Incorporating multiple ecological criteria in classical zero one selection algorithms

D. P. Memtsas, P. G. Dimitrakopoulos and A. Y. Troumbis


Avifauna on Greek wetland sites is used as a model for the implementation of the Set Covering Problem in selecting nature reserves. Three site conservation values, which depend on species presence, are used as selection criteria. Their calculation is based upon species richness, species rarity and species-danger status. The conservation values must be inserted in the linear programming problems objective function by the form of weighting factors. Optimal solutions according to the three ecological criteria are produced. These solutions belong to the set of alternative optimal solutions of the basic Set Covering Problem with no other criterion taken into account except that of the whole species-list coverage. The set of alternative optimal solutions is generated by the explicit exclusion method. The relative value of goal programming and weighing up-criteria methods in producing a unique solution based on the three criteria simultaneously is assessed. Both methods coincide with the same alternative solution that is thus regarded as the final optimal one incorporating all the three ecological criteria.

D. P. Memtsas (dmem@env.aegean.gr), P. G. Dimitrakopoulos and A. Y. Troumbis, Biodiversity Conservation Laboratory, Dept of Environmental Studies, Univ. of the Aegean, GR-81100 Mytilene, Greece.

Selection of nature reserves is a common problem decision makers face all over the world in order to establish nature reserve networks for the protection of biodiversity (e.g. the NATURA 2000 network in Europe, Directive 92/43/EU, GAP project in the USA). Reserve selection methods have advanced beyond simple scoring procedures (Margules and Usher 1981), where a numerical score reflecting reserve’s “conservation value” was used to detect top-ranked reserves in need of protection, to more complicated heuristic selection algorithms and mathematical programming methods (Margules et al. 1988, Kershaw et al. 1995, Dobson et al. 1997, Freitag et al. 1997).

Three criteria based upon species presence and absence data have been proposed for setting priorities in the reserve selection problem (Primack 1993): a) “Distinctiveness” that is higher priority is given to sites including rare endemic species in contrast to those including common ones, b) “Endangerment” that is priority is given to species in danger of extinction, c) “Utility” when sites with species valued more highly by people are given a higher priority for conservation. These species richness criteria apply to single sites. Prioritisation of sites on them leads to the determination of the high-value top-ranked sites according to the criterion set. Such a process has been applied for the determination of the richness, rarity and threat to hotspots (Kershaw et al. 1995, Williams et al. 1996, Reid 1998, Troumbis and Dimitrakopoulos 1998).

On the other hand, specific criteria for whole-system evaluation have been proposed for the selection process so as to produce the best total reserve system (Primack 1993,
Church et al. 1996): a) “Complementarity” that is each new site added should complement previously selected ones by contributing as many as possible species not represented yet in the reserve system; b) “Representativeness” that is the selected reserves should be as representative as possible of the biological diversity; c) “Irreplaceability” that is sites including rare species or communities that are not present anywhere else have to be included in any network representing all species; d) “Flexibility” that is alternative solutions should be generated so as the manager could take into account other external factors not included in the problem initially.

Heuristic reserve selection algorithms have been criticised because they are unlikely to produce optimal solutions (Underhill 1994, Curti et al. 1996, Pressey et al. 1996). Selection of nature reserves by the use of operations research techniques may answer the problem of non-optimality of heuristic methods. By these methods the optimal reserve system can be selected, either minimising the set of selected reserves representing all the species or maximising the number of species represented for a fixed total number of reserves (Camm et al. 1996, Church et al. 1996, Pressey et al. 1997, Ando et al. 1998). Nowadays these methods are more usually applied as easy-to-use optimization software products are developed (LINDO, CPLEX, LP-SOLVE etc).

