Environmental Modeling and Assessment **9:** 179–187, 2004.

A scenario optimization model for dynamic reserve site selection

Stephanie A. Snyder [a], Robert G. Haight [a] and Charles S. ReVelle [b]

[a] *USDA Forest Service, North Central Research Station, 1992 Folwell Ave., St. Paul, MN 55108, USA*
E-mail: stephaniesnyder@fs.fed.us

[b] *Johns Hopkins University, Department of Geography and Environmental Engineering, 313 Ames Hall, Baltimore, MD 21218, USA*

Conservation planners are called upon to make choices and trade-offs about the preservation of natural areas for the protection of species in the face of development pressures. We addressed the problem of selecting sites for protection over time with the objective of maximizing species representation, with uncertainty about future site development, and with periodic constraints on the number of sites that can be selected. We developed a 0–1, linear optimization model with 2 periods to select the sites that maximize expected species coverage subject to budget constraints. The model is based on the idea that development uncertainty can be characterized by a set of scenarios, each of which is a possible second-period development outcome for the set of sites. We also suggest that our 2-period model can be used in a sequential fashion that is consistent with adaptive planning. Results are presented for the Fox River watershed in Chicago.

Keywords: scenario optimization, robust optimization, reserve site selection, integer programming, incomplete information, maximal covering problem, adaptive management

1. Introduction

Metropolitan areas in the United States are experiencing escalating rates of land conversion from open space to developed uses [1] with concomitant reductions in the amount and quality of natural areas and the species they support [2]. The establishment of a system of protected natural areas is one means of reducing the loss of biodiversity associated with land conversion and development [3,4]. Because conservation planners are constrained by land costs, they must make wise selections of parcels to create a system of reserves that yields the most conservation protection under a given budget [5]. Quantitative models for reserve site selection can inform decision makers about the consequences of their choices [6].

A variety of tools have been developed to help planners determine efficient site selection strategies and understand the goal trade-offs, and excellent reviews of reserve design goals and modeling techniques are available [6–8]. A commonly expressed goal is maximizing species representation, where a species is represented or covered if it is present in one or more protected sites. This maximal species covering problem, formulated as a linear-integer programming model, has counterparts in the location science literature [9] and has been widely applied in conservation planning [5,10–17]. These applications assume that the land protection decisions are made all at once. In practice, however, the decisions take place sequentially as funds and political support become available. Further, land availability is dynamic: sites currently available may be developed if protection is delayed, or sites not immediately available may be open for protection later. Methods are needed to address this sequential site selection problem and account for the uncertainties in budget and site availability.

We formulated a 2-period site selection model that maximizes the expected number of species represented while accounting for uncertainty in site development between periods and constraints on the number of sites selected. The model is based on the idea that development uncertainty can be characterized by a set of scenarios, each of which is a possible second-period development outcome for the set of sites. Those scenarios, together with estimates of their probabilities of occurrence, are incorporated into an optimization model to determine the best set of sites to protect now and the best set of sites to protect in the second period under each scenario. The formulation is a linear-integer program applying logic from robust or scenario optimization [18–20].

Researchers are just beginning to address multi-period site selection problems with uncertainty in site development. Costello and Polasky [21] used a dynamic programming algorithm to solve problems with up to 10 candidate sites and 6 periods. Because computational burden increased exponentially with the number of sites, they investigated the performance of simple heuristics and found that sequential application of a 1-period look-ahead heuristic provided solutions almost as good as those from the multi-period algorithm. Our scenario optimization model limits the time dimension to 2 decision periods and employs a linear-integer formulation, which allows solution of problems with hundreds of sites using commercial software. Solving these large problems can still be computationally intensive because the number of possible development scenarios increases exponentially with the number of sites. Therefore, we investigated problems in which development uncertainty was approximated with a limited number of scenarios.

We first present the optimization model and then describe its application to a problem of acquiring natural areas for protection in the Fox River watershed in the Chicago metropolitan area. The Chicago area is the third largest metropolitan region in the United States. The region experienced rapid population growth and land conversion in the 1990s [22], and the size of the metropolitan area could double in the next 30 years [23]. In response, county forest preserve districts evaluate and acquire privately-owned open space for protection [24]. Our application used 146 sites containing 116 rare species and analyzed protection strategies when the budget was limited and development was uncertain.

