FINDING ECOLOGICAL RESERVES: A DECISION MAKING APPROACH

Jorge Argáez, J. Andrés Christen and Miguel Nakamura

Comunicación Técnica No I-07-05/08-03-2007
(PE/CIMAT)

Finding Ecological Reserves: a Decision Making Approach

Jorge Argáez [a,*], J. Andrés Christen[b] and Miguel Nakamura[b]

[a] Faculty of Mathematics, UADY, Anillo Periférico Norte Tablaje Cat. 13615 Colonia Chuburná Hidalgo Inn, Mérida, Yucatán, México (999) 942 31 40

[b] Center for Mathematical Research, Apdo. Postal 402 Guanajuato, Guanajuato, C.P. 36000, México.

E-mail address: jac@cimat.mx (J.A. Christen), nakamura@cimat.mx (M. Nakamura).

[*] Corresponding author: Tel (52) 999 9423140. E-mail address: argasosa@uady.mx (J. Argáez)

Finding Ecological Reserves: a Decision Making Approach

ABSTRACT

In this paper we consider the problem of optimally finding a region for protection, which could be proposed as an ecological reserve. By adopting a Decision Theoretic approach we propose a suitable loss function which allows for the consideration of certain factors and/or restrictions arising in practice, and that have not been considered in the methods actually in use. The proposed loss function can be used in the context of the well-known *Single Large or Several Small* ecological debate, related to the form and connexity of the region to be proposed as reserve. The elements that define the loss function possess a clear interpretation from the biologist's point of view. Due to the nature of the space of actions, the searching process for the solution cannot be made exhaustively, and so we resort to numerical methods. A case study is presented, using 12 endangered species native from the peninsula of Yucatan, Mexico.

Keywords: Species conservation, Loss function, Species valuation, Budget restriction.

1.- Introduction

In recent years a great diversity of species has become extinct, mainly due to human activity. Accordingly, several species have now been labeled as endangered species, because their population and habitat have been evidently decreased. Several lists are currently maintained, like those generated by the International Union for the Conservation of Nature (IUCN), the Convention of International Trade in Endangered Species of Wild Fauna and Flora (CITES) and in Mexico, the Norma Oficial Mexicana (NOM 059). These contain the names of endangered species, and are frequently updated in accordance with the experience of experts. These lists alone are indication that the extinction of species is a topic of great concern for ecologists.

There is no standard definition of what protection means. The World Foundation for Environment and Development states that a protected area is "a geographically defined area which is designed or regulated and managed to achieve specific conservation objectives" (www.wfed.org/resources/glossary). On the other hand, the International Union for the Conservation of Nature and Natural Resources (IUCN) states that a protected area is "An area of land and/or sea especially dedicated to the protection and maintenance of biological diversity, and of natural and associated cultural resources, and managed through legal or other effective means" (www.unep-wcmc.org/protected_areas/data/sample/iucn_cat.htm). Agencies committed to the study and protection of biodiversity deal with the problem of creating ecological reserves, where certain species (ecosystems, in general) are to be protected. In practice, these agencies are to determine which region is more suitable to be protected not only from a purely ecological point of view, but also taking into account practical issues such as costs, land use and other economical factors. In this paper we focus on detecting an optimal geographical region to be put forward as a protected area, considering certain restrictions that are commonly encountered.

As is customary in this setting, we consider the region to be covered by a grid of nodes, so that the problem at hand is to decide which subset of nodes is to be protected. Let

$S = \{s_1,...,s_N\}$ be the set of nodes linked to a regular, square grid covering the region under study. The region A^* to be proposed as protected area will be a subset of S, $A^* \subset S$. Several ways exist to formulate this problem, depending on the restrictions and/or assumptions considered. When the target is to find A^* in such a way that the number of protected species is maximum and the number of nodes to be protected is known in advance, a common goal is to maximize the probability that each species is represented in at least one $s \in A^*$ (Camm et al., 2002). Malcolm (2001) assumes that the presence or absence of each species is known in advance in each node of the region. This author also assumes that the selection of nodes is made in two stages, each with a fixed budget assigned. The goal is to maximize the number of preserved species. Polasky et al. (2001) propose to find A^* in such a way that each species is represented in at least one of its nodes, under similar assumptions as those declared in Malcolm (2001).

In these formulations, the problem is tackled using linear programming. Costello and Polasky (2004) use a dynamic approach, assuming that the selection of nodes is made in stages. At each stage, the degree of development of each node is known reserved or not reserved- and the target is to maximize the number of protected species in A^*.

McDonnell et al. (2002) take into consideration the connexity of A^*, also using a linear programming approach. Their goal is for the reserved nodes to represent, at least, a certain percentage of the area occupied by each species. Therefore, they also assume that the distribution area of each species is known.