These problems are limited by the fact that they include only one objective function. More advanced methods have to be implemented if the selection is based upon multiple criteria. In such cases the decision-maker may be confused by conflicting solutions that result from the application of the algorithm on each criterion individually. Goal programming or multiobjective methods have to be additionally used so as a solution that adequately satisfies all criteria is found. In goal programming (Dykstra 1984, Gass 1985), the best compromise is found from among the alternative solutions that minimise deviation of the objective function from the optimal score it could achieve on each criterion separately. In multiobjective methods (Cohon 1978, Rothley 1999) all alternative optimal solutions (if any) are initially produced and afterwards they are extensively investigated and the best compromise solution is selected among them. In recent years multicriteria decision making techniques have been developed as a special part of operations research to help the decision maker in such complicated multiple criteria selection problems (Salminen et al. 1998, Klimberg and Cohen 1999, Shields et al. 1999, Hayashi 2000, Joerin and Musy 2000, Zopounidis and Doumbos 2000).

The purpose of this paper is to assess nature reserve selection algorithms by linear programming methods when different ecological selection criteria are taken into account. In particular we wish to 1) investigate how different conservation values upon sites or species may differentiate the solution produced by the algorithms, 2) to compare the solutions produced based upon different conservation values with the corresponding hot-spot areas upon the same criteria, 3) to investigate how goal programming methods can be used to produce an unique solution when multiple criteria are taken into account.

**Materials and methods**

**Data base**

Seventy one wetland sites belonging in the CORINE-biotopes database for Greece (Fig. 1) were selected as a model to apply operations research techniques to select subsets of them according to different priorities or criteria (Anon. 1991). All selection algorithms were run according to wetlands’ bird species richness. One hundred and ninety three different birds (breeding and non-breeding migratory and wintering birds) are recorded in these wetland sites.

**Conservation values**

Three site conservation values based on species presence were estimated: 1) species richness score, 2) species rarity score, and 3) species-threat score. Species richness score (A) was just the number of bird species recorded on each site j. Rarity was measured as follows. For each species the number of its occurrences on all sites was measured and first single-species rarity score was evaluated as the inverse of the previous measurement. The species rarity score of each site j (B) was then calculated by adding up all single-species scores for the species present on the site (Usher 1986, Williams et al. 1996):

\[ B_j = \sum \frac{1}{d_{ij}}, \]

where \( d_{ij} \) is the number of total occurrences of species i for all species i that exist on site j.

The degree of species-threat was calculated by taking into account the number of species falling into each one of the categories in the Red Data Book of threatened vertebrates in Greece (Karandinos 1992), namely “endangered”, “vulnerable”, “rare” and of “unidentified status”. A weight of 1, 0.75, 0.5 and 0.25 was assigned to each species falling into each one the four categories respectively (Freitag and van Jaarsveld 1997). The species threat score of each site j (C) was finally evaluated by summing up the products of the number of species in each category multiplied by the corresponding weight.

\[ C_j = 1 \cdot n_{e} + 0.75 \cdot n_{v} + 0.5 \cdot n_{r} + 0.25 \cdot n_{u}, \]

where \( n_{e}, n_{v}, n_{r}, n_{u} \) are the numbers of endangered, vulnerable, rare and unidentified bird species that exist on site j.
Set Covering Problem (SCP)

The Set Covering Problem has a similar goal with heuristic selection algorithms looking for the best subset of reserves where all species are represented. Through it the optimal subset of reserves is found simultaneously instead of sequentially selecting new reserves, as heuristic algorithms do. The object of the Set Covering Problem is to choose the minimum set of reserves containing all species at least once. The problems can be mathematically expressed as follows: assume that \( J = \{j/j = 1,…, n\} \) denotes the index set of candidate reserves from which to select, and \( I = \{i/i = 1,…, m\} \) denotes the index set of the species to be covered. The general linear programming problem can then be formulated as (Camm et al. 1996):

\[
\begin{align*}
\text{Minimise} & \quad \sum_{j \in J} x_j \\
\text{subject to} & \quad \sum_{i \in N_j} x_j \geq 1, \quad \text{for all } i \in I \\
\end{align*}
\]

where \( N_i = \{ j / \text{species } i \text{ exists in site } j \} \)
\[ x_j = (0,1), \quad \text{for all } j \in J \]

