THE ROYAL SOCIETY

Received 14 December 2002
Accepted 13 March 2003
Published online 23 May 2003

Selection of a minimum-boundary reserve network using integer programming

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

[1]*Department of Agricultural and 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*

In the conservation literature, heuristic procedures have been employed to incorporate spatial considerations in reserve network selection with the presumption that computationally convenient optimization models would be too difficult or impossible to formulate. This paper extends the standard set-covering formulation to incorporate a particular spatial selection criterion, namely reducing the reserve boundary to the extent possible, when selecting a reserve network that represents a set of target species at least once. Applying the model to a dataset on the occurrence of breeding birds in Berkshire, UK, demonstrated that the technique resulted in significant reductions in reserve boundary length relative to solutions produced by the standard set-covering formulation. Computational results showed that moderately large reserve network selection problems could be solved without issue. Alternative solutions may be produced to explore trade-offs between boundary length, number of sites required or alternative criteria.

Keywords: minimum boundary; reserve network; linear integer programming

1. INTRODUCTION

Selection of efficient conservation reserves has been an important focus of the biological conservation literature in the past decade owing to the increasing need to protect species from anthropogenic habitat loss. The problem has been approached using either heuristic, rule-based approaches to site selection (e.g. Margules *et al.* 1988; Vane-Wright *et al.* 1991; Nicholls & Margules 1993; Pressey *et al.* 1993, 1996, 1997; Csuti *et al.* 1997; Williams 2000; Polasky *et al.* 2001*a*) or formal optimization, specifically linear integer programming (IP) (e.g. Church *et al.* 1996; Ando *et al.* 1998; Polasky *et al.* 2001*b*; Önal & Briers 2002; Rodrigues & Gaston 2002). Heuristic procedures may occasionally yield optimum solutions, but more often they lead to significantly suboptimal outcomes, which may deviate from optimum solutions by as much as 10–15% (Church *et al.* 1996; Pressey *et al.* 1996; Rodrigues & Gaston 2002). Although formal optimization guarantees the most efficient use of conservation resources, relatively few studies have employed this approach. This is motivated primarily by two reasons. First, IP models may not be computationally tractable due to excessive processing time. For instance, Pressey *et al.* (1996) reported computational difficulties (or no solution at all) experienced when working with large-scale reserve selection problems. However, Church *et al.* (1996) and Rodrigues & Gaston (2002) reported that fairly large IP models could be solved easily. A general conclusion cannot be made based on these observations, because computational performance of IP solvers varies depending on the problem structure and data characteristics. However, Önal (2003) showed that near-optimum solutions, which are considerably better than heuristic solutions, could always be obtained from IP solvers within a reasonable computation time. Second, and more importantly, optimization models are presumed to have limited ability to reflect some important and realistic aspects of reserve selection, particularly when spatial criteria, such as compactness of the reserve or clustering of sites, are involved in the selection process (Pressey *et al.* 1996; Possingham *et al.* 2000; McDonnell *et al.* 2002; Nalle *et al.* 2002*a*). Owing to the perceived difficulties of modelling, spatial issues have been addressed in only a few studies using IP formulations (Williams 1998; Williams & ReVelle 1998; Nalle *et al.* 2002*b*). Önal & Briers (2002) demonstrated the use of an IP method based on minimizing intersite distances to incorporate spatial criteria into the selection process. An alternative spatial criterion that has been used in other studies (e.g. Possingham *et al.* 2000; McDonnell *et al.* 2002) is to design a reserve network with the minimum boundary length. Reducing the boundary length of a reserve minimizes edge effects that may influence the persistence of species within the protected area and also reduces economic costs, which are likely to scale more closely with boundary length than area (Possingham *et al.* 2000). Minimizing the boundary will also result in greater clustering of sites, which may be critical to the long-term persistence of species by allowing interpopulation dispersal and colonization of adjacent sites (Önal & Briers 2002). Pressey *et al.* (1996) claimed that this problem cannot be modelled as a linear IP and requires a nonlinear formulation, which would be impossible to solve in practical applications (see also Possingham *et al.* 2000). The purpose of this paper is to show that a computationally convenient linear IP formulation of this problem is possible and demonstrate the utility of this approach in the selection of reserves for conservation.

