THE ROYAL SOCIETY

Received 20 June 2002
Accepted 3 September 2002
Published online 21 October 2002

Incorporating spatial criteria in optimum reserve network selection

Hayri Önal[1*] and Robert A. Briers[2]

[1]*Department of Agricultural & Consumer Economics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[2]*Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3DA, UK*

Considering the spatial location of sites that are to be selected for inclusion in a protected reserve network may be necessary to facilitate dispersal and long-term persistence of species in the selected sites. This paper presents an integer programming (IP) approach to the reserve network selection problem where spatial considerations based on intersite distances are taken into account when selecting reserve sites. The objective is to reduce the fragmentation of preserved sites and design a compact reserve network. Two IP formulations are developed which minimize the sum of pairwise distances and the maximum intersite distance between all sites in the reserve network, respectively, while representing all species under consideration. This approach is applied to a pond invertebrate dataset consisting of 131 sites containing 256 species in Oxfordshire, UK. The results show that significant reductions in reserve fragmentation can be achieved, compared with spatially unrestricted optimum reserve selection, at the expense of a small loss in reserve efficiency.

Keywords: integer programming; reserve selection; spatial fragmentation; intersite distance

1. INTRODUCTION

In the light of increasing awareness of the scale of global biodiversity loss (Ehrlich & Wilson 1991; Sala *et al.* 2000), the problem of designing efficient networks of protected reserves for the conservation of biodiversity has attracted significant attention from conservation biologists over the past two decades (Kirkpatrick 1983; Margules *et al.* 1988; Vane-Wright *et al.* 1991; Nicholls & Margules 1993; Pressey *et al.* 1993, 1996, 1997; Church *et al.* 1996; Ando *et al.* 1998; Polasky *et al.* 2001). In its simplest form, the problem is stated as determining the minimum number of reserve sites that represent a given set of species at least once. This is a special case of a prototype problem known as the 'set covering problem' or SCP (Camm *et al.* 1996). Limitations on the availability of resources for conservation may not allow the protection of all species under consideration, in which case the problem is stated as finding a subset of reserve sites that maximizes the number of species represented under a given budget constraint or a restricted number or area of sites that can be conserved. This is a special case of the prototype 'maximal covering problem' or MCP (Church & ReVelle 1974; Camm *et al.* 1996; Church *et al.* 1996).

Both the SCP and MCP can easily be represented as integer programming (IP) problems and solved using commercial optimization software (Underhill 1994; Camm *et al.* 1996; Rodrigues & Gaston 2002*b*). However, many studies have used heuristic, rule-based algorithms to guide site selection (e.g. Nicholls & Margules 1993). The standard 'greedy heuristic' procedure selects at each step a reserve site that adds the largest number of species to the set of represented species (the complementarity principle (Pressey *et al.* 1993)). IP approaches are guaranteed to find an optimal solution to the problem (providing it is analytically tractable), whereas heuristic methods give approximate solutions that may occasionally be optimal but in some cases can be significantly suboptimal (Underhill 1994; Camm *et al.* 1996; Önal 2002*b*). The use of heuristic methods in reserve site selection is motivated primarily by the perceived computational difficulty of IP approaches, i.e. an excessive processing time may be required to solve large IP problems. A second major criticism of exact optimization approaches is the difficulty of modelling some practical problems that cannot be adequately represented in the standard SCP and MCP frameworks (Pressey *et al.* 1996, 1997). It has been shown that the first argument is largely invalid (Rodrigues & Gaston 2002*b*; Önal 2002*b*). The remaining challenge is to extend the conventional SCP and MCP formulations to address more complex conservation problems (Rodrigues *et al.* 2000; Rodrigues & Gaston 2002*a*; Önal 2002*a*).

A potentially important constraint on reserve site selection, which has been generally neglected in the literature, is the spatial location of selected sites (Nicholls & Margules 1993; Briers 2002). It is possible to envisage situations where the close proximity of sites would be undesirable, such as spatially correlated environmental fluctuations or the spread of disease between sites (Possingham *et al.* 2000; Shafer 2001). More generally, however, clustering of reserve sites may enhance the long-term persistence of species by allowing dispersal and colonization of adjacent sites. Potential reserve sites are unlikely to be contiguous and in such fragmented habitats intersite dispersal may be important to species' regional persistence (Harrison 1994; Hanski & Simberloff 1997).