The binary variables \( x_j \) equal one when its corresponding site \( j \) is selected otherwise they equal zero. The constraints correspond to each one of the total \( m \) species present in all sites and each one safeguards the inclusion of a specific species in the final solution. The summation in the left part of each constraint contains the binary variables \( x_j \) corresponding to the sites including the particular species \( i \). The summation

\[ \sum_{j \in J} x_j \]

is the so-called objective function \( f(x_1, x_2, \ldots, x_J) \). Optimisation problems are associated with such a function that determines how good a solution is.

Site characteristics such as area, cost, etc. may also be used in the formation of the objective function when the minimisation of the final system’s area, cost, etc. is the final goal. By a similar approach, any ecological characteristic may also be included in the formation of the objective function by expressing it as a conservation value. Thus a numerical score \( w \), reflecting a conservation value attributed to each site may be used as a weighting factor in the formation of the objective function. In these cases the minimisation of the objective function objective function takes the form:

\[ \text{Minimise } \sum_{j \in J} w x_j \]

Problems solved in the analysis

By the use of site conservation values the Set Covering Problem was solved with the following objective functions:

1. Minimise the number of reserve sites so that each species is included in at least one chosen site,

  \[ \text{Minimise } \sum_{j \in J} x_j \]
2. Minimise the number of reserve sites so that each species is included in at least one chosen site, but at the same time maximise reserve richness score,

$$\text{Minimise } \sum_{j \in J} a_j x_j,$$

where $a_j$ is an appropriately defined weighting factor reflecting the richness score of site $j$.

3. Minimise the number of reserve sites so that each species is included in at least one chosen site but at the same time maximise reserve rarity score,

$$\text{Minimise } \sum_{j \in J} b_j x_j,$$

where $b_j$ is an appropriately defined weighting factor reflecting the rarity score of site $j$.

4. Minimise the number of reserve sites so that each species is included in at least one chosen site but at the same time maximise reserve threat score.

$$\text{Minimise } \sum_{j \in J} c_j x_j,$$

where $c_j$ is an appropriately defined weighting factor reflecting the threat score of site $j$.

The three objective functions of the set-covering problem including a conservation value as a weighting factor might seem irrational as they would have incorporated two opposite objectives, a minimisation of the number of reserves and a maximisation of the site conservation value score. A simple trick was used to overcome this problem. The coefficients $a_j$ were evaluated by the use of their corresponding site conservation values $A_j$, by applying the formula:

$$a_j = \frac{\max(A_1, \ldots, A_n) - A_j}{\max(A_1, \ldots, A_n) - \min(A_1, \ldots, A_n)} + 0.0001,$$

$$j = 1, 2, \ldots, m.$$

The same formula was used for the evaluation of the coefficients $b_j$ and $c_j$ by the use of their corresponding $B_j$ and $C_j$ site conservation values respectively. By these calculations an inverse rank of the sites according to their conservation values was produced as high-conservation valued sites were assigned low weights. These calculations have also the advantage that all the three weighting conservation factors were scaled to a 0–1 scale and so all the numerical results upon them are comparable. Finally in order to avoid the problem of a zero weighting factor (as a zero coefficient results in the exclusion of its corresponding site from the final solution) a trivial quantity of 0.0001 was added to each weighting factor. The constraints were exactly the same in all the four problems as they are presented in the formulation of the general problem. Mixed integer program solver LINGO was used for the solution of the single objective problems (Anon. 1999).

**Simplification analysis**

The Set Covering Problem may be simplified that so large data sets may be processed by the following procedure. First all species occurring on only one site are recorded; we call them the mono-site species, and we also call the corresponding sites on which they exist mono-specific sites. The mono-specific sites have to be included in any final solution since all species have to be represented. All other species occurring on the mono-specific sites may thus be eliminated from the species’ list since their occurrence in the final solution is certain thanks to the mono-specific sites inclusion.