In the papers mentioned above there are several elements assumed to be known, but that in general are not easy to specify. These include the nodes of presence of the species, the nodes of absence or the number of nodes that should be protected. Even if these elements were known, the assumptions and the method used may noticeably restrict the space of solutions. The common target of these papers is to find a set of nodes that maximizes (or minimizes) a suitable objective function, subject to some restrictions. Moreover, none of these papers consider the case when the studied species feature different risks of extinction.

In this paper we do not assume that the distribution area for each species is known exactly. Instead, we assume that for each node we are provided with a probability of presence for each considered species (that may be obtained by a variety of methods, in particular see Argáez et al., 2005). We tackle the problem from a Decision Theoretical point of view, defining an appropriate loss function which allows for the consideration of certain constraints arising in practice.

This paper is organized as follows. Section 2 introduces the foundations of the decision theory for this problem, including an appropriate loss function. Section 2.1 deals with restrictions that are imposed to the solution, by means a modification of the loss function. Section 2.2 addresses some practical considerations that are useful for the determination of the loss function. A case study is presented in Section 3, where 12 species under different scenarios are considered in order to select regions (sets of nodes) to be proposed for protection.

2.- Method

2.1.- Decision theoretical framework

We follow a Decision Theoretical approach, as explained, for example, in Berger (1985, chap. 4). For greater clarity in the description of our method we proceed in three steps: First we address the case when a single node, s, is considered, and that a decision to protect or not to protect that node must be reached. This loss function is then used to construct the general case where the goal is to find an unrestricted region (set of nodes) for protection. Finally, we generalize the previous loss function to the case when practical restrictions are imposed.

Suppose for the moment that our purpose is to decide if a single node $s \in S$ should be protected. The decision (or action) space can be specified by means of the binary variable a(s), which takes the value 0 if the decision is not to protect s and 1 if the decision is to protect s. The action space is thus $A_s = \{0,1\}$, and $a(s) \in A_s$ indicates a possible action at node s. For the i-th species we consider the random binary variable $u_i(s)$, which takes on the value 0 if the i-th species is absent at s, and the value 1 if the i-th species is present at s. Let $p_i(s) = P(u_i(s) = 1)$. Let $\Theta_s^i = \{0,1\}$ be the set of states of nature on s for the i-th species, and consider the loss function, $L_s^i(u_i(s), a(s)) : \Theta \times A_s \to \Re$, where $u_i(s) \in \Theta_s^i$ and $a(s) \in A_s$. This loss function is summarized in Table 1. The quantities $x_i(s)$, $y_i(s)$, $z_i(s)$ and $t_i(s)$ in Table 1 need to be measured on the same scale (USD or Mexican Pesos, for example). In Section 2.2 we entertain a technique for interpretation and definition of these quantities, by means of practical considerations.

Table 1: Loss function for the species i at node $s \in R$.

Θ_s \ A_s	$a(s) = 0$	$a(s) = 1$
$u_i(s) = 0$	$x_i(s)$	$y_i(s)$
$u_i(s) = 1$	$z_i(s)$	$t_i(s)$

To denote the presence ($u_i(s) = 1$) or absence ($u_i(s) = 0$) of each one of the I species at node s, let $U_s = (u_1(s),...,u_I(s))$ and let $\Theta_s = \Theta_s^1 \times \cdots \times \Theta_s^I$. This last set is formed by all possible I-dimensional binary vectors, such that the value 0 (or 1) in the i-th position denotes the absence (or presence) of the i-th species at node s. Our proposed loss function, $L_s(U_s, a(s)) = \Theta_s \times A_s \to \Re$, where $U_s \in \Theta_s$ and $a(s) \in A_s$, for making the decision of protecting or not protecting node s, is

$$L_s(U_s, a(s)) = \sum_{i=1}^{I} w_i L_s^i(u_i(s), a(s)) \quad (1)$$

with $w_i \in [0,1]$, $\sum_{i=1}^{I} w_i = 1$, and $L_s^i(u_i(s), a(s)$ defined as above. The weight w_i is interpreted as the degree of importance assigned to the i-th species to be protected. If all the species are considered equally important, it would be sensible to postulate $w_i = 1/I$ for all i. Polasky et al. (2001) suggest a quantity that weights the importance of a species to be protected, but it is not actually accounted for.

