EXPLORING THE SOLUTION SPACE OF SEMI-STRUCTURED SPATIAL PROBLEMS USING GENETIC ALGORITHMS

David A. Bennett, Department of Geography, Southern Illinois University, Carbondale, IL 62901-4514

Greg A. Wade, Department of Computer Science, Southern Illinois University, Carbondale, IL 62901-4514

Marc P. Armstrong, Department of Geography and Program in Applied Mathematical and Computational Sciences, The University of Iowa, Iowa City, IA 52242

ABSTRACT

The resolution of semi-structured spatial problems often requires consensus building and compromise among stakeholders as they attempt to optimize their own set of criteria. The union of these sets form a criteria space that constrains the set of viable solutions that may be adopted by decision-makers. Knowledge about the criteria space, the solution space, and the relation between the two is normally incomplete and this lack of understanding places real limits on the ability of decision-makers to solve complex spatial problems. This research explores new approaches that are designed to establish a link between criteria space and solution space and to provide a mechanism that competing stakeholders can use to identify areas of conflict and compromise.

1.0 INTRODUCTION

Spatial problem solving often requires collaboration among multiple decision-makers because the effects of spatial decisions often cut across traditional bounds of discipline, jurisdiction, and ownership. Because different decision-makers will have different views of a problem, the evaluation of alternative solutions to it is complicated since: 1) a collection of spatial models and analytical tools is needed to evaluate how well each alternative meets stated criteria; 2) multicriteria evaluation tools are needed to integrate the results of these models and tools; 3) the set of all possible solutions (the solution space) is often intractable (theoretically infinite for field-based problems); and 4) not all criteria are well articulated or even known at the beginning of an analysis (*i.e.*, spatial problems are often semi-structured). Furthermore, the resolution of semi-structured spatial problems often requires consensus building and compromise among decision-makers because as individuals attempt to optimize their own set of criteria they will often come into conflict with others. The union of these criteria sets forms a criteria space that constrains the set of viable solutions that may be adopted by decision-makers. Understanding the relation between criteria space and solution space is a key element in the successful resolution of spatial problems.

The integration of spatial decision making, spatial models and geographic information systems (GIS) has been an active area of research and advances have been made in loosely coupled systems (He *et al.*, 1995), tightly coupled systems (Bian *et al.*, 1996), and fully integrated systems (Bennett, in press; Wesseling *et al.*, 1996). At the same time researchers have been investigating techniques designed to integrate multicriteria analysis into GIS (Carver, 1991; Jankowski, 1995). What has not yet been investigated are tools to explore, analyze, and visualize the solution space of a problem with respect to multiple models and criteria. Providing such tools has several benefits: 1) new and unique solutions can be identified; 2) unarticulated criteria can be identified and incorporated into an analysis; 3) the spatial implications of specific criteria can be visualized; and 4) areas of agreement and conflict can be identified and discussed. As suggested above, the set of all possible solutions can be very large. The time required to create, model, and evaluate such large sets of possible solutions is prohibitive and heuristic tools are needed to guide and expedite this effort.

In this paper a two-dimensional genetic algorithm (Bennett *et al.*, 1996) is used to evolve landscapes that meet stated criteria based on a set of spatial models. An initial population of random landscapes is created. Each landscape is represented as a raster file in which cells are assigned a particular land cover. The fitness of a landscape is evaluated by intelligent agents that act as surrogates for decision-makers that represent competing stakeholders. Agents implement a multicriteria evaluation scheme that models the success of each landscape in meeting the stakeholders' stated criteria. Agents rank the competing landscapes and a mediating agent uses these rankings to calculate an overall fitness value for each landscape. Those landscapes deemed most "fit" by this process are used to propagate new landscapes. Thus, the solution space is heuristically expanded and explored. Delta maps derived from those landscapes that were ranked high (*e.g.*, the top three alternatives) by individual agents illustrate areas of consensus, conflict, and potential compromise. When highly ranked alternatives fail to meet the expectations of a stakeholder then the criteria space should be reevaluated.

