vol. 165, no. 6 the american naturalist june 2005

Ⓔ

Reserve Selection Using Nonlinear Species Distribution Models

Atte Moilanen[*]

Metapopulation Research Group, Department of Biological and Environmental Sciences, P.O. Box 65, FI-00014 University of Helsinki, Finland

Submitted June 30, 2004; Accepted February 15, 2005; Electronically published May 10, 2005

Online enhancements: appendixes.

abstract: Reserve design is concerned with optimal selection of sites for new conservation areas. Spatial reserve design explicitly considers the spatial pattern of the proposed reserve network and the effects of that pattern on reserve cost and/or ability to maintain species there. The vast majority of reserve selection formulations have assumed a linear problem structure, which effectively means that the biological value of a potential reserve site does not depend on the pattern of selected cells. However, spatial population dynamics and autocorrelation cause the biological values of neighboring sites to be interdependent. Habitat degradation may have indirect negative effects on biodiversity in areas neighboring the degraded site as a result of, for example, negative edge effects or lower permeability for animal movement. In this study, I present a formulation and a spatial optimization algorithm for nonlinear reserve selection problems in grid-based landscapes that accounts for interdependent site values. The method is demonstrated using habitat maps and nonlinear habitat models for threatened birds in the Netherlands, and it is shown that near-optimal solutions are found for regions consisting of up to hundreds of thousands grid cells, a landscape size much larger than those commonly attempted even with linear reserve selection formulations.

Keywords: spatial reserve design, reserve selection, site selection algorithm, habitat model, stochastic optimization, genetic algorithm.

Reserve selection is concerned with efficient use of conservation resources (see Pressey 1999; Margules and Pressey 2000; Noss 2003; Cabeza et al. 2004*b*; Williams et al. 2004 for recent reviews). Biological reserves can also hold economic or recreational values, but from the perspective of conservation biology, the question is essentially about the allocation of resources so that long-term biodiversity persistence is ensured to a satisfactory degree (Margules and Pressey 2000; Cabeza and Moilanen 2001). One aspect of this problem is a decision about which particular parcels of land (e.g., areas, sites, patches, cells, selection units) one should purchase to achieve a given conservation goal. This study concentrates on a single critical assumption commonly made by reserve selection algorithms, that selection units are independent of each other; that is, the biological value of a selection unit does not depend on the structure of the rest of the landscape.

The assumption of independence of selection units is false when selection units are of the size commonly used in landscape planning (ha, km^2) because at these scales, spatial population dynamics influence species distribution patterns strongly. The idea that landscape structure influences species distribution patterns is the fundamental tenet of metapopulation biology, landscape ecology, and spatial ecology in general (e.g., MacArthur and Wilson 1967; Levin 1974; Hanski 1998; King and With 2002). The habitat pattern of the selected reserve network and the habitat around it could influence the biological value of the network. Nevertheless, this fact is in general ignored by reserve selection formulations (Cabeza and Moilanen 2003) possibly because of computational difficulties implied by the interdependence of sites. In this study, I emphasize that if the biological value contained by a reserve network is calculated in a realistic manner, one ends up with a nonlinear formulation for the reserve selection problem not readily solvable by standard reserve selection approaches. I show how such a problem can be solved, even for large landscapes, using a combination of a genetic algorithm variant tailored for spatial optimization and a local improvement method acting on detailed reserve structure.

A reserve selection problem can be divided into two parts: first, definition of the conservation goal (target, objective, objective function) and constraints (available resources, money, etc.) and, second, an analytical or numerical solution. Typical goals include the following: find the cheapest solution that gives at least one population per species with a given probability (Haight et al. 2000; Polasky et al. 2000; Williams and Araújo 2000, 2002; Ar-

---

[*] E-mail: atte.moilanen@helsinki.fi.

Am. Nat. 2005. Vol. 165, pp. 695–706.

thur et al. 2002), find the cheapest solution that includes a given proportion of the distribution of the species (ReVelle et al. 2002; Cabeza 2003; Cabeza et al. 2004*a*, 2004*b*; Sarkar et al. 2004), find the cheapest solution that includes a given proportion of each land cover type in the region (Pressey and Tully 1994; Pressey et al. 1997; McDonnell et al. 2002; Leslie et al. 2003), and maximize the number of species covered adequately by the available resource (the maximum coverage problem; Csuti et al. 1997; Pressey et al. 1997; Snyder et al. 1999; Camm et al. 2002).

There are three general optimization frameworks that have been used for solving reserve selection problems. The oldest of these is a complementarity-based deterministic greedy stepwise heuristic (e.g., Kirkpatrick 1983; Margules et al. 1988; Csuti et al. 1997; Pressey et al. 1997), which adds to the solution, one site at a time, so that at each step one achieves maximal increase in the value (number of species represented) of the solution. Stepwise heuristics are not guaranteed to find the global optimum of a reserve selection problem (Cocks and Baird 1989; Underhill 1994; Camm et al. 1996; Önal 2004). Linear and integer programming are so-called exact (or complete or exhaustive) optimization techniques, which are guaranteed to find the global optimum of an optimization problem given that both the objective and the constraints are linear functions (Underhill 1994; see Williams et al. 2004 for review). Certain objective functions, such as the "probability of having at least one population," may also be modeled linearly via linearization techniques (Haight et al. 2000; Camm et al. 2002). Although these studies have a nonlinear objective function, the model is linear in the sense that the probabilities of occurrence do not depend on the pattern of the selected habitat. A third and less precisely defined class of optimization methods used for reserve selection is stochastic global optimization, including simulated annealing (Possingham et al. 2000; McDonnel et al. 2002; Westpahl and Possingham 2003) and genetic algorithms (Moilanen and Cabeza 2002). These techniques use intelligent randomization to guide the search process to near-optimal solutions. Stochastic global optimization methods can, in principle, be applied to any problem structure. However, the global optimum is not guaranteed, and the performance of the optimization in terms of convergence speed will be highly dependent on the details and implementation of the optimization technique.

All of these objectives and solution methods depend on a particular way of calculating the value of a site. In particular, direct application of linear or integer programming methods (see Williams et al. 2004) requires that the biological value of a site does not depend on the spatial pattern of the selected habitat. Even taking into account that integer programming can capture many nonlinear relationships via linearization techniques or piecewise approximation (Hof and Bevers 2002), more complex (and possibly more realistic) relationships can be captured if the optimization method allows direct use of nonlinear formulations. The structure of the landscape does affect the distributions of species there. If the landscape changes, one should expect the distribution of species within it to change. For example, if a suburb is built next to a piece of old-growth forest, the biological value of the forest edge can be expected to decline because of negative edge effects: increased disturbance, invasive species, and changes in abiotic conditions at the edge (see, e.g., Debinski and Holt 2000; Gaston et al. 2002).

