

# Optimising Nature Reserve Design: A Fuzzy Logic Approach

## Abstract

This paper describes an approach using fuzzy sets to help solve the real world optimization problem of nature reserve design: the objective is to preserve and protect a maximum number of species and habitat for a given amount of resources. The search for the optimal solution hinges on data that is characteristically unreliable and generally makes no allowance for the rarity of species under consideration. Using a fuzzy logic approach, this research was able to handle the uncertainty in input data, incorporate a measure of species rarity and also encapsulate a set of reserve heuristics into the solution. The fuzzy approach offers a more realistic real-world practical solution than does traditional approaches using 0/1 data values and the probabilistic approach.

## Introduction

Optimization has been extensively covered in the computer science literature; most work done in optimization assumes that the data used is reliable. However, for some real world problems, such as nature reserve design, this is not the case. In the reserve design problem, wildlife resource managers attempt to preserve the maximum number of species and viable habitat with a limited amount of resources. To do this, they must assess alternative reserve designs. Such evaluation ultimately hinges on species' presence or absence data for each design, and, as widely acknowledged in the literature, this data is notoriously unreliable (Margules & Nicholls, 1987; Pressey et al., 1995; Polasky et al., 2000; Mowrer, 2000). In reality, the presence/absence value is an approximation of a species' residency in a particular location. It is estimated from the number of visits to the site and evidence of the species having been recorded on these visits. Particularly for species with restricted ranges and most often at a high risk of extinction, the uncertainty in presence/absence records could lead to poor reserve selection and subsequent dire consequences.

One approach developed to cope with the recognized paucity of species presence/absence data is to use probability. Several researchers (Margules & Nicholls, 1987; Margules & Stein, 1989; Pressey et al., 1995) centred their investigations on how well individual species were represented in optimal reserve designs when probabilities were converted to presence/absence data. Later work by Polasky et al. (2000) looked at the effects on reserve choice when probability was used to compute the total expected number of species likely to be present in any one design.

This paper describes a different and novel approach that uses fuzzy sets to cope with the uncertainty of species' presence/absence data that is most commonly used in the optimization problem of reserve design. In addition, fuzzy sets are used to describe the World Conservation Union (International Union for the Conservation of Nature, IUCN) classifications of each species of interest, that are used together with presence/absence data in the selection of the optimal reserve design.

## The Approach

Reserve design is an optimization problem where the objective is to preserve and protect a maximum number of species and habitat for a given amount of resources. The optimal solution is approached by computing a value for each of the several candidate reserve systems and comparing their respective costs with the number of species protected. To calculate the costs and value of a reserve, it is useful to view a reserve as a set of sites. In fact, it is practical to look at all candidate reserve systems simply as various combinations and subsets of all available sites. To find the best subset of sites, i.e. an optimal reserve system, it becomes necessary to assign a site value to each location under consideration.

## Site Value

Site value can be evaluated as a function of cost, location and overall biodiversity. Cost and location are easily measured, however overall biodiversity is more difficult to determine. Previous approaches and software have attempted to compute a site's biodiversity value as a function of the summation of the number of species present at a site. Although, biologically it could be argued that the whole is greater than the component parts, the sum is a good comparative indicator to the ecological worth of a site. Thus, assuming that across sites the species presence/absence data is equally reliable, a site supporting ten different species is viewed as more worthwhile to a reserve system than a site supporting nine species. On first appearances, this would seem a reasonable statement, but in reality, a site visited by fewer species where one is a rare or endangered species needs to be considered more valuable when compared to a site hosting a multitude of non-threatened and more widely dispersed species. So, although the most

common scenario described in the literature is that of selecting reserves on presence/absence data (Gaston & Rodrigues, 2003), the measure of a site's biodiversity needs to include some rules or representation of the species' ecological worth.

### **Fuzzy Logic**

The research reported in this paper is an attempt to cope with both the uncertainty in species' presence/absence data and the incorporation of a measure of rarity, or risk of extinction, for each species when computing site and reserve values. To achieve this, fuzzy logic was used; the method was first proposed by Zadeh (Zadeh, 1965). Fuzzy logic, or fuzzy set theory, attempts to calibrate vagueness and uncertainty in sets of objects or variables that are not adequately described by sharp distinctions such as 0 or 1, endangered and not endangered. Fuzzy logic allows 'shades of gray', a continuum of values between 0 and 1, or between endangered and not endangered. In fuzzy logic, it is possible for objects to simultaneously maintain membership in more than one fuzzy set, thus describing their location in these continuums. For instance, a species' presence at a site can be stated as both "slightly unlikely" and as "just possible" and this statement can have meaning in the reserve design computations.

