Environmental Modeling and Assessment **7:** 81–89, 2002.

Analysis of the threshold and expected coverage approaches to the probabilistic reserve site selection problem

Jeffrey L. Arthur [a], Robert G. Haight [b], Claire A. Montgomery [c] and Stephen Polasky [d]

[a] *Department of Statistics, Oregon State University, Corvallis, OR 97331, USA*
[b] *USDA Forest Service, North Central Forest Experiment Station, 1992 Folwell Avenue, St. Paul, MN 55108, USA*
[c] *Department of Forest Resources, Oregon State University, Corvallis, OR 97331, USA*
[d] *Department of Applied Economics, University of Minnesota, St. Paul, MN 55108, USA*

Two approaches to formulating the reserve site selection problem when species occurrence data is probabilistic were solved for terrestrial vertebrates in a small set of potential reserve sites in Oregon. The expected coverage approach, which maximizes the sum of the occurrence probabilities, yielded solutions that covered more species on average in Monte Carlo simulations than the threshold approach, which maximizes the number of species for which the occurrence probability exceeds some threshold.

Keywords: reserve site selection, incomplete information, species distributions

1. Introduction

Pressure to convert natural habitat for human use will intensify in the foreseeable future as the human population continues to grow, perhaps by another several billion over the next fifty years. The resulting losses of habitat pose severe threats to biological diversity. The question of how best to conserve biological diversity in the face of habitat loss is an urgent one. Given that not all habitat will be protected, which areas of habitat are the most important to protect in order to conserve biological diversity?

One important conservation strategy is to establish a biological reserve network composed of a set of natural reserves or parks (Noss and Cooperrider [1], Pimm and Lawton [2]). If it is not possible to reserve all potential sites in such a network, it is desirable for a conservation agency to choose the subset of sites that would protect the most biological diversity. This problem is known as the reserve site selection problem and is generally posed as follows. Given a constraint on the number of sites that may be selected as reserves, choose the set of sites that covers the maximum number of species, where a species is defined as covered if it is present in at least one selected site. A species that is not present in any selected site is not covered. If the presence or absence of each species at each site is known with certainty, the reserve site selection problem is a maximal coverage problem (Church and ReVelle [3]). Applications of the maximal coverage problem to reserve site selection include Church et al. [4], Kiester et al. [5], Csuti et al. [6], Pressey et al. [7], Ando et al. [8] among others.

However, the presence or absence of a species at a site is rarely known with certainty. Except for a few well-studied or easily observed species, only sketchy information about species distributions typically exists. In the absence of certain knowledge, it is sometimes possible to estimate the probability of species presence at a site as a function of environmental characteristics (e.g., climate, soil, topography). For examples of this approach, see Austin et al. [9], Margules and Nicholls [10], Margules and Stein [11], Nicholls [12], and Margules and Austin [13].

In the face of uncertainty, the decision-maker must use a surrogate for the maximal coverage problem – the ultimate objective still being to choose the set of sites that do, in fact, cover the maximum number of species. Two methods have been proposed in the literature for solving the reserve site selection problem with probabilistic data.

One approach, which we call the expected coverage approach, maximizes the expected number of species covered, which is the sum over all species of the probability of coverage for each species (Polasky et al. [14]). This approach addresses the ultimate objective of maximal coverage most directly, but because the objective function involves sums of probabilities, each of which is a nonlinear function of the decision variables, it cannot be transformed into a linear form. Camm et al. [15] demonstrate that the expected coverage problem is a nonlinear binary integer program that is NP hard. Hence, finding optimal solutions to this problem cannot be guaranteed for problems of reasonably large size. While we utilized complete enumeration for the purposes of identifying optimal solutions in our analysis to follow, heuristic approaches can be used to find good though not necessarily optimal solutions for larger problems. Among these are the greedy addition and substitution approach (Daskin [16], Daskin et al. [17]) developed for a closely related expected covering problem. Others have reported on the successful application of various metaheuristics such as simulated annealing (Liu and Wang [18], Murray and Church [19], Righini [20], and Ernst and Krishnamoorthy [21]), genetic algorithms (Hosage and Goodchild [22], Houck et al. [23], Kumar et al. [24], and George [25]) and tabu search (Crainic et al. [26], Rolland et al. [27], Ohlemuller [28], and Gendron et al. [29]) on facility location types of problems. In addition, developments on a separable programming approximation approach to this problem have shown promising results (Camm et al. [15]).

A second approach, which has been applied more often in the conservation literature, is the threshold approach. The goal in the threshold approach is to maximize the number of species covered, where a species counts as covered only if the probability of coverage reaches a specified threshold, e.g., 95% (Haight et al. [30], Margules and Nicholls [10], Margules and Stein [11]). The threshold approach is structurally similar to the maximal coverage problem because continuous probabilities are converted to dichotomous 0,1 variables and then counted. This approach has the advantage that it is computationally tractable because it can be transformed into an equivalent linear integer-programming problem. Consequently, problem sizes of practical significance can be solved to optimality (Haight et al. [30]). The threshold approach is similar to certain environmental laws that require meeting a specified target level of risk. It is also similar to the "safe minimum standard" approach to conservation (Ciriacy-Wantrup [31], Bishop [32]) in which there is a threshold that can be identified as the minimum acceptable probability of success (or conversely the maximum acceptable level of risk of failure) with respect to the conservation objective. Under this approach, any species that meets the threshold counts as contributing to the success of the conservation plan. Any species that fails to achieve the threshold does not count as contributing to the success of the conservation plan. This approach, however, begs the question of how to identify a safe minimum standard that embodies a socially acceptable level of risk and how to evaluate alternative conservation programs that exceed that standard.