If the probability of a species j being present at any site i, p_{ij}, depends on the distribution of the species elsewhere in the landscape, one may write

$$p_{ij} = f(\mathbf{h}_{ij}, \mathbf{p}_j(\mathbf{x})), \tag{1}$$

in which $\mathbf{h}_{ij}$ is a vector of habitat quality variables for species j at site i, $\mathbf{p}_j(\mathbf{x}) = \{p_{ij}(\mathbf{x})\}$ is the vector of probabilities of occupancy for the species j given a particular landscape structure, and $\mathbf{x} = \{x_i\}$ is a selection vector with element $x_i = 1$ if site i is selected and 0 if not. Equation (1) is an implicit nonlinear equation solvable via iteration. It states that the probability of the species being present at a site depends on the quality of the habitat at the location and the probability of presence of the species at other locations in the landscape. Most spatial population models would acknowledge such a structure because in such models, the probability of a species being present at a location would depend on (among other things) immigration into the site, which would be affected by the distribution of the species elsewhere in the landscape (e.g., MacArthur and Wilson 1967; Hanski 1998; Moilanen and Nieminen 2002). Another reason why p_{ij} might depend on $\mathbf{x}$ is anthropogenic threat. Habitat near human activity may be in danger of being degraded, which could be accounted for in reserve planning via nonlinear reserve selection formulations.

Nonlinear models have rarely been used in reserve selection, probably because of the computational difficulty involved, and when they have been considered, the landscapes have been small (see Moilanen and Cabeza 2002; Cabeza 2003; Westpahl et al. 2003; van Teeffelen et al. 2005). Because equation (1) is implicit, it needs to be iterated until convergence for each species separately for each habitat configuration $\mathbf{x}$ evaluated. Such iteration might be computationally very expensive for even moderately large landscapes, which are difficult to handle even with linear problem formulations. Williams et al. (2004) state that integer programming reserve selection formulations tend to reach their limit at 10,000 landscape elements because of combinatorial explosion of the search space. Westpahl and Possingham (2003) use simulated annealing on a large landscape, "select 3,000 out of 80,000 sites," in the context of habitat restoration for birds in South Australia. Their model is nonlinear in the sense that $p_{ij} = f(\mathbf{x})$ and that probabilities depend on a logistic regression that accounted for measures of habitat structure in the neighborhood of the focal cell. This work differs from that of Westpahl and Possingham (2003) in that the implementation of equation (1) here assumes an autologistic structure: p_{ij} depends on $f(\mathbf{x}, \mathbf{p}_j(\mathbf{x}))$. The use of an autologistic model structure comes from population dynamical reasoning; landscape structure affects occupancy at a location both directly (via local habitat quality and neighborhood habitat structure) and indirectly (via occupancy of the rest of the landscape; migrants come only from occupied sites).

Note that any reserve selection formulation might be nonlinear even if the species distribution model is linear. This happens, quite realistically, if a nonlinear utility function for the value of the representation of a species is used (Arponen et al. 2005). Typically, and in this study as well, a species is counted as represented if the species representation level exceeds a given target, such as a number of occurrences or a proportion of the distribution. Thus, the value of a species representation is essentially a step function of representation with the step located at the species-specific target. However, for reserve selection purposes, it might be better to be able to differentiate between how much over or under the target the representation of the species is. For example, 100 populations would be better than 50, but using a step function with a target of 20 would not differentiate between these two cases. Equally, something just barely under the target is in reality much better than zero representation. Valuing of over- and underrepresentation can be achieved via the use of a nonlinear monotonically increasing value function for species representation.

In this study, I present an optimization framework for nonlinear reserve selection problems on large grids. The technique is based on a genetic algorithm (GA) variant specially tailored for spatial problems. Nonlinearity in the proposed framework arises from population-dynamical connectivity calculated using a species-specific dispersal kernel (Hanski 1994; Moilanen and Nieminen 2002), which allows for long-distance interactions. A critical component of the proposed framework is calculation of the connectivity via the use of fast Fourier transforms (FFT; Brewster and Allen 1997), which enable such calculations for landscapes of the order of hundreds of thousands of grid cells. To improve search efficiency, the GA is combined with a local search method, which optimizes the fine details of the solutions. The proposed method is introduced in the context of nonlinear habitat models, but the algorithm can also be used for optimizing spatial pattern using linear species distribution models or simple presence-absence data.

Methods

Overview

The objective function of the nonlinear reserve selection problem studied here is explained first below. The optimization solution for this problem has two components: a stochastic global search method (a spatial GA with self-adaptation of control parameters) and a local improvement method. The GA searches for large-scale reserve configurations that give good performance, and its most relevant feature is recombination, the way two different parent solutions (reserve structures) are combined to produce a new, hopefully improving, solution. The local search manipulates the details of the reserves to converge to locally optimal solutions. Most successful GA applications have often used local search (Blum and Roli 2003). The solution method used for solving the nonlinear reserve selection problem is described in detail in appendix A in the online edition of the *American Naturalist*. Appendix B in the online edition of the *American Naturalist* describes a method for computing a lower bound for the optimal solution size of the nonlinear optimization problem. Having a lower bound estimate changes the present algorithm from a heuristic to an approximation in the sense that the maximum level of suboptimality of the proposed solution is known. Table 1 gives a summary of the symbols used in the study.

Objective Function of Optimization

The optimization problem used here is a proportional coverage variant of the minimum set covering problem, where the objective is to cover a given proportion of the distribution of the species with minimal cost:

$$\min F(\mathbf{x}) = \sum_{i=1}^{N} x_i c_i, \tag{2}$$

given that $R_j(\mathbf{x}) \geq T_j$ for all j.

In equation (2), c_i is cost of site i; further, T_j is the representation target level for species j, and $R_j(\mathbf{x})$ is the representation of species j in solution $\mathbf{x} = \{x_i\}$. The targets T_j were calculated as a proportion of the number of populations predicted by the habitat model to be present in the full original landscape.

In itself, there is not necessarily anything nonlinear

Table 1: Explanations for symbols used in problem formulation, the objective function, and the genetic algorithm

Symbol	Definition
i	Index for site i
(k, m) or (x, y)	Indexes for site i on a grid; (x, y) is for focal cell, and (k, m) is for another cell
N	Number of sites in landscape
x_i	Equals 1 if site i is included in solution $\mathbf{x}$, 0 if not
$\mathbf{x}$	$\{x_i\}$, solution vector
c_i	Cost of site i
j	Index for species
p_{ij}	Probability of presence of species j in site i
T_j	Target for species j
$R_j(\mathbf{x})$	Representation level of species j in solution $\mathbf{x}$
$\mathbf{h}_{ij}$	Vector of habitat variables relevant for species i in site j
α	Parameter of negative exponential dispersal kernel; small α indicates that the species has good dispersal ability
B_{ij}	Buffer connectivity of habitat suitable for species j in the immediate neighborhood of site i
I_{ij}	Incidence function model connectivity for species j in site i (eq. [4])
b	Penalty for the boundary length of solution
$B_1(T)$, $B_2(T)$	Two different lower bounds for smallest possible solution size for target T (defined in app. B in the online edition of the *American Naturalist*)
$\mathrm{GA_M}$	Population size
$\mathrm{GA_G}$	Number of generations
$\mathrm{GA_{XN}}$	Maximum number of crossover blocks in recombination
$\mathrm{GA_{XS}}$	Maximum size of a crossover block as a fraction of edge length of study area
$\mathrm{GA_{MN}}$, $\mathrm{GA_{MS}}$	As $\mathrm{GA_{XN}}$ and $\mathrm{GA_{XS}}$ but for the mutation operator
$\mathrm{LS_{rnds}}$	Number of local search rounds used by the algorithm