2.0 GENETIC ALGORITHMS

Genetic algorithms are modeled after those processes that drive biological evolution and the evaluation of fitness values provides an effective heuristic for the exploration of problems that may otherwise be intractable. Alternatives in the solution space of such problems represent individuals in an evolving population. Characteristics that can be used to evaluate the relative success of individual solutions are stored in classifiers which are often implemented as bit-strings that document when a specific solution possesses a given characteristic

(Booker *et al.*, 1989; Armstrong and Bennett, 1990). Fitness in this context is proportional to how well a particular solution meets stated criteria. Three genetic operators are used to evolve a large number of new alternatives from existing alternatives: cross-over, mutation, and inversion. Mutation is a unary operator that makes random changes in a linear sequence of characteristics. Inversion, also a unary operator, flips values in a linear sequence of characteristics. Cross-over, the most powerful of these operators (De Jong, 1990), is a binary operator that generates two new offspring by duplicating two individuals (parents) and swapping "genetic code" beyond some randomly selected cross-over point. New and innovative solutions are created through random cross-overs, mutations, and inversions.

A more formal description of the genetic algorithm is as follows (after De Jong, 1990; Koza, 1994):

1. Generate an initial population, P_0, of potential solutions. These individual solutions are often created as random combinations of identified characteristics.
2. For each individual, I_m, in the current population, P_i, calculate a fitness, $f(I_m)$. Select n individuals from P_i that will be used to generate n new solutions for P_{i+1} via cross-over. The probability, p, that individual I_m will be used to create new alternatives for population, P_{i+1}, is a function of its fitness, $f(I_m)$:

$$p(I_m) = \frac{f(I_m)}{\sum_{k=1}^{n} f(I_k)} \tag{1}$$

where:

$f(I_m)$ = fitness of individual m

$p(I_m)$ = probability of individual m producing offspring in the next generation

3. Remove x individuals from P_{i+1} (based on user defined criteria).
4. Add n new individuals to P_{i+1} by applying cross-over, mutation, and inversion operators.
5. If an acceptable solution exists then stop; else advance to generation $i+1$ and return to step 2.

Although geographical applications of this approach are rare, Dibble and Densham (1993) illustrate the utility of genetic algorithms in the solution of location-allocation problems. Zhou and Civco (1996) use neural networks that employ genetic algorithms as a learning mechanism to conduct land use suitability analyses. These projects do not, however, apply genetic algorithms to two dimensional landscapes.

3.0 AGENT-DIRECTED GENETIC ALGORITHMS FOR ENVIRONMENTAL PROBLEM SOLVING

In order to manage environmental resources in privately owned landscapes it is necessary to understand how individual decisions effect environmental processes across space and through time. The tools used by resource managers to promote environmental objectives in a privately owned landscape depend largely on education, incentive-based policy initiatives (*e.g.*, conservation reserve program) and quasi-regulatory compliance programs (*e.g.*, commodity programs). Furthermore, private and public concern about the environmental ramification of land management decisions is only one of many competing issues that must be addressed by land managers. To develop feasible and politically acceptable solutions to environmental problems generated by the cumulative impact of multiple decision-makers it is often necessary to foster compromise and consensus among a diverse set of special interest groups who possess overlapping objectives; some quantifiable, and some not. Thus, environmental management, like many spatial problems, is often a semi-structured problem that requires a collaborative effort among multiple stakeholders.

3.1 Genetic Algorithms for Two Dimensional Space

Traditional genetic algorithms operate on a finite set of well-defined characteristics that are easily mapped to a linear data structure. When this approach is adapted to the generation of alternative landscapes it is necessary to extend the notion of a linear sequence of genetic code to a two dimensional representation. The linearization of space is a well-studied problem. Mark and Lauzon (1984) illustrate how to accomplish this task using a two dimensional run-length encoding (2DRE) scheme based on a Morton index of a raster-based geographical data set. Using 2DRE and two randomly selected cross-over points, two new landscapes that possess characteristics of two parent landscapes can be created (Figure 1).

