Discovering future landscapes under Climate Change Scenarios using Genetic Algorithms (Ref: POCTI/MGS/37970/2001).

1. Introduction

Land use models have been developed for the last decades to answer land use planning issues, such as: which types of land can be used, how much each can be utilized, for which activities at a given amount of resources, as well as the conditions of uses within particular circumstances. Several methodological approaches have been adopted to produce land use scenarios for a wide range of time scales, with emphasis on the next cycle of decision making policy (i.e. next years). In the last years, the need to reason in a longer time scale has appeared, in part motivated by the climate change issue. It is the case of Rousenvell *et al* [1] who presented scenarios of future agricultural land use in Europe for 2080. Within this time context, it is proposed herein that the land use scenarios should be approached as a goal-oriented process, providing a framework for identifying concerns, assessing trends and spatial configurations and producing knowledge for future actions, instead of a solution-based process derived from scenario writing. The former approach relies on the discovering of expected behaviors of a landscape to key drivers changes (mostly environmental and morphological drivers), after which policy scenarios should be designed, while the later considers, from the beginning, different policy scenarios to which a landscape should be adapted. For example, discovering long term future landscapes (understood as a spatial configuration of land uses) under reduced water availability and increased temperature climate, in a goal-oriented process, should result in different expected behaviors (solutions) to which technology, social and economic scenarios should be designed in order to prevent or accommodate them. On the contrary, discovering long term future landscapes in a solution-based process requires firstly some scenarios assumptions, as from the SRES narratives [2], to which adapted landscapes are determined in response to a set of allocation rules, for example.

It is proposed in this paper that reasoning on the very long term land use change should be approached as an optimization exercise, where several instantiations of the objective function result in different landscape solutions in order to facilitate the identification of emergent spatial patterns. Landscapes can be understood as complex systems, considering their characteristics of spatial self-organizing, and non-linear behavior to long term drivers. However, one may assume that a landscape under a specific climate and morphological conditions, for example a Mediterranean region, will evolve in a specific way, if (i) one considers no policy assumptions, and (ii) a set of constraints is respected in order to keep spatial coherence and physical feasibility in the future. For example, no trees could be considered in an area with a soil depth less than 30 cm, or no landscape which exceeds a specific fragmentation index can exist.

In this paper, the formulation of long term future landscape generation is proposed as an optimization problem, with more than one objective function and, at least, two sets of constraints. There is a wide application of techniques to land use optimization problems [3]. However, most of these techniques were not developed to allow objectives that require spatial data and are unable to handle spatial objectives. Ducheyne [4] presents an overview of the disadvantages of classic optimizing procedures to handle spatial data and spatial tailored objective functions. Linear programming, for example, uses continuous variables and this is not suitable when spatial integrity is of concern. Integer and mixed integer programming overcome this problem, but in order to explicitly formulate the spatial requirements, they have computing power problems. Heuristics have also been proposed to handle complex optimizing problems, but they are essentially single objective optimizers. Usually, they require that multiple objectives have to be reformulated into a single objective function and this hampers the search for the trade-off front between these objectives.

This work proposes the approach of evolutionary computation to deal with land use generation as an optimization task, accommodating its explicit spatial dimension and multiple objective functions. Specifically, the proposed algorithm, named GeneticLand, accommodates two objective functions - minimization of soil erosion, and maximization of carbon sequestration - and a set of physical feasibility and suitability constraints to specific land uses, as well as landscape structure and spatial organization constraints indices.

2. Methodology

Modeling land use change is addressed in our work as a multi-objective optimization problem in which landscapes (land uses) are generated by means of Evolutionary Computation (EC), a field that contains a number of techniques that have been applied successfully in search and optimization problems across a variety of domains.

Evolutionary Algorithms are based on the principles of Natural Selection. The idea is that fit individuals survive and propagate their traits to future generations and unfit individuals have a tendency to die. In this type of algorithms, an individual corresponds to a possible solution for a particular problem that we are interested in solving. In the case of GeneticLand, an individual is a land use assignment for a grid of 100x100 cells. The task of the EA is to search for a good land use assignment.

As opposed to classical optimization methods, which often combine multiple objectives into a single objective by assigning different weights to each of them, the field of Evolutionary Computation has developed various methods that allow us to evolve a diverse set of solutions which incorporate the tradeoffs that are intrinsic to the problem at hand. It is left up to the decision-makers which of the different alternative solutions should be chosen, something that is usually done based on higher level information [5].