By using the loss function summarized in Table 1 for each species, we obtain that the expected losses for the actions not to protect ($a(s) = 0$) and protect ($a(s) = 1$) node s are

$$
\begin{aligned}
L_s(0) &= \sum_{i=1}^{I} w_i \{x_i(s) - p_i(s)x_i(s) + p_i(s)z_i(s)\} \\
L_s(0) &= \sum_{i=1}^{I} w_i \{y_i(s) - p_i(s)y_i(s) + p_i(s)t_i(s)\}
\end{aligned}
\tag{2}
$$

respectively. As usual, the decision to protect node s will be reached if $L_s(1) \le L_s(0)$. In the postulation of (1) and in the computation of quantities (2) it is assumed that the species make themselves present over the study area independently, conditioned, perhaps, on the values of some covariates (see Argáez et al., 2003). In other words, given the vector of covariates at each node of the region, the sites where each species makes itself present are selected independently by each species. This assumption is also adopted by the methods currently used to tackle this problem (Malcolm, 2001; Polasky et al., 2001; Camm et al., 2002; McDonnell et al., 2002; Costello and Polasky, 2004).

Turning now to the issue of determining a region to protect, we note that the decision space is now given by the power set of S, that is, $\boldsymbol{A} = \boldsymbol{P}(S)$. The loss function $L(U, A) : \Theta \times A \to \Re$, with $\Theta = \prod_{s \in S} \Theta_s$, $U = \prod_{s \in S} U_s$ and $A \in \boldsymbol{A}$, that is now proposed is

$$
L(\boldsymbol{U}, A) = \sum_{s \in S} L_s(\boldsymbol{U}_s, a(A, s)) \tag{3}
$$

where $a(A,s) = 0$ if $s \notin A$, and $a(A,s) = 1$ if $s \in A$, and $L_s(\boldsymbol{U}_s, a(A,s))$ is defined as in (1). The $a(A,s)$ notation generalizes $a(s)$, and is in fact the indicator function $I_A(s)$. By calculating the expected value for (3) and using expressions (2), we obtain the expected loss for a set $A \in \boldsymbol{A}$:

$$
\begin{aligned}
L^*(A) = \sum_{s \notin A} \sum_{i=1}^{I} w_i \{x_i(s) - p_i(s)x_i(s) + p_i(s)z_i(s)\} \; + \\
\sum_{s \in A} \sum_{i=1}^{I} w_i \{y_i(s) - p_i(s)y_i(s) + p_i(s)t_i(s)\}
\end{aligned}
\tag{4}
$$

Let $A^{*} = \arg\min_{A \in \boldsymbol{A}} \{L^{*}(A)\}$, that is to say, A^{*} is the set of nodes that minimizes $L^{*}(A)$. This is the region to be protected if no restrictions apply. But as mentioned earlier, in real applications, there are some restrictions that could be imposed to the solution. This is considered in the following sections.

2.2- Imposing Restrictions

In real applications it is natural to assume that the region to be put forth for protection satisfies given restrictions which arise due to economical and/or ecological reasons. In what follows, we consider two types of restrictions commonly adopted in real case studies. The first restriction is called Budget Restriction, arising because there usually exists a fixed quantity of money (budget) assigned to the protection of species. The second restriction is called Restriction for Connectivity, which originates in biological and/or economical considerations, and gives rise a debate about the dispersion of the nodes that form the region to be protected (Section 2.1.2). In the following sections we discuss how to properly introduce these restrictions in the decision process.

2.2.1.- Restriction for budget

Let B be the budget assigned for the protection of species. The protection of a node s involves a cost $c(s)$, which in this paper is assumed as monetary, and is interpreted as the quantity to be invested if node s is selected for protection. We do not strive here to give a general approach for defining $c(s)$. However, note that $c(s)$ may take into consideration several factors like economic activity, accessibility, land costs or even political factors, depending on the type of "protection" entertained by the study. For example, $c(s)$ could be interpreted as the mean cost of the land represented by node s or it could be any indicator of its value. Moreover, $c(s)$ could represent the cost involved in buying, fencing, and managing the land represented by s. In any case, if the region $A \in \boldsymbol{A}$ is proposed to be protected, the quantity $\sum_{s \in A} c(s)$ will be committed, and this sum should not exceed B. We introduce this restriction by means of considering the (restricted) decision space

$\boldsymbol{A}_B = \{A \in \boldsymbol{A} : \sum_{s \in A} c(s) \leq B\}$, which only considers regions that meet the budget B (note that $A = A_\infty$. In this case, $A^* = \arg\min_{A \in \boldsymbol{A}_B} \{L^*(A)\}$ is the region to be protected.

2.2.2.- Restriction for Connectivity

In the literature on species protection there exist a debate driven by the question: Is it best to invest in a Single Large or Several Small reserves? This debate is known as the SLOSS debate, and for both options there are biological (Baz and García-Boyero, 1996) and economical (Drechsler and Wätzold, 2001) arguments put forth.