### **Species' Rarity Value**

Therefore in this research, the first task was to decide upon the input and output variables that would be used to evaluate individual reserve designs, i.e. collections of sites. As each site value was seen as a function of the number and rarity values of the species of interest, the problem was further decomposed to an evaluation of each species at the site. Each species' contribution (or lack thereof) to a site value depends on its local presence as well as its inherent risk of extinction or rarity value over its range. A species' rarity value represented by an IUCN classification is based on various field collections of data accumulated over differing locales, countries and continents and "usually, species are assigned to different categories based on the experience and intuition of researchers rather than on quantitative analyses" (Todd & Bergman, 1998). There is a history of using fuzzy logic to deal with the inherent uncertainty in the assigned risk classification value for a species. Todd & Bergman (1998) suggested a fuzzy set approach would be useful in assessing a species risk of extinction and further work by Akcakaya et al. (2000) to proposed a method for assigning species to a single IUCN category.

### **Choice of Fuzzy Variables and Fuzzy Rules**

Two input variables for each species were chosen and stored as fuzzy variables: the presence/absence value and the rarity measure. The fuzzy variable **Presence** was represented by the fuzzy sets **unlikely**, **possible**, **probable** and **highly likely**, which in turn give a linguistic measure to the likelihood of a particular species' presence at a particular site. The fuzzy variable **Rarity** was represented by the fuzzy sets **critically endangered**, **endangered**, **vulnerable** and **not threatened/least concern**, which paralleled the IUCN classifications for species vulnerability or risk of extinction. As the reserve design problem is an optimization problem, a single fuzzy output variable **Reserve Status** was defined by the fuzzy sets **poor**, **fair**, **good** and **best**. It is important to realize that although fuzzy variables and sets are used to represent the problem, the input values to the model are single crisp values that are converted to fuzzy values. For example, the likelihood of finding a species at a particular site may be 40%. This crisp value would be converted to a fuzzy variable and then represented by several membership values; in this case being 0.2 membership in **unlikely Presence** and 0.5 in **possible Presence**. The range of values or universe of discourse for membership, is between 0 and 1.

Having decided upon the fuzzy variables and their constituent fuzzy sets to describe the possible presence and rarity values for each species at a particular site, a set of sixteen fuzzy rules were created to incorporate some well known heuristics in reserve design. These fuzzy rules represented various relationships between the input fuzzy variables **Presence** and **Rarity** and the output **Reserve Status**. For example, one rule was

IF species **Presence** is highly likely  
AND species **Rarity** is critically endangered  
THEN **Reserve Status** value is best

and another rule

IF species **Presence** is likely  
AND species **Rarity** is not threatened  
THEN **Reserve Status** value is poor

These sixteen rules not only establish the relationship between all fuzzy variables, but are needed to compute a single crisp value for a species' **Reserve Status**. To do this, all sixteen rules are checked sequentially against the input values for the degree of membership, or truth, in the values of the IF components, or antecedents, of the rules. If there is any truth values for the antecedents, then the rule is said to fire the consequent to the same degree of membership as the combined antecedents. Using the first rule as an example, if a species has a degree of membership equal to 0.6 in the fuzzy set of highly likely for **Presence** and it also has a degree of membership equal to 0.4 in the fuzzy set of critically endangered for the fuzzy variable **Rarity**, then the antecedent value for the conjunction by fuzzy join is 0.4. A fuzzy join between antecedents is simply done by taking the minimum value of 0.6 and 0.4. Thus the antecedent value is 0.4 and the consequent of the rule is said to fire to a degree of 0.4 for the fuzzy set of best in **Reserve Status**. So, for any individual species at a particular site, it is possible for none to sixteen of the fuzzy rules to partially fire and generate several different degrees of membership in the fuzzy sets of **Reserve Status**. Finally, these values are converted, by finding the centroid, to a single crisp output representing a particular species at the site under investigation. This species' output value is then combined with other species' outputs and used to score the overall site value.

