Biological Conservation 70 (1994) 85–87

0006-3207/94/$07.00

ELSEVIER

SHORT NOTE

OPTIMAL AND SUBOPTIMAL RESERVE SELECTION ALGORITHMS

L. G. Underhill

Avian Demography Unit, Department of Statistical Sciences, University of Cape Town, Rondebosch 7700, Republic of South Africa

(Received 29 May 1993; revised version received 28 October 1993; accepted 4 November 1993)

Abstract

This paper criticises some reserve selection algorithms that have recently been published in Biological Conservation *and have rapidly become enshrined in the* principle of complementarity. *These algorithms are shown, by means of a counter-example, to be suboptimal. Integer programming techniques, available for 30 years, provide optimal solutions to the reserve selection problem. The paper appeals for closer co-operation between biologists and mathematicians in the development of algorithms.*

Keywords: nature reserve selection, integer programming methods, greedy algorithms, optimality.

A recent series of papers purports to present algorithms that minimize the number of reserves required to conserve every species in an area (Margules *et al.*, 1988; Rebelo & Siegfried, 1990; Vane-Wright *et al.*, 1991; Nicholls & Margules, 1993; Pressey *et al.*, 1993). Although the details differ, the algorithms are essentially similar, and are manifestations of heuristic procedures known to mathematicians as 'greedy algorithms' (e.g. Nemhauser & Wolsey, 1989). Although greedy algorithms may occasionally deliver the correct solution, there is no certainty that they will, and they may also produce grossly suboptimal results. The algorithms proposed in the above papers have become popular, to the extent that they appear to have been elevated to the status of a principle, *the principle of complementarity*, a term coined by Vane-Wright *et al.* (1991). In this note, I present a simple counter-example that demonstrates that the greedy algorithm is not optimal, and show that the problem of minimizing the number of reserves to conserve every species is a straightforward application of a standard technique in operational research.

A feature of the above series of papers (and others in the area) is the poor use of mathematical terms such as 'optimal' 'minimum' and 'maximum'. Especially disconcerting is the appearance of the solecism 'more optimal' (Vane-Wright *et al.*, 1991, p. 245), on a par with 'more unique'. It needs to be borne in mind that if an algorithm is claimed to be optimal, or that it produces a maximum or a minimum, then this needs to be proved, and that a single counter-example, however simple, represents disproof (Solow, 1982).

The counter-example is provided by considering the hypothetical sites by species matrix in Table 1. No species is endemic to a single site, so no sites are automatically included in the reserve system. The algorithms of all the above authors would agree that the first site to be chosen is the one with the highest species richness, namely site 1. Eliminating the species conserved at site 1 from further consideration, site 2 is then the best of the rest, followed by site 3 (or 5). Thus the 'greedy algorithm' selects three sites for reserves. However, it is obvious by inspection that only two sites, sites 2 and 3, are needed to conserve all the species. The greedy algorithm is suboptimal because it does not allow a site to be dropped from the priority set once it has been selected at an earlier iteration.

The primary purpose of this paper is not to denigrate the principle of complementarity. Rather, I point out that it is wrong to equate the principle with a suboptimal algorithm, and point to a better solution. The following formulation of the problem leads to the optimal solution of the choice of a set of priority sites. Let $x_{ij} = 1$ if site i ($i = 1,...,I$) conserves species j ($j = 1,...,J$), and $x_{ij} = 0$ otherwise, as in Table 1. Then the rows of the $I \times J$ matrix $X = (x_{ij})$ refer to sites and the columns to species, so that the ith row of X shows the species that would be conserved if site i were in the reserve network, and column j shows the sites at which species j occurs in sufficient abundance that the creation of a re-

Table 1. Sites × species data matrix producing a counter-example to the greedy algorithm. The algorithm chooses sites 1, 2 and 3, in this sequence, as priority sites, while it is obvious that only sites 2 and 3 are required to preserve all species

	Species							
	1	2	3	4	5	6	7	8
Site 1	0	0	1	1	1	1	1	0
Site 2	1	1	1	1	0	0	0	0
Site 3	0	0	0	0	1	1	1	1
Site 4	1	1	1	0	0	0	0	0
Site 5	0	0	0	0	0	1	1	1

serve at one of these sites would conserve the species.

Let Y_i be a zero-one variable that takes on the value 1 if site i is included in the reserve network, and the value 0 otherwise. In order to have species j in at least one reserve, the I variables Y_i must satisfy the constraint

$$\sum_{i=1}^{I} Y_i x_{ij} \geq 1$$

This constraint must be satisfied for each species, yielding the J constraints

$$\sum_{i=1}^{I} Y_i x_{ij} \geq 1 \quad j = 1, \ldots, J \tag{1}$$

We wish to minimize the number of reserves, i.e. the objective is to minimize

$$Z = \sum_{i=1}^{I} Y_i,$$

subject to the constraints in eqn (1). This is a standard integer programming problem, to which solutions have been available for over three decades, and can be solved by software such as LINDO (Schrage, 1989) which uses the branch-and-bound search process (Land & Doig, 1960; see also Zionts, 1974; Garfinkel & Nemhauser, 1972; Nemhauser & Wolsey, 1989). Examples of the use of integer programming in reserve selection algorithms are Cocks and Baird (1989) and Sætersdal *et al.* (1993). From a mathematical perspective, these papers represent the strand of progress in reserve selection algorithms that ought to be pursued.

Vane-Wright *et al.* (1991) considered that the order in which the greedy algorithm selects reserves to be important, in that it gives a priority ranking for each site, rather than a set (= an unordered group) of priority sites. However, because the set of reserves recommended by the greedy algorithm is itself likely to be suboptimal, the concept of 'priority sequence' falls away. Arguments for the ranking of priorities within optimal sets will need to be based on a theoretically sound technique, and not on the selection order produced by a suboptimal algorithm.

It is not clear how the greedy algorithms perform when recommendations are required for nature reserves such that each species is conserved in more than one reserve. If each species is required to be conserved in p reserves, then, with the integer programming approach, only the constraints in eqn (1) need to be changed, to become

$$\sum_{i=1}^{I} Y_i x_{ij} \geq p \quad j = 1, \ldots, J \tag{2}$$

However, this presumes that the occurrence of each species is such that it may be conserved at p or more sites. If some of the column sums are less than p, no feasible solution exists to the problem. Thus let $s_j = \sum_{i=1}^{I} x_{ij}$, the number of sites at which species j may be conserved. Then to ensure that a feasible solution to the integer programming problem exists, the constraints need to be modified to

$$\sum_{i=1}^{I} Y_i x_{ij} \geq \min \{p, s_j\} \quad j = 1, \ldots, J \tag{3}$$

If some sites are already reserves, then the Y_i associated with these sites can be fixed to be equal to one, thus allowing decisions to be taken on supplementing a current reserve system in an optimal way. Other criteria and limitations that the analyst wishes to place on the reserve selection can be expressed as additional constraints to the standard integer programming software. Examination of the iteration-by-iteration output from the branch-and-bound search process gives insight into flexibility and irreplaceability, the two other 'guiding principles' of nature reserve selection (Pressey *et al.*, 1993), and provides guidance on which sites should enjoy top priority.

A more realistic problem than selecting the minimum number of reserves to conserve all species is to maximize the number of species that can be conserved within a fixed financial budget or a given total area. In this problem, the species may be divided into groups that are not necessarily mutually exclusive, such as genera, waterbirds, economically important species, insectivores, pollinators, etc., and the constraints could be requirements to conserve at least p_g species from the gth group. This, and many other approaches, can be formulated as integer programming problems for which computer software is readily available. The advantage of using standard software is that it allows the biologist flexibility in concentrating on biological issues, while retaining the mathematical security of the optimal properties of a well-understood algorithm. Research is currently in progress on the application of integer programming to the reserve selection problem, using the database of the Southern African Bird Atlas Project (Harrison, 1992).

Besides integer programming, another technique conservation biologists could borrow from operational research is multiple-criteria decision making (Keeney & Raiffa, 1976; Haimes & Chankong, 1984; Stewart, 1992). Margules (1986) considered that six criteria needed to be considered to judge the conservation value of a site, and presented an *ad hoc* approach to a problem for which an extensive tool kit already exists.

Butterworth (1989) made the suggestion that recommendations about the management of biological resources would be handled better by mathematicians and engineers than by biologists. This may be an overstatement, but mathematicians can certainly play a key role in the development of methods for the efficient choice of systems of nature reserves, and other problems in conservation biology that require an algorithmic solution.

ACKNOWLEDGEMENTS

Support from the Foundation for Research Development and the University of Cape Town Research

Committee is acknowledged. T. J. Stewart and J. H. Webb assisted with references. J. A. Harrison and T. J. Stewart commented on an earlier draft.

REFERENCES

Butterworth, D. S. (1989). The Benguela Ecology Programme: successful and appropriate? *S. Afr. J. Sci.*, **85**, 633–43.

Cocks, K. D. & Baird, I. A. (1989). Using mathematical programming to address the multiple reserve selection problem: an example from the Eyre Peninsula, South Australia. *Biol. Conserv.*, **49**, 113–30.

Garfinkel, R. S. & Nemhauser, G. L. (1972). *Integer programming*. Wiley, New York.

Haimes, Y. Y. & Chankong, V. (1984). *Decision making with multiple objectives*. Springer-Verlag, Berlin.

Harrison, J. A. (1992). The Southern African Bird Atlas Project databank: five years of growth. *S. Afr. J. Sci.*, **88**, 410–13.

Keeney, R. L. & Raiffa, H. (1976). *Decisions with multiple objectives: preferences and value tradeoffs*. Wiley, New York.

Land, A. H. & Doig, A. G. (1960). An automatic method of solving discrete programming problems. *Econometrica*, **28**, 497–520.

Margules, C. R. (1986). Conservation evaluation in practice. In *Wildlife conservation evaluation*, ed. M. B. Usher. Chapman and Hall, London, pp. 297–314.

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. (1989). Integer programming. In *Handbooks in operations research and management science, I. Optimization*, ed. G. L Nemhauser, A. H. G. Rinnooy Kan & M. J. Todd. North Holland, Amsterdam, pp. 447–527.

Nicholls, A. O. & Margules, C. R. (1993). An upgraded reserve selection algorithm. *Biol. Conserv.*, **64**, 165–69.

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. Evolut.*, **8**, 124–8.

Rebelo, A. G. & Siegfried, W. R. (1990). Protection of fynbos vegetation: ideal and real world options. *Biol. Conserv.*, **54**, 15–31.

Sætersdal, M., Line, J. M. & Birks, H. J. B. (1993). How to maximize biological diversity in nature reserve selection: vascular plants and breeding birds in deciduous woodlands, western Norway. *Biol. Conserv.*, **66**, 131–8.

Schrage, L. (1989). *User's manual for linear, integer, and quadratic programming with LINDO*. The Scientific Press, San Francisco, California.

Solow, D. (1982). *How to read and do proofs: an introduction to mathematical thought process*. Wiley, New York.

Stewart, T. J. (1992). A critical survey on the status of multiple-criteria decision making theory and methods. *OMEGA Int. J. Mgmt Sci.*, **20**, 569–86.

Vane-Wright, R. I., Humphries, C. J. & Williams, P. H. (1991). What to protect? Systematics and the agony of choice. *Biol. Conserv.*, **55**, 235–54.

Zionts, S. (1974). *Linear and integer programming*. Prentice-Hall, Englewood Cliffs, New Jersey.