We do not adopt a particular point of view of the SLOSS debate. Instead, we consider both options of the debate by analyzing regions with different degrees of fragmentation. In order to measure the fragmentation of the zones we consider the perimeter of the region, which in our context, is defined as the length of the contour of the surface determined by the nodes conforming the region. It is clear that a highly fragmented area will have a larger perimeter than a connected area containing the same number of nodes. Let $H(A)$ be the perimeter of the region A. To take into account the SLOSS debate we propose to use $H(A)$ as an additional term in the loss function. The influence that this term will have on the solution will be modulated by a parameter $\beta \in [0, \infty)$ associated with $H(A)$. The suggested loss function is

$$L_\beta(\boldsymbol{U}, A) = L(\boldsymbol{U}, A) + \beta H(A) \tag{5}$$

For a fixed β, the region proposed for protection is $A^*_\beta = \arg\min_{A \in A_\beta} \{L^*_\beta(A)\}$, where $L^*_\beta(A) = E\big[L_\beta(\boldsymbol{U}, A)\big]$.

We expect a great deal of work in matching the units in which $x(s), y(s), z(s)$ and $t(s)$, in Table 1, are measured. On the other hand, matching these units with β seems to be an insurmountable and rather controversial task. We propose a practical and perhaps more useful procedure, instead. Allow the parameter β to be manipulated by the user in order to observe different geographical configurations of the region proposed to be protected. This procedure permits the exploration of different fragmentation levels for the region. If $\beta = 0$, then the perimeter has no role in the loss function and $A_0^* = A^*$ ($L_0^*(A) = L^*(A)$, for all *A*). As the value for β increases, a greater importance to the protection of non-fragmented areas is imposed.

It is possible to compare the corresponding solutions A_β^* by means of the corresponding expected losses $L^*(A_\beta^*)$, that is, by comparing the loss of A_β^* without the term $\beta H(A_\beta^*)$; it is clear that $L^*(A^*) \le L^*(A_\beta^*)$, for all $\beta \ge 0$. Taking a set of values $\beta_0, \beta_1, \ldots \beta_G$, with $\beta_0 = 0$, we consider a table displaying the percentage of change in the expected loss of the region obtained with $\beta_g, g \ge 1$, compared with the loss obtained with β_0. The comparison can be made by means of a table with the β value in one column and the percentage $L_\beta = \left\{ L^*(A_\beta^*) / L^*(A_0^*) - 1 \right\} \cdot 100$ in another. The quantity L_β may be interpreted as the percentage loss of considering the connexity imposed by β_g (in Section 3, Table 4, we use these ideas to compare regions with different degrees of fragmentation).

2.3.- Further considerations about the loss function

The general form of $L_s^i(\cdot,\cdot)$ is stated in Table 1. Here we make further considerations for simplifying this loss function. Let $s \in S$ and suppose for the moment that only the i-th species is considered. When the correct decision of not protecting the node s is taken (the i-th species is not present at s), there is no loss, and we postulate $L_s^i(0,0) = 0$, that is, $x_i(s) = 0$. On the other hand, if it is decided not to protect the node but the species is present at s, then $L_s^i(1,0) = z_i(s)$. The quantity $z_i(s)$ is interpreted as the cost (loss) involved in "misprotecting" species *i* at node *s*. We find it reasonable to assume that the value $z_i(s)$ depends on the species and not on the geographical position of the node, thus $z_i(s) = z_i$ for all $s \in R$.

If it is decided to protect the node s and the species is not present at s, the quantity $c(s)$ will be expended. In this case $c(s)$ is considered a loss, because this quantity will be wrongly used. That is, we postulate $L_s^i(0,1) = c(s)$, or $y_i(s) = c(s)$, assuming that this cost is the same for all species. Finally, if we decide to protect node s and the species is present at s, the quantity $c(s)$ invested should not be considered a loss, since it will be correctly used. Moreover, when the correct decision is taken at node s, a gain should be considered, resulting from the protection of species *i*. Therefore, we find it sensible to postulate that the gain is in fact the quantity that would be lost in the case of misprotecting the species, that is, $L_s^i(1,1) = -z_i$. Applying these considerations, the simplified loss function obtained for the i-th species, is summarized in Table 2. Indeed, it is assumed that the quantities z_i and $c(s)$ are measured in a common unit. Using (5) and taking expected values as in (2) and (4) we obtain

$$L_\beta^*(A) = \sum_{s \notin A} \sum_{i=1}^{I} w_i p_i(s) z_i + \sum_{s \in A} \left\{ c(s) - c(s) \sum_{i=1}^{I} w_i p_i(s) - \sum_{i=1}^{I} w_i p_i(s) z_i \right\} + \beta H(A) \qquad (6)$$

Minimizing $L_\beta^*(A)$ is not trivial since the search space has $2^{|S|}$ elements, $|S|$ usually being in the several hundreds. In the Appendix we present a simple numerical solution to this problem. A comprehensive simulation study, taking many different scenarios into consideration, may be found in http://www.cimat.mx/~jac/areasim.html.

Table 2: Loss function for species I at node $s \in R$ (particular case).