2. Methods

2.1. Site selection model

To address the planner's problem, we formulated a 2-period, 0–1 integer optimization model to select the set of sites that maximizes the expected number of species conserved at the end of the second period subject to an upper bound on the number of sites that can be protected each period. The model assumes that we have a list of sites, each of which is available for protection at the beginning of the first period. If a site is not selected for protection at the beginning of the first period, there is a probability that it becomes developed and unavailable for protection at the beginning of the second period. The model also assumes that we have a list of the species present in each site and that a species is conserved if it is present in at least one site selected for protection.

To handle uncertainty about the development of unprotected sites, we created a set of scenarios of site development. Each scenario is a list of the 146 sites in the Fox River watershed identifying whether each is developed or undeveloped in period 2 and represents one possible development outcome. Associated with each scenario is a probability of occurrence.

The model has 2 sets of 0–1 site selection decision variables. The first set of decision variables are the yes–no choices for the sites to be selected for protection in the first period. The model assumes that the protection decisions in the second period are made after the decisions in the first period are implemented and the site development scenario is revealed. The second set of decision variables are the yes-no choices on sites that are to be selected for protection in the second period under each of the development scenarios. The model includes logic from scenario optimization problems [20] and maximal covering problems in reserve selection science [9] and is expressed with the following notation:

i, I = index and set of species,

j, J = index and set of potential reserve sites,

s, S = index and set of site development scenarios,

p_s = probability that scenario s occurs,

b_1 = upper bound on number of reserve sites selected in period 1,

b_2 = upper bound on number of reserve sites selected in each scenario in period 2,

d_{js} = 0, 1 parameter; 1 if site j is undeveloped in period 2 scenario s, 0 otherwise,

N_i = set of sites j that contain species i,

x_{j1} = 0, 1 variable; 1 if site j is selected for protection in period 1, 0 otherwise,

x_{j2s} = 0, 1 variable; 1 if site j is selected for protection in period 2 scenario s, 0 otherwise,

y_{is} = 0, 1 variable; 1 if species i is represented in one or more of the parcels chosen for protection in periods 1 or 2, given that scenario s occurs, 0 otherwise.

The model was formulated as follows:

$$\text{Maximize} \sum_{s \in S} \left(p_s \sum_{i \in I} y_{is} \right), \tag{1}$$

subject to:

$$x_{j1} + x_{j2s} \leqslant 1 \qquad \text{for all } j \in J \text{ and } s \in S, \tag{2}$$

$$x_{j2s} \leqslant d_{js} \qquad \text{for all } j \in J \text{ and } s \in S, \tag{3}$$

$$\sum_{j \in J} x_{j1} \leqslant b_1, \tag{4}$$

$$\sum_{j \in J} x_{j2s} \leqslant b_2 \qquad \text{for all } s \in S, \tag{5}$$

$$y_{is} \leqslant \sum_{j \in N_i} (x_{j1} + x_{j2s}) \qquad \text{for all } i \in I \text{ and } s \in S, \tag{6}$$

$$x_{j1}, x_{j2s}, y_{is} \in \{0, 1\}. \tag{7}$$

The objective (1) maximizes the expected number of species represented by the set of selected sites in period 1 and by the selected sites in each scenario in period 2. Constraint (2) specifies that site j can at most be selected for protection in either period 1 or period 2, but not both, over all scenarios. Constraint (3) specifies that site j can only be selected for protection in period 2 in scenario s if site j is undeveloped in that scenario. Constraint (4) limits the number of sites selected for protection in period 1 to at most b_1. Constraint (5) limits the number of sites selected in each second period scenario to at most b_2. Constraint (6) defines the conditions under which a species i is protected. This constraint stipulates that in order for a species to be protected if scenario s occurs, at least one site that contains that species must be selected for protection either in the first period or in scenario s in the second period. The result is a set of sites for protection in period 1 and a set of sites for protection in period 2 under each scenario.