Spatial criteria are generally incorporated into heuristic approaches through rules to ensure that in the event of a choice between sites, the site that is closest to an existing site or which minimizes the total distance between all sites is selected (Nicholls & Margules 1993; Briers 2002).

* Author for correspondence (h-onal@uiuc.edu).

Here, we demonstrate how this type of spatial consideration can be incorporated into an IP formulation of the reserve site selection problem, through the minimization of a linear objective function based on intersite distances.

2. MATERIAL AND METHODS

(a) *Model specification*

The notation used in the models is as follows: $I = \{1,\ldots,N\}$ is the set of reserve sites; $S = \{1,\ldots,K\}$ is the set of species under consideration; X_i is a binary variable where $X_i = 1$ indicates that reserve i is in the network, otherwise $X_i = 0$; Z_{ij} is a binary variable where $Z_{ij} = 1$ indicates that both reserve $i \in I$ and reserve $j \in I$ are in the network, otherwise $Z_{ij} = 0$; d_{ij} is the distance between reserve $i \in I$ and reserve $j \in I$ (from centre to centre or edge to edge); δ_{is} is the Kronecker delta where $\delta_{is} = 1$ indicates that reserve $i \in I$ includes species $s \in S$, otherwise $\delta_{is} = 0$.

The SCP formulation that minimizes the number of selected sites while representing each species s at least k_s times ($k_s \geqslant 1$) is as follows.

$$\text{Min } \sum_i X_i$$

such that:

$$\sum_i \delta_{is} X_i \geqslant k_s \quad \text{for all } s \in S.$$

In this formulation, the only criterion for reserve site selection is the requirement that each species will be represented by selecting at least k_s sites including that species, without any spatial consideration. In the empirical application that will be presented below we assume that $k_s = 1$ for all s (solution A).

The reserve site selection problem considered here is to reduce the fragmentation of preserved sites and design a compact reserve network while representing all species under consideration. One way in which this can be achieved is by minimizing the sum of the distances between all pairs of sites included in the reserve network. The following IP model determines an optimum reserve network with minimum total distance between all pairs of reserves in the network (solution B):

$$\text{Min } \sum_{j>i}^{N} \sum_{i=1}^{N} d_{ij} Z_{ij}$$

such that:

$$Z_{ij} \geqslant X_i + X_j - 1 \quad \text{for all } i \in I \text{ and } j \in I, \text{ with } j > i$$
$$\sum_{i=1}^{N} \delta_{is} X_i \geqslant k_s \quad \text{for all } s \in S$$
$$X_i = 0{,}1 \text{ for all } i \in I,\ Z_{ij} = 0{,}1 \text{ for all } i \in I \text{ and } j \in I, \text{ with } j > i.$$

The first constraint plays a key role in the workings of this model. Consider the case where both $X_i = 0$ and $X_j = 0$ for some i and $j > i$. In this case, $Z_{ij} \geqslant -1$. The zero lower bound for Z_{ij} and minimization of the objective function together imply that $Z_{ij} = 0$, because otherwise a positive value d_{ij} would be added to the objective function value. Similarly, $Z_{ij} = 0$ when $X_i = 0$ and $X_j = 1$ or $X_i = 1$ and $X_j = 0$. Finally, if both $X_i = 1$ and $X_j = 1$, then the first constraint implies that $Z_{ij} \geqslant 1$. As the upper bound for Z_{ij} is 1, then $Z_{ij} = 1$. This implies that $Z_{ij} = 1$ only if both reserve i and reserve j are included in the network, in which case the distance between the two reserves will be accounted for in the objective function. Therefore, the objective function value represents the sum of the distances between all pairs of selected sites in the network, as desired. The second constraint is the usual species-covering constraint.

Although being theoretically adequate, the above model would not be a practical tool in empirical applications because of the number of integer variables involved. The main reason for this is the number of Z_{ij} variables, which is $N(N-1)/2$. Defining the Z_{ij} variables as non-negative continuous variables, rather than binary variables, circumvents this problem. Despite this relaxation, the optimum solution will always be binary for all Z_{ij}. This is ensured by the first constraint and minimization of the objective function. To see this, consider the case where both $X_i = 0$ and $X_j = 0$ for a pair of reserves i and $j > i$. Again $Z_{ij} \geqslant -1$, but $Z_{ij} \geqslant 0$ overrides this restriction, and minimization of the objective function implies that $Z_{ij} = 0$. This would again be the case for $X_i = 0$ and $X_j = 1$ or $X_i = 1$ and $X_j = 0$. If both $X_i = 1$ and $X_j = 1$, then the first constraint implies that $Z_{ij} \geqslant 1$, but minimization of the objective function would always imply that $Z_{ij} = 1$.