Θ_s \ A_s	$a(s) = 0$	$a(s) = 1$
$u_i(s) = 0$	0	$c(s)$
$u_i(s) = 1$	z_i	$-z_i$

3.- Results

We use an example from the Yucatán Peninsula, in Mexico. Twelve endangered plant species are listed in Table 3, along with the corresponding values w_i and z_i. These values where genuinely elicited by experts from the Research Center of Yucatan. For more details on this on other aspects of this example see Argáez (2003).

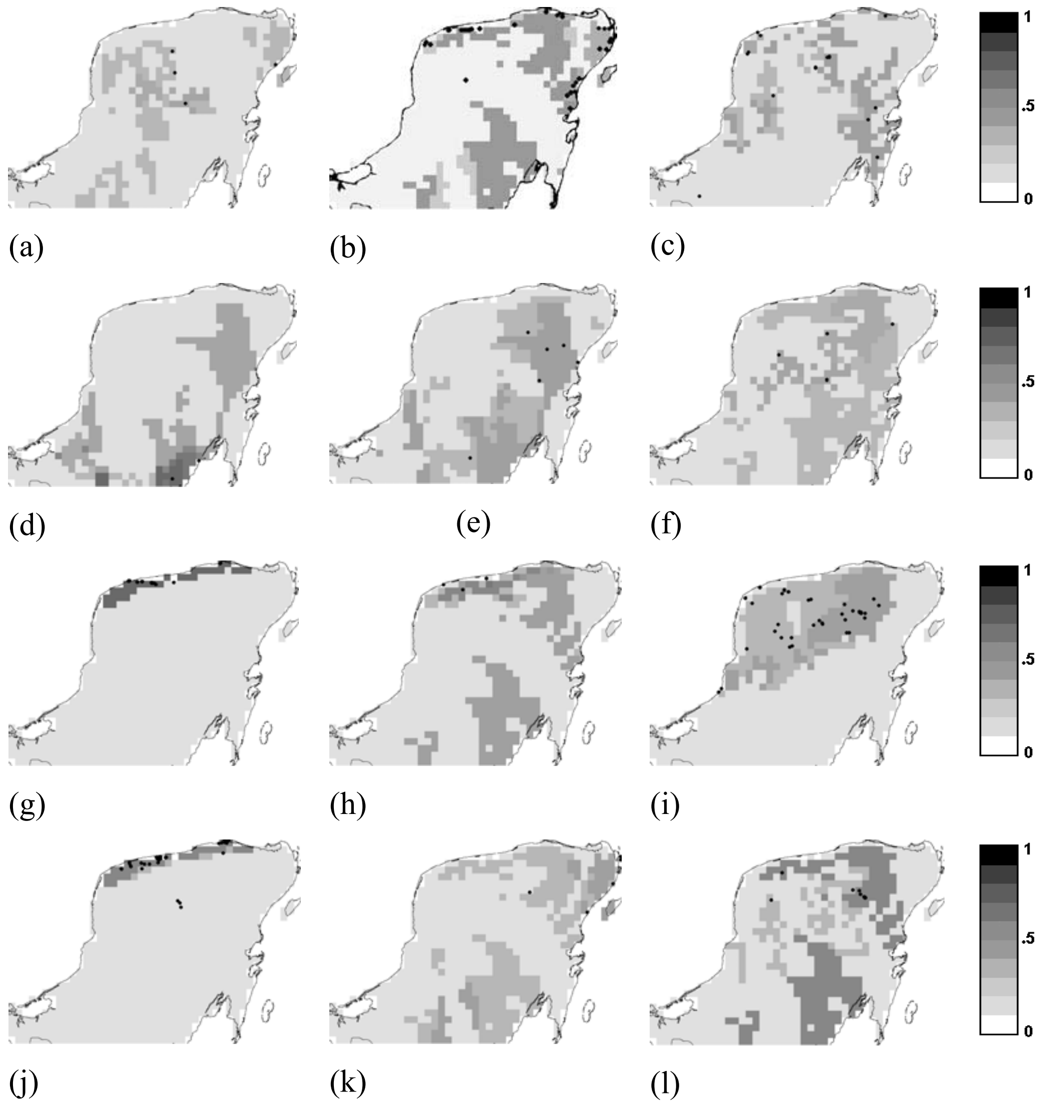
According to w_i values, the species considered of greatest importance for protection are C. myriantha, G. maya, and M. aenea, whereas the species considered less important are C. readii, L. longistylus, M. yucatanensis, and X. yucatanense. On the other hand, by observing the values z_i, in the last column of Table 3, the species considered more valuable are M. yucatanensis and X. yucatanense, whereas the species less valuable are G. maya and S. nanum.

Table 3: Species under study and quantities of interest.