Perhaps the major difference between the threshold and expected coverage approaches is in how they handle species coverage probabilities. The threshold approach truncates the coverage probabilities by assuming that a species is covered if it meets the threshold and is otherwise not covered. As a result, the threshold approach fails to use all of the information contained in the species occurrence probabilities. For example, there is no difference between a species with a coverage probability that is just below the threshold (say 94%) and one that has a zero coverage probability. Both species would not be counted as covered. In contrast, the expected coverage approach uses all the information in the coverage probabilities because those probabilities enter directly in the objective function. We view this as a major advantage of the expected coverage approach over the threshold approach.

The purpose of this paper is to investigate how the differences in the threshold and expected coverage approaches affect the choice of reserve sites. Given the ultimate objective of maximal coverage, we compared their performance in terms of the number of species covered in the reserve sites chosen. Because the actual number of species covered cannot be observed, we used Monte Carlo simulations to generate distributions of realizations drawn from the occurrence probabilities associated with the chosen sites. If the results are similar, the more computationally tractable threshold approach may be preferred. If the results differ, preference will be given to the approach that yields the highest level of coverage on average.

We formulated and solved threshold and expected coverage problems using probabilistic species occurrence data on terrestrial vertebrates in Oregon. We restricted the set of potential reserve sites to twenty so that optimal solutions for both the threshold and expected coverage approaches could be found via complete enumeration. We examined the extent to which the sets of selected sites overlapped using the two approaches. Performance was compared using estimates of the means and standard deviations of the number of species covered from Monte Carlo simulations.

In section 2 of the paper, we describe the threshold and expected coverage approaches. The Oregon terrestrial vertebrate data set is described in section 3. We describe solution methods including the Monte Carlo simulation in section 4. We show the results of applying the threshold and expected coverage approaches to the Oregon terrestrial vertebrate data in section 5. We discuss the results and offer concluding comments in section 6.

2. The threshold and expected coverage approaches

Let I represent the set of species under consideration and let J represent the set of potential reserve sites. From the set J, k sites may be chosen as the conservation reserve network. Define the binary variable X_j for all $j \in J$ as follows:

$$X_j = \begin{cases} 1, & \text{if site } j \text{ is selected to be part} \\ & \text{of the reserve network,} \\ 0, & \text{if site } j \text{ is not selected.} \end{cases}$$

Let p_{ij} be the probability that species $i \in I$ exists at site $j \in J$. These probabilities may be derived from expert opinion or from statistical methods such as logistic regression. We assume that p_{ij} and p_{mn} are independent for $i \neq m$ or $j \neq n$. Under these assumptions, the probability that species i is represented in a reserve network is:

$$P_i = 1 - \prod_{j \in J}(1 - p_{ij}X_j). \tag{1}$$

2.1. Expected coverage approach

With the model defined above, the reserve site selection problem using the expected coverage approach is:

$$\begin{aligned} &\max \sum_{i \in I} P_i, \\ &\text{s.t.} \quad \sum_{j \in J} X_j \leqslant k. \end{aligned} \tag{2}$$

Because the objective function is the sum of terms that involve the products of the decision variables X_j, the problem cannot be reduced to a linear integer-programming problem. Nevertheless, good solutions to problems of reasonably large size can be found using a linear approximation approach (Camm et al. [15]) or a greedy adding heuristic (Polasky et al. [14]).

2.2. The threshold approach

With the threshold approach, a binary variable, Y_i, is defined for each species i to represent whether or not the species is covered in the selected set of sites with the required probability. Letting $\overline{P}$ represent the selected threshold probability,

$$Y_i = \begin{cases} 1, & \text{if } P_i \geqslant \overline{P}, \\ 0, & \text{otherwise.} \end{cases}$$

Under the threshold approach, the objective of the reserve site selection problem is to maximize the number of species whose coverage probabilities exceed the specified threshold probability (Haight et al. [30]):

$$\begin{aligned} &\max \sum_{i \in I} Y_i, \\ &\text{s.t.} \quad \sum_{j \in J} X_j \leqslant k, \\ &\qquad \prod_{j \in J} (1 - p_{ij})^{X_j} \leqslant \left(1 - \overline{P}\right)^{Y_i} \quad \forall i \in I, \end{aligned} \tag{3}$$

where, as before, X_j is the binary decision variable for site j and p_{ij} is the probability that species i exists at site j.

The second set of constraints in the threshold problem defines the conditions under which species are considered covered. This constraint set stipulates that to cover species i, the probability that species i is absent from the selected set of sites, $\prod_{j \in J}(1 - p_{ij})^{X_j}$, must be less than the specified threshold probability of absence $(1 - \overline{P})$. If

$$\prod_{j \in J} (1 - p_{ij})^{X_j} > (1 - \overline{P}),$$

then the corresponding Y_i on the right-hand-side of the constraint must equal zero, indicating that the selected sites do not cover species i with the required probability. If

$$\prod_{j \in J} (1 - p_{ij})^{X_j} \leqslant (1 - \overline{P}),$$

then $Y_i = 1$, indicating that species i is covered with the required probability. This constraint set can be replaced with an equivalent set of linear constraints by means of a log transformation:

$$\sum_{j \in J} X_j \ln(1 - p_{ij}) \leqslant Y_i \ln(1 - \overline{P}) \quad \forall i \in I. \tag{4}$$

As a result, the threshold approach can be formulated as a linear integer-programming problem, and reasonably sized problems can be solved using exact optimization methods (Haight et al. [30]).