Note: See appendix A in the online edition of the *American Naturalist* for GA details.

about equation (2). Rather, the nonlinearity arises from the way $R_j(\mathbf{x})$ is calculated:

$$R_j(\mathbf{x}) = \sum_{i=1}^{N} x_i p_{ij}(\mathbf{x}),$$

$$p_{ij}(\mathbf{x}) = f(\mathbf{h}_{ij}, \mathbf{n}_{ij}, \mathbf{p}_j(\mathbf{x})). \quad (3)$$

Equation (3) states that the probability of species j being present at site i, p_{ij}, depends on the set of selected sites, that is, the spatial structure of the reserve network (Moilanen and Cabeza 2002; Cabeza 2003; van Teeffelen et al. 2005). Here it is assumed that nonselected sites will eventually be lost, and thus they will not contribute to any connectivity measures calculated for the species (Cabeza and Moilanen 2003), but it is possible to include a statistical model for the changes in the quality of nonselected sites if such a model exists. The probability p_{ij} would also typically depend on local species-specific habitat quality variables ($\mathbf{h}_{ij}$) and the habitat quality of the immediate neighborhood of the site ($\mathbf{n}_{ij}$). The actual species distribution models used here are logistic regression habitat models with connectivity-like autocovariates (Augustin et al. 1996; Ferrier et al. 2002; Cabeza 2003; van Teeffelen et al. 2005):

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = a + \boldsymbol{\beta}_j \mathrm{h}_{ij} + \gamma_j B_{ij}(\mathbf{x}) + \varphi_j I_{ij}(\mathbf{x}) + \text{interaction terms}, \quad (4)$$

in which $\boldsymbol{\beta}_j$ is a vector of coefficients for local habitat variables and φ_j and γ_j are coefficients for two different autocovariates; $B_{ij}(\mathbf{x})$ is a measure over the (eight) cells neighboring i, which means that it essentially measures neighborhood habitat quality and structural habitat connectivity; and $I_{ij}(\mathbf{x})$ is an incidence function model-type connectivity measure (IFM; Hanski 1994; see Moilanen and Nieminen 2002) calculated for the species over the entire landscape. The IFM connectivity measure relates to the number of immigrants arriving to the focal location—the occupancy probability of the species at the site would depend in part on that quantity. Marking by (x, y) the coordinates of cell i on a grid of dimensions x-dim × y-dim, the IFM connectivity can be calculated for any cell in the landscape as a sum over those selected cells in the rest of the landscape $x_{(k,m)}$ that have a positive occupancy probability $p_{(k,m),j}$ taking into account the distance between the cells $d_{x-k,y-m}$ and the habitat quality (surrogate for population size and number of emigrants leaving the location) at source locations, $h_{(k,m),j}$:

$$I_{ij}(\mathbf{x}) = I_{(x,y),j}(\mathbf{x}) = \sum_{k=1}^{x\text{-dim}} \sum_{m=1}^{y\text{-dim}} x_{(k,m)} p_{(k,m),j} \times \exp(-\alpha d_{x\text{-}k,\,y\text{-}m}) h_{(k,m),j}. \quad (5)$$

The problem with equation (5) is that it needs to be calculated multiple times for all cells and for each species when evaluating a single reserve structure. The multiple evaluations are needed because equations (4) and (5) need to be iterated until probabilities in all cells converge (this is the nonlinearity caused by the landscape structure). Connectivity for all cells, even those not in **x**, is needed because a local improvement method requires information on the potential connectivity of each cell when evaluating whether a cell should be joined to **x**.

Direct calculation of equation (5) for all cells is essentially an $O(N^2)$ operation, which becomes impossibly slow for even moderately large N (order of 10^4). Luckily, there is a computational method that facilitates the calculation of equation (5). It can be solved in $O(N \log N)$ time as a two-dimensional convolution in the spectral domain using the FFT algorithm to facilitate the transition to spectral domain and back. The same technique has been employed successfully in the context of integrodifference models of spatial population dynamics (see, e.g., Brewster and Allen 1997). For N in the order 10^7, direct calculation of equation (5) would take weeks or months, but it can be solved in seconds using an efficient FFT subroutine library (Frigo and Johnson 1998). See appendix A for a description of the optimization method used for solving the nonlinear reserve selection problem.

Testing Algorithm Performance

Appendix B describes an algorithm for computing lower bounds for the smallest solution size of the nonlinear reserve selection problem. The method is based on the analysis of the degree of overlap between species distributions. Two different bounds are computed; the B_1 bound assumes that loss of connectivity does not affect probabilities of occurrence. Bound B_2 even allows that loss of connectivity could in some cases increase probabilities of occurrence. Both of these bounds are underestimates for the true solution size required because they assume that connectivity loss never decreases the probabilities of occurrence and they assume that complementarity occurs in an optimal manner in the data so that the highest overlap of species distributions is at locations where the highest species-specific probabilities also occur. Appendix C in the online edition of the *American Naturalist* gives information concerning algorithm performance with artificially enlarged landscapes.

Material

This study is based on independent habitat suitability indexes and presence-absence data for seven threatened birds in a 55 × 40-km study area in the West Brabant province in the Netherlands (Rejnen et al. 2001). The statistical habitat models with connectivity components used here (eqq. [4], [5]) are the ones fitted by van Teeffelen et al. (2005). Connectivity-based autocovariates were statistically highly significant and biologically important for all species in these data. This indicates that habitat structure had an effect on the distribution of all of the species at the 1-km^2 scale; a suitable but isolated plot was less likely to be occupied than a suitable and well-connected one. Average dispersal distances for the species varied from 4 to 12 km (see van Teeffelen et al. 2005 for details).

Results

Figure 1 shows a summary of 10 optimization runs for the seven study species first separately and then together with two different target levels. The best solutions for species are quite different, reflecting differences in the pattern of suitable habitat in the landscape (fig. 1). The best solution for all species together shows a pattern that seemingly captures the overlap regions of distributions of different species in a highly connected manner.