Species	w_i	z_i
Carlowrightia myriantha	0.213	92.27
Coccothrinax readii	0.021	516.98
Furcraea cahum	0.051	92.27
Gaussia maya	0.213	24.69
Gonolobus yucatanensis	0.044	83.71
Lonchocarpus longistylus	0.022	422.07
Mammilaria gaumeri	0.092	85.06
Matelea aenea	0.150	90.54
Matelea yucatanensis	0.008	1220.97
Pterocereus gaumeri	0.064	137.19
Stenandrium nanum	0.102	79.57
Xanthosoma yucatanense	0.020	1026.74

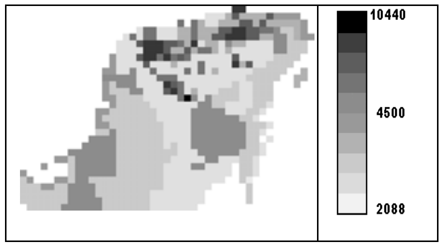
Maps of probability of presence for each species are depicted in Figure 1. These where obtained using historical recording sites for each species, a set of climatic and topographic covariates and prior information, using the methods in Argáez (2003, chap. 1) and Argáez et al. (2005).

Figure 1: Map of probability of presence for (a) C. myriantha (b) C. readii (c) F. cahum (d) G. maya (e) G. yucatanensis (f) L. longistylus (g) M. gaumeri (h) M. aenea (i) M. yucatanensis (j) P. gaumeri (k) S. nanum (l) X. yucatanense.



A map of $c(s)$ (land protection cost) used in this application is depicted in Figure 2, that was obtained using a combination of the map of Economical Regionalization (MER; see García and Alonzo, 1999) and the map of Regions of Productive Specialization (MRPS; see Aké et al., 1999). For more details see Argáez (2003, Chap. 3). The quantities z_i were defined as the sum of costs of those nodes contained in the zones where an expert substantiates that the species under study may be present with high probability. Thus, z_i is measured in Mexican Pesos, so that the quantities $c(s)$ and z_i are comparable.

Figure 2: The c(s) (land protection cost) used in the example.



Regarding the budget, we consider the quantities $B = .05C_T$ (low budget) and $B = .15C_T$ (high budget), where $C_T = \sum_{s \in S} c(s)$. In this application $C_T = 2{,}703{,}840$. So, we consider $B = 135{,}192.7$ and $B = 405{,}578.1$, respectively. For *β* the values *β*=0, *β*=5, and *β*=10 were selected, after inspection of $L_\beta(A)$. The values postulated for each one of the factors produces a total of 6 scenarios, whose results are summarized in Figure 3. In this figure we observe that the regions obtained by using the value *β*=10 are less fragmented than the corresponding regions obtained using the values *β*=0 or *β*=5, as it is expected. Compare for example, Figures 3(a), (b), and (c), which corresponds to the budget B=135 192.7, and Figures 3(d), (e), and (f), which corresponds to the budget B=405 578.1.

By comparing Figures 3 and 1, we observe that regions that contain nodes of high probability for the majority of the species are included in A^*_β. The species which is less represented in A^*_β in each case is F. cahun. From Table 1, we see that this species is considered relatively less important for protection, since it has the value $w_3 = .051$, and compared with the other species, it has an average biological value, with $z_3 = 92.27$. G. maya has the greater influence in the region to be protected, since $w_4 = .213$ is specified for this species. By comparing the maps to be protected depicted in Figure 3 with the map of probability of presence for this species (Figure 1(d)), we observe that regions proposed include, in each case, nodes with high probability of presence for the species.

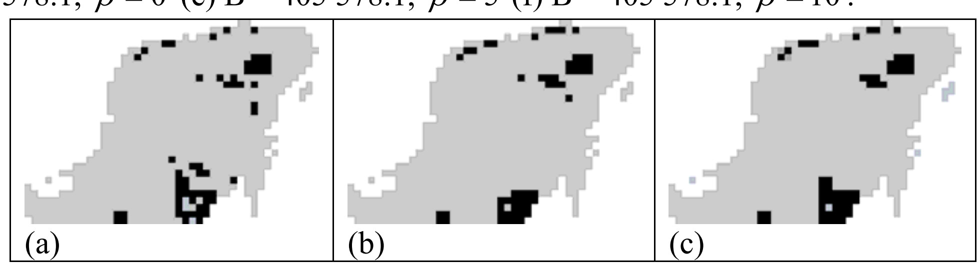
In order to determine which region (in Figure 3) will be proposed, we proceed to compare the corresponding loss functions, as explained above. For the budgets B=135 192.7 and B=405 578.1 the results may be found in Table 4.