3. Terrestrial vertebrate data for Oregon

We used the same data set on terrestrial vertebrate distributions in the state of Oregon as used in Polasky et al. [14]. A hexagonal grid (635 km^2) divided the state into 441 sites. Data on the geographic distribution of 426 terrestrial vertebrate species that breed in the state of Oregon was collected as part of a national program to map biodiversity. For each of the 441 sites, the likelihood of each species occurrence was assessed and placed into one of four categories:

(a) *confident* – a verified sighting of the species in the site has occurred in the past two decades (probability of 0.95–1);

(b) *probable* – the site contains suitable habitat for the species, there have been verified sightings in nearby sites, and in the opinion of a local expert, it is highly probable that the species occurs in the site (probability of 0.8–0.95);

(c) *possible* – no verified sightings have occurred in the site, the habitat is of questionable suitability for the species, and in the opinion of a local expert, the species might occur in the site (probability of 0.1–0.8);

(d) *not present* – habitat is unsuitable for the species (probability of 0–0.1).

Detailed descriptions of the basis for category assignment for each species at each site are given in Master et al. [33].

For most of the analysis, we used the lower bound probability for each category to assign species occurrence probabilities (0.95, 0.8, 0.1, 0), which we refer to as the base case probabilities. If a species was rated as confident at a site, we assigned an occurrence probability of 0.95. If a species was rated as probable, we assigned an occurrence probability of 0.8, and so on. To investigate the effects of uncertainty, we also performed analyses using an alternative set of probabilities. In this set, the probability for the "probable" category was decreased to 0.6 and the probability of the "possible" category was increased to 0.4. This increased the effects of uncertainty by increasing the number of midrange probability estimates.

4. Solution methods

To perform the analysis, we randomly selected two separate sets of 20 potential reserve sites, which we refer to as data sets 1 and 2. For each set of potential reserve sites and each set of species occurrence probabilities, we formulated reserve site selection problems using the threshold and expected coverage approaches with upper bounds on the number of reserved sites (k) increasing from 1 to 10 and a threshold probability $\overline{P} = 0.95$. An optimal solution to each problem was obtained by evaluating all possible sets of k reserved sites and picking the best set. With the threshold approach we often found multiple optimal solutions, particularly as the value of k increased. In such instances, we selected the optimal set of sites for our analysis that also gave the highest expected coverage value. We also investigated the effects of lowering the threshold probability $\overline{P}$.

We compared the two approaches with respect to the stated objective of the reserve site selection problem, which

Table 1
Comparing threshold and expected coverage approaches using data set 1 and base case probabilities (0.95, 0.8, 0.1, 0) with threshold probability of 0.95.

Number of sites selected	Expected coverage approach		Threshold approach		Percent of shared selected sites
	Expected coverage	Threshold value	Expected coverage	Threshold value	
1	209.00	58	184.95	87	0
2	277.99	184	261.30	227	0
3	311.72	244	299.68	267	66.7
4	330.88	289	326.24	292	75.0
5	340.70	307	331.89	310	60.0
6	348.04	316	341.41	324	66.7
7	352.62	327	348.77	332	57.1
8	355.67	332	351.05	336	62.5
9	357.29	336	353.92	339	66.7
10	358.69	339	358.13	341	90.0

is to maximize the number of species actually covered in the chosen sites. Although actual species coverage cannot be observed, it is possible to calculate how well each approach achieved species coverage using Monte Carlo simulation. For each optimal set of reserved sites obtained using the threshold and expected coverage formulations, we estimated the probability distribution of the number of species covered in Monte Carlo simulation. Each simulation consisted of 10,000 replicates. In each replicate, choosing a random number between zero and one and comparing this to the probability of species presence determined the presence or absence of a species at a given site. If the random number was less than the probability, the species was deemed to be present, otherwise it was deemed absent from the site. As a result, each replicate represented one possible outcome of species occurrences in the reserved sites. The simulations were conducted using a common random number stream obtained from a prime modulus multiplicative linear congruential generator (Law and Kelton [34], p. 227). The outcomes of each simulation were plotted as a histogram of number of species covered by the reserved sites, and the mean and standard deviation of the outcomes were computed for the performance evaluation.

5. Results

In table 1, we compared the results using threshold and expected coverage approaches for the first data set using the base case set of probabilities (0.95, 0.8, 0.1, 0). We solved both for the collection of sites that maximizes expected coverage and for the collection of sites that maximizes the number of species achieving the threshold probability. We reported both the expected coverage value and the threshold value for each solution. We assumed a threshold probability ($\overline{P}$) of 0.95. We varied the constraint on the number of sites that may be selected (k) from one to ten.