Table 2 demonstrates that the level of spatial information used in the model strongly impacts the structure of the optimal reserve as measured by the ratio of reserve boundary length to area (BL/A). Use of presence-absence data results in a reserve structure for all species where BL/A = 1.96, meaning that on average around half of the edge of each cell is on the boundary of the reserve. Using probabilities of occurrence based on local habitat quality results in an optimal reserve structure with BL/A = 1.68, which still is highly fragmented. If connectivity components are included in the models, the optimal reserve structure becomes much more compact and connected (which naturally is to be expected if probability of occurrence is positively affected by connectivity). The ratio BL/A = 0.82 for models with local habitat quality + buffer connectivity, and BL/A = 0.70 for the full model structure with both buffer and long-distance connectivity.

A similar trend is seen though optimizations for individual species (table 2). Figure 2 illustrates strong effects of connectivity with one of the present study species; using a nonlinear model with effects of connectivity (fig. 2*B*, 2*C*) results in a much more aggregated reserve structure than the use of a (linear) model that is based on local habitat quality only (fig. 2*A*). Use of connectivity added significant clustering also for species 6, for which the BL/A ratio dropped from 1.53 to 0.93 when buffer connectivity was

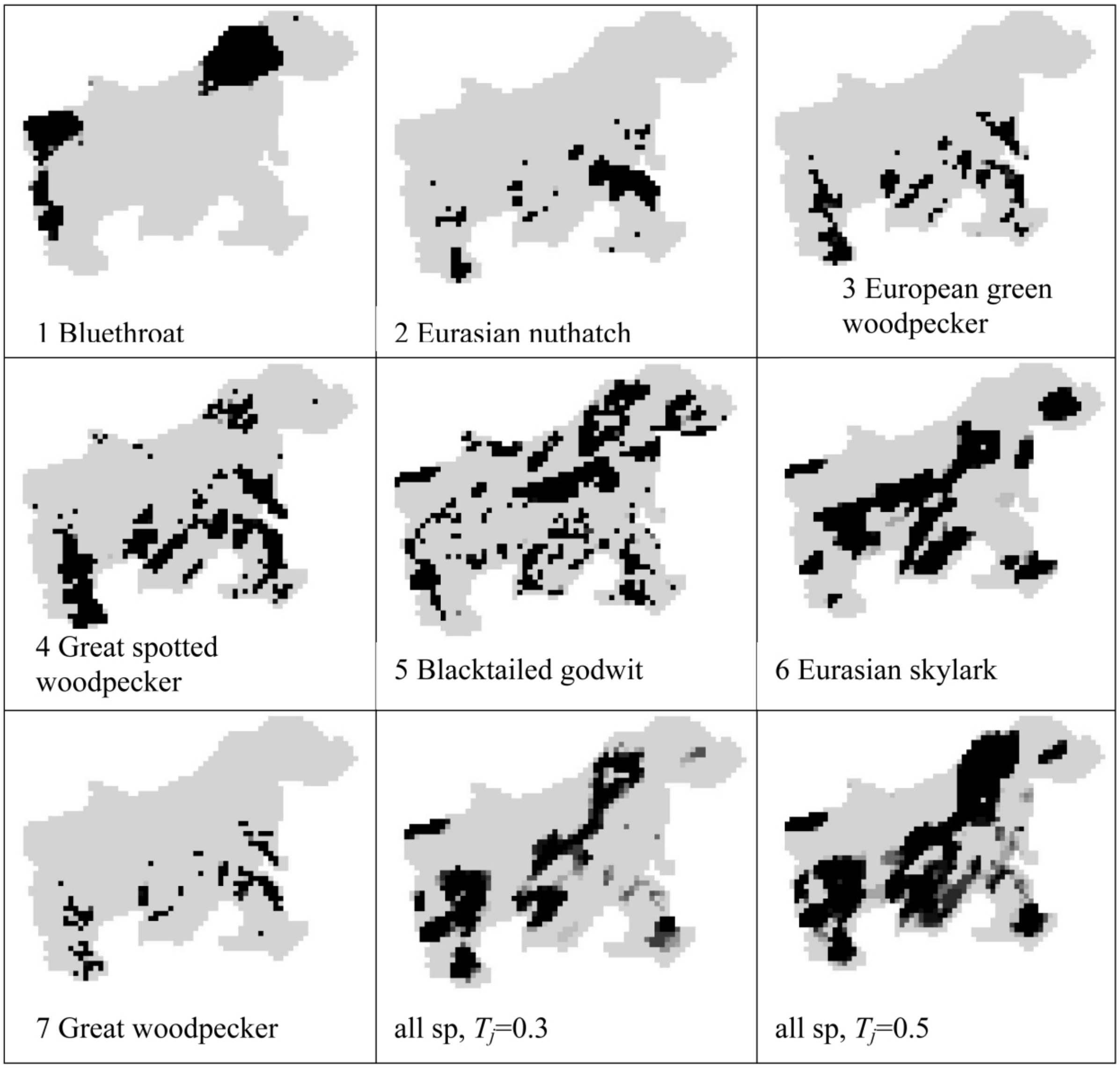

Figure 1: Optimal reserve selections for the species separately and together; summary of 10 replicate optimizations, $T_j = 0.5$ for individual species. White indicates areas outside the study area, and light gray indicates unselected sites potentially suitable for the species. Not all optimizations converge to the same spatial pattern, which is indicated by darkness of color: black indicates selection always, and shades of darker gray indicate less frequent inclusion in the optimal result of an optimization run. Variation in the selected regions indicates multiple spatially different solutions with equal biological value.

added to the model on top of habitat quality. Other species were affected varyingly less by connectivity, and one species (number 7) may even favor fragmentation in its habitat choice, as shown by a slight increase in BL/*A* with the addition of connectivity components into models (table 2).

Figure 3 is more relevant for conservation because it concerns the multispecies optimizations. The reserve structures produced using presence-absence data (fig. 3*A*) or local habitat quality in a linear manner (fig. 3*B*) are visibly highly fragmented and thus not particularly suitable as a basis for conservation planning. Going from presence-absence data to the full spatial model (fig. 3*D*) shows a major drop in BL/*A* from 1.96 to 0.70. It is worth emphasizing that this large difference in reserve structures is caused by the inclusion of connectivity components in the statistical model fit (van Teeffelen et al. 2005). It is well known that so-called neighborhood measures are often

Table 2: Effects of the level of spatial information on the optimal reserve structure

Species	PA	M_L	M_B	M_F
1	2.06	1.47	.93	.60
2	2.37	1.72	1.47	1.44
3	2.30	1.64	1.43	1.52
4	1.88	1.45	1.34	1.34
5	2.76	1.92	1.57	1.60
6	1.46	1.53	.93	.78
7	2.46	2.02	2.34	2.13
E(sp.)	2.18	1.68	1.43	1.34
All spp.	1.96	1.65	.82	.70

Note: Shown are presence-absence data (PA), probabilities of occurrence based on local habitat quality used in a linear nonspatial manner (M_L), a spatial model with local habitat quality and buffer-type connectivity included in the model (M_B), and a full model with habitat quality and both buffer and long-distance connectivity components (M_F). The ratio of reserve boundary length to area is given for each species individually and then for a joint estimation with all species included. *E*(sp.) is the average over the single-species estimations. $T_j = 0.5$.

important variables in habitat models (Guisan and Zimmermann 2000). If neighborhood measures or connectivity measures are significant in the habitat model for any species, then effects such as those shown in figures 2 and 3 can be expected when going from linear reserve selection models to nonlinear reserve selection models.