3.2 Multicriteria Decision Space

The set of relevant criteria and the relative importance of specific criteria vary with the goals and objectives of the stakeholder. To integrate the concerns and objectives of multiple competing stakeholders we recast the genetic algorithm fitness function into a modified multicriteria evaluation function. To construct a composite fitness value for a given alternative and set of criteria, criterion-specific fitness values must be standardized since each analyses will not use the same units or, perhaps, even the same scale of measurement (*e.g.*, nominal, ordinal, interval, ratio). Furthermore, decision-makers must provide a subjective weighting scheme that documents the relative importance of each criterion. Several compositing schemes exist (for a review of MCE techniques in GIS see Carver 1991 and Jankowski 1995). For most of these schemes the final fitness score is a linear function that takes as input the standardized score and associated weight of each criterion.

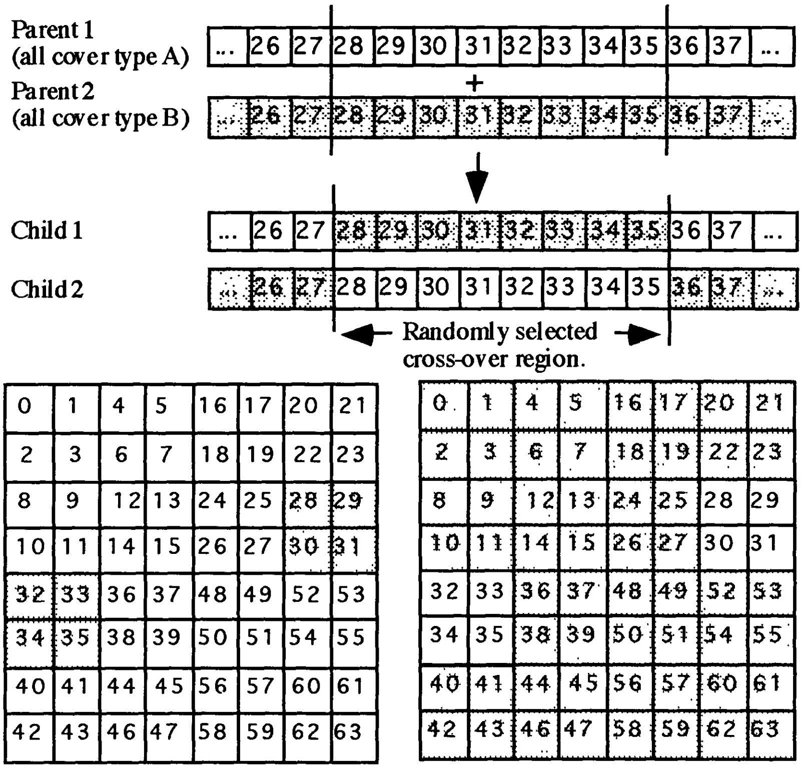

Figure 1. The results of a two dimensional cross-over.

As suggested by Breeden (1995) and determined empirically in the context of this work, the linear combination of fitness scores can be problematic in the context of genetic algorithms. If the distribution of the standardized scores for a particular criterion is positively skewed then landscapes that perform well for that criterion will be unduly favored as the next generation of solutions is created. Furthermore, outliers in the distribution of the standardized scores restrict the variance of more "typical" scores and, thus, mask potentially significant differences among alternative solutions. As a result, the probability that an individual landscape will be propagated into the next generation will be approximately equal for many individuals in the population. One way to overcome the impact of outliers and skewed distributions is to construct a composite fitness value based on ranked order. The fitness value for a particular landscape is then calculated as (Breeden, 1995):

$$f(I_{kd}) = \sum_{j=1}^{n} w_{jd} r_{kjd} \qquad (2)$$

where:

r_{kjd} = rank score of alternative *k* given criterion *j* and decision maker *d*.

w_{jd} = weighted value of criterion j given decision maker d.
$f(I_{kd})$ = fitness value for alternative k given decision maker d.