**Production of alternative optimal solutions**

The optimal solution of the basic Set Covering Problem with no weighting coefficients in the objective functions and containing all species at least once in the minimum number of reserves (objective function 1), is not necessarily unique, as alternative optimal sets may exist. However, the incorporation of some conservation values as weighting factors in the objective function (objective functions 2, 3, 4) results in specific alternative solutions safeguarding optimality according to a criterion. Besides that, a systematic approach is necessary in order to produce the complete set of the alternative optimal solutions to the basic problem.

The complete set of the alternative solutions to the basic set-covering problem may be produced by the Explicit Exclusion method (Arthur 1997). This method involves the repeated solution of the problem by: 1) evaluating the minimum number of sites, be it $k$ out of $n$, containing all species at least once, 2) repeating the same process by adding each time an additional constraint guaranteeing in turn the exclusion of the optimal solution found in the previous step. This process may be continued until no other feasible solution can be found.

**Goal programming**

A unique solution to the Set Covering Problem, upon the multiple selection criteria may be finally determined by goal programming. By the single-objective functions’ runs where each criterion was utilised separately, the optimum values of the objective functions according to each criterion are determined. These optimum values are afterwards used as the achievement levels in a goal programming analysis by which the best compromise solution is found. Goal programming was formulated as a one-way goal problem (Dykstra 1984) permitting an overcoming of the optimum single objective function performance values (as the objective function was in all three weighted cases expressing a minimisation). The object of the problem was to
determine the solution “minimising” the sum of the deviations from the optimum values according to the three different criteria. Assuming that the optimum values of the objective function were \( M_1, M_2, \) and \( M_3 \) according to the three different criteria the objective function can be formulated as:

\[
\text{Minimise } d_1 + d_2 + d_3
\]

where \( d_1, d_2, \) and \( d_3 \) express the possible exceeding in each single objective function’s value for the three criteria respectively. Three additional constraint equations incorporating \( d_1, d_2, \) and \( d_3 \) were necessary in the formulation of the problem:

\[
\sum_{j \in J} w_j x_j - d_i \leq M_i, i = 1, 2, 3
\]

where \( M_i, i = 1, 2, 3 \) was the optimum performance score that was evaluated by the single objective function solution of the problem for each one of the selection criteria.

**Results**

The minimum number of reserves containing all species at least once was 26 reserves in all cases either weighted according to a criterion (species richness, rarity or threat) or not (Fig. 2). The minimum set of 26 wetlands representing the whole Greek wetlands’ avifauna constitute 37% of the 71 CORINE wetland sites that were included in the study. The SCP optimal solutions based upon the three criteria were not identical, having some minor differences. They can be grouped as follows: 1) species richness as a selection criterion, 2) species rarity or threat as a selection criterion. These solutions are members of the complete alternative optimal solutions set. Fifteen alternative optimal solutions were found by the explicit exclusion method. The corresponding site richness, rarity and threat scores are presented in a pay-off table (Table 1) according to the SCP objective functions at each single-objective optimal and hot-spot solution. The complete set of the alternative optimal solutions as well as their conservation value scores is presented in Table 2.

Simplification analysis resulted in 47 mono-site species occurring on 23 mono-specific sites that have to be included in any SCP solution (Fig. 2). A number of 142 other species were also present on these sites that could be eliminated. The problem could thus be simplified by including only four species occurring on nine sites, instead of 193 species on 71 sites. This means of simplifying the problem may be useful in case involving large data sets.

Goal programming produced a final single solution as a compromise between the alternative solutions to the SCP that were found when weighted by a site conservation value. The solution found was identical with the particular one that resulted according to the site rarity or threat criterion. Thus the minimisation of the objective functions scores’ deviations from their optimal performance scores was achieved with a deviation of zero for the rarity- and threat-weighted objective functions and an excess of 0.04 units for the richness weighted one.