Even if the landscape used here is not very large (~2,000 cells), the nonlinear reserve selection problem is not trivially solvable computationally. The optimization method described in appendix A produced consistent performance over all optimizations tried in this study (table 2; app. C). Best solutions found by different optimization runs were highly consistent for the single-species optimizations, and a near-optimal result is on average obtained in some tens of seconds of computation on a fast desktop PC. The lower bound estimates (app. B) show that maximum suboptimality of solutions varies from 0.7% to 15.8%. This suboptimality could in principle be a result of convergence problems with the algorithm. However, the lower bound estimate assumes no negative effects of connectivity loss, which means that the lower bound estimates are optimistic to a degree depending on the response of the particular species to habitat loss. No improvement for the smallest solution size could be found for any of the individual species even with a computational effort 100 times greater than that required to find one of the results in figure 1. A stochastic search algorithm that incorporates a gradient-type local search and that starts from random patterns can be expected to behave like this only if the solution at this point is one of the global optima or a locally optimal solution with a very large basin of attraction.

Results are different when optimizing for all species simultaneously. Now convergence is slower, and there is some variation between the spatial patterns of the best solutions found by different optimization runs. Figure 1 shows that many regions (marked in black) are always selected when optimizing for all species, whereas some areas are selected only occasionally. Because minimal solutions for the optimization replicates (table 3) were within a couple of area units from the smallest result found (400–403 and 714–716 for the $T_j = 0.3$ and $T_j = 0.5$ problems, respectively) and lower bound estimates indicate maximum suboptimality of 9.8% and 11.2%, it appears that there are multiple almost equal near-global optima. Such a result is not surprising because there can easily exist multiple spatial patterns with essentially equal biological value, which is helpful for conservation decision making because there are options from which to choose.

The habitat models used by van Teeffelen et al. (2005) include interaction components that allow connectivity

Figure 2: Comparison of optimal reserve structures obtained for the bluethroat using different models. *A*, Local habitat quality only. *B*, Local habitat quality + buffer connectivity. *C*, Local habitat quality + long distance (IFM) connectivity. Ratios of boundary length to area for *A–C* are 1.47, 0.93, and 0.60, respectively.

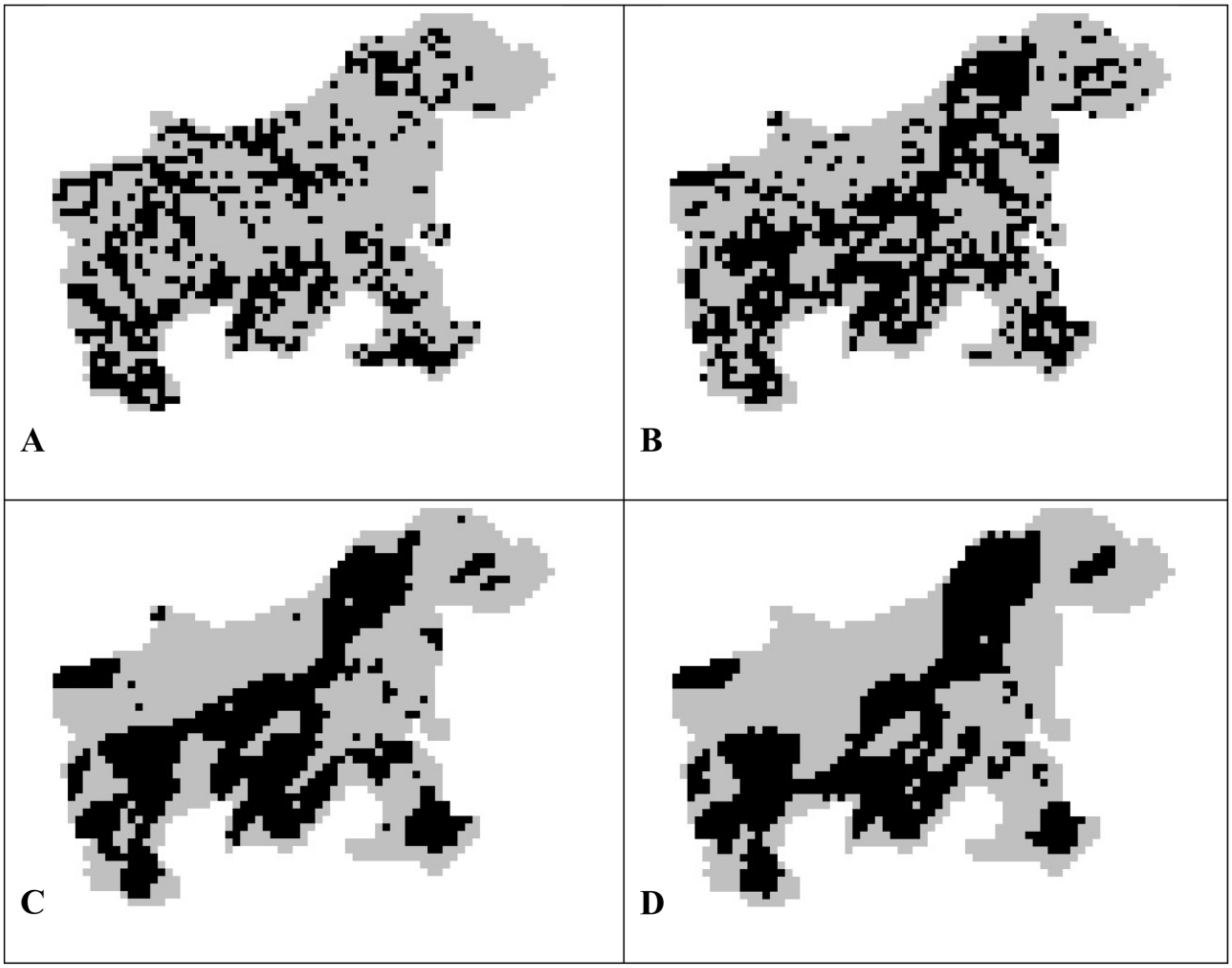

Figure 3: Optimal reserve structure for all species ($T_j = 0.5$) using different models as in table 3. *A*, Presence-absence data. *B*, Only local habitat quality used in linear nonspatial manner. *C*, Local habitat quality + buffer connectivity. *D*, Local habitat quality + buffer connectivity + long-distance connectivity. Ratios of boundary length to area for *A–D* are 1.96, 1.65, 0.80, and 0.70, respectively.

loss to have positive effects on probabilities of occurrence for a minority of cells for some species (table 3). To get an improved understanding of the optimality properties of the algorithm, the negative interaction terms were dropped from the models, leaving only positive effects of connectivity and local habitat quality (now the B_2 bound equals the B_1 bound, and the lower bound estimate is more precise). For these models, the species-specific suboptimality varied from 0.0024% to 13.8%, with an average of 3.85%. Maximum suboptimality for the $T_j = 0.5$ problem with all species was only 4.7%.