### Implementation

The fuzzy approach described above was coded as a set of Java classes using a library of fuzzy Java classes available with the Java Expert System Shell (JESS) tool available from Sandia National Laboratories, California (JESS, 2004). An instance of the main computational class, FuzzySpecies, collects the two crisp inputs for a particular species' presence at a site and their IUCN rarity value. This object converts these inputs to the relevant degree of membership values of the fuzzy variables **Presence** and **Rarity**. Next, all the fuzzy rules are processed with these fuzzy values and the outputs are collected from the fired rules for conversion to a single crisp value. Finally, this value is returned as a representation of the individual species at the location concerned. Another class FuzzySite represents a collection of species at one location. It was coded to create instances of FuzzySpecies and sum their return values for use in ranking one site against another.

The above approach has been evaluated using an artificial data set of twenty different species located across ten possible sites (Table 1). The data set used by Polasky et al. (2000) in their investigation of how probability predicts the expected number of species in a reserve system, given likelihood values for species presence at component sites. A comparison of Polasky et al. (2000) results for this artificial data will be made with the results from the fuzzy approach in this report. It should be noted that, the "expected number of species at the site" is simply a summation of the individual probabilities of species at each site. Polasky et al. (2000) converted these individual probabilities to presence/absence data by assuming that a species was present if the probability was 0.5 or greater and the total "number of species recorded as present at the site" reflects this conversion.

### Results

A reserve system can consist of one or many sites. Using the artificial data set, it was possible by complete enumeration, to find the optimal solution for reserve systems composed of a single site through to reserves of ten sites.

The values in Table 1 provide the likelihood of presence values for each species at each site for input into the fuzzy model, however a computation was needed to decide the IUCN value for each of the artificial A to T for use in the fuzzy approach. A summation across the rows of Table 1 for each species gave the following values (Table 2). The boundaries between classifications was chosen arbitrarily as the research being undertaken was to test proof of concept stage and, besides whatever the designation, a species could still have membership in several fuzzy sets of the **Rarity** value.

Firstly, reserve networks consisting of one site were run to find the optimal solution. The reserve system with the highest return value was site 1. An inspection of the data set shows that site 1 is the only site to have a likely population of the critically endangered species T. In addition, site 1 also has the highest total value for the number of species expected at the site, based on the likely presence values. Not surprisingly, this result agrees with that of Polasky et al. (2000) probabilistic method which only considered the overall expected number of species at a site, not the rarity values of the component species. The fuzzy choice of site 1 is at odds with the result of site 6 computed using 0/1 presence/absence values. Site 6 was chosen by this method since it has five species as coded as present, and no conservation value is given to sites 1 to 5 (understandably no consideration was given to the rarity of the species at these locations).

Sites	1	2	3	4	5	6	7	8	9	10
Species A	0.4	0.4	0.4	0.4	0.4	0.6	0	0	0	0
B	0.4	0.4	0.4	0.4	0.4	0.6	0	0	0	0
C	0.4	0.4	0.4	0.4	0.4	0.6	0	0	0	0
D	0.4	0.4	0.4	0.4	0.4	0.6	0	0	0	0
E	0.4	0.4	0.4	0.4	0.4	0.6	0	0	0	0
F	0.4	0.4	0.4	0.4	0.4	0	0.6	0	0	0
G	0.4	0.4	0.4	0.4	0.4	0	0.6	0	0	0
H	0.4	0.4	0.4	0.4	0.4	0	0.6	0	0	0
I	0.4	0.4	0.4	0.4	0.4	0	0.6	0	0	0
J	0.4	0.4	0.4	0.4	0.4	0	0	0.6	0	0
K	0.4	0.4	0.4	0.4	0.4	0	0	0.6	0	0
L	0.4	0.4	0.4	0.4	0.4	0	0	0.6	0	0
M	0.4	0.4	0.4	0.4	0.4	0	0	0	0.6	0
N	0.4	0.4	0.4	0.4	0.4	0	0	0	0.6	0
O	0.4	0.4	0.4	0.4	0.4	0	0	0	0	0.6
P	0.4	0.4	0.4	0.4	0.4	0	0	0	0	0
Q	0.4	0.4	0.4	0.4	0	0	0	0	0	0
R	0.4	0.4	0.4	0	0	0	0	0	0	0
S	0.4	0.4	0	0	0	0	0	0	0	0
T	0.4	0	0	0	0	0	0	0	0	0
Expected number of species at the site	8	7.6	7.2	6.8	6.4	3.0	2.4	1.8	1.2	0.6
Number of species recorded as present at the site	0	0	0	0	0	5	4	3	2	1

Table 1: Artificial data of likelihood of occurrence as used by Polasky et al.(2000, *Biological Conservation* 94, pg 4) in their probabilistic approach.