2.1. The GeneticLand Algorithm formulation

GeneticLand is an evolutionary algorithm for land use generation, working on a region represented by a bidimensional array of cells. For each cell, there is a finite number of possible land uses. The task of the algorithm is to search for an optimal assignment of these land uses to the cells, evolving the landscape patterns that are most suitable for the various objectives satisfying the set of restrictions. The objective functions and the restrictions of the GeneticLand algorithm were designed and instantiated from an application case. Spatial analysis were performed over the landscape under study, including land use and geomorphological assessment, and landscape ecology, in order to conclude for the appropriate restrictions. A Mediterranean landscape, located at Southern Portugal, was considered for analysis due to its vulnerability to climate change, which fulfils the motivation for this work.

2.1.1. The objective functions

GeneticLand considers two goals: minimization of soil erosion and maximization of carbon sequestration. Minimization of soil erosion was selected due to its importance in the Mediterranean landscapes and the perspective of its development in climate scenarios characterize by annual reduced rainfall patterns and increased flash floods [6]. Each landscape solution, provided by GeneticLand algorithm, is validated by applying the USLE (Universal Soil Loss Equation), with the best solution being the one that minimizes both the global landscape soil erosion value and lowers the local erosion values below a manageable threshold (10 $ton.ha^{-1}.y^{-1}$). The USLE predicts the long term average annual rate of erosion on a field slope based on rainfall pattern, soil type, topography, crop system and management practices. Global soil loss for the whole landscape is derived from the following expression:

$$\text{Global Soil loss} = \Sigma_i \left[R * LS * K * C * \text{Landscape}_i\right],$$

being i each pixel of the landscape

R is the erosivity factor due to rainfall; The greater the intensity and duration of the rain storm, the higher the erosion potential. LS refers to the slope length-gradient factor; The steeper and longer the slope, the higher is the risk for erosion. K is soil erodibility factor; K is a measure of the susceptibility of soil particles to detachment and transport by rainfall and runoff. C is the crop/vegetation and management factor. It is used to determine the relative effectiveness of soil and crop management systems in terms of preventing soil loss.

Each landscape (land uses) solution is multiplied by the USLE factors resulting in a long term average annual soil loss in tons per acre per year. The global landscape soil erosion value is the sum of average annual soil loss calculated for all pixels, except for those with values less than 10 ton/ha/year, which are negligible. The GeneticLand algorithm is implemented in a way that all factors may be changed in order to consider different data sets. For example, by using different rainfall data, future climate scenarios can be accommodated and landscapes generated according to these scenarios. Table 1 presents the values for the C factor for the Mediterranean landscape [7] under study.

Table 1: C factors of land use classes to soil loss

LAND USE CLASSES	SUSCEPTIBILITY FACTOR OF LAND USES TO SOIL LOSS
Forest	0,1
Shrubs	0,02
Permanent agriculture	0,1
Annual agriculture	0,3
Mixed agriculture	0,3

Maximization of carbon sequestration was considered due to its importance under the carbon cycle and climate change issues; each solution, provided by GeneticLand algorithm, is validated by applying atmospheric CO_2 carbon uptake estimates, according to the following expression

$$\text{Global Carbon uptake} = \Sigma_i \text{ C uptake} * \text{Landscape}_i$$

being i each pixel of the landscape

The landscape (land uses) solution is multiplied by the average carbon uptake indicators [8] for each land use (Table 2), with the best solution being the one that maximizes the global landscape carbon uptake. The global landscape carbon uptake value is the sum of carbon uptake calculated for all pixels.

Table 2: Susceptibility factors of land uses to carbon uptake

LAND USE CLASSES	CARBON UPTAKE INDICATORS (t C/ha/YEAR)
Forest	1,6
Shrubs	0,4
Permanent agriculture	0,5
Annual agriculture	0
Mixed agriculture	0,1

2.1.2. The set of restrictions

As a guideline for the placement of land cover classes in future landscapes by the GeneticLand algorithm, two types of constraints were considered: physical constraints, concerning geomorphological variables, and landscape ecology indices at the patch, land use and landscape levels. The physical constraints were developed by analyzing the 1990 Corine land cover for the Portuguese Guadiana watershed (11 600 km^2) regarding the distribution of land cover against four different variables:

(i) Soil type suitability, considering nine different soil types;

(ii) Slope, at 90m spatial resolution, divided in to fifteen classes from plane (1) to very steep (15);

(iii) Aridity Index, consisting in a ratio between Precipitation and Thornthwaite's Potential Evapotranspiration (P/PET), and comprehending seven classes raging from very dry (1) to very humid (7);

(iv) Topographic Soil Wetness Index (TSWI) [9] consisting of thirteen classes varying from very low soil humidity to very high soil humidity

Table 3 illustrates the constraints stated for each land cover in soil type 1. For example, annual agriculture only occur in areas with slopes ranging from 1 to 4, P/PET classes from 2 to 3 and TSWI values from 6 to 28, while forests occur in every slope classes, in areas where P/PET ranges from 2 to 7, and TSWI classes from 4 to 16. This procedure was performed for every land cover in every soil type, providing a complete set of geomorphological constraints to feed GeneticLand.