Relaxation of the binary nature of Z_{ij} is a crucial step. This leaves the X_i variables as the only binary variables in the model and therefore reduces the number of binary variables from $N(N-1)/2$ to N. The computational difficulty when working with IP models is determined primarily by the number of integer variables, rather than the column or row size of the problem, because the size of the branch and bound tree depends directly on the number of integer variables (see Nemhauser & Wolsey 1988).

This model would be tractable in most empirical applications involving up to a few hundred reserve sites. The processing time required to obtain an optimal solution for larger problems may be unacceptably long (but see Önal 2002*b*; Rodrigues & Gaston 2002*b*). If this is the case then the Z_{ij} variables could be restricted to those reserve pairs that are in a specified proximity, which would effectively restrict the selection of sites that could be potentially included in the reserve. The model can be solved by specifying a maximum distance between any two selected reserves and defining a variable only for those pairs within that proximity. The solutions obtained with alternative proximity specifications could be analysed further for determining a final network design.

An alternative way to incorporate pairwise distances when selecting an optimum reserve network is to minimize the largest distance between selected reserves, instead of the total distance. This can also be modelled conveniently using IP. Define a new variable $D \geqslant 0$, which denotes the largest pairwise distance. The 'mini-max' formulation is as follows (solution C):

$$\text{Min } D$$

such that:

$$D \geqslant d_{ij}(X_i + X_j - 1) \quad \text{for all } i \in I \text{ and } j \in I, \text{ with } j > i$$
$$\sum_{i=1}^{N} \delta_{is} X_i \geqslant k_s \quad \text{for all } s \in S$$
$$X_i = 0{,}1 \text{ for all } i \in I,\ D \geqslant 0.$$

To see how this model works, consider first the case where $X_i = 0$ and $X_j = 0$ for a pair of reserves i and $j > i$. The first constraint implies that $D \geqslant -d_{ij}$, but because of the non-negativity restriction $D \geqslant 0$. The case is similar when $X_i = 1$ and $X_j = 0$ or $X_i = 0$ and $X_j = 1$. In all these cases, the first constraint will not have any impact on the optimum solution. When both $X_i = 1$ and $X_j = 1$, then $D \geqslant d_{ij}$. Therefore, D must be greater than the distances between all pairs included in the reserve network. The minimization of D ensures that it is exactly equal to the largest of those distances. Thus, the optimum value of D is the distance between reserves i and j that are farthest apart and can be interpreted as the largest diagonal of the reserve network (viewed as a polytope). In general, the reserve site selection problem is a multi-criteria decision-making problem since usually more than one attribute is involved when designing a reserve network. The above formulations all assume a single objective. Goal programming techniques can be employed to incorporate multiple objectives by optimizing the primary objective while restricting the remaining attributes in the model constraints by specifying a goal for each (see Romero 1989). For instance, one may restrict the number of sites to be included in the network or impose an upper bound on pairwise distances between selected sites while minimizing the total distance. The conventional SCP and MCP formulations usually have multiple optimum solutions (Camm *et al.* 1996). By incorporating additional relevant constraints, such as the number or cost of sites selected, into the model framework, one may be able to identify the preferable solution(s) among the multiple optimum solutions. Such flexibility presents another powerful aspect of using IP models when determining optimum reserve networks. This approach is used in the empirical application presented below, where the sum of pairwise distances between selected sites is minimized subject to the constraint that the number of sites selected must be no greater than the number required by the standard SCP formulation (solution D).

(b) *Dataset used*

The dataset to which these approaches are applied is the pond invertebrate dataset used by Briers (2002). This consists of 131 pond sites in Oxfordshire, UK, which contained a total of 256 species of invertebrate. The data were derived from the Oxfordshire Pond Survey carried out between 1989 and 1990 by Pond Action. Further details of the sites and survey methodology are given in Pond Action (1994*a*,*b*). Not all pond sites in Oxfordshire were surveyed, but here it is assumed that there are no intervening sites available for selection. As a result of this the intersite distances are considerably larger than would be expected for most ponds.