2.2. Study area and data

The study area was the Fox River watershed, which covers more than 4,000 km^2 in parts of ten counties in northeastern Illinois, USA (figure 1). The topography is flat with elevations of 150–300 m. The climate is continental with hot, humid summers, cold winters, and precipitation throughout the year. The watershed covers prairie, savanna, and woodland ecosystems and includes 1,389 plant and animal species, 44% of the species in Illinois. More than 100 of

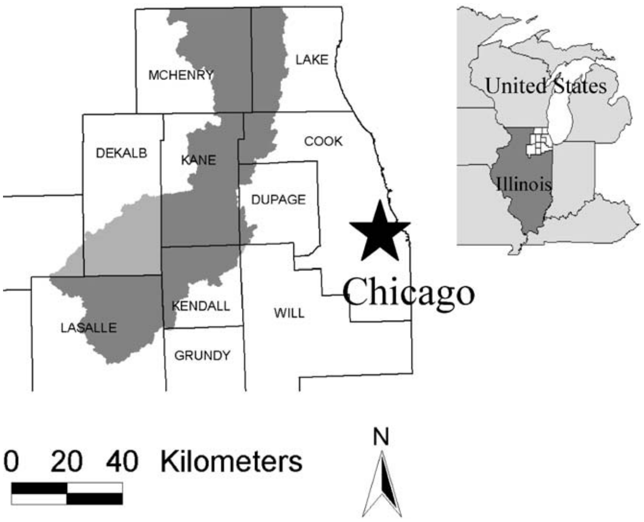

Figure 1. Fox River watershed in counties of northeastern Illinois, USA. The application used 146 sites in the dark shaded area covering 7 counties.

those species are listed as threatened or endangered in the state.

Another feature of the watershed is its proximity to the city of Chicago, which is located in Cook County on the shore of Lake Michigan (figure 1). The Chicago metropolitan area includes Cook County and nine surrounding counties. In the 2000 census, the region had more than eight million people, making it the third largest metropolitan region in the United States [22]. Of this eight million population, about 64% of the people lived in Cook County, and 36% lived in the nine surrounding suburban counties. While the regional population increased 11.6% in the period 1990–2000, the population growth in the nine suburban counties was more than twice the population growth in metropolitan Cook County.

In response to population growth and conversion of open space to housing and commercial development, county governments in the Chicago metropolitan area have acquired open space for a variety of public goals, including protecting the habitat of rare animals and plants and providing equitable public access to recreation and educational opportunities [24]. In 1995–2000, the voters in 6 counties approved bond referenda, backed by property tax increases, to finance more than $400 million of open space acquisition.

Our analysis was conducted using information from 365 natural areas in 7 counties of the Fox River watershed (figure 1). The information was obtained from the Fox River Watershed Biodiversity Inventory, which was conducted in the 1990s under the direction of Chicago Wilderness, a coalition of organizations dedicated to the survival of the natural ecosystems of the Chicago area. The Nature Conservancy shared the information with us. The natural areas were identified using a variety of criteria. Some sites contained high quality natural communities or habitat for rare animal or plant species. Other sites were significant open spaces that contained potentially restorable natural communities, special geological or archaeological features, rare species, or large grasslands. Each site was described by a list of rare plants and animals living in the site. Collectively, 116 rare species were found in 146 sites. We used this subset of 146 sites in our analysis. To demonstrate our model, we assumed that none of the natural areas were already protected and that each site had a 50% chance of development, although local government or planning agencies may well have specific knowledge of the probability of site development.

2.3. Generating scenarios

Solving the optimization model can be computationally intensive because the number of possible development scenarios increases exponentially with the number of sites. For example, if the development of any particular site is independent of the development of any other, then with 146 sites, each with a 50% probability of development, there are 2^{146} possible development scenarios, each with a 0.50^{146} probability of occurrence. Because even modest sized 0–1 integer programming problems can quickly become unwieldy and difficult to solve to optimality, we needed to explore ways to reduce the number of development scenarios in the model while adequately representing the development uncertainty.

We reduced the number of scenarios by randomly selecting a subset of n scenarios to include in the model. Each scenario d_{js}, $j = 1, \ldots, 146$ was a vector of 0–1 parameters where $d_{js} = 1$ meant that site j was undeveloped under scenario s. The value of each parameter d_{js} was determined by comparing a uniform 0–1 random number from a prime modulus multiplicative linear congruential generator [25, p. 227] with the probability of development (0.50). A single seed was used to generate the subset of scenarios. The probability of occurrence of each scenario was $1/n$.

2.4. Testing model performance

Because we included only a subset of possible development scenarios in the model, our primary research question was how many randomly selected scenarios provided an adequate representation of development uncertainty. To determine the impacts of increasing the number of scenarios, we formulated and solved two problems, one in which 2 parcels were selected in each period and the other in which 4 parcels were selected in each period. We solved 3 sets of 50 replicates of each problem, where each set of replicates included a unique set of 1, 10, or 100 scenarios. If the optimal first-period site selections showed little variation across a set of 50 replicates, then we would conclude that the number of scenarios used in that set adequately represented the site development outcomes.