Table 3 - Example of the constraints feeding the GeneticLand algorithm, concerning for land covers occurring in Soil Type 1 (Cambisols) for the case study.

Cambisols	**Slope**		**P/PET**		**TSWI**	
	Min.	Max.	Min.	Max.	Min.	Max.
Annual agriculture	1	4	2	3	6	28
Permanent agriculture	1	7	2	5	4	16
Mixed agriculture	1	6	2	7	4	26
Forest	1	15	2	7	4	16
Shrubs	5	8	4	6	4	4

Landscape ecology restrictions were used to insure the spatial coherence of the landscape. Landscape ecology provides indices that help the characterization and quantification of landscape structure, function and change [10]. These metrics were calculated with the support of the Fragstats 3.3 software [11]. Two indices are selected from a previous landscape ecology analysis supported by a set of different indices: the patch size for each land use, and the adjacency index, named contagion.

A patch is defined as a non-linear surface that differs in the appearance, shape and complexity, and includes a single pixel or a set of adjacent pixels of the same land use class. The patches vary in size, form, type, heterogeneity and characteristics of edge. The size is an important aspect of a patch, since it governs the circulation of nutrients through the landscape and the amount of species in a region, and thus can be assumed as a fundamental characteristic in a specific landscape. GeneticLand considers a range for the patch size (min and max) as a characteristic of each land use class that, for simplicity, will be respected in order to preserve the spatial structure of each land use.

Contagion is a landscape ecology index that measures the probability of "adjacency" of cells (pixels) of the same land use class. This index measures the degree of dispersion or aggregation of the landscape elements where high values (Max. 100) are from landscapes with few patches of great dimension, while low values (Min. 0) show landscapes with many dispersed units. This index considers all patch types present on an image, and considers similar adjacencies (i.e., cells of a patch type adjacent to cells of the same type). The numeric expression of contagion index can be consulted in [11].

2.2. The GeneticLand Algorithm implementation

Multi-objective evolutionary algorithms (MOEAs) rely on the concept of non-dominated solutions, which state that two solutions are only comparable if one of them is better than the other in all the objectives. This idea has been incorporated successfully in a number of MOEAs, and the main reason for that is due to the fact these algorithms work with a population of solutions rather than with single solutions.

For the evolutionary algorithm, the Pareto Archived Evolution Strategy (PAES) developed by [12], was used. This algorithm is perhaps the simplest evolutive scheme for multi-objective optimization and it is based on an extension of the (1+1) evolution strategy [13]. In starts with a random solution (one parent) and then it generates one offspring by means of a mutation operator. If the offspring is a better solution than the parent, it replaces the parent for the next generation (iteration). In the opposite case, the offspring is discarded. This process is repeated a number of times until a specified stopping criterion is satisfied.

The PAES algorithm maintains an archive of non-dominated solutions, each of which cannot be said to be better that the other. The PEAS algorithm was selected, rather than a more sophisticated evolutionary algorithm, due to the very large problem dimension that we are facing. Current applications of evolutionary algorithms have in general no more than a few thousand decision variables, while the problem that we are modeling can go up to over 10 thousand decision variables, which correspond to a landscape with 100 x 100 cells.

The mutation operator that was implemented in GeneticLand changes a cell to a different land use. In addition to doing that, it also changes a number of surronding cells to the new land use. We did that because otherwise the mutation operator would generate invalid solutions with a very high frequency. When constraints are violated by means of the mutation operator, the fitness of the solution is penalized by a certain amount. The more a constraint is violated, the more the solution is penalized.

The algorithm was run for a total of slightly over 1 million iterations. There are theoretical results (Muhlenbein, 1992) that say that for a simple unimodal function, the average number of iterations that the (1+1) EA needs to find the optimal solution is exp(1)*N*log(N/2), where N denotes the number of decision variables of the problem.

That result was derived for solutions coded as binary decision variables. In our case, we have 5 possible values for each decision variable. Considering that the most difficult step is to optimize the last gene (which means not mutating the correct genes and mutation the incorrect one to the right value), we decided to multiply the number of iterations by a factor of 5. Note however that this is only a rough approximation.

The various simulations were run on a Linux cluster. A single run took several days to complete. Recall that the evaluation of a single solution requires several arithmetic operations on a grid of 100x100 cells. A possible improvement to speed up the computational time is to come up with an incremental fitness function. This way, a new solution would not have to be reevaluated from scratch. Instead, it's fitness could be derived based on the fitness of the parent and on the cells that were affected by mutation. This scheme could be easily implemented but some care would have to be taken regarding the possible constraint violations introduced.