(c) *Evaluation of solutions*

All of the IP models detailed above were solved using GAMS (Brooke *et al.* 1992), incorporating OSL (Optimization Subroutine Library, SC23-0519-1, IBM Corporation) as the IP solver. The GAMS code for implementing the models is available from the corresponding author. The solutions to the reserve site selection problem were evaluated on the basis of two criteria, namely the number of sites required to represent all species (a measure of efficiency (Pressey & Nicholls 1989)) and reserve fragmentation. The latter is measured by the total distance between all pairs of selected sites and the maximum distance between any pair of selected sites.

3. RESULTS

Table 1 summarizes the results of the alternative solutions, and figures 1 and 2 illustrate the spatial locations of selected sites for two of the solutions.

The unrestricted SCP formulation (solution A in table 1) required 30 out of 131 potential sites to represent all 256 species at least once (see figure 1). The next two IP formulations minimize the total distance and the largest pairwise distance between all selected reserves (solutions B and C, respectively, in table 1). Both formulations required 31 reserves to be included in the network, although the sites selected are marginally different. Owing to the similarity of these two solutions, only solution B is illustrated (see figure 2). The last IP solution (solution D in table 1) minimizes the total distance between selected reserves while restricting the number of selected reserves not to exceed the minimum number of reserves obtained in solution A (i.e. 30). Again this solution differed only slightly from solution A and hence is not considered further.

The unrestricted SCP solution shows a relatively fragmented reserve network (figure 1) compared with the alternative reserve networks obtained in solutions B (figure 2) and C. Solution B (minimizing the sum of distances) differs chiefly in that one site in the upper central section of the region is replaced with two sites in the centre, resulting in a visible increase in site clustering (figures 1 and 2). This results in a significant reduction in the sum of pairwise distances and the maximum intersite distance, despite including an extra site (see table 1). Solution C (minimizing the maximum pairwise distance) differs from B by one site and has the same maximum pairwise distance as solution B, but has a larger sum of pairwise distances. Hence, in this case it is less preferable to solution B. The existence of multiple optima with respect to a given objective function specification (in this case minimization of the maximum pairwise distance) occurs quite frequently when determining optimum reserve site selection using IP. Incorporating multiple objectives in the model constraints, by specifying a goal for each, while optimizing a given objective can determine such alternative optima (if they exist) and provide valuable policy choices when working with the IP approach. This approach can also be used to analyse the trade-off between conflicting objectives, such as the optimum reserve size versus the cost of conservation.

4. DISCUSSION

IP formulations of the reserve selection problem which incorporate spatial objectives along with the more familiar representation constraints result in significant reductions in reserve fragmentation (in terms of the total and maximum pairwise distances) compared with the spatially unrestricted SCP solution. The reduction in fragmentation comes at a cost in terms of efficiency (*sensu* Pressey & Nicholls 1989), although the cost in this case was small (one extra site required). For the pond invertebrate dataset used here, minimizing the sum of the pairwise distances performed better than minimizing the maximum pairwise distance. These two solutions represent alternative ways of minimizing reserve site fragmentation and their relative performance may vary depending on the application. In practice, implementation of one or the other solution would depend on the trade-off between economic and ecological costs and benefits of having more, but spatially closer, reserve sites versus fewer, distant sites in the network. The approaches developed here incorporate spatial considerations as an additional relevant dimension and extend the conventional optimization methods in reserve network selection. Furthermore, as shown in the last column of table 1, the two models presented here have remarkable compu-

Table 1. Performance of the alternative solutions produced by different formulations of the reserve site selection problem. (See § 2 for details of the formulations.)

solution	number of sites selected	sum of pairwise distances between sites (km)	maximum pairwise distance (km)	solutiontime (s)
A	30	11 320	111.1	0.1
B	31	9750	63.1	6.4
C	31	10 344	63.1	0.9
D	30	11 181	111.1	18.6

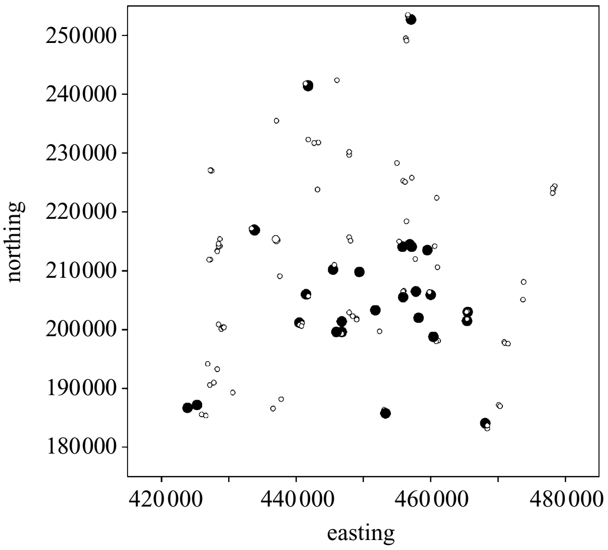

Figure 1. The spatial distribution of the pond sites selected by spatially unrestricted IP SCP formulation of the reserve selection problem. See § 2 for details of the formulation. The black circles indicate the selected sites. The figures on the axes are the British National Grid References (in metres). The sizes of the sites are not to scale and the symbols of the sites that are not selected have been reduced in size to aid clarity.