As shown in table 1, the outcomes differed between the threshold and expected coverage approaches. In some cases, the expected coverage and threshold values were dramatically different. For example, for $k = 1$, the expected coverage value was 11.5% lower in the threshold approach (185 compared to 209) while the threshold value was 33.3% lower in the expected coverage approach (58 compared to 87). Given $\overline{P} = 0.95$ and $k = 1$, the threshold approach chooses the site with the maximum number of species rated as "confident." The expected value approach chooses a site with many species rated as "probable" as well as "confident." For $k = 2$ and $\overline{P} = 0.95$, the objective functions of the two approaches were more similar because a species rated as "probable" ($p_{ij} = 0.80$) at two sites could pass the threshold with a joint occurrence probability of $P_i = 0.96$ if both sites were selected. This fact also explains the large increase in the threshold value between $k = 1$ and 2. As the number of sites increases, the scores of the two approaches tended to converge. At $k = 10$, the expected coverage values were almost identical for the two approaches.

In the last column of table 1, we showed the percentage of sites chosen in common between the two approaches. For $k = 1$ and 2, no sites were chosen in common. For $k > 2$, the majority (but never all) of the sites were chosen in common by the two approaches.

In table 2, we compared the results using the threshold and expected coverage approaches for the same set of 20 sites using the alternative set of probabilities (0.95, 0.6, 0.4, 0). As in table 1, with a large number of sites selected, the scores of the two approaches were similar. With a small number of sites selected, however, the pattern was more erratic. For $k = 2$, both approaches selected the same two sites. This result is somewhat surprising because, with $k = 2$, only those species that were rated as "confident" were counted as covered under the threshold approach while "probable" ($p_{ij} = 0.6$) and "possible" ($p_{ij} = 0.4$) categories could factor in heavily under the expected coverage approach. There was little or no similarity in sites selected and relatively large differences in scores for $k = 1$ and 4. At $k = 4$, species rated only as "probable" could be counted as covered, which they could not be for $k < 4$. Hence, there was a premium for selecting four sites that all rate a species as "probable" under the threshold approach but not under the expected coverage approach. There was a large jump in the threshold value in going from $k = 3$ to 4, similar to the jump from $k = 1$ to 2 in table 1. These jumps occurred because

Table 2
Comparing threshold and expected coverage approaches using data set 1 and alternative case probabilities (0.95, 0.6, 0.4, 0) with threshold probability of 0.95.

Number of sites selected	Expected coverage approach		Threshold approach		Percent of shared selected sites
	Expected coverage	Threshold value	Expected coverage	Threshold value	
1	179.30	58	164.25	87	0
2	251.68	113	251.68	113	100.0
3	293.74	119	290.29	123	66.7
4	317.61	188	286.20	223	25.0
5	331.90	228	320.42	253	80.0
6	340.35	255	324.45	280	66.7
7	347.22	269	338.11	287	57.1
8	351.13	285	349.30	296	87.5
9	354.32	293	351.70	308	66.7
10	357.09	311	354.58	317	80.0

Table 3
Comparing threshold and expected coverage approaches using data set 2 and base case probabilities (0.95, 0.8, 0.1, 0) with threshold probability of 0.95.

Number of sites selected	Expected coverage approach		Threshold approach		Percent of shared selected sites
	Expected coverage	Threshold value	Expected coverage	Threshold value	
1	192.50	82	182.45	87	0
2	273.20	179	223.28	220	0
3	313.72	239	300.52	261	33.3
4	329.74	281	321.63	286	75.0
5	338.18	303	333.63	306	80.0
6	343.90	314	338.78	327	66.7
7	348.89	327	346.72	335	42.9
8	351.80	331	348.86	339	75.0
9	354.58	337	350.87	342	77.8
10	356.34	340	352.77	344	80.0

Table 4
Comparing threshold and expected coverage approaches using data set 2 and alternative case probabilities (0.95, 0.6, 0.4, 0) with threshold probability of 0.95.

Number of sites selected	Expected coverage approach		Threshold approach		Percent of shared selected sites
	Expected coverage	Threshold value	Expected coverage	Threshold value	
1	169.70	82	162.05	87	0
2	246.82	119	240.99	127	50.0
3	288.97	121	288.62	149	33.3
4	311.79	198	240.24	218	0
5	325.24	237	311.05	258	60.0
6	334.24	252	313.33	274	50.0
7	341.48	266	327.03	284	57.1
8	347.04	282	336.73	293	50.0
9	350.34	294	338.51	302	55.6
10	352.79	306	348.00	310	70.0

of the discrete character of our probability data, where all species within a class have the same probability. With more refined probability estimates we expect that the jumps would disappear. It is interesting to note that under the threshold approach the expected coverage actually falls in going from $k = 3$ to 4.

In tables 3 and 4 we reported the results comparing the expected coverage and threshold approaches for data set 2, analogous to tables 1 and 2. Though some details differed, similar patterns emerged in tables 3 and 4 as were found in tables 1 and 2, indicating that the patterns are more general and not a function of the particular set of 20 sites in the data set. However, it should be noted that all of our results are generated using data on terrestrial vertebrates in Oregon and do not constitute proof that the patterns we found generalize. As before, the scores of the two methods tended to be more similar as the number of sites increased. There were big increases in the threshold value from $k = 1$ to 2 in table 3

Table 5
Means and standard deviations of numbers of species covered by sets of reserved sites using expected coverage and threshold approaches for data set 1, base case and alternative probabilities and threshold probability of 0.95.