We continued the analysis of model performance by exploring the computational impacts of incrementally increasing the upper bound on the number of sites selected each period from 1 to 34. Each problem maximized the expected number of species covered subject to an upper bound on the sites selected and included 100 randomly selected development scenarios. We solved 50 replicates of each problem, and a common set of random numbers was used to create the scenarios across the problems to allow for comparison.

Another test of model performance was a comparison of results from the 2-period site selection model with those from the sequential application of a 1-period maximal species coverage model that does not account for development uncertainty. In the latter case, the 1-period model was used to select sites in the first period, and then the 1-period model was used again in the second period under a given development scenario while accounting for the sites selected in the first period. This sequential procedure was repeated for each development scenario, and the objective function values were averaged to obtain the expected number of species covered. The difference in expected numbers of species covered obtained from the two types of models represented the gain from accounting for development uncertainty in the 2-period model. The gains were estimated for problems in which 1–10 sites were selected each period.

2.5. Estimating the cost of delaying site protection

To demonstrate how the model could be used in policy analysis, we estimated the cost of delaying protection when a total of 8 sites could be protected over 2 periods. We computed the optimal site selections for problems in which 0–8 sites could be selected in the first period. The differences in the expected numbers of species covered were estimates of the costs of delay. To examine the impacts of increasing the likelihood of development, we computed and compared the optimal site selections for problems in which the second period development probability was increased from 50–75% for every site. Each problem was solved using the same set of 100 randomly selected scenarios.

2.6. Solution method

All of the problems were solved on an IBM Pentium™4 personal computer, using the integrated solution package GAMS/OSL 2.25 [26], which was designed for large and complex linear and mixed integer programming problems. The input files were created using GAMS (General Algebraic Modeling System), a program designed to generate data files in a standard format that optimization programs can read and process. The models were solved using IBM's OSL (Optimization Subroutine Library), a Fortran-based subroutine library designed to solve optimization problems. The revised primal simplex algorithm, in conjunction with the branch and bound algorithm for integer-variable problems, were used to solve the models.

3. Results

3.1. Number of scenarios

The impacts of increasing the number of development scenarios in the 2-period site selection problem are listed in table 1. In the base case, we maximized the expected number of species covered while selecting 2 sites in each period. We found that increasing the number of scenarios used in the model reduced the variability in the first-period optimal solutions. From the 50 replicates with 1 scenario, 6 different pairs of sites were found in the first period solutions. With 10 scenarios, 3 different pairs of sites were selected for the first period with frequencies ranging from 26–42% of the replicates. With 100 scenarios, the same 3 pairs of sites were selected for the first period solutions, and one pair emerged 76% of the time: sites 52 and 146. Individually, site 52 was selected in 90% of the replicates and site 146 in 86% for the first period solution. The implication of finding a particular first-period solution 76% of the time is that this solution is likely to perform well as a solution to a problem in which uncertainty is fully represented.

Increasing the number of scenarios reduced the mean of the expected number of species covered (table 1). The mean decreased 4% from 62.44 species with one scenario to 59.93 with 100 scenarios. We expected this decrease because including more scenarios in the model increased the uncertainty about site development. With one scenario, the best set of sites in the first and second periods was selected for a single development outcome known with certainty. With 100 scenarios, the best set of sites in the first period had to be determined with 100 possible outcomes of site development in the second period.

The mean solution time and the number of branch and bound nodes increased with the number of scenarios in the model (table 1). With up to 10 scenarios, problems were solved in less than 5 seconds, and most solutions required no branch and bound nodes. With 100 scenarios, the mean solution time for 50 replicates was 1,675 seconds with an average of 6,002 nodes. About 20% of the replicates with 100 scenarios had very long solution times: up to 22,000 seconds and 81,000 nodes. The other replicates solved in less than 300 seconds with no nodes. We hypothesize that the reason for the inordinate amount of branch and bound in these problems with very few sites being chosen is that a large number of alternate optima exist in these problems. Branch and bound is known in cases of alternate optima to take very many nodes of exploration since all alternate optima need to be generated in order to declare an optimal solution.