	Species	Sum of likelihood values across sites 1 to 10 of Table 1
CR Critically Endangered	T	0.4
EN Endangered	S	0.8
	R	1.2
VU Vulnerable	Q	1.6
	P	2.0
NT Not threatened, least concern	A thru O	2.6

Table 2: Generated rarity values for artificial species A thru to T derived from Table 1.

Next, an iteration of all the possibilities for reserve systems of two sites was undertaken. The optimal solution using the fuzzy approach were sites 1 and 6. This result is at variance with both the probabilistic selection of sites 1 and 2 and the presence/absence result of sites 6 and 7. As pointed out by Polasky et al. (2000), the presence/absence method chooses site 7 as the next site to be added to site 6 in a two-site reserve system because this model recognizes that site 7 adds four more species to the reserve. They argue however that the greatest increase in the expected number of species' presence is achieved by the addition of site 2 to site 1 (Table 3). Although the probabilistic method and the fuzzy approach agree that site 1 is the most important site, albeit for different reasons, the fuzzy approach indicates that the addition of site 6 to site 1 is the optimal solution.

Number of sites in reserve system	Fuzzy Reserve system	Probabilistic reserve system (Polasky et al., 2000)	Presence/absence reserve system (Polasky et al., 2000)
1	1	1	6
2	1,6	1,2	6,7
3	1,6,7	1,2,3	6,7,8
4	1,6,7,8	1,2,3,4	6,7,8,9
5	1,6,7,8,9	1,2,3,4,5	6,7,8,9,10
6	1,6,7,8,9,10	1,2,3,4,5,6	1,6,7,8,9,10
7	1,2,6,7,8,9,10	1,2,3,4,5,6,7	1,2,6,7,8,9,10
8	1,2,3,6,7,8,9,10	1,2,3,4,5,6,7,8	1,2,3,6,7,8,9,10
9	1,2,3,4,6,7,8,9,10	1,2,3,4,5,6,7,8,9	1,2,3,4, 6,7,8,9,10
10	1,2,3,4,5,6,7,8,9,10	1,2,3,4,5, 6,7,8,9,10	1,2,3,4,5, 6,7,8,9,10

Table 3: Optimal site selection using the fuzzy, probabilistic and presence/absence approaches. The probabilistic and presence/absence selections are those reported by Polasky et. Al (2000).

### Discussion

The disparity in these two site groupings between the fuzzy approach and probabilistic methods is a result of the join method used in combining sites. As an aside, it was possible to have the fuzzy approach behave in exactly the same manner as Polasky's method by coding a probabilistic join between sites rather than use the fuzzy (maximum value) join. Although Polasky's method looks at the sum of the probabilities of representation and the possible increase to be expected if a particular site is added, for explanation purposes a simplified comparison between the behaviours of probability and fuzzy joins can be made using probabilities.

Suppose one species has a likelihood of being found equal 0.4 at two independent sites A and B. To compute the likelihood of union for the species found at either site or both sites simultaneously:

Using probability

$$\text{Probability from union of two sites, A and B} = \text{Probability of site A} + \text{Probability of site B} - (\text{Probability of site A} * \text{Probability of site B})$$

in this case

$$\text{Probability from union of two sites, A and B} = 0.4 + 0.4 - (0.4 * 0.4) = 0.64$$

For this species, this represents an increase from 0.4 in a single site system to 0.64 in a two site system.

Using fuzzy join

$$\text{Fuzzy union of two sites, A and B} = \text{maximum value (site A, site B)}$$

in this case

$$\text{Fuzzy union of two sites, A and B} = \max(0.4, 0.4) = 0.4$$

For this species, there is no expected increase in likelihood of the species being found by the addition of a second site.

So the question now is whether a fuzzy join has meaning in the reserve design problem. It is an appropriate join when viewing the likelihood of a species being found at a particular site as an indicator of overall habitat quality at that location. Adding two locations with the same habitat quality equates to a larger reserve system, where the likelihood of finding a particular species across a larger area remains the same throughout. If joining two sites of different likelihoods, say 0.4 at site A and 0.6 at site B, then the fuzzy join returns the maximum value of 0.6 for the combined site reflecting an improvement in the reserve system of habitat quality. Thus examining the fuzzy optimal two-site solution, the addition of site 6 to site 1 improves the likelihoods of five different species already present at site 1 from 0.4 to 0.6.