Number of sites selected	Base case probabilities (0.95, 0.8, 0.1, 0)		Alternative probabilities (0.95, 0.6, 0.4, 0)	
	Expected coverage	Threshold	Expected coverage	Threshold
1	209.05	184.93	179.30	164.24
	6.02	5.03	7.32	6.10
2	277.88	261.28	251.66	251.66
	5.29	4.18	6.76	6.76
3	311.65	299.65	293.77	290.25
	4.72	3.81	5.86	6.19
4	330.85	326.24	317.62	286.33
	3.88	3.71	5.32	4.25
5	340.66	331.92	331.88	320.43
	3.54	3.25	4.73	4.37
6	348.00	341.41	340.33	324.49
	3.37	3.03	4.45	3.99
7	352.56	348.77	347.19	338.13
	3.13	2.92	4.08	3.85
8	355.62	351.06	351.10	349.30
	3.00	2.67	3.68	3.59
9	357.23	353.93	354.32	351.73
	2.80	2.58	3.55	3.39
10	358.65	358.11	357.12	354.61
	2.65	2.55	3.27	3.12

using the base case probabilities and from $k = 3$ to 4 in table 4 using the alternative probabilities. In table 4, as in table 2, there was a decrease in the expected coverage under the threshold approach in moving from $k = 3$ to 4. At $k = 4$, no sites were chosen in common between the two approaches.

The results of the Monte Carlo simulations are shown in table 5, which shows the means (top line in each row) and standard deviations (bottom line in each row) of numbers of species covered by optimal sets of reserved sites obtained using the expected coverage and threshold approaches. Because the results were qualitatively similar between data sets 1 and 2, we only report results for data set 1. The histograms of numbers of species covered by optimal sets of reserved sites obtained using the expected coverage and threshold approaches and $k = 1$, 5, and 10 are shown in figures 1–3.

In all cases, the mean value for the expected coverage approach was at least as great as the mean value for the threshold approach. The standard deviations for the two approaches were similar, slightly smaller for the threshold approach than for the expected coverage approach in almost all cases, and generally decreasing as k increased. Because the mean was higher and the standard deviation was similar for each pair, the probability distribution of species covered under the expected coverage approach is shifted towards higher levels of coverage. The shift between the distributions is larger for small k; in figure 3 for $k = 10$, the shift is small and there is a good deal of overlap between the two distributions while in figure 1 for $k = 1$, the shift between the two distributions is far greater and the overlap is less. These results suggest that there is a cost to using the threshold ap-

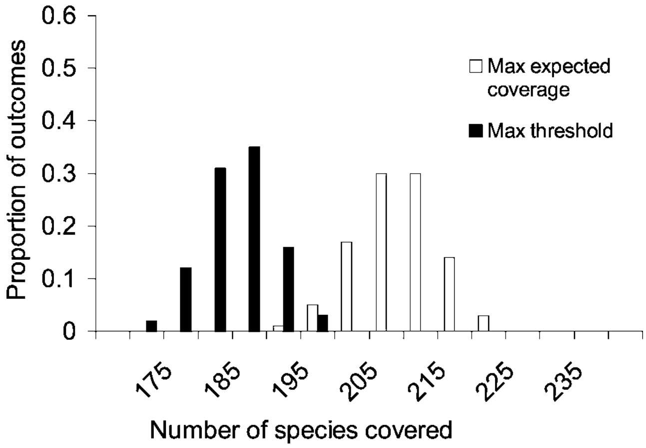

Figure 1. Number of species covered by optimal sets of reserved sites obtained using the expected coverage and threshold approaches and base case probabilities. Number of sites selected was 1.

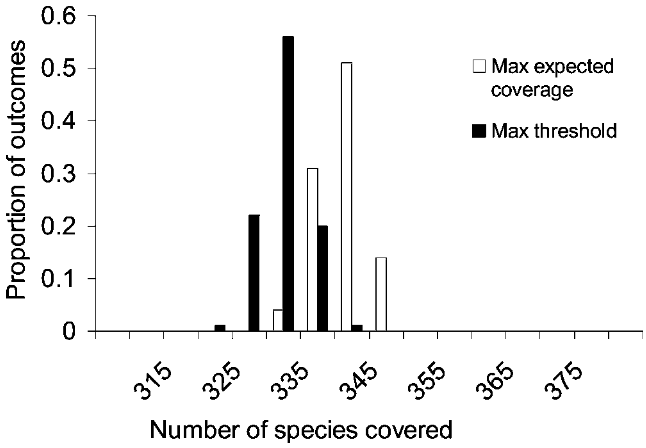

Figure 2. Number of species covered by optimal sets of reserved sites obtained using the expected coverage and threshold approaches and base case probabilities. Number of sites selected was 5.

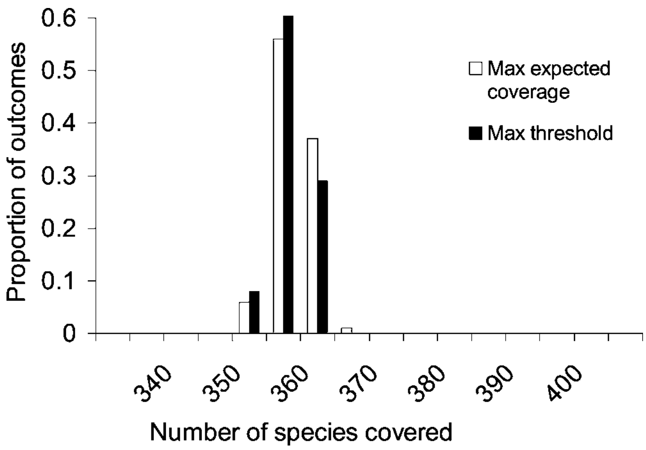

Figure 3. Number of species covered by optimal sets of reserved sites obtained using the expected coverage and threshold approaches and base case probabilities. Number of sites selected was 10.