2. THE MODEL

Suppose a potential conservation reserve area is partitioned into square parcels (see figure 1). Although square

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

DOI 10.1098/rspb.2003.2393

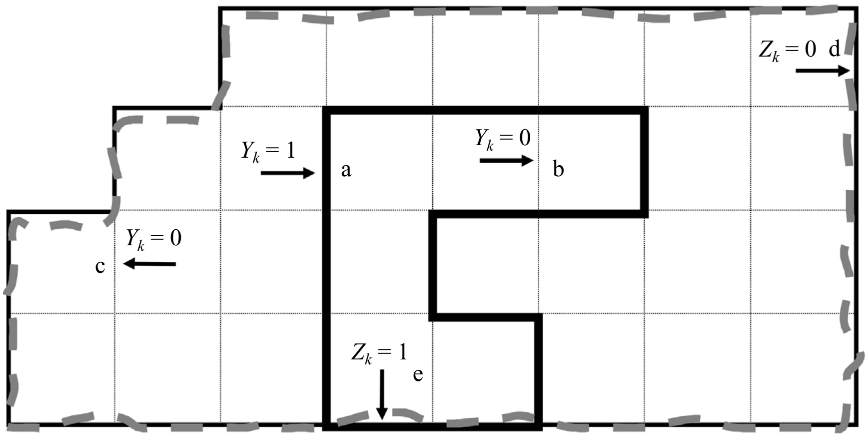

Figure 1. A reserve network with six parcels. The thick grey dashed curve outlines the potential reserve area. The thin dashed grid lines define the edges of individual parcels. The thin solid lines represent the boundary of the partition, while the thicker solid lines represent the boundary edges of the reserve network. An edge in the partition is a boundary edge of the network if $Y_k = 1$ or $Z_k = 1$ (e.g. edges a and e). If $Y_k = 0$ or $Z_k = 0$, then edge k is either not part of the network (e.g. edges c and d), or it is an interior edge of the network (e.g. edge b).

parcels are used in the present application, the approach developed here can be applied without any difficulty to other geometric forms such as rectangles or hexagons. The problem is to select a subset of those parcels in such a way that the resulting network will have the smallest boundary length among all selections which cover a given set of target species at least once (i.e. each species must be present in at least one selected site). This problem will be formulated below as an extension of the 'set covering' problem (SCP) (Church & ReVelle 1974).

Consider the set of all edges of the parcels that make up the potential conservation area, here termed the partition. Each edge belongs to either only one parcel, in which case it is a boundary edge (d and e in figure 1), or two adjacent parcels, in which case it is an interior edge (a, b and c in figure 1). When a subset of the parcels is selected to form a reserve network, we can define the same concepts relative to the network. Explicitly, each edge in the selected reserve network is either an interior edge of the network (if shared by two parcels in the network, e.g. edge b in figure 1) or it is part of the network boundary (if it belongs to only one parcel in the network, e.g. edges a and e in figure 1). Thus, a boundary edge of the network is either a boundary edge of the partition, if the parcel containing that edge is in the network (such as edge e in figure 1), or only one of the two adjacent parcels belongs to the network (such as edge a in figure 1). The main difficulty in modelling the selection problem is to determine whether or not one of these two cases occurs. The model developed below uses binary variables to accomplish this.

The notation used in the model is as follows: I is the set of all parcels in the partition and $i, j \in I$ denote individual parcels; K is the set of all edges and $K' \subset K$ is the set of boundary edges in the partition; S is the set of species considered for protection; δ_{si} is a parameter where $\delta_{si} = 1$ if parcel i contains species s, and $\delta_{si} = 0$ otherwise; λ_{ki} is a parameter, where $\lambda_{ki} = 1$ if edge k belongs to parcel i, and $\lambda_{ki} = 0$ otherwise; X_i is a binary variable, where $X_i = 1$ if parcel i is in the selected network, and $X_i = 0$ otherwise; Y_k is a binary variable, where $Y_k = 1$ if edge k is an interior edge of the partition and a boundary edge of the selected network, and $Y_k = 0$ otherwise; Z_k is a binary variable, where $Z_k = 1$ if edge k is a boundary edge of the partition and included in the selected network, and $Z_k = 0$ otherwise.

The following model minimizes the total length of the reserve network boundary:

$$\text{minimize} \quad \sum_{k \in K} Y_k + \sum_{k \in K'} Z_k,$$

such that:

$$Y_k \geqslant X_i - X_j \quad \text{for all } k \in K, \text{ where } \lambda_{ki} = \lambda_{kj} = 1, \tag{2.1}$$

$$Y_k \geqslant X_j - X_i \quad \text{for all } k \in K, \text{ where } \lambda_{ki} = \lambda_{kj} = 1, \tag{2.2}$$