We repeated the analysis using a problem in which 4 sites could be selected in each period and found the same pattern of results (table 1). While 34 different solutions were found in 50 replicates of the problem with one scenario, one solution was found in all 50 replicates of the 100-scenario formulation (sites 6, 52, 11, and 146). As a result, a decision maker could be reasonably confident that this set of sites is the optimal first-period solution to a problem in which uncertainty is fully represented. Increasing the number of scenarios produced a slight decrease in the mean of expected species coverage. The mean solution time increased from less than two seconds for the replicates with up to 10 scenarios to 214 seconds for the replicates with 100 scenarios. Very few of the replicates required branch and bound nodes.

Table 1
The impacts of increasing the number of development scenarios in the 2-period site selection problem. Results are based on 50 replicates of each problem.

Scenarios	Expected number of species covered		Mean solution	Mean number of	Optimal first period solutions	
	Mean	Std. dev.	time (sec)	nodes	Sites	Freq. (%)
			Select 2 sites each period			
1	62.44	1.15	0.11	0.12	6, 52	28
					6, 134	18
					134, 146	16
					52, 134	14
					52, 146	12
					6, 146	12
10	60.43	0.88	2.50	6.00	52, 146	42
					6, 52	32
					6, 146	26
100	59.93	0.25	1675.00	6002.00	52, 146	76
					6, 52	14
					6, 146	10
			Select 4 sites each period			
1	83.36	1.32	0.09	0.00	34 solutions	
10	80.23	0.69	1.71	0.60	6, 52, 111, 146	74
					6, 52, 77, 146	8
					6, 52, 134, 146	4
					6, 8, 52, 111	4
					6, 8, 111, 146	2
					6, 8, 52, 98	2
					6, 52, 98, 146	2
					6, 8, 52, 146	2
					6, 77, 111, 146	2
100	79.54	0.19	213.83	0.18	6, 52, 111, 146	100

Based on these results and the relatively small variability in the first-period solutions over many replicates, we concluded that using 100 randomly selected site development scenarios in problems with up to 4 sites selected each period adequately represented all possible development outcomes for our data and problem specification. Further, we concluded that most problems with 100 scenarios could be solved in less than 5 minutes, but a few problems could take much longer. We therefore decided to use 100 scenarios in the remainder of the analyses discussed.

3.2. Number of sites selected

Increasing the number of sites selected each period increased the variability of the optimal solutions found in 50 replicates of each problem (table 2). Solutions to problems with up to 6 selected sites each period were relatively stable: fewer than 3 first-period solutions were found to each problem over all replicates, and one of those solutions was obtained most of the time. As the number of sites selected in each period increased beyond 6, the stability of solutions deteriorated: there were many first-period solutions to each problem, and none was dominant. The means of the expected species coverage increased with the number of sites selected in each period, and the standard deviations were less than 1% of the means. Solutions to problems with more than 4 sites selected were usually obtained without branch and bound nodes.

The results in table 2 show that sites selected in the first period depend on the subset of scenarios used. How will a first-period solution perform if a scenario other than one in the subset used to obtain that solution occurs? We addressed this question using the eight first-period solutions listed in table 2 for 8 sites selected. Recall that each of those solutions was obtained with a different set of 100 randomly selected scenarios. We forced each first-period solution into the optimization model and computed the best second-period solution and total number of species covered for each of 100 new randomly-generated scenarios. In terms of mean coverage, there was very little difference in the performance of the first-period solutions using new sets of randomly generated scenarios. Across the eight first-period solutions, the averages of the number of species covered ranged from 95.01 to 95.18. In terms of the range of coverage, there was no difference in the performance of the first-period solutions using new sets of randomly generated scenarios. For each first-period solution, the minimum and maximum number of species covered were 90 and 98, respectively, across the 100 new scenarios. These results suggest that each of the first-period solutions obtained with a subset of randomly selected scenarios is likely to perform well in terms of expected number of species covered in a problem in which uncertainty is fully represented.

It is interesting to note that the first-period solutions obtained with different sets of randomly selected scenarios had much overlap in terms of sites selected (table 2). For exam-

Table 2
The impacts of changing the number of sites selected each period in a 2-period problem. Results are based on 50 replicates of each problem.