3. Application

The GeneticLand algorithm was designed and tested in an area located in southern Portugal, within the Guadiana watershed, as illustrated in Figure 1. The physical characteristics of this region, from where the constraints were derived, are shown in Figure 2, both the corresponding maps and the histograms.

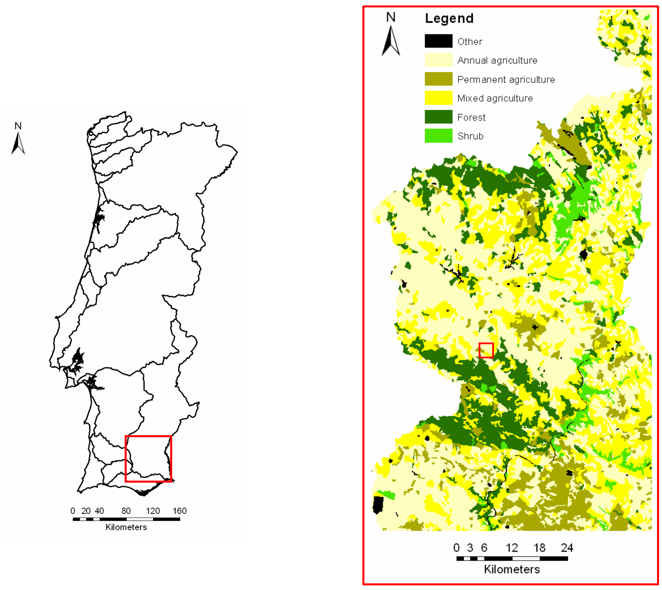
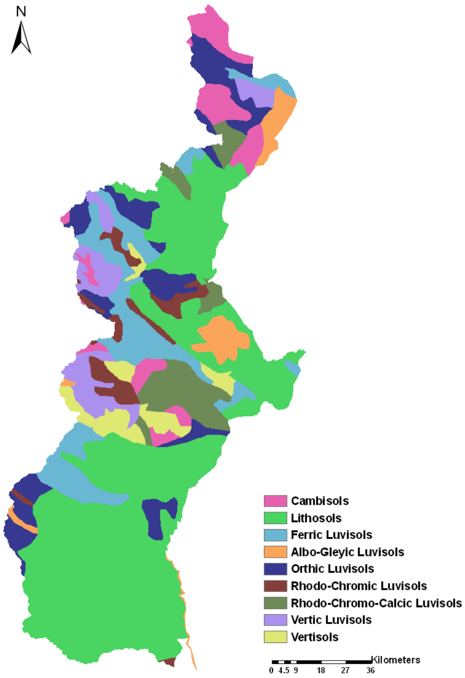
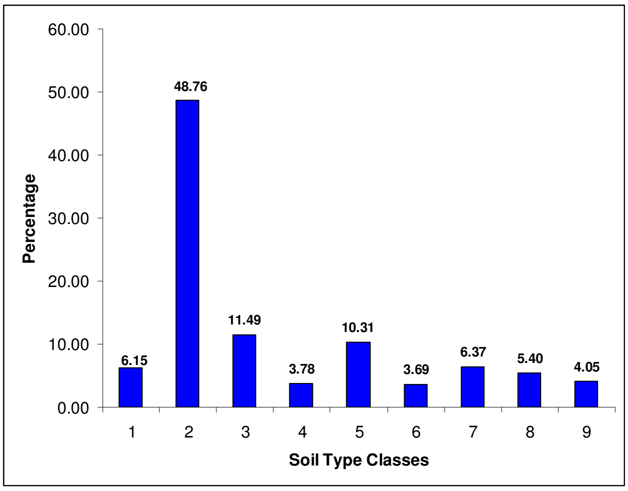


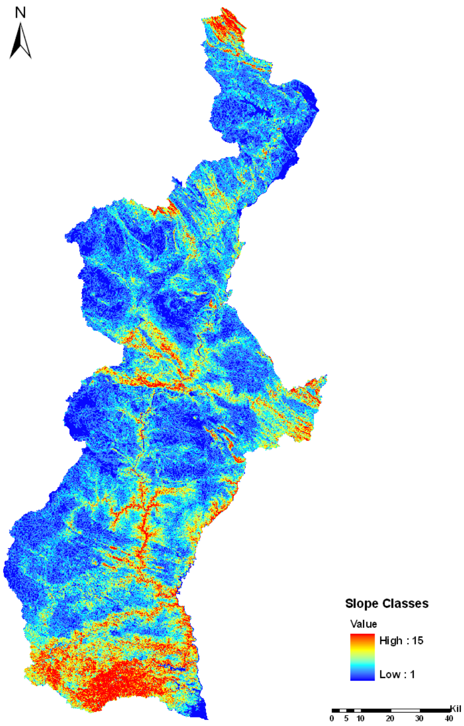
Figure 1: Location of the study landscape supporting the design and test of GeneticLand