Criteria weights reflect a qualitative assessment of a single decision-maker, a class of decision-makers or, perhaps, a set of decision-makers who have reached a consensus. If a consensus has been reached then, in many situations, compromise among competing decision-makers has been reached in criteria space. This may or may not be possible. An alternative approach is to allow individual decision-makers, or sets of decision-makers that represent specific classes of stakeholders, to define criteria independently and attempt to construct a compromise in the solution space. To accomplish this using a genetic algorithm a global fitness value for each landscape must be calculated to represent all decision-makers. Here this global fitness value is calculated as the mean of the independent fitness values:

$$F(I_k) = \frac{\sum_{d=1}^{n} f(I_{kd})}{n} \tag{3}$$

where:
$F(I_k)$ = Global fitness value for individual k.
$f(I_{kd})$= Fitness value for individual k given decision-maker d.
n = Total number of decision-makers.

3.3 Intelligent Agents

As geoprocessing software becomes more sophisticated it is able to support the analysis of an increasingly rich set of problems. This richness, however, has a downside: software has become increasingly complex and, thus, more difficult to use. Furthermore, decision-makers often represent several interests and bring to the negotiation table different types of training, levels of education, experience with computing technologies, and familiarity with the problem that is being addressed. Though such differences can prove valuable since distributed expertise may allow for decision making procedures that are less prone to errors attributable to a lack of domain specific knowledge, this differential in knowledge can also have interaction effects that complicate the decision making process. Because of the number of analytical tools available and the disparate backgrounds of individual decision-makers, users may not always understand the implications of particular analytical methods. In many cases, additional knowledge may be required to support informed use.

One way to provide a more common level of support to decision-makers is to create intelligent software agents equipped with knowledge about how and when to implement specific analytical tools. At this point, two classes of intelligent agents have been implemented, mediating agents and user agents (see Shoham, 1993 for a discussion on intelligent agents). User agents acting on behalf of specific decision-makers, calculate $f(I_{kd})$ (equation 1) for each individual landscape k using applicable analytical tools and user supplied criteria weights, and returns these fitness values to the mediating agent. Using this information the mediating agent calculates $F(I_k)$ (equation 2), selects individuals for propagation and builds consensus among competing interests.

4.0 A CASE STUDY

A multidisciplinary research team from Southern Illinois University at Carbondale is investigating the impact of alternative resource policy and management scenarios on the economy, hydrology, and ecology of the Cache River (IL) watershed. The goal of this research effort is to develop a land use management plan that is generally acceptable to all stakeholders. A small study site within this watershed was selected to test the utility of two-dimensional genetic algorithms in the resolution of semi-structured spatial problems. The study site is approximately 3.69km^2 captured as a grid with a 30m cell resolution (64 rows and 64 columns). This site was selected because it possesses considerable spatial variability within a manageable area. Alternative landscapes are comprised of corn, soybean, double crop (winter wheat then soybean), wheat, grassland, and forest.

Stakeholders within the region were generalized into three classes:

1. Farmers who want to maximize farm revenue.
2. Conservationists interested in reducing soil loss and non-point pollution and agricultural productivity.
3. Wildlife enthusiasts, local entrepreneurs, and recreational hunters interested in the maintenance and enhancement of wildlife populations.

To assess how well alternative landscapes meet the concerns of these stakeholders models were developed that evaluate agricultural income, soil erosion, and the interspersion and juxtaposition of land cover types. To support these models spatial databases were developed that capture the topographic and edaphic characteristics of the study area.

Agricultural income is generated from corn, soybean, wheat, and hay (grasslands). No income is attributed to forest land. For each alternative landscape net agricultural return is calculated for each 30x30m cell by considering land cover, the expected productivity of that cover type given the associated soil, the market value of that crop, and the expected costs of producing that crop. Market prices and production costs for agricultural produce are based on ten year averages for the state of Illinois. Soil productivity values are derived from the Union County, IL soil survey (USDA, 1979). An estimate of the rate of erosion that is associated with each cell is calculated using the universal soil loss equation. The estimated value for soil erodibility (K) was derived from the Union County soil survey. The cropping factor (C) is an estimate based on cover type. A 7.5 minute DEM was used to estimate the slope of each cell. This information was, in turn, used to estimate the LS factor of the universal soil loss equation. Land management practices (P) were assumed to be the same on all cells. Interspersion is an index of the "intermixing of units of different habitat types" (Giles 1978:156). It is assumed that interspersion is desirable for wildlife but can lead to inefficiencies in agricultural production. Juxtaposition, as used here, is a measure of adjacency among cover types. The value of an edge between land cover types depends on the objectives of the land manager and the cover types involved.