Table 6
Results using alternative threshold probabilities using data set 1 and alternative case probabilities (0.95, 0.6, 0.4, 0).

k	Expected coverage approach	Threshold approach with $\overline{P} = 0.9$		Threshold approach with $\overline{P} = 0.8$		Threshold approach with $\overline{P} = 0.7$	
		Expected coverage	Threshold value	Expected coverage	Threshold value	Expected coverage	Threshold value
1	179.30	164.25	87	164.25	87	164.25	87
2	251.68	251.68	113	244.33	227	239.82	238
3	293.74	273.49	230	282.86	267	293.74	279
4	317.61	308.38	263	313.20	300	313.20	307
5	331.90	320.42	285	322.31	319	328.11	326
6	340.35	326.76	301	335.42	332	336.50	339
7	347.22	331.02	315	345.49	339	345.49	346
8	351.13	343.59	325	351.02	342	349.30	350
9	354.32	352.65	331	354.25	345	354.25	352
10	357.09	353.73	335	357.09	348	357.09	354

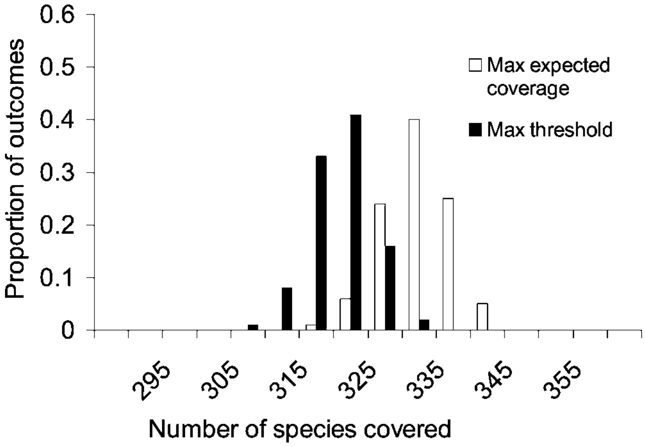

Figure 4. Number of species covered by optimal sets of reserved sites obtained using the expected coverage and threshold approaches and alternative case probabilities. Number of sites selected was 5. The threshold probability was 0.95.

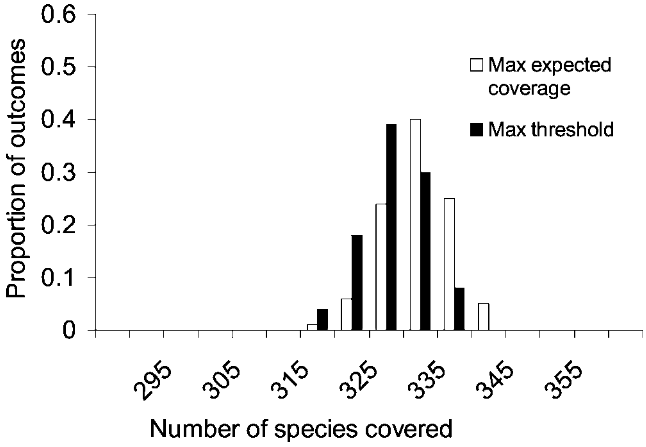

Figure 5. Number of species covered by optimal sets of reserved sites obtained using the expected coverage and threshold approaches and alternative case probabilities. Number of sites selected was 5. The threshold probability was 0.70.

proach; on average, actual species coverage will be lower using the sites selected by the threshold approach than it would be using the sites selected by the expected coverage approach. But this cost becomes smaller as the number of selected sites increases.

To this point, we used a threshold probability of $\overline{P} = 0.95$. The threshold probability, however, can be set at varying levels to represent different levels of willingness by the conservation agency to accept risk. Lowering the threshold probability allows more species the potential to contribute to the objective function at each site. To investigate the effect of altering the threshold probability, we solved the problem using the threshold approach for data set 1 and the alternative case probabilities using different threshold probabilities: $\overline{P} = 0.9, 0.8$ and 0.7. The results for different threshold probabilities for $k = 1$–10 are shown in table 6. As the threshold probability falls, it is easier to count species as covered, which is shown by the results in table 6. What is of greater interest is how the actual number of species covered is likely to change under the threshold approach for different values of the threshold. With one exception (for $k = 2$), the lower values of the threshold probability ($\overline{P} = 0.7$ or 0.8) yielded higher expected coverage than did the higher threshold probability ($\overline{P} = 0.95$, as reported in table 2, or $\overline{P} = 0.90$ in table 6). There is no reason to expect this result to hold generally. The best threshold probability, in terms of choosing sites that have the highest expected coverage, depends on the interplay between the species occurrence probabilities, the threshold probability, and the number of sites that may be selected; together these determine whether certain combinations of species occurrence exceed the threshold or not.

In figures 4 and 5 are shown results for two pairs of Monte Carlo simulations for $k = 5$ using different threshold probalities, $\overline{P} = 0.95$ and 0.7, respectively, for data set 1 and the alternative case probabilities. As illustrated, the shift in probability mass towards higher levels of coverage for the expected coverage approach over the threshold approach is smaller for lower $\overline{P}$. This suggests that the advantage of using the expected coverage approach over the threshold approach may decrease as the conservation agency is willing to recognize the value of less certain outcomes.