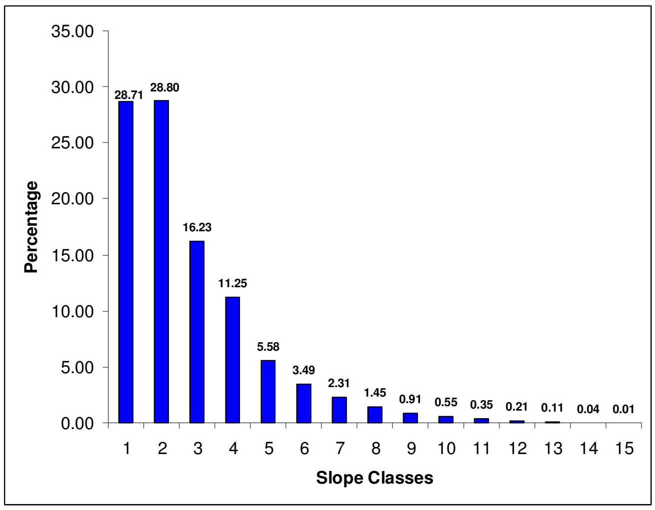
Soil Type





Slope





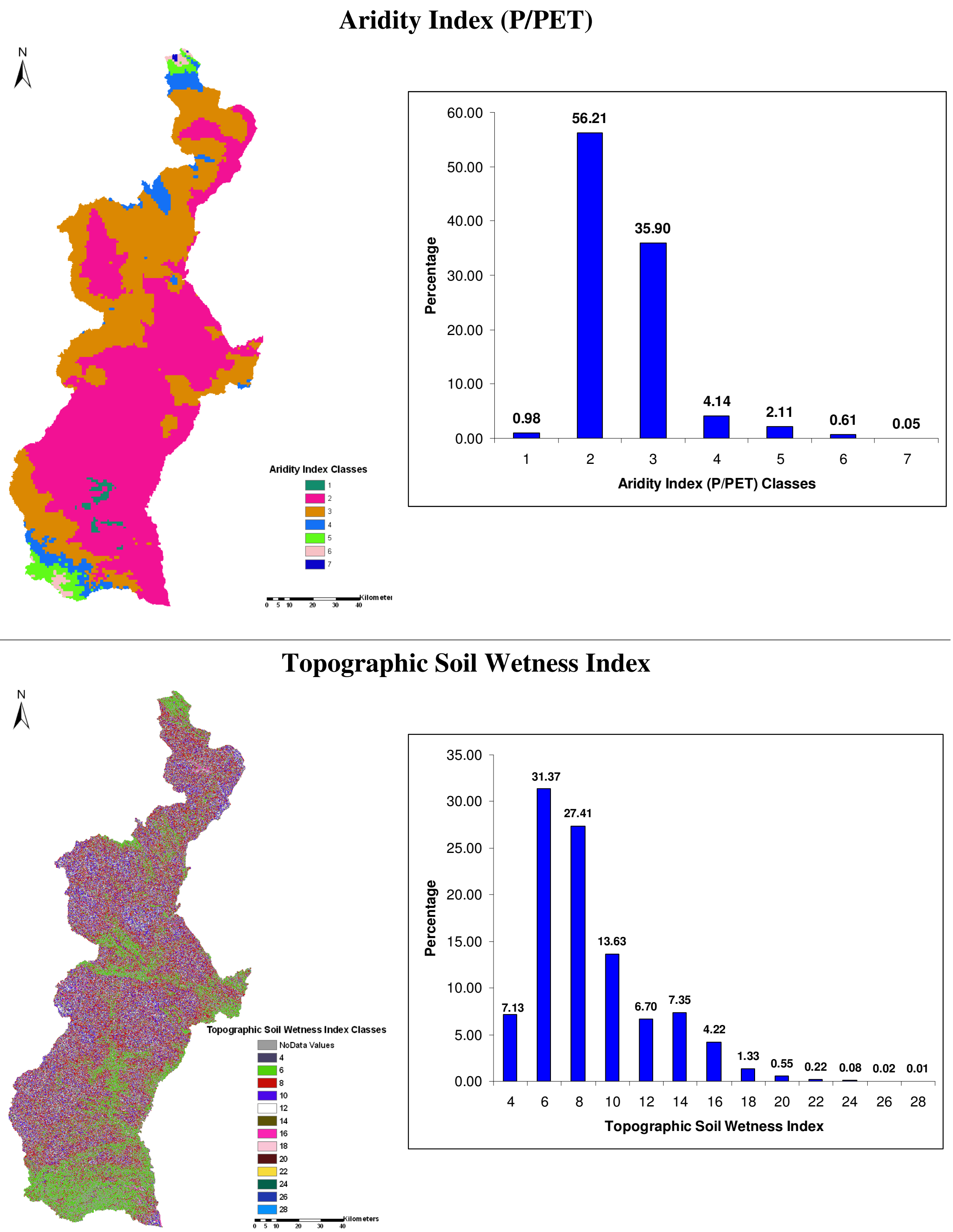


Figure 2: Selected physical characteristics of the landscape under study, from where the constraints were designed to feed the GeneticLand algorithm.