Sites selected	Expected number of species covered		Optimal first period solutions	
	Mean	Std. dev.	Sites	Freq. (%)
1	38.47	0.18	146	100
2	59.93	0.25	52, 146	76
			6, 52	14
			6, 146	5
4	79.54	0.19	6, 52, 111, 146	100
6	89.54	0.16	6, 8, 52, 98, 111, 146	86
			6, 8, 52, 77, 111, 146	10
			6, 8, 52, 111, 134, 146	4
8	95.30	0.12	6, 8, 52, 77, 98, 111, 134, 146	68
			6, 8, 11, 52, 77, 98, 111, 146	8
			6, 8, 52, 98, 105, 111, 134, 146	6
			6, 8, 29, 52, 77, 98, 111, 146	4
			6, 8, 52, 57, 98, 111, 134, 146	4
			6, 8, 52, 77, 98, 111, 134, 146	4
			6, 8, 29, 52, 98, 111, 134, 146	4
			6, 8, 52, 57, 77, 98, 111, 146	2
10	100.35	0.09	6, 8, 11, 29, 52, 77, 98, 111, 134, 146	12
			17 other solutions	88
15	108.83	0.19	Many solutions	
20	111.52	0.15	Many solutions	
25	113.90	0.11	Many solutions	
30	115.41	0.05	Many solutions	
34	116.00	0.00	1, 4, 6, 8, 11, 17, 20, 22, 23, 25, 28, 29, 31, 32, 35, 52, 57, 64, 68, 69, 71, 75, 77, 85, 92, 93, 98, 100, 105, 111, 114, 118, 134, 146	

ple, six sites (6, 8, 52, 98, 111, and 146) appeared in solutions to all 50 replicates of the problem in which 8 sites were selected each period, and sites 134 and 77 occurred in solutions to 86% of the replicates. This overlap helps explain why any of the different first-period solutions are likely to perform well and suggests sites that are likely to be selected in the optimal solution to a problem in which uncertainty is fully represented.

The variability of the first period solutions over many replicates suggested that more than 100 randomly selected scenarios were needed to adequately represent uncertainty in site development in problems where many sites can be selected; e.g., choosing 10 or more each period. To see if we could increase the stability of the solutions, we solved 50 replicates of a problem in which 10 sites can be selected each period using 500 randomly selected scenarios. Compared to the formulation with 100 scenarios which produced 18 different solutions, only 10 different sets of sites were produced by the formulation with 500 scenarios. The 500-scenario problem produced a set of 8 sites (6, 8, 52, 77, 98, 111, 134, and 146) that appeared in the solutions to all 50 replicates. The mean objective function value from the 50 replicates of the 500-scenario problem was only 0.08% less than the mean obtained with the 100-scenario formulation, and the standard deviation was less than 1% of the mean. The average solution time increased from 152 seconds with 100 scenarios to 4,386 seconds with 500 scenarios, and none of the replicates required branch and bound nodes. From these results we concluded that increasing the number of scenarios beyond 500 would further reduce the number of solutions obtained in replicates of the problem and increase the number of sites common to those solutions. However, the value of increasing solution stability by increasing the number of scenarios must be weighed against the cost of the additional computational burden.

3.3. Gain from accounting for uncertainty

Results from the 2-period site selection model and the sequential application of a 1-period model are listed in table 3. The 2-period model was solved using 100 development scenarios, and the sequential application of the 1-period model was repeated using the same 100 scenarios. The gain in the expected number of species covered from using the 2-period model was up to 2% of the expected coverage using a 1-period model sequentially. In addition, the best set of sites to select in the first period was different when taking into consideration site availability in the second period. For example, when 2 sites were selected each period, the 2-period model selected sites 52 and 146 in period 1 while sites 6 and 146 were selected by the 1-period model. While sites 6 and 146 maximized species coverage given that no more sites could be protected, sites 52 and 146 were optimal given that a pair of available sites could be protected later. This result illustrates that the 2-period model was not simply making the most greedy choice in the first period regardless of what can be protected in the second period. While the gain in expected coverage from using a 2-period model was small in our case, the magnitude of the gain may be a function of the data, and greater gains may be found with different data.