$$Z_k = \sum_i \lambda_{ki} X_i \quad \text{for all } k \in K', \tag{2.3}$$

$$\sum_i \delta_{si} X_i \geqslant 1 \quad \text{for all } s \in S, \tag{2.4}$$

$$X_i, Y_k, Z_k = 0, 1.$$

The objective function represents the total number of boundary edges in the network, because each Y_k or Z_k equals unity only if edge k is on the network boundary. The model uses the total number of boundary edges in the network as a measure of the boundary length, which implicitly assumes that all edges are of equal length. If individual edges are of different length (which would be the case, for instance, with rectangular parcels), then the Y_k and Z_k variables can be multiplied by the respective edge lengths. Constraint (2.4) is the usual set covering constraint, which ensures that every species s must be present in at least one parcel containing that species. The model could be extended easily to incorporate multiple representation targets for individual species, by specifying the right-hand side of this constraint as a positive integer, instead of unity. The first three constraints form the heart of the model and determine whether an edge is an interior edge or a boundary edge of the network. Constraint (2.3) does this for boundary edges of the partition. For any $k \in K'$, there is only one parcel i for which $\lambda_{ki} = 1$ and for $j \neq i$ we have $\lambda_{kj} = 0$. Thus, equation (2.3) can be simplified as $Z_k = \lambda_{ki} X_i$, which implies that $Z_k = 1$ only if $X_i = 1$, i.e. if parcel i is in the network in which case edge k is on the boundary of the network, otherwise $Z_k = 0$. The first two constraints jointly determine whether or not an interior edge of the partition is on the boundary of the network. To see this, consider the adjacent parcels i and j that share edge k (i.e. $\lambda_{ik} = \lambda_{jk} = 1$). If both parcels are excluded in the network, i.e. $X_i = X_j = 0$, or both parcels are included, i.e. $X_i = X_j = 1$, constraints (2.1) and (2.2) can be reduced to $Y_k \geqslant 0$. However, minimization of $\Sigma_k Y_k$ ensures that $Y_k = 0$, as required. If, however, either parcel i or parcel j is included in the network, but not both, then we have $X_i = 1$ and $X_j = 0$, or $X_j = 1$ and $X_i = 0$. In both cases, constraints (2.1) and (2.2) together imply that $Y_k \geqslant 1$, but as Y_k is a binary variable we have $Y_k = 1$, as required. Figure 1 illustrates alternative situations where each of these cases may occur.

The above model can be simplified by eliminating equation (2.3) and substituting the expression for Z_k into the objective function (which is not done here for clarity of model development). Furthermore, because both Z_k and Y_k are related to the binary X_i variables through binary coefficients, their values will always be binary even if they are defined as continuous non-negative variables. This relaxation leaves the X_i variables as the only binary variables in the model. This offers an important computational advantage when working with medium- or large-scale reserve selection problems, because the computation time needed to solve IP models is sensitive to the number of binary variables.

3. AN EMPIRICAL APPLICATION

To demonstrate the workings of the minimum-boundary model and also investigate its computational efficiency, the model was applied to a dataset for the occurrence of breeding bird species in the county of Berkshire, UK (Standley *et al.* 1996). Between 1987 and 1989, a survey of all 391 tetrads (2 km × 2 km squares) that fall within the administrative boundary of Berkshire was undertaken to record breeding distribution of bird species. A total of 121 species were recorded as breeding within the county; our analyses are based on the distribution of all species except the feral pigeon or rock dove (*Columba livia* Gmelin), which was excluded owing to doubt over the domesticated status of many of the records, the Chukar (*Alectoris chukkar* Gray), an introduced species which does not form self sustaining wild populations, and the stone curlew (*Burhinus oedicnemus* (L.)), a nationally endangered species whose distribution was not mapped in Standley *et al.* (1996). The standard SCP solution to the reserve selection problem was also calculated to provide a comparison with the minimum-boundary model solution.

The SCP solution to the reserve selection problem required nine cells to represent all 118 species, with a total boundary length of 36 units (one unit equals the edge length of the square parcels) (figure 2*a*). Although this approach gives the smallest (and hence least expensive) reserve, it resulted in a highly fragmented reserve structure, where every parcel was isolated from the other selected parcels. The minimum-boundary solution however, required 17 cells with a total boundary length of 30 units to cover all species (figure 2*b*). The solution was not