The fifteen alternative solutions were ranked using a simple multi-attribute rating technique (Rothley 1999) where equal weights (1/3 for each one) were attributed to the three site criteria. The top solution found by that procedure according to the function score ranking coincides with the goal programming solution. The corresponding additive value function scores for each one of the alternative optimal solutions is also presented in Table 2.

**Discussion**

The 26 top-ranked hot-spot sites of richness, rarity and threat failed to represent the whole avifauna of wetlands, as they included a portion < 75% of it. Thus hotspot analysis, although a useful tool in setting priorities for conservation planning, is subordinate in comparison to selection algo-

<table>
<thead>
<tr>
<th>Species richness</th>
<th>Species rarity score</th>
<th>Site richness score</th>
<th>Site rarity score</th>
<th>Site threat score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted by site richness</td>
<td>193</td>
<td>75.43</td>
<td>13.32</td>
<td>15.68</td>
</tr>
<tr>
<td>Weighted by site rarity</td>
<td>193</td>
<td>75.43</td>
<td>13.36</td>
<td>15.61</td>
</tr>
<tr>
<td>Weighted by site threat</td>
<td>193</td>
<td>75.43</td>
<td>13.36</td>
<td>15.61</td>
</tr>
<tr>
<td>Richness hot-spots</td>
<td>140</td>
<td>55.84</td>
<td>9.66</td>
<td>15.73</td>
</tr>
<tr>
<td>Rarity hot-spots</td>
<td>144</td>
<td>57.57</td>
<td>11.60</td>
<td>14.85</td>
</tr>
<tr>
<td>Threat hot-spots</td>
<td>140</td>
<td>55.84</td>
<td>9.87</td>
<td>16.00</td>
</tr>
</tbody>
</table>
rithms since top-ranked sites although hot-spots of richness, rarity or threat fail to constitute a complementary solution of species diversity. On the other hand a rather high percent of 37% of sites is necessary to represent all species, indicating that the 5% level that is widely used in the determination of hot-spots (e.g. Prendergast et al. 1993, Williams et al. 1996, Freitag et al. 1997) is too low and represents only a small portion of the whole species’ list.

Table 2. Alternative optimal solutions, their corresponding site-richness, rarity and threat scores as well as their additive value function scores. The best solution according to the additive value function coincides to the particular one according to site rarity or threat. The same solution was also found by the application of the goal programming method. Bold values are the optimum scores found according to site richness, site rarity, site threat and additive value function scores respectively.

<table>
<thead>
<tr>
<th>Alternative solution</th>
<th>Site richness score</th>
<th>Site rarity score</th>
<th>Site threat score</th>
<th>Additive value function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.83</td>
<td>15.87</td>
<td>16.09</td>
<td>15.26</td>
</tr>
<tr>
<td>2</td>
<td>13.66</td>
<td>15.75</td>
<td>16.06</td>
<td>15.16</td>
</tr>
<tr>
<td>3</td>
<td>13.78</td>
<td>15.79</td>
<td>16.10</td>
<td>15.22</td>
</tr>
<tr>
<td>4</td>
<td>13.66</td>
<td>15.78</td>
<td>15.81</td>
<td>15.08</td>
</tr>
<tr>
<td>5</td>
<td>13.49</td>
<td>15.66</td>
<td>15.78</td>
<td>14.98</td>
</tr>
<tr>
<td>6</td>
<td>13.61</td>
<td>15.69</td>
<td>15.82</td>
<td>15.04</td>
</tr>
<tr>
<td>7</td>
<td>13.56</td>
<td>15.78</td>
<td>15.66</td>
<td>15.00</td>
</tr>
<tr>
<td>8</td>
<td>13.39</td>
<td>15.66</td>
<td>15.64</td>
<td>14.90</td>
</tr>
<tr>
<td>9</td>
<td>13.51</td>
<td>15.69</td>
<td>15.68</td>
<td>14.96</td>
</tr>
<tr>
<td>10</td>
<td>13.49</td>
<td>15.80</td>
<td>15.78</td>
<td>15.03</td>
</tr>
<tr>
<td>11</td>
<td>13.32</td>
<td>15.68</td>
<td>15.76</td>
<td>14.92</td>
</tr>
<tr>
<td>12</td>
<td>13.44</td>
<td>15.72</td>
<td>15.80</td>
<td>14.99</td>
</tr>
<tr>
<td>13</td>
<td>13.53</td>
<td>15.73</td>
<td>15.58</td>
<td>14.95</td>
</tr>
<tr>
<td>14</td>
<td>13.36</td>
<td>15.61</td>
<td>15.56</td>
<td>14.84</td>
</tr>
<tr>
<td>15</td>
<td>13.48</td>
<td>15.64</td>
<td>15.59</td>
<td>14.91</td>
</tr>
</tbody>
</table>