3.4. Cost of delaying protection under different probabilities of development

The impacts of delaying site protection when a total of 8 sites can be protected over 2 periods are listed in table 4. If all 8 sites were protected in period 1, 84 species were covered. Expected coverage dropped with a decrease in the number of sites selected in period 1 because delaying protection made sites vulnerable to development. Although the reduction in the expected coverage was 18 species if no sites were protected in period 1, the reduction was fewer than 4 species if the delay involved fewer than 4 sites. The reduction in the expected coverage was small because many of the species were present in more than one site. If a particular site was developed, an alternative site was selected to compensate. With a 75% probability of development, the impacts of delaying site protection had the same trends as with a 50% probability of development (table 4), although losses in expected coverage were greater. Again, we might observe more dramatic changes in coverage with different data sets.

Table 3
Comparison of results from a 2-period site selection model that accounts for development uncertainty and the sequential application of a 1-period maximal covering model that does not account for development uncertainty.

Number of sites selected each period	Expected number of species covered		Optimal first period solution	
	2-period model	1-period model	2-period model	1-period model
1	38.37	38.37	146	146
2	59.57	59.22	52, 146	6, 146
4	80.18	78.95	6, 52, 111, 146	6, 52, 134, 146
6	89.53	88.63	6, 8, 52, 98, 111, 146	6, 52, 58, 111, 134, 146
8	95.18	95.18	6, 8, 52, 77, 98, 111, 134, 146	6, 8, 52, 77, 98, 111, 134, 146
10	100.21	100.06	6, 8, 29, 52, 57, 77, 98 111, 134, 146	4, 6, 8, 11, 52, 77, 98 77, 98 111, 134, 146

Table 4
The impacts of delaying site protection when 8 sites can be selected in 2 periods.

Number of sites selected		Expected number	Sites selected
Period 1	Period 2	of species covered	first period
		Probability of development = 0.50	
8	0	84.00	6, 8, 52, 77, 98, 111, 134, 146
6	2	82.12	6, 8, 52, 98, 111, 146
4	4	80.18	6, 52, 111, 146
2	6	74.16	6, 52
0	8	66.23	none
		Probability of development = 0.75	
8	0	84.00	6, 8, 52, 77, 98, 111, 134, 146
6	2	80.60	6, 8, 52, 111, 134, 146
4	4	76.90	6, 52, 111, 146
2	6	67.39	6, 146
0	8	52.48	none

4. Discussion

We addressed the problem of selecting sites for protection over time with the objective of maximizing species representation, with uncertainty about future site development, and with periodic constraints on the number of sites that can be selected. We showed that a 2-period problem could be formulated as a discrete, 0–1 integer optimization model using logic from the scenario optimization and reserve site selection literature [9,20]. This is important because the model can be solved using commercial software to determine the best set of sites to protect immediately and the best set of sites to protect next period depending on the observed development.

While the model incorporates development uncertainty using a probability distribution of development scenarios for the set of sites, the model can be computationally intensive because the number of possible scenarios increases exponentially with the number of sites. We reduced the number of scenarios by randomly selecting a subset to include in the model. Recognizing that a model with a subset of possible scenarios could bias the optimal solution, we investigated the impacts of the number of randomly selected scenarios on solutions to problems with 146 sites and 116 species. We found that the optimal solutions varied little across 50 replicates of a problem with 100 randomly selected development scenarios and an upper bound of 2 sites selected each period. However, the stability of the optimal solutions to problems with 100 scenarios decreased as the number of sites selected each period increased beyond 6. Despite their variability, solutions to the replicates of a given problem had almost the same expected coverage and contained many of the same sites, suggesting that any one of the solutions would perform well as a solution to the problem in which uncertainty in site development was fully represented. We also found a gain in the expected number of species covered when using the 2-period model versus using a 1-period, deterministic model solved sequentially. Although the gains were not extremely large, they nonetheless signify that a planner is able to make better near-term decisions by taking into consideration what the development landscape may look like in the near future. Finally, we found that when we delayed site protection for the case in which a total of 8 sites can be protected over 2 periods, the expected species' coverage dropped with the proportion of the sites selected in the second period. Thus, there is a premium to being able to protect sites sooner rather than later. As we wait to make protection decisions, sites become vulnerable to development and opportunities are lost. The same trend was found with development probabilities of both 50% and 75%, although, as would be expected, the drop in expected species' coverage was greater when the development threats were higher.

We found that our scenario optimization models did not take long to solve, despite their size. While problems with 146 sites, 116 species, and 100 development scenarios included 26,346 discrete 0/1 variables for site selection and species representation and 26,301 structural constraints, the majority of problems solved without any branch and bound nodes in a matter of seconds to a few minutes. We believe our dataset is representative in size of real-world reserve site selection applications, and that our model and solution approach would be applicable and readily able to handle many other reserve site applications.