4. Results and Discussions

A small subsection of the landscape under analysis, representing an area with about 9 km^2 (100 columns and 100 rows), was selected to test the implementation and performance of the GeneticLand algorithm. As previously explained, and according to the evolutionary methods, a set of solutions were generated for the multi-objective functions. In this case, two sets of 15 landscape solutions were generated, with each set evolving from an initial random image and the two images being generated with different random seeds. The solutions' tradeoff between reducing soil erosion and increasing carbon uptake follow the traditional Pareto curve, as can be seen in Figure 3. Although there are some differences among the 30 landscape solutions, they are less than 2%, and all the solutions comply in more than 91% in the area with the set of restrictions. It can be seen that more carbon uptake corresponds to more soil erosion. A detailed analysis of each solution reveals that the algorithm prefers to increase the carbon uptake instead of decrease soil erosion. This is a consequence of the constraints imposed on this landscape since the choice of shrubs, the land use more appropriate to prevent soil erosion, is restricted in part of this area.

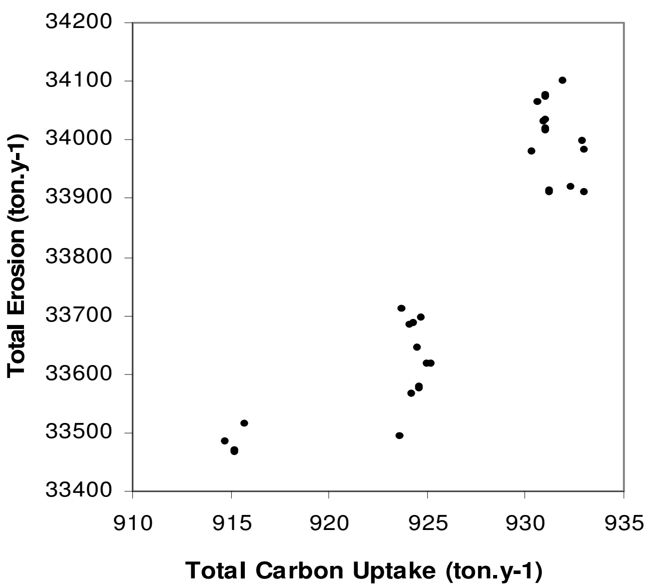


Figure 3: Pareto curve for the 30 landscape solutions generated by the multi-objective GeneticLand algorithm

Figure 4 presents the current land use in the test area, and one of the landscape solutions generated by the multi-objective GeneticLand algorithm. It can be seen that forest was chosen over annual agriculture, which fulfils the carbon uptake goal. Permanent agriculture was also generated due to its higher carbon uptake values and lower susceptibility to soil erosion when compared to other land uses. Shrubs were also selected due to their ability to prevent soil loss. Figure 5 shows soil erosion patterns comparing the current land uses and those derived from the landscape generated by the multi-objective GeneticLand algorithm. It is visible that the algorithm succeeds in reducing the overall soil erosion patterns.

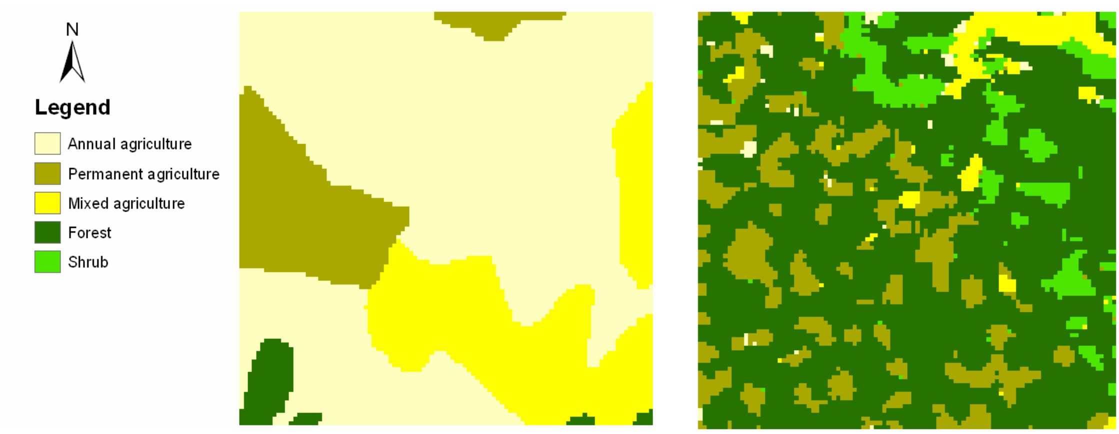


Figure 4: Current land use and landscape generated by the multi-objective GeneticLand algorithm.