Using the fuzzy approach, reserve systems of three to ten sites were also enumerated and the optimal reserve set reported as column 2 in Table 3: columns 3 and 4 of this table were supplied by Polasky et al.(2000). Examining all the results in this table, it is apparent that the traditional presence/absence approach to reserve design selected quite different optimal reserve solutions to those chosen by methods that allow for data uncertainty. The probabilistic approach, while allowing for uncertainty, assumed that each species is of equal contribution to a reserve system and

attempted to optimize the number of expected species at a site. This method used a probabilistic join to combine sites. This may not be the best representation of species' expected presence in natural reserves.

The fuzzy logic approach coped with the uncertainty inherent in the presence/absence of data records without the necessity of assuming that each species is of equal weight to the calculations. Quite the contrary, a rarity value was incorporated into the computation and allowance was made for uncertainty in the rarity measurement. Additionally, reserve heuristics or rules were used in the computations and fuzzy joins were used to combine sites. The advantage of the fuzzy join, is that it equates a species' expected presence over several sites as a function of habitat quality.

### **Conclusion and Future Work**

The fuzzy approach shows considerable promise in finding optimal solutions to the reserve design problem during testing on artificial data, particularly if a species' presence at a site is considered indicative of the quality of habitat.

Further experimentation to explore the behaviour of the fuzzy model under the influence of rarity compared with compared presence. The fuzzy approach described can further be elaborated using linguistic qualifiers known as hedges. For instance, a species rarity would be described as somewhat threatened or very, very threatened and these would have quite different computational meanings in the model. Additionally, it is possible to introduce meta-rules to the system. These could be used to state priorities amongst rule firings, chose particular sites in a system , or introduce other heuristics based on geographic locations and biological factors. Most importantly, the performance of fuzzy approach developed here needs to be tested using real-world biological data.

### **References**

- Akcakaya, H. Resit., Ferson, Scott, Burgman, Mark A., Keith, David A., Mace, Georgina M., Todd, Charles R. (2000) Making Consistent IUCN Classifications under Uncertainty. *Conservation Biology*. Volume 14. No. 4 pp 1001-1013
- Gaston, Kevin J. & Rodrigues, Ana S. L. (2003) Reserve Selection in Regions with Poor Biological Data. *Conservation Biology*. Volume 17. No 1. pp 188-195
- JESS (2004) The Rule Engine for the Java Platform. Sandia National Laboratories, Livermore, California. USA., as accessed 14<sup>th</sup> March, 2004 <http://herzberg.ca.sandia.gov/jess/>
- Margules, C. R & Nicholls, A. O. (1987) Assessing the conservation value of remnant habitat 'islands': mallee patches on the western Eyre Peninsula, South Australia. *Nature Conservation: The Role of Remnants of Vegetation*. Saunders, D.A., Arnold, G.W., Burbidge, A.A., Hophkins, A. LM (Eds), Surrey Beatty and Sons, Sydney. pp 89-102
- Margules, C. R. & Stein, J.L. (1989) Patterns in the distribution of species and the selection of nature reserves: an example from Eucalyptus forests in southeastern New South Wales. *Biological Conservation*. **50**. pp 219-238
- Mowrer, H.Todd. (2000) Uncertainty in natural resource decision support systems: sources, interpretation, and importance. *Computers and Electronics in Agriculture*. 27. pp 139-154
- Polasky, Stephen, Camm, Jeffery D., Solow, Andrew R., Csuti, Blair, White, Denis & Ding, Rugang. (2000) Choosing reserve networks with incomplete species information. *Biological Conservation*. 94. pp 1-10.
- Pressey R.L., Ferrier, S., Hutchinson, C.D, Sivertsen, D.P., Manion, G. (1995) Planning for negotiation: using an interactive geographic information system to explore alternative protected area networks. *Nature Conservation: The Role of Remnants of Vegetation*. Saunders, D.A., Arnold, G.W., Burbidge, A.A., Hophkins, A. LM (Eds), Surrey Beatty and Sons, Sydney. pp 89-102
- Todd, Charles R. & Burgman, Mark A. (1998) Assessment of Threat and Conservation Priorities under Realistic Levels of Uncertainty and Reliability. *Conservation Biology*. Vol 12, no 5. pp 966-974
- Zadeh, L. (1965) Fuzzy sets. *Information and Control*. 8. pp 338-353