We suggest that our 2-period formulation is consistent with the uncertain nature of future budget and site availability. These uncertainties are often too great to reliably specify site selection decisions for years in the future. Instead, results from a 2-period model inform managers about the impacts of their current site selection decisions taking into consideration near-term uncertainties, which can be specified relatively easily and accurately.

We also suggest that our 2-period model can be used in a sequential fashion that is consistent with adaptive planning. The model solution includes a set of sites to protect in the current period and another set of sites to protect in the next period for each of the development scenarios. Only the set of decisions in the current period would be implemented. Then, once the development outcome is revealed, the 2-period model can be used again to determine the best sites to protect under the new conditions while accounting for uncertainty in the subsequent period. This recursive procedure gives the decision maker a tool to deal with an uncertain future as it unfolds allowing adaptation to the vagaries of development. While we cannot say from our analysis how different or better the solutions from our model might be if we included a third or fourth period, we suggest that including more than 2 planning periods will provide diminishing returns. Three period models may be able to be constructed, although they may not be solvable by the scenario approach because the expansion of scenarios rapidly increases the dimensions of the problem. Limited useful information would be expected to result from three periods, however, because the first period's decisions are conceived of as the only decisions that would ever be implemented. Costello and Polasky [21] found that sequential application of a 2-period model with just one decision period performed almost as well as solving a full dynamic program to the end of the planning horizon all at once.

The scenario optimization model can be extended in at least four ways to enhance its applicability. First, we could add site costs as a function of the site size and quality. Then, a 2-period problem could be formulated to maximize the expected species' representation subject to periodic budget constraints. Next, sites could be included that are not available for protection in the first period, and the uncertainty about the availability of those sites for protection in the second period could be represented with scenarios. Then, a 2-period problem could be formulated to maximize the expected species' representation subject to periodic budget constraints and an allowance for using surplus first-period funding in the second period. The third extension would treat the second period funding as uncertain and represent this budget uncertainty with a set of scenarios. Then, a 2-period problem could be formulated to maximize the expected species' representation subject to periodic budget constraints that are conditioned on the budget scenarios. While each of these extensions can be formulated as a linear-integer program, their solvability cannot be predicted a priori. We recognize that replacing a constraint on the number of selected parcels with one restricting the total area or cost of selected parcels may create a more difficult 0–1 integer programming model to solve. Finally, a regret-based approach could also be developed in which sites are selected such that the average or maximum regret associated with sub-optimal choices is minimized. We are actively exploring these models.

Acknowledgement

The authors thank Jane Rullifson of the Department of Applied Economics at the University of Minnesota for providing the species' distribution data and maps used in this research. The authors also thank the Associate Editor and two anonymous reviewers for their helpful comments. The USDA Forest Service North Central Research Station supported this research.

References

[1] P.H. Gobster, R.G. Haight and D. Shriner, Landscape change in the Midwest: An integrated research and development program, Journal of Forestry 98(3) (2000) 9–14.

[2] L.E. Hollis and W. Fulton, *Open Space Protection: Conservation Meets Growth Management* (The Brookings Institution Center on Urban and Metropolitan Policy, Washington, DC, 2002.

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

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

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

[6] S.E. Kingsland, Creating a science of nature reserve design: perspectives from history, Environmental Modeling and Assessment 7 (2002) 61–69.

[7] R.L. Pressey, C.J. Humphries, C.R. Margules, R.I. Vane-Wright and P.H. Williams, Beyond opportunism: key principles for systematic reserve selection, Trends in Ecology and Evolution 8 (1993) 124–128.

[8] C.R. Margules and R.L. Pressey, Systematic conservation planning, Nature 405 (2000) 243–253.

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

[10] K.D. Cocks and I.A. Baird, Using mathematical programming to address the multiple reserve selection problem: an example from the Eyre Peninsula, South Australia, Biological Conservation 49 (1989) 113–130.

[11] J. Saetersdal, J.M. Line and H.J. Birks, How to maximize biological diversity in nature reserve selection: vascular plants and breeding birds in deciduous woodlands, western Norway, Biological Conservation 66 (1993) 131–138.

[12] J.D. Camm, S. Polasky, A. Solow and