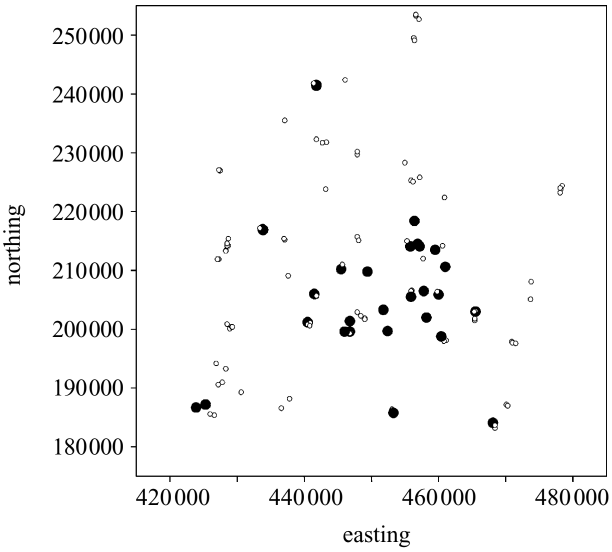

Figure 2. The spatial distribution of the pond sites selected by IP solution B (minimizing the sum of intersite distances). See § 2 for details of the formulation. The black circles indicate the selected sites. The figures on the axes are the British National Grid References (in metres). The sizes of the sites are not to scale and the symbols of the sites that are not selected have been reduced in size to aid clarity.

tational efficiency. The processing times in all four solutions were under 20 s. This indicates that the models can be used conveniently in much larger empirical applications without any serious computational difficulty. In the present application, the sites selected by the alternative solutions differ only marginally from each other. This is due to the particular characteristics of the dataset. Specifically, 23 of those 30 and 31 reserves selected in the optimum solutions given in table 1 contain a species present at only that site (which is the main reason for using the standard SCP formulation, which requires each species to be represented at least once rather than a larger occurrence) and so have to be selected in order to satisfy the species representation constraint. Variation in the degree of reserve fragmentation can occur only through selecting the remaining seven or eight reserve sites differently, significantly restricting the model's flexibility. Whilst the particular results obtained here are due to the characteristics of the dataset, to some extent the problem may generalize to other applications. Most natural ecosystems contain many species which are present at very few sites (Gaston 1994) and hence the extent to which incorporating spatial criteria reduces reserve fragmentation will depend upon the number and distribution of such species.

In conclusion, the spatial distribution of reserve sites is likely to be of considerable importance in maintaining viable populations of the species which it is aimed at protecting. Following from previous extensions of IP models that have incorporated additional criteria in the selection process (Rodrigues *et al.* 2000; Önal 2002*a*; Rodrigues & Gaston 2002*b*), the approach developed here demonstrates that IP formulations have the flexibility to address more complex and realistic reserve selection problems.

REFERENCES

Ando, A., Camm, J. D., Polasky, S. & Solow, A. 1998 Species distribution, land values and efficient conservation. *Science* **279**, 2126–2128.

Briers, R. A. 2002 Incorporating connectivity into reserve selection procedures. *Biol. Conserv.* **103**, 77–83.

Brooke, A., Kendrick, D. & Meeraus, A. 1992 *GAMS, general algebraic modeling system, release 2.25, a user's guide*. Washington, DC: GAMS Development Corporation.

Camm, J. D., Polasky, S., Solow, A. & Csuti, B. 1996 A note on optimal algorithms for reserve site selection. *Biol. Conserv.* **78**, 353–355.

Church, R. L. & ReVelle, C. 1974 The maximum covering location problem. *Pap. Regional Sci. Assoc.* **32**, 101–118.

Church, R. L., Stoms, D. M. & Davis, F. W. 1996 Reserve selection as a maximal covering location problem. *Biol. Conserv.* **76**, 105–112.