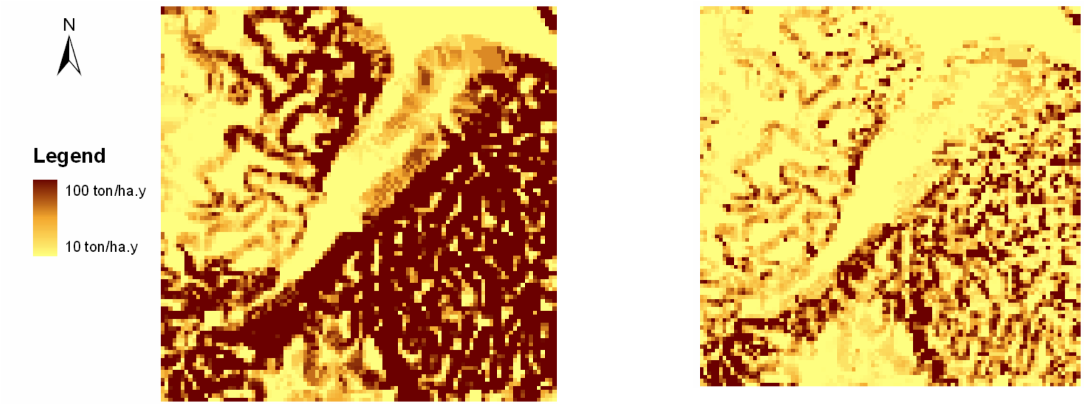


Figure 5: Current soil erosion patterns and optimized soil erosion patterns derived from the landscape generated by the multi-objective GeneticLand algorithm.

Finally, Figure 6 shows the comparison of global results concerning soil erosion and carbon uptake, between current landscape and one generated by the multi-objective GeneticLand algorithm. On average, for the set of landscape solutions, the GeneticLand algorithm increases carbon uptake up to more than 740% when compared with current land uses, at the same time that reduces soil erosion by about 65%. Also, it should be noticed that areas of serious soil erosion (>100 $ton.ha^{-1}.y^{-1}$) were reduced about 77%, when compared with current land uses. This is a very interesting result, because the reduction of serious soil erosion areas was not a requested goal, but the algorithm chose to improve its solution in these pixels.

A final assessment of the landscape generated by the GeneticLand algorithm refers to the type of spatial distribution of land uses. Compared to the current landscape, the landscape solution is characterized by a higher spatial heterogeneity of the land uses. This type of spatial distribution has been proposed as the better approach to promote ecosystem sustainability in general, and preventing soil erosion, in particular.

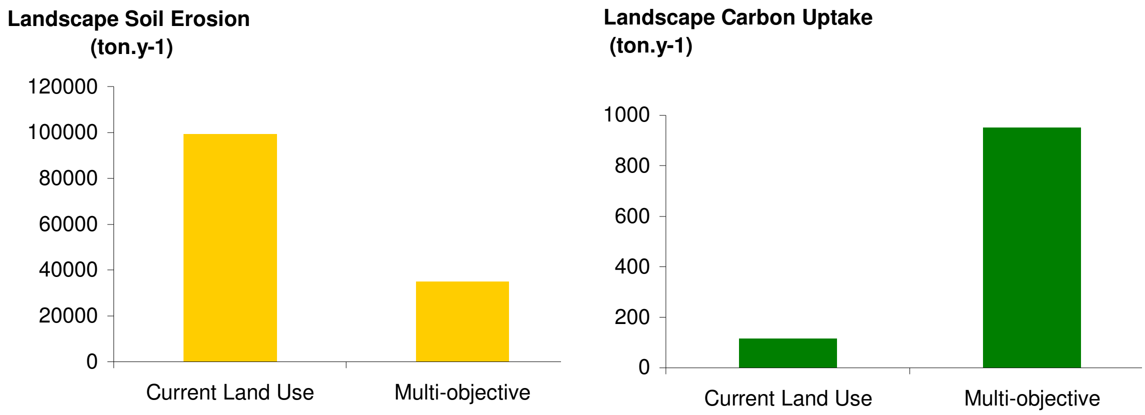


Figure 6: Comparison of global results, concerning soil erosion and carbon uptake, between current landscape and that generated by the GeneticLand algorithm