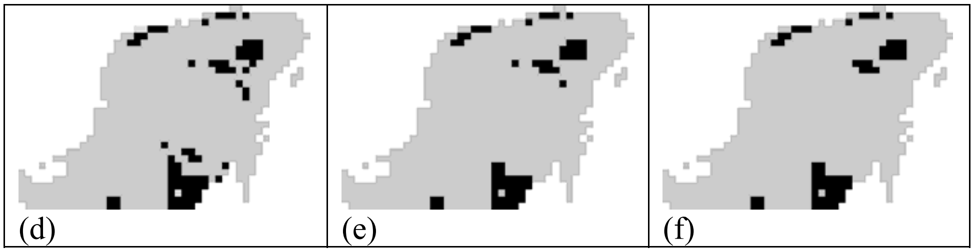
Table 4: Relative expected loss L_{β_h} for budgets B = 135 192.7 (middle column) and B = 405 578.1 (right column), for example.

β	B=135 192.7	B=405 578.1
0	0%	0%
5	69.89%	-13.45%
10	0.47%	-15.70%

For a fixed budget let A_0^*, A_5^*, and A_{10}^* be the regions obtained to be protected with the values β=0, β=5, and β=10, respectively. If the low budget (B=135 192.7) is considered, from Table 4 we observe that the expected loss of region A_5^* is 69.89% less than the expected loss obtained with A_0^*. On the other hand, the expected loss obtained with region A_{10}^* is 0.47% less than the expected loss obtained with A_0^*. From these results we conclude that the region A_5^* is to be proposed for protection. In contrast, if the high budget (B=405 578.1) is considered, from Table 4 we observe that the expected loss of the region A_5^* represents a -13.45% more than the expected loss of the region A_0^*. In the same table, we observe that the expected loss for the region A_{10}^* is -15.70% greater than the expected loss obtained with A_0^*. In this case, the region that would be proposed for protection is A_0^*.

Figure 3: Suggested regions for protection for the 12 species considered for the various scenarios (a) B = 135 192.7, $\beta = 0$ (b) B = 135 192.7, $\beta = 5$ (c) B = 135 192.7, $\beta = 10$ (d) B = 405 578.1, $\beta = 0$ (e) B = 405 578.1, $\beta = 5$ (f) B = 405 578.1, $\beta = 10$.





Regarding the regions obtained to be protected, in Espadas et al. (2003) the so-called endemism areas over the Peninsula of Yucatan were obtained, considering several species, including the 12 species used in the case study of this paper. The endemism areas are relevant due to the fact that in these areas several species are located, and then the conservation of the areas allows for the protection of several species at once. The regions obtained to be protected considering the 12 species, which are observed in Figure 3, are located in the areas of endemism described by Espadas et al. (2003). This fact provides an empirical test that the methodology proposed in this paper produces relevant results.

4.- Discussion

Methodology introduced in this paper does not automatically give regions that consist of nodes of high probability of presence for each species under study. It takes into account (1) the probability of presence of each species in each node, $p_i(s)$, (2) a biological value for each species, z_i, (3) a level of importance regarding protection that each species has, w_i and (4) the cost involved in protecting each node, $c(s)$. As mentioned in the Introduction, none of the existing methods consider all these elements in its procedures.

On the other hand, the Decision Theory approach allows us to consider these four elements simultaneously, in contrast to the other methods where certain strong restrictions need to be imposed in order to be able to obtain a solution. Those restrictions produce a solution that is in general very restrictive, for example, in the number of nodes that are obtained.

The region that will be proposed to be protected is formed by nodes corresponding to species for which the probability of presence is high, species more highly valuated or species considered more important to be protected. The user has the option of observing different solutions, depending on the importance assigned to the preference for connected areas (given by $\beta H(A)$ in the loss function).

If additional restrictions are relevant, the approach developed here allows their inclusion in at least two ways: (1) It is possible to consider a more restrictive decision space, or (2) it is possible to consider an additional term in the loss function. For example, we could restrict the number of nodes k that can be protected, by restricting the action space to $\left\{A \in \mathbf{A} : \sum_{s \in A} c(s) \leq B \text{ and } |A| \leq k\right\}$.

The numerical minimization of the expected loss requires some care when this method is implemented. It will be necessary to verify that the iterative procedure is reaching convergence. In order to verify this we recommend to run code several times and to compare the resulting zones. If these zones are quite similar, it is reasonable to conclude that the resulting zone corresponds to the optimal solution (see details in Appendix).

Acknowledgments

J.A. Christen is partially founded by CONACYT grant 40233-F. This work was completed while J. Argáez was a PhD student of CIMAT, supported by grant 115344 from CONACYT. We would like to thank the Centro de Investigación Científica de Yucatán, for the data and Celene Espadas for helping in the postulation of values regarding the valuation of species and nodes. We thank Fernando Avila and Eloisa Diaz-Frances for reading preliminary versions of the manuscript.

Appendix: Minimizing $L_\beta(A)$.