Discussion

Landscape structure affects the distributions of species. Inclusion of realistic complications, such as spatial population dynamics, connectivity, edge effects, or spatially correlated threats, can make species distribution models nonlinear. A reserve selection formulation also becomes nonlinear if any nonlinear transformation from species representation to value of representation is used. Nonlinear spatial population models have not been widely used in reserve selection studies probably because of substantial computational difficulties involved. This study describes and tests a novel reserve selection algorithm that is applicable in the context of suitably formulated nonlinear species distribution models, including those based on deterministic metapopulation models (Ovaskainen 2002; Ovaskainen and Hanski 2004) or statistical habitat models (e.g., Guisan and Zimmermann 2000) with dispersal kernel–based autocovariates (Augustin et al. 1996; Ferrier et al. 2002; Cabeza 2003).

Table 3: Optimal results and timings for individual species using $T_j = 0.5$ targets

Species (α)	Range (suboptimal)	B_1	B_2	$t(s)$
1 (.23)	276–276 (+14.5%)	247	241 (.981)	38
2 (.15)	139–139 (+0.7%)	138	111 (.976)	21
3 (.23)	205–205 (+15.8%)	218	177 (.858)	19
4 (.23)	391–391 (+2.9%)	390	380 (.842)	16
5 (.08)	529–529 (+2.3%)	518	517 (.998)	15
6 (.23)	536–538 (+5.1%)	538	510 (.940)	46
7 (.08)	118–118 (+8.3%)	112	109 (.985)	22

Note: The table gives the cost (area) of the optimal (smallest) solution for each species. Optimal result ranges give highest and lowest value of the objective function in 10 replicate optimization runs using parameters $\mathrm{GA_G} = 30$ and $\mathrm{GA_M} = 300$. Bound B_1 is the lower bound for optimal solution size on the assumption that loss of connectivity does not affect probabilities of occurrence (app. B in the online edition of the *American Naturalist*); B_2 is the lower bound when taking into account that loss of connectivity may even slightly increase probabilities of occurrence for some cells (because of negative interaction terms involving connectivity; see app. B). The number in parentheses following B_2 is the proportion of cells (for that species) for which the condition of a nonnegative derivative of connectivity is satisfied (app. B). Maximum suboptimality of the best result with respect to the B_2 bound is given in parentheses. Time is the average required by a small population genetic algorithm (GA; $\mathrm{GA_M} = 50$) to find a result equal in size to the smallest one found.

Inclusion of connectivity in the models had variable effects for different species. Some were strongly affected (fig. 2), whereas for others, inclusion of connectivity caused only a minor increase in reserve compactness. It is worth noting that inclusion of species with strong effects of connectivity apparently causes the reserve structure for all species to be strongly aggregated (figs. 1, 3); some species are relatively indifferent to fragmentation, but others are not, and those species drive the aggregation in the all-species solutions. Consequently, one could hypothesize that if fragmentation-sensitive and fragmentation-indifferent species have at least partially overlapping distributions of suitable habitat and unless some species are actually hurt by reserve compactness, an optimal solution for many species would be aggregated because of the presence of the fragmentation-sensitive species in the data. In figure 1, BL/*A* is only 0.99 and 0.70 for the $T_j = 0.3$ and $T_j = 0.5$ problems, respectively. For comparison, optimizations for the same area using presence-absence data or habitat models with only effects of local habitat quality produced "optimal" reserve selections that were much more fragmented, with BL/*A* ratios close to 2 (see also van Teeffelen et al. 2005).

Table 4 summarizes qualitative effects of the use of different kinds of data and models with and without connectivity components on reserve selection. Optimization based on presence-absence data or on probabilities of occurrence based on local habitat quality is likely to produce very scattered reserve structures (e.g., Cabeza et al. 2004*b*; van Teeffelen et al. 2005). Basing reserve selection on any statistical habitat model with a positive effect of connectivity (or a positive effect of the habitat neighborhood, such as the amount of forest within a radius) will necessarily result in more compact optimal spatial reserve designs because in such models, compactness mathematically translates into higher probabilities of occurrence. (The quantitative effects of the inclusion of connectivity will be species and landscape specific.) In general, the use of buffer-type connectivity measures alone may induce more compact reserve structures than the use of IFM-type connectivity, which allows connectivity not only to the neighboring cells but also over longer distances. Even without explicit consideration of quantitative effects of connectivity, aggregation may be induced into reserves with potentially little perceived loss of biological value via the use of a penalty for reserve boundary length (not analyzed here, but see Possingham et al. 2000; Nalle et al. 2002; Önal and Briers 2002*a*, 2002*b*; Fischer and Church 2003; Cabeza et al. 2004*a*).

Qualitatively, the optimal reserve structure will be affected by the aggregation level of good-quality habitat, species-specific dispersal distances, and the strength of the effects of connectivity in the species models. The shorter the dispersal distances, the stronger the effects of connectivity on occupancy, and the more aggregated the habitat is, the more aggregated the optimal reserve structure can be expected to be.

Concerning the optimization method, a genetic algorithm (as any stochastic search method) is a heuristic algorithm in the sense that the global optimum is not guaranteed and it is not known how good the solution is compared with the global optimum. However, for the present reserve selection formulation, it was possible to formulate a lower bound for the optimal (minimum) solution size, which changes the method from a heuristic to an approximation. The solutions found in this study were from 0.024% to 16% suboptimal (depending on the problem) as compared with an optimistically calculated lower bound for the solution size. This combined with consistent convergence to a particular solution over multiple stochastic optimization runs shows the ability of the algorithm to consistently find near-optimal solutions. The proposed algorithm has good time-scaling properties, $O(N \log N)$, as a function of problem size, which indicates that the method should be suitable for solving increasingly large problems with the continuing increase of computing power. Even with current PC computers, the algorithm can be used with landscapes in the order of 10^4–10^5 cells.