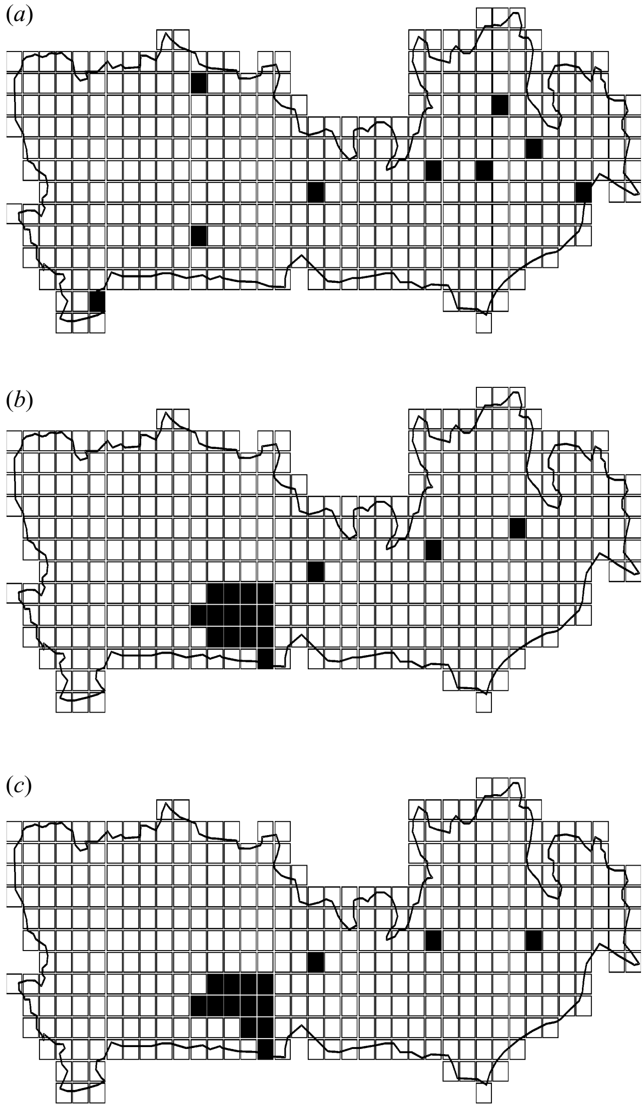

Figure 2. Map of tetrads (2 km × 2 km squares) within the administrative boundary of Berkshire, UK (bold irregular line) required to represent all species of breeding birds at least once within selected sites. Shaded squares indicate selected sites. (*a*) Sites required by the set covering formulation with no spatial criteria; (*b*) sites required by the minimum-boundary solution; and (*c*) an alternative minimum-boundary solution with fewer sites. See text for details of models.

unique, however. Multiple optimum solutions with the same boundary length could be found by directing the model. This was accomplished by excluding the selected parcels one at a time (this can be done by imposing $X_i = 0$ if parcel i was in the base solution) or by reducing the network size (through an added constraint $\Sigma_i X_i \leqslant n$, where n is the maximum number of parcels that can be included in the network). Both approaches resulted in several optimum solutions, containing a different number of parcels in the network or including the same number of parcels, but placed in a different configuration. One of these solutions is shown in figure 2*c*. This selection included 15 parcels, again with a boundary length of 30 units, and four separated reserves. Thus, it is preferable to the first minimum-boundary solution if the number of parcels is to be considered as an additional selection criterion. Using the techniques detailed in § 2, further reductions in the number of parcels were possible, while maintaining the boundary length, at the expense of greater reserve fragmentation, i.e. more disconnected parcels. One of those alternative optimum solutions contained only 12 parcels, for instance, but it consisted of five separate conservation areas instead of four.

As has been previously observed (Possingham *et al.* 2000), reductions in boundary length come at a cost in terms of efficiency (i.e. the number of sites required to achieve a given objective). The minimum-boundary solution did not produce a completely connected network of parcels. The degree of connectedness that could be achieved is highly dependent on the data characteristics. A large number of endemic or rare species limits the model's ability to produce connected reserve networks and is a problem inherent to all spatial criteria used in reserve selection (Önal & Briers 2002).

Computational efficiency of IP formulations has always been an important concern in conservation reserve selection, as mentioned at the outset. As elaborated by Camm *et al.* (1996), being able to model a reserve selection problem using IP may not necessarily mean that the problem can be solved. With the dataset used in this particular application, the minimum-boundary reserve selection model included 1684 equations and 1320 variables, 391 of which were binary. In all cases, the solutions were obtained in less than 2 min of computation time using GAMS (Brooke *et al.* 1992) interfaced with OSL (a linear IP solver developed by IBM), which suggests that larger models can be solved conveniently. The number of reserves sites, rather than species, is the critical factor, as it determines the number of binary variables in the model. Computational complexity is also closely related to the data structure, particularly the distribution of species and the number and location of sites that contain endemic species (in which case the site will be irreplaceable) or rare species.