A user agent was created to represent each stakeholder class. Each agent maintains a weight and a ranked list of alternatives for each criteria/model. Criteria weights used by each agent are listed in Table 1. An attempt is made to maximize all criteria except soil loss which is minimized. Note that these values were used only for "proof-of-concept" and, thus, are not intended to be representative of the groups identified.

	Ag. Production	Soil Loss	Interspersion	Juxtaposition
Farmer	1	0	0	0
Conservationist	0.5	0.5	0	0
Wildlife Enthus.	0.25	0	0.5	0.25

Table 1. Agent Weights

Figure 2 illustrates the two most influential landscape characteristics, soil erodibility and soil productivity. The dominant landscape feature within the study area is a floodplain that runs from the northeast to the southwest. A somewhat smaller tributary enters the study area in the northeast corner and continues south until it meets the larger floodplain. As can be seen in Figure 2A, the side slopes of these valleys are highly erodible and the uplands are moderately erodible. Most of the highly productive soils are located within the floodplain of the two streams (Figure 2B). However, in the southwest the floodplain is too wet to provide a reliable crop. An initial set of random landscapes were created and this "population" was allowed to evolve for more than 200 generations. The results of this experiment are presented in Figure 3. The spatial patterns that evolved through this process are logical given the character of landscape and represent reasonable compromise solutions given the objectives of the stakeholders (highly erodible and low producing soils in forest, moderately erodible soils with reasonable productivity in wheat, slightly erodible productive soils in soybeans).

5.0 CONCLUSION

To support effective resource management practices new tools are needed that allow decision-makers to build consensus among multiple stakeholders and to investigate the cumulative impact of individual actions. This research investigates two technologies that offer promise for such collaborative spatial decision making processes, agent-oriented programming and genetic algorithms. Genetic algorithms are used here to evolve landscapes that meet predetermined criteria. Intelligent agents provide a means of evaluating the fitness of these landscapes based on weighted criteria. Through this interaction between intelligent agents and genetic algorithms management strategies can evolve in ways that begin to meet the goals of multiple stakeholders.

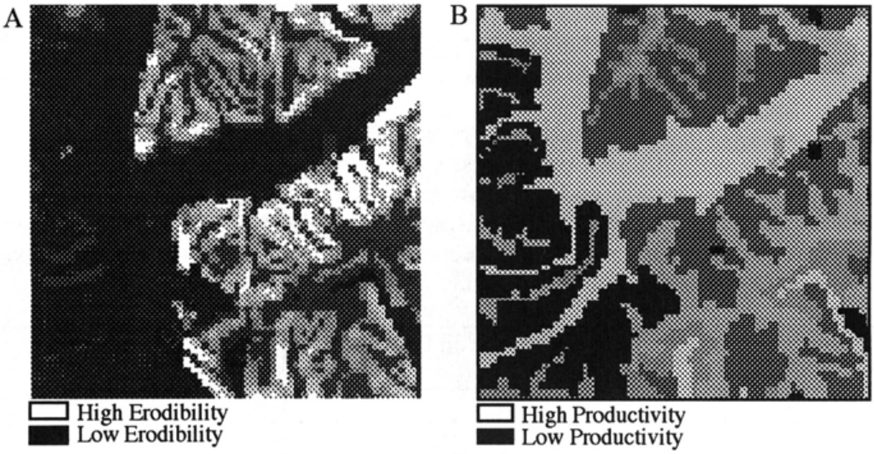

Figure 2. Soil erodibility and agricultural productivity maps for the study site.