6. Discussion

The results reported in the previous section using data on terrestrial vertebrates in Oregon show that the expected coverage approach and the threshold approach select different sets of sites to compose a reserve network. As shown in the Monte Carlo simulations using this data, the expected coverage approach generally selected a set of sites that yields greater numbers of species actually covered than does the threshold approach. In fact, the expected coverage approach appeared to stochastically dominate the threshold approach in the sense that the probability distribution is shifted to higher levels of actual number of species in the simulations. Given advances in the ability to solve the expected coverage approach for optimal or near optimal solutions in large problems (Camm et al. [15]), these results suggest that the expected coverage approach appears preferable to the threshold approach for choosing reserve sites to cover the maximum number of species.

The two approaches differ not only in the mean number of species actually covered but also in the attributes of those species. The threshold approach counts the number of species that are fairly certain to be present, with the threshold serving as the standard for certainty. In contrast, with the expected coverage approach, any nonzero occurrence probability in a selected site contributes to the objective function. To see the difference in attributes of species chosen by the two approaches, consider the solutions in table 1 for $k = 1$. The site selected using the expected coverage approach has 274 species present with nonzero probability, with 247 rated "probable" or "confident", but relatively few rated as "confident" (58 species). The site selected using the threshold approach has only 228 species present with nonzero probability, of which 213 are rated "probable" or "confident", but has 87 species rated "confident". In other words, the threshold approach chooses the reserve site that maximizes the number of species meeting some minimum risk standard but contains fewer species with positive probabilities lower than the threshold probability and, consequently, a lower expected number of species covered.

It may be that there is a subset of species for which a high degree of confidence is desirable, perhaps because they have been listed as threatened and endangered under the Endangered Species Act or are recognized to fill particularly important ecological roles. Both the expected coverage and the threshold approaches can be modified to accommodate such a circumstance. One method would be to add a set of constraints that requires those species to be covered at specified minimum occurrence probabilities. Haight et al. [30] formulated the threshold approach with such a set of constraints for a subset of vegetation communities covered in Research Natural Areas in the Superior National Forest. Alternatively, high priority species can be assigned greater weight so that they contribute more to the value of the objective function when they are covered than other species, thereby increasing the likelihood that they will be covered in the set of selected sites.

In this paper, we have assumed that the objective is to maximize the number of species covered by a reserve system. Because of uncertainty, however, the actual number of species covered by selected reserve sites is unknown at the time sites are selected. Therefore, we have compared different site selections on the basis of the mean number of species covered. A reasonable alternative objective function might factor in variance in coverage as well as the mean. For example, a decision-maker might prefer having 100 species covered for sure rather than being subject to a gamble with a 50% probability of getting 90 species and a 50% probability of getting 110 species. Risk aversion in the number of covered species would arise with a concave objective function in the number of species covered, i.e., with declining marginal benefit of covered species. Introducing a concave function in number species covered introduces further non-linearity in the problem making optimal solutions more difficult to find. According to the Monte Carlo simulation results, however, the introduction of risk aversion appears unlikely to lead to a change in the preference between the two approaches compared here because the expected coverage approach stochastically dominated the threshold approach in the simulations.

While the reserve site selection problem of the type considered in this paper can yield important insights, it abstracts from a number of important conservation issues. Other potential extensions to the reserve site selection problem include the following:

(1) One could specify the resource constraint as a budget limit rather than as a limit on the number or area of selected sites (Ando et al. [8], Polasky et al. [35]). This would allow the problem to identify cost-effective reserve systems.

(2) The quality of the solutions to the reserve site selection problem using probabilistic data depends on the quality of the probabilistic data. The reliability of estimated occurrence probabilities will improve as methodologies for estimating them are developed and refined (again, see Austin et al. [9], Margules and Nicholls [10], Margules and Stein [11], Nicholls [12], and Margules and Austin [13]). In addition, the simplifying assumption that the occurrence probabilities are independent across species and sites is unrealistic. For example, it is likely that the occurrence probability for a species at a site is influenced by the presence of predator or prey species at that site or by proximity to other sites at which the species occurs.

(3) As currently defined, presence of a species at a site does not necessarily imply the long run viability of the species at the site. Another way in which a probabilistic approach is important is to consider species survival probabilities as a function of the reserve network chosen. Dealing with survival probabilities rather than occurrence probabilities is a much more complex problem. Its formulation would necessarily include assumptions about the continued presence or absence of species

and habitat outside the reserve network, possible links between populations and different sites, and interactions between species within the network (e.g., keystone species or predator/prey relationships).

The reserve site selection problem is still evolving and the models are simple relative to complex real world conservation problems to which they are applied. Incorporating uncertainty into the reserve site selection problem is, however, an important step toward increasing the realism and relevance of the analysis. Methods to solve the probabilistic reserve site selection problem can be used to provide useful information to conservation agencies and others involved in making conservation decisions. Used appropriately, they can help guide decision-makers as they attempt to understand how best to use limited resources to achieve conservation objectives.

References

[1] R.F. Noss and A.Y. Cooperrider, *Saving Nature's Legacy* (Island Press, Washington, DC, 1994).

[2] S.L. Pimm and J.H. Lawton, Planning for biodiversity, Science 279 (1998) 2068–2069.

[3] R.L. Church and C.S. ReVelle, The maximal covering location problem, Papers of the Regional Science Association 32 (1974) 101–118.

[4] R.L. Church, D.M. Stoms and F.W. Davis, Reserve selection as a maximal coverage location problem, Biological Conservation 76 (1996) 105–112.