Alternative approaches to the present reserve selection problem include exhaustive search using nonlinear programming and other stochastic search algorithms such as simulated annealing (SA). Currently, exhaustive search us-

Table 4: Qualitative effects of different levels of inclusion of connectivity into reserve selection

Model and data	Qualitative characteristics of optimal reserve selection
1. P-A data, no effects of connectivity	Small solution sizes, with potentially very scattered/perforated spatial structures. Perceived value of solution not affected by the spatial structure of the solution (no effects of habitat loss). Number of persisting populations always less than the number of presences in data (presence does not indicate long-term persistence, especially when using small selection units).
2. Probabilities of occurrence based on local habitat quality, no connectivity	Much as with presence-absence data but with slightly more continuous solution structures because of autocorrelation of good habitat. Probabilities are used in a linear manner; perceived value of solution is not affected by the spatial structure of the solution.
3. P-A or probability data, clustering via penalty for boundary length	Arbitrary level of clustering can be achieved by varying the magnitude of a penalty given for the boundary length of a solution. Quantitative effects of clustering to the number of populations is unknown. A good default strategy, a cost-benefit analysis may reveal a level of clustering that may be achieved with minor cost.
4. Habitat models with neighborhood measures or buffer connectivity measures	Quantitative effects of connectivity and habitat loss. Potentially very strong clustering because connectivity is between adjacent cells only, which induces the selection of continuous blocks of habitat. No species-specific effects of distance in the connectivity measure. Computationally yet moderately light.
5. Habitat models that include long-distance (IFM-type) connectivity (and possibly also buffer connectivity)	Species-specific dispersal kernels, quantitative effects of connectivity and habitat loss/deterioration. Solution not necessarily as aggregated as with buffer measures only because connectivity can occur over longer distances, not only to neighbors. No explicit effects of habitat matrix structure (only distance) but still computationally heavy.
6. Population viability analysis simulation models with explicit effects of habitat matrix structure on dispersal	Very detailed effects of connectivity but models hard to parameterize for many species. Computationally very heavy. Has been done in the context of comparison of conservation scenarios for a few species but not as part of large-scale multispecies optimal reserve selection.

Note: P-A = presence-absence. IFM = incidence function model.

ing integer programming becomes practically impossible for linear reserve selection problems when the number of landscape elements is around 10^4 (Williams et al. 2004), which suggests that prospects are not promising for using nonlinear programming on even larger landscapes. There is a particular reason why local search was combined with a GA but not SA. The performance of the proposed search method is very much dependent on the use of an efficient local search operator in the optimization process. Local search combines naturally with a GA, in which selection is constantly improving the average performance of the solutions in the search point population and where local search can be seen as a way of helping the convergence of the population to good solutions. In contrast, SA uses a single search point, which is roaming around the search space. Importantly, SA allows nonimproving changes to the search point so as to enable the search point to escape local optima. Adding local search to SA would have the effect of repeatedly forcing the single search point back to a nearby local optimum, which fights against the need to allow the point escape local optima. Thus, combining SA and local search is conceptually harder than combining a GA with local search.

Although there are an increasing number of exceptions, reserve selection algorithms have not been extensively used in real conservation planning, probably for three reasons (see Margules and Pressey 2000; Faith et al. 2003). Algorithm formulations have not allowed for important aspects of planning (already included reserves, considerations of connectivity and population persistence, considerations of cost, threat, availability, scheduling, etc.), the algorithms have been unable to handle landscapes at the scale and resolution commonly used in planning, and the algorithms have not been implemented in a decision-making framework available for conservation planners. This work contributes to the first two reasons, allowing reserve selection using comparatively realistic habitat models with long-distance connectivity components at scales that are relevant for practical applications.

Acknowledgments

I thank H. Sierdesema for the use of the habitat maps for the species. M. Burgman, C. Flather, and three anonymous reviewers are thanked for helpful comments on the manuscript. This study was funded by the Academy of Finland project 1206883 and the Finnish Center of Excellence Programme 2000–2005, grant 44887.

Literature Cited

Arponen, A., R. Heikkinen, C. D. Thomas, and A. Moilanen. 2005. The value of biodiversity in reserve selection: representation, species weighting and benefit functions. Conservation Biology (forthcoming).

Arthur, J. L., R. G. Haight, C. A. Montgomery, and S. Polasky. 2002. Analysis of the threshold and expected coverage approaches to the probabilistic reserve site selection problem. Environmental Modeling and Assessment 7:81–89.

Augustin, N. H., M. A. Mugglestone, and S. T. Buckland. 1996. An autologistic model for the distribution of wildlife. Journal of Applied Ecology 33:339–347.

Beyer, H.-G., and K. Deb. 2000. On the desired behaviors of self-adaptive evolutionary algorithms. Pages 59–68 *in* M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H.-P. Schwefel, eds. Parallel problem solving from nature. PPSN VI. Lecture notes in computer science. LNCS 1917. Springer, Heidelberg.

Blum, C., and A. Roli. 2003. Metaheuristics in combinatorial optimization: overview and conceptual comparison. ACM Computing Surveys 35:268–308.

Brewster, C. C., and J. C. Allen. 1997. Spatiotemporal model for studying insect dynamics in large-scale cropping systems. Environmental Entomology 26:473–482.

Cabeza, M. 2003. Habitat loss and connectivity of reserve networks in probability approaches to reserve design. Ecology Letters 6:665–672.

Cabeza, M., and A. Moilanen. 2001. Design of reserve networks and the persistence of biodiversity. Trends in Ecology & Evolution 16: 242–248.

———. 2003. Site-selection algorithms and habitat loss. Conservation Biology 17:1402–1413.

Cabeza, M., M. B. Araujo, R. J. Wilson, C. D. Thomas, M. J. R. Cowley, and A. Moilanen. 2004*a*. Combining probabilities of occurrence with spatial reserve design. Journal of Applied Ecology 41:252–262.

Cabeza, M., A. Moilanen, and H.-P. Possingham. 2004*b*. Metapopulaton dynamics and reserve network design. Pages 541–564 *in* I. Hanski and O. Gaggiotti, eds. Metapopulation ecology, genetics, and evolution. Academic Press, London.

Camm, J. D., S. Polasky, A. R. Solow, and B. Csuti. 1996. A note on optimal algorithms for reserve design. Biological Conservation 78: 353–355.

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

Cocks, K. D., and I. A. Baird. 1989. Using mathematical programming to address the multiple reserve selection problem: an example from the Eyre Peninsula, South Australia. Biological Conservation 78:113–130.

Csuti, B., S. Polaski, P. H. Williams, R. L. Pressey, J. D. Camm, M. Kershaw, R. Kiester, et al. 1997. A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon. Biological Conservation 80:83–97.

Debinski, D. M., and R. D. Holt. 2000. A survey and overview of habitat fragmentation experiments. Conservation Biology 14:342–355.

Faith, D. P., G. Carter, G. Cassis, S. Ferrier, and L. Wilkie. 2003. Complementarity, biodiversity viability analysis and policy-based algorithms for conservation. Environmental Science and Policy 6: 311–328.

Ferrier, S., G. Watson, J. Pearce, and M. Drielsma. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. Biodiversity and Conservation 11:2275–2307.

Fischer, D. T., and R. L. Church. 2003. Clustering and compactness in reserve site selection: an extension of the biodiversity management area selection model. Forest Science 49:555–565.