Although the basic SCP solution has the complementarity property that any satisfactory selection process should incorporate (Primack 1993, Church et al. 1996), it does not meet the criterion of flexibility since only one solution is found by a single run of the problem. That is, there are not alternative solutions for the manager in case the network found by a single run of the algorithm cannot be established due to external factors. Besides that, the solu-

![Fig. 2. All wetland sites appearing in any one of the alternative 26-site optimal solutions. In all cases (regardless the criterion set) the minimum number of reserves containing all species at least once were 26 sites. A circle indicates each one of the 23 monospecific sites included in any solution, while a diamond indicates the rest alternate sites by which the additional three are selected.](image-url)
tion found in such a case may be just a random one (produced just by the random ranking of species or sites in the formulation of the problem) among a set of possible equivalent alternative optimal solutions. The inclusion of a weighting factor in the objective function may answer the problem of randomness since it takes into account a prioritisation of sites according to some numerical score of them and selects the particular optimal solution achieving the best objective function's performance according to the selection criterion. The weighting factor may be either a conservation value based upon species abundance (richness, rarity, vulnerability) or any other site characteristic that the manager may take into account (area, cost etc.). By a similar approach a weighting factor referring to species and not to sites, might also be used in the objective function of the other basic operation research algorithm (Maximal Coverage Problem) that maximises species richness when a given number of reserves are selected (Church and Revelle 1974, Camm et al. 1996).

Besides that, slight differences observed in the criteria scores (Table 2) also underline the necessity of producing alternative optimal solutions. In such a case, a solution achieved by a weighted objective function according to a criterion may in practice be equivalent or only slightly superior to other alternative solutions that could perhaps be easier applied. Goal programming, although producing a compromise between alternative solutions to the criteria, may in practice be equivalent or only slightly superior to other alternative solutions that could perhaps be easier applied. Goal programming, although producing a compromise between alternative solutions to the criteria, produces only a single solution also and thus fails to meet the flexibility property.

Furthermore, additional attributes based upon other site characteristics may also be incorporated as selection criteria in the investigation of the alternative optimal solutions according to the initial criteria, as they may contribute additional columns in the pay-off table and help the manager to make a stronger decision. The manager may carry out this process by using a multicriteria selection method to rank the alternative solutions according to a scoring procedure and finally select the best one satisfying all the criteria set.

Solutions based upon different criteria or databases are generally different (Prendergast et al. 1993, Kershaw et al. 1995, Williams et al. 1996, Reid 1998, Van Jaarsveld et al. 1998) and a further analysis is necessary in order to arrive at a single solution. On solution of the basic Set Covering Problem, alternative optimal solutions have to be found as they are all equivalent and they may constitute alternative choices to a manager. When no single alternative optimal solution exists, sub-optimal alternative solutions may be found by the constraint method, the parametric programming, etc. (Cohon 1978, Dylstra 1984, Gass 1985).

References