5. Conclusion

It is proposed in this paper that reasoning on the very long term land use change should be approached as an optimization exercise. An optimization problem was formulated based on two objective functions (soil erosion minimization and carbon uptake maximization), subject to a set of physical and spatial constraints. This paper proposes an algorithm, name dGeneticLand, in which landscapes (land uses) are generated by means of Evolutionary Computation (EC), based on the principles of natural selection, a field that contains a number of techniques that have been applied successfully in search and optimization problems across a variety of domains. In this type of algorithms, an individual corresponds to a possible solution for a particular problem that we are interested in solving. The task of the EA is to search for a good land use assignment. The GeneticLand algorithm was designed and tested in a small area (100x100 cells) located in southern Portugal, within the Guadiana watershed. All the landscape solutions comply in more than 91% in the area with the set of restrictions, and fulfill the stated objective goals. In fact, the algorithm increases carbon uptake up to more than 740% when compared with current land uses, at the same time that reduces soil erosion by about 65%. Also, the areas of serious soil erosion (>100 $ton.ha^{-1}.y^{-1}$) were reduced about 77%, when compared with current land uses. The landscape solution is characterized by a higher spatial heterogeneity of the land uses, which is appointed as the better approach to promote ecosystem sustainability in general, and preventing soil erosion, in particular.

Although the results are very promising concerning the feasibility of landscape generation through the use of evolutionary algorithms, there are a set of limitations that should be considered for further development. Some of them refer to: (i) the physical constraints were derived from CORINE 1987 spatial analysis and it should be derived from a time series analysis in order to accommodate land use change dynamics; (ii) the algorithm is allowed to choose the same land use classes as the current land use map, but other classes should be allowed; (iii) physical suitability of land uses, stated in the set of constraints, are maintained constant in the future, but some should be different as it is the case of forests that could adapt for other P/PET classes than 2 to 7; (iv) improvement of the implementation strategy to run images larger than 100*100, in order to consider landscapes with more complex spatial variability patterns.

REFERENCES

[1] Rousenvell, M. D. A, F. Ewert, I. Reginster, R. Leemans, T.R. Carter, Future Scenarios of European agricultural land use. II.Projecting changes in cropland and grassland, Agriculture Ecosystems & Environment, Article in Press, 2005, Available online at www.sciencedirect.com

[2] Nakicenovic, N. Alcamo, J, Davis, G., de Vries, B., Fehann, J., Gaffin, S., Gregory, K., Grubber, A., Jung, T.Y., Kram, T., Emilio, la Rovere, E., Michaelis, L., Mori, S., Morita, T., Pepper, W., Pitcher, H., Price, L., Riahi, K., Roehrl, A., Rogner, H., Sankovski, A., Schelesinger, M.E., Shukla, P.R., Smith, S., Swart, R.J., van Rooyen,S. Victor, N., Dadi, Z., 2000, Special Report, on Emissions Scenarios, Cambridge University Press, Cambridge. 2000.

[3] Pongthanapanich, T., Review of Mathematical Programming for Coastal Land Use Optimization University of Southern Denmark, Department of Environmental and Business Economics IME WORKING PAPER 52/03, ISSN 1399-3224, 2003.

[4] Ducheyne, Els, Multiple Objective Forest Management using GIS and Genetic Optimization Techniques, Ph.D. Thesis, University of Gent, Gent, 2003.

[5] Deb, K., Multi-Objective Optimization using Evolutionary Algorithms. John Wiley and Sons, Ltd, Chichester, 2001.

[6] McCarthy, O. Canziani, N. Leary, D. Dokken, K. White, Eds. Climate Change 2001: Impacts, Adaptation, and Vulnerability, Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 1042 pages.

[7] Pimenta, M.T. 1998. Directrizes para a aplicação da Equação Universal de Perda dos Solos em SIG: factor de cultura C e factor de erodibilidade do solo K. INAG, Lisbon, 12 pp. (available online at http://snirh.inag.pt/, under "reports").

[8] Instituto do Ambiente, Faculdade de Ciências e Tecnologia, CEEETA, Programa Nacional para as Alterações Climáticas, 2003, Vol. 8: Florestas e Produtos Florestais.

[9] Beven, K., Rainfall-Runoff Modelling – The Primer, John Wiley & Sons, Chichester, 2000.

[10] McGarigal, K., and B. J. Marks. 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. USDA For. Serv. Gen. Tech. Rep. PNW-351.

[11] http://www.umass.edu/landeco/research/fragstats/fragstats.html

[12] Knowles & Corne, 1999. Knowles, J.D., Corne, D.W. (1999), The Pareto Archived Evolution Strategy: A New Baseline Algorithm for Pareto Multiobjective Optimisation. In Proceedings of the 1999 Congress on Evolutionary Computation (CEC'99), Volume 1, pages 98-105.

[13] Schwefel, H.-P., 1981, Numerical Optimization of Computer Models. John Wiley & Sons, Chichester, 1981.