We consider non-exhaustive searching iterative algorithms to minimize $L_\beta(A)$ in (6). In particular we use the greedy algorithm (McDonnell et al. 2002) and the simulated annealing method (Bertsimas and Tsitsiklis, 1993). In McDonnell et al. (2002) both algorithms are compared in the context of designing spatially connected ecological reserves. After an extensive simulation study, they conclude that, in general, the greedy algorithm was quicker to run, but the simulated annealing gave better results in terms of the loss function. In this paper we use the solution obtained with the greedy algorithm as the initial solution in the simulated annealing algorithm. That is, the solution obtained by the greedy algorithm is improved upon by simulated annealing to obtain a global minimum.

We first use a greedy algorithm (McDonnell et al. 2002). Let $A^{(t)} \in A_B$ denote the current solution at iteration t. At iteration t+1 all the neighboring solutions of $A^{(t)}$ are considered and the solution producing the greater decrement in the expected loss is taken as the new solution, $A^{(t+1)}$. A neighboring solution A' of $A^{(t)}$, is defined as the current solution with an additional node (not necessarily connected to some node of $A^{(t)}$). The iterative process is repeated until none of the neighboring A''s has a lower expected loss that the solution at hand. It is clear that a node that has been selected as part of the current solution will be part of the final solution and thus a local minimum could be reached.

Secondly, we consider a simulated annealing type algorithm (see, for example, Bertsimas and Tsitsiklis, 1993). Given a candidate solution $A^{(t)}$ in iteration t, the simulated annealing algorithm considers again a neighboring solution A' of $A^{(t)}\}$, which is obtained by selecting a node s at random from S. If $s \in A^{(t)}$, then s is removed from $A^{(t)}$, and if $s \notin A^{(t)}$, then s is added to $A^{(t)}$, to form a proposal solution A'. Solution A' is accepted as the current solution, $A^{(t+1)} = A'$, with probability $\rho = \min\{1, \exp(-\left[L^*_\beta(A') - L^*_\beta(A^{(t)})\right] / T(t)\}$, where $T(t)$ is a non-increasing function called the cooling schedule. Many types of cooling schedules have been suggested (see Cohn and Fielding, 1998). In this paper we use the logarithmic cooling schedule defined by

$T(d) = d / \log(1+t)$, where d is a constant related with the quantity of energy necessary for escaping from the local minima.

REFERENCES

Aké, A., Jiménez, J., Ruenes, M., 1999. El solar Maya, in: A. García and J. Córdoba (cords), Atlas de Procesos Territoriales de Yucatán. Facultad de Arquitectura:UADY.

Argáez, J., 2003. Estimación de la probabilidad de presencia de una especie con base en mediciones de covariables y diseño de zonas para proteger especies. PhD Thesis, CIMAT: Mexico.

Argáez, J., Christen, J.A., Nakamura, M., Soberón, J., 2005. Prediction of potential areas of species distributions based on presence-only data. Environmental and Ecological Statistics 12, 27-44.

Baz, A., García-Boyero, A., 1996. The SLOSS dilemma: a butterfly case study. Biodiversity and Conservation 5, 493-502.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer-Verlag.

Bertsimas, D., Tsitsiklis, J., 1993. Simulated annealing. Statistical Science 8(1), 10-15.

Camm, J.D., Norman, S.K., Polasky, S., Solow, A.R., 2002. Nature reserve site selection to maximize expected species covered. Operations Research 50, 946-955.

Cohn, H., Fielding, M., 1998. Simulated annealing: searching for an optimal temperature schedule. SIAM Journal of Optimization 9(3), 779-802.

Costello, C., Polasky, S., 2004. Dynamic reserve site selection. Resource and Energy Economics 26, 157-174.

Drechsler, M., Wätzold, F., 2001. The importance of economic cost in the development of guidelines for spatial conservation management. Biological Conservation 97, 51-59.

Espadas, C., Durán, R., Argáez, J., 2003. Phytogeographic analysis of taxa endemic to the Yucatan peninsula using geographic information systems, the domain heuristic method and parsimony analysis of endemicity. Diversity and Distributions 9(4), 313-330.

García, A., Alonzo, A., 1999. Regionalización Económica, in: A. García and J. Córdoba (cords). Atlas de Procesos Territoriales de Yucatán. Facultad de Arquitectura: UADY.

Malcolm, S.A., 2001. Sequential land acquisition decision for nature reserves under acquisition and population uncertainty. Operational Research. Department of food and resource economics, College of Agriculture and Natural Resources. University of Delaware.

McDonnell, M., Possingham, H.P., Ball, I.R., Cousins, E., 2002. Mathematical methods for spatially cohesive reserve design. Journal of Environmental Modeling and Assessment 7(1), 107-114.

Polasky S., Camm, J.D., Garber-Yonts, B., 2001. Selecting biological reserves cost-effectively: an application to terrestrial vertebrate conservation in Oregon. Land Economics 77(1), 68-78.