[5] A.R. Kiester, J.M. Scott, B. Csuti, R.F. Noss, B. Butterfield, K. Sahr and D. White, Conservation prioritization using GAP data, Conservation Biology 10(5) (1996) 1332–1342.

[6] B. Csuti, S. Polasky, P.H. Williams, R.L. Pressey, J.D. Camm, M. Kershaw, A.R. Kiester, B. Downs, R. Hamilton, M. Huso and K. Sahr, A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon, Biological Conservation 80 (1997) 83–97.

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

[8] A. Ando, J. Camm, S. Polasky and A. Solow, Species distributions, land values, and efficient conservation, Science 279 (1998) 2126–2128.

[9] M.P. Austin, R.B. Cunningham and R.M. Fleming, New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures, Vegetatio 55 (1984) 11–27.

[10] C.R. Margules and A.O. Nicholls, Assessing the conservation value of remnant habitat "islands": mallee patches on the western Eyre Peninsula, South Australia, in: *Nature Conservation: The Role of Remnants of Native Vegetation*, eds. D.A. Saunders, G.W. Arnold, A.A. Burbidge and A.J.M. Hopkins (Surrey Beatty and Sons, Sydney, 1987) pp. 89–102.

[11] C.R. Margules and J.L. Stein, Patterns in the distribution of species and the selection of nature reserves: an example from Eucalyptus forests in South-Eastern New South Wales, Biological Conservation 50 (1989) 219–238.

[12] A.O. Nicholls, How to make biological surveys go further with generalized linear models, Biological Conservation 50 (1989) 51–75.

[13] C.R. Margules and M.P. Austin, Biological models for monitoring species decline: the construction and use of data bases, Philosophical Transactions of the Royal Society of London Biological Sciences 345 (1994) 69–75.

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

[15] J.D. Camm, S.K. Norman, S. Polasky and A. Solow, Nature reserve selection to maximize expected species covered, Operations Research (2001), in press.

[16] M.S. Daskin, A maximum expected covering location model: formulation, properties and heuristic solution, Transportation Science 17 (1983) 48–70.

[17] M.S. Daskin, K. Hogan and C.S. ReVelle, Integration of multiple, excess, backup and expected covering models, Environment and Planning B 15 (1988) 15–35.

[18] C.-M. Liu and A.-H. Wang, Solving location-allocation problems with rectilinear distances by simulated annealing, Journal of the Operational Research Society 45 (1994) 1304–1315.

[19] A.T. Murray and R.L. Church, Applying simulated annealing to location-planning models, Journal of Heuristics 2 (1996) 31–53.

[20] G. Righini, Annealing algorithms for multisource absolute location problems on graphs, Computational Optimization and Applications 7 (1997) 325–337.

[21] A.T. Ernst and M. Krishnamoorthy, Solution algorithms for the capacitated single allocation hub location problem, Annals of Operations Research 86 (1999) 141–159.

[22] C.M. Hosage and M.F. Goodchild, Discrete space location-allocation solutions from genetic algorithms, Annals of Operations Research 6 (1986) 35–46.

[23] C.R. Houck, J.A. Joines and M.G. Kay, Comparison of genetic algorithms, random restart and two-opt switching for solving large location-allocation problems, Computers & Operations Research 23 (1996) 587–596.

[24] A. Kumar, Y. Gupta, M. Gupta and C. Sundaram, A genetic algorithm-based approach to cell composition and layout design problems, International Journal of Production Research 34 (1996) 447–482.

[25] F.A.W. George, Hybrid genetic algorithms with immunisation to optimise networks of retail outlets, Studies in Locational Analysis 8 (1996) 59–80.

[26] T.G. Crainic, M. Toulouse and M. Gendreau, Synchronous tabu search parallelization strategies for multicommodity location-allocation with balancing requirements, Operations Research Spektrum 17 (1995) 113–123.

[27] E. Rolland, D.A. Schilling and J.R. Current, An efficient tabu search procedure for the p-median problem, European Journal of Operational Research 96 (1997) 329–342.

[28] M. Ohlemuller, Tabu search for large location-allocation problems, Journal of the Operational Research Society 48 (1997) 745–750.

[29] B. Gendron, J.-Y. Potvin and P. Soriano, Tabu search with exact neighbour evaluation for multicommodity location with balancing requirements, INFOR 37 (1999) 255–270.

[30] R.G. Haight, C.S. ReVelle and S.A. Snyder, An integer optimization approach to the probabilistic reserve site selection problem, Operations Research 48(5) (2000) 697–708.

[31] S.V. Ciriacy-Wantrup, *Resource Conservation: Economics and Policies*, 3rd ed. (University of California Press, Berkeley, CA, 1968).

[32] R.C. Bishop, Endangered species and uncertainty: the economics of the safe minimum standard, American Journal of Agricultural Economics 60 (1978) 10–18.

[33] L. Master, N. Clupper, E. Gaines, C. Bogert, R. Solomon and M. Ormes, *Biodiversity Research Consortium Species Database Manual* (The Nature Conservancy, Boston, MA, 1995).

[34] A.M. Law and W.D. Kelton, *Simulation Modeling and Analysis* (McGraw-Hill, New York, 1982).

[35] S. Polasky, J.D. Camm and B. Garber-Yonts, Selecting biological reserves cost effectively: an application to terrestrial vertebrate conservation in Oregon, Land Economics 77(1) (2001) 68–78.