4. CONCLUSIONS

Conservation reserve selection is a typical example of multi-objective decision making. Typically, multiple criteria simultaneously govern the reserve selection process, such as the area, boundary length, connectivity, and even the reserve configuration (Possingham *et al.* 2000; Siitonen *et al.* 2002). Reducing the boundary length of a reserve network relative to its area is an important concern in conservation planning owing to its influence on the economic costs of establishing and maintaining the reserve, and on the likely persistence of species within the selected sites (Possingham *et al.* 2000; Briers 2002; Önal & Briers 2002; Siitonen *et al.* 2002).

While the model presented here minimizes the boundary length of the reserve, it does not guarantee a fully connected reserve network. When applied to a real world dataset on the distribution of breeding birds, the minimum-boundary solution was not fully connected as a result of the particular distribution of species in the dataset used. Two sites were irreplaceable as they contained species not found elsewhere and those sites were far from the cluster of sites which made up the remainder of the reserve. Although having disconnected parcels usually implies an increased number of boundary edges, this may still be preferable because producing a fully connected network may require a large number of sites, and hence may not be economically viable. The trade-off between boundary length and number and configuration of sites included in the network can be explored by searching alternative optimum solutions or incorporating additional selection constraints, such as limiting the reserve size (number of parcels included) as detailed above. However, imposing too tight an upper bound for the number of parcels that can be included in the network is likely to result in increasingly fragmented reserve networks.

In conclusion, the present paper demonstrates that it is possible to develop computationally convenient linear IP reserve selection models that incorporate spatial criteria, in addition to considering species' complementarity, when designing efficient reserve networks. While a particular spatial criterion, namely minimization of the reserve boundary, is considered here, alternative formulations may also be possible for some other widely used spatial criteria (e.g. Önal & Briers 2002). In general, incorporating spatial considerations in an optimization framework is more complicated than the standard set covering formulation because of modelling difficulties and computational complexity. The latter can be restrictive, especially when dealing with a large number of parcels, which would require numerous additional constraints and, more importantly, additional binary variables in the model. However, such models more closely represent real world conservation planning problems and offer a significant level of flexibility which can be used to explore alternative reserve configurations, while retaining the ability to produce optimal solutions. Several optimum solutions which have identical boundary length, but differ in reserve size and location, could be generated for the particular problem studied here (two of which are presented in figure 2). A practical issue which needs to be explored further is how to generate all such solutions in a systematic way so that the planning decisions can be based on a full range of optimum reserve configurations.

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 2.50, a users 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.

Csuti, B. (and 10 others) 1997 A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon. *Biol. Conserv.* **80**, 83–97.

McDonnell, M. D., Possingham, H. P., Ball, I. R. & Cousins, E. A. 2002 Mathematical methods for spatially cohesive reserve design. *Environ. Model. Assess.* **7**, 107–114.

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

Nalle, D. J., Arthur, J. L. & Sessions, J. 2002*a* Designing compact and contiguous reserve networks with a hybrid heuristic algorithm. *Forest Sci.* **48**, 59–68.

Nalle, D. J., Arthur, J. L., Montgomery, C. A. & Sessions, J. 2002*b* Economic and spatial impacts of and existing reserve network on future augmentation. *Environ. Model. Assess.* **7**, 99–105.

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

Önal, H. 2003 First best, second best, and heuristic solutions in conservation reserve selection. *Biol. Conserv.* (In the press.)

Önal, H. & Briers, R. A. 2002 Incorporating spatial criteria in optimum reserve selection. *Proc. R. Soc. Lond.* B **269**, 2437–2441. (DOI 10.1098/rspb.2002.2183.)

Polasky, S., Csuti, B., Vossler, C. A. & Meyers, S. M. 2001*a* A comparison of taxonomic distinctness versus richness as criteria for setting conservation priorities for North American Birds. *Biol. Conserv.* **97**, 99–105.

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

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., 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 Optimisation in reserve selection procedures: why not? *Biol. Conserv.* **107**, 123–129.

Siitonen, P., Tanskanen, A. & Lehtinen, A. 2002 Method for selection of old-forest reserves. *Conserv. Biol.* **16**, 1398–1408.

Standley, P., Bucknell, N. J., Swash, A. & Collins, I. D. 1996 *The birds of Berkshire*. Reading, UK: Berkshire Atlas Group.

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.

Williams, J. C. 1998 Delineating protected wildlife corridors with multiple-objective programming. *Environ. Model. Assess.* **3**, 77–86.

Williams, J. C. & ReVelle, C. S. 1998 Reserve assemblage of critical areas: a zero-one programming approach. *Eur. J. Oper. Res.* **104**, 497–509.

Williams, P. H. 2000 Complementarity. In *Encyclopedia of biodiversity* (ed. S. A. Levin), pp. 813–829. San Diego, CA: Academic.