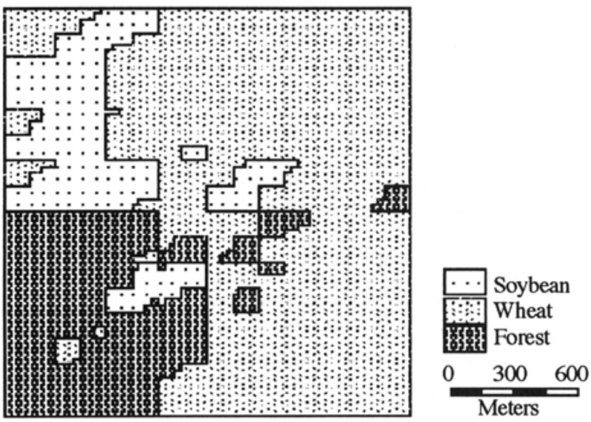

Figure 3. Most "fit" landscape after 200 generations.

6.0 REFERENCES

Armstrong, M.P. and Bennett, D.A. (1990). A bit-mapped classifier for groundwater quality assessment. *Computers and Geosciences*, 16 (6): 811-832.

Bennett, D.A. (in press). A framework for the integration of geographic information systems and modelbase management. *International Journal of Geographical Information Systems*.

Bennett, D.A., Armstrong M.P. and Wade, G. A. (1996). Agent mediated consensus-building for environmental problems: A genetic algorithm approach. *Third International Conference on Environmental Modeling and Geographic Information Systems*, Santa Barbara, CA: National Center for Geographical Analysis, CD and WWW.

Bian, L., Sun, H., Blodgett C., Egbert, S., Li, W., Ran, L., and Koussis A. (1996). An integrated interface system to couple the SWAT model and ARC/INFO. *Third International Conference on Environmental Modeling and Geographic Information Systems*, Santa Barbara, CA: National Center for Geographical Analysis, CD and WWW.

Booker, L.B., Goldberg, D.E., and Holland, J.H. (1989). Classifier Systems and Genetic Algorithms. *Artificial Intelligence*, 40, pp. 235-282.

Breeden, J.L. (1995). Optimizing stochastic and multiple fitness functions. In *Evolutionary Programming*, edited by McDonnell, J.R., Reynolds, R.G., and Fogel, D.B. MIT Press, Cambridge, MA.

Carver, S.J. (1991). Integrating multi-criteria evaluation with geographical information systems. *International Journal of Geographical Information Systems*, 5(3), pp. 321-339.

De Jong, K. (1990). Genetic-algorithm-based learning. In *Machine Learning* edited by Y. Kodratoff and R. Michalski. Morgan Kaufmann, San Mateo, CA.

Dibble, K., and Densham P.A. (1993). Generating interesting alternatives in GIS and SDSS using genetic algorithms. In *Proceedings of GIS/LIS '93 ,Volume 1*. Bethesda, MD: American Congress on Surveying and Mapping, pp. 180-189.

Giles, R.H. (1978). *Wildlife Management*. Freeman, San Francisco, CA.

He, C., J.F. Riggs and Kang, Y.T. (1993). Integration of geographic information systems and a computer model to evaluate impacts of agricultural runoff on water quality. *Water Resources Bulletin* 29, pp. 891-900.

Koza, J.R. (1994). Introduction to genetic programming. In *Advances in Genetic Programming* edited by Kinnear, K.E. MIT Press, Cambridge, MA.

Jankowski, P. (1995). Integrating geographical information systems and multiple criteria decision-making methods. *International Journal of Geographical Information Systems*, 9(3), pp. 251-275.

Mark, D.M., and Lauzon, J.P. (1984). Linear quadtrees for geographic information systems, *Proceedings of the International Symposium on Spatial Data Handling*, Zurich, pp. 412-430.

Shoham, Y. (1993). Agent-oriented programming. *Artificial Intelligence*, 60, pp. 51-92.

USDA (1979). *Soil Survey of Unions County, Illinois*. National Cooperative Soil Survey.

Wesseling, C.G, Karssenberg, D., Burrough, P.A., and van Deursen, W.P. (1996) Integrating dynamic environmental models in GIS: The development of a dynamic modelling language. *Transactions in GIS*, 1(1), pp. 40-48.

Zhou, J. and Civco, D. (1996). Using genetic learning neural networks for spatial decision making in GIS. *Photogrammetric Engineering and Remote Sensing*, 62(11), pp. 1287-1295.