Ehrlich, P. R. & Wilson, E. O. 1991 Biodiversity studies: science and policy. *Science* **253**, 758–762.

Gaston, K. J. 1994 *Rarity*. London: Chapman & Hall.

Hanski, I. & Simberloff, D. 1997 The metapopulation approach, its history, conceptual domain and application to conservation. In *Metapopulation biology: ecology, genetics and evolution* (ed. I. Hanski & M. E. Gilpin), pp. 5–26. London: Academic.

Harrison, S. 1994 Metapopulations and conservation. In *Large-scale ecology and conservation biology* (ed. P. J. Edwards, R. M. May & N. R. Webb), pp. 111–128. Oxford: Blackwell Scientific.

Kirkpatrick, J. B. 1983 An iterative method for establishing priorities for the selection of nature reserves—an example from Tasmania. *Biol. Conserv.* **25**, 127–134.

Margules, C. R., Nicholls, A. O. & Pressey, R. L. 1988 Selecting networks of reserves to maximise biological diversity. *Biol. Conserv.* **43**, 63–76.

Nemhauser, G. L. & Wolsey, L. A. 1988 *Integer and combinatorial optimization*. New York: Wiley Interscience.

Nicholls, A. O. & Margules, C. R. 1993 An upgraded reserve selection algorithm. *Biol. Conserv.* **64**, 165–169.

Önal, H. 2002*a* Preservation of species and genetic diversity. *Am. J. Agricultural Econ.* (In the press.)

Önal, H. 2002*b* First best, second best, and heuristic solutions in conservation reserve site selection. *Biol. Conserv.* (In the press.)

Polasky, S., Camm, J. D. & Garber-Yonts, B. 2001 Selecting biological reserves cost-effectively: an application to terrestrial vertebrate conservation in Oregon. *Land Econ.* **77**, 68–78.

Pond Action 1994*a* . *The Oxfordshire pond survey: a report to the World Wide Fund for Nature* (*WWF-UK*), vol. 1. Oxford: Pond Action.

Pond Action 1994*b* . *The Oxfordshire pond survey: a report to the World Wide Fund for Nature* (*WWF-UK*), vol. 2 (appendices). Oxford: Pond Action.

Possingham, H., Ball, I. & Andelman, S. 2000 Mathematical methods for identifying representative reserve networks. In *Quantitative methods for conservation biology* (ed. S. Ferson & M. Burgman), pp. 291–306. New York: Springer.

Pressey, R. L. & Nicholls, A. O. 1989 Efficiency in conservation evaluation: scoring versus iterative approaches. *Biol. Conserv.* **50**, 199–218.

Pressey, R. L., Humphries, C. J., Margules, C. R., Vane-Wright, R. I. & Williams, P. H. 1993 Beyond opportunism: key principles for systematic reserve selection. *Trends Ecol. Evol.* **8**, 124–128.

Pressey, R. L., Possingham, H. P. & Margules, C. R. 1996 Optimality in reserve selection algorithms: when does it matter and how much? *Biol. Conserv.* **76**, 259–267.

Pressey, R. L., Possingham, H. P. & Day, J. R. 1997 Effectiveness of alternative heuristic algorithms for identifying indicative minimum requirements for conservation reserves. *Biol. Conserv.* **80**, 207–219.

Rodrigues, A. S. L. & Gaston, K. J. 2002*a* Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biol. Conserv.* **105**, 103–111.

Rodrigues, A. S. L. & Gaston, K. J. 2002*b* Optimisation in reserve selection procedures—why not? *Biol. Conserv.* **107**, 123–129.

Rodrigues, A. S. L., Cerdeira, J. O. & Gaston, K. J. 2000 Flexibility, efficiency and accountability: adapting reserve selection algorithms to more complex conservation problems. *Ecography* **23**, 565–574.

Romero, C. 1989 *Multiple criteria analysis for agricultural decisions*. Amsterdam: Elsevier.

Sala, O. E. (and 18 others) 2000 Global biodiversity scenarios for the year 2100. *Science* **287**, 1770–1774.

Shafer, C. L. 2001 Inter-reserve distance. *Biol. Conserv.* **100**, 215–227.

Underhill, L. G. 1994 Optimal and suboptimal reserve selection algorithms. *Biol. Conserv.* **70**, 85–87.

Vane-Wright, R. I., Humphries, C. J. & Williams, P. H. 1991 What to protect? Systematics and the agony of choice. *Biol. Conserv.* **55**, 235–254.