Frigo, M., and S. G. Johnson. 1998. FFTW: an adaptive software architecture for the FFT. Pages 1381–1384 *in* Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, Piscataway, NJ.

Gaston, K. J., R. L. Pressey, and C. R. Margules. 2002. Persistence and vulnerability, retaining biodiversity in the landscape and in the protected areas. Journal of Biosciences 27:361–384.

Goldberg, D. E. 1989. Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Reading, MA.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147–186.

Haight, R. G., C. S. Revelle, and S. A. Snyder. 2000. An integer optimization approach to a probabilistic reserve site selection problem. Operations Research 48:697–708.

Hanski, I. 1994. A practical model of metapopulation dynamics. Journal of Animal Ecology 63:151–162.

———. 1998. Metapopulation dynamics. Nature 396:41–49.

Hof, J., and M. Bevers. 2002. Spatial optimization in ecological applications. Columbia University Press, New York.

Holland, J. H. 1975. Adaption in natural and artificial systems. University of Michigan Press, Ann Arbor.

Jaramillo, J. H., R. Bhadury, and R. Batta. 2002. On the use of genetic algorithms to solve location problems. Computers and Operations Research 29:761–779.

King, A. W., and K. A. With. 2002. Dispersal success on spatially structured landscapes: when do spatial pattern and dispersal behavior really matter? Ecological Modelling 147:23–39.

Kirkpatrick, J. B. 1983. An iterative method for establishing priorities for the selection of natural reserves: an example from Tasmania. Biological Conservation 25:127–134.

Korte, B., and J. Vygen. 2000. Combinatorial optimization: theory and algorithms. Springer, Berlin.

Leslie, H., M. Ruckelshaus, I. R. Ball, S. Andelman, and H. P. Possingham. 2003. Using siting algorithms in the design of marine reserve networks. Ecological Applications 13:185–198.

Levin, S. A. 1974. Dispersion and population interactions. American Naturalist 108:207–228.

MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, NJ.

Margules, C. R., and R. L. Pressey. 2000. Systematic conservation planning. Nature 405:243–253.

Margules, C. R., A. O. Nicholls, and R. L. Pressey. 1988. Selecting networks of reserves to maximize biological diversity. Biological Conservation 43:63–76.

McDonnell, M. D., H. P. Possingham, I. R. Ball, and E. A. Cousins. 2002. Mathematical methods for spatially cohesive reserve design. Environmental Modeling and Assessment 7:107–114.

Moilanen, A., and M. Cabeza. 2002. Single species dynamic site-selection. Ecological Applications 12:913–926.

Moilanen, A., and M. Nieminen. 2002. Simple connectivity measures for metapopulation studies. Ecology 84:1131–1145.

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

Noss, R. F. 2003. A checklist for wildland networks design. Conservation Biology 5:1270–1275.

Önal, H. 2004. First-best, second-best, and heuristic solutions in conservation reserve site selection. Biological Conservation 115: 55–62.

Önal, H., and R. S. Briers. 2002*a*. Incorporating spatial criteria in optimum reserve network selection. Proceedings of the Royal Society of London B 269:2437–2441.

———. 2002*b*. Selection of a minimum-boundary reserve network using integer programming. Proceedings of the Royal Society of London B 270:1487–1491.

Ovaskainen, O. 2002. The effective size of a metapopulation living in a heterogeneous patch network. American Naturalist 160:612–628.

Ovaskainen, O., and I. Hanski. 2004. Metapopulation dynamics in highly fragmented landscapes. Pages 73–104 *in* I. Hanski and O. Gaggiotti, eds. Metapopulation ecology, genetics, and evolution. Academic Press, London.

Polasky, S., J. D. Camm, A. R. Solow, B. Csuti, D. White, and R. Ding. 2000. Choosing reserve networks with incomplete species information. Biological Conservation 94:1–10.

Possingham, H. P., I. R. Ball, and S. Andelman. 2000. Mathematical methods for identifying representative reserve networks. Pages 291–305 *in* S. Ferson and M. Burgman, eds. Quantitative methods for conservation biology. Springer, New York.

Pressey, R. L. 1999. Systematic conservation planning for the real world. Parks 9:1–6.

Pressey, R. L., and S. L. Tully. 1994. The cost of ad hoc reservation: a case study in the Western Division of New South Wales. Australian Journal of Ecology 19:375–384.

Pressey, R. L., H. P. Possingham, and J. R. Day. 1997. Effectiveness of alternative heuristic algorithms for identifying indicative minimum requirements for conservation reserves. Biological Conservation 80:207–219.

Rejnen, R., R. Jochem, M. de Jong, M. de Heer, and H. Sierdesema. 2001. Larch vogels nationaal: een expertsystem voor het beoordelen van de ruimtelijke samenhang en de duurzaamheid van broedvogelpopulaties in Nederland. Alterra Rapport 235, Alterra, Research Instituut voor de Groene Ruimte, Waningen, The Netherlands.

ReVelle, C. S., J. C. Williams, and J. J. Boland. 2002. Counterpart models in facility location science and reserve selection science. Environmental Modeling and Assessment 7:71–80.

Sarkar, S., C. Pappas, J. Garson, A. Aggarwal, and S. Cameron. 2004. Place prioritization for biodiversity conservation using probabilistic surrogate distribution data. Diversity and Distributions 10: 125–133.

Smith, P. G. R., and J. B. Theberge. 1986. A review of criteria for evaluating natural areas. Environmental Management 10:715–734.

Snyder, S. A., L. E. Tyrell, and R. G. Haight. 1999. An optimization approach to selecting research natural areas in national forests. Forest Science 45:458–469.

Underhill, L. G. 1994. Optimal and suboptimal reserve selection algorithms. Biological Conservation 105:103–111.

van Teeffelen, A. J. A., M. Cabeza, and A. Moilanen. 2005. Habitat loss, connectivity and probabilities: comparing reserve selection strategies. Biodiversity and Conservation (forthcoming).

Westpahl, M. I., and H. P. Possingham. 2003. Applying a decision-theory framework to landscape planning for biodiversity: follow up to Watson et al. Conservation Biology 17:327–329.

Westpahl, M. I., M. Pickett, W. M. Getz, and H. P. Possingham. 2003. The use of stochastic dynamic programming in optimal landscape reconstruction for metapopulations. Ecological Applications 13: 543–555.

Williams, J., C. S. ReVelle, and S. A. Levin. 2004. Using mathematical optimization models to design nature reserves. Frontiers in Ecology and the Environment 2:98–105.

Williams, P. H., and M. B. Araújo. 2000. Using probabilities of persistence to identify important areas for biodiversity conservation. Proceedings of the Royal Society of London B 267:1959–1966.

———. 2002. Apples, oranges, and probabilities: integrating multiple factors into biodiversity conservation with consistency. Environmental Modeling and Assessment 7:139–151.

Editor: Jonathan B. Losos
Associate Editor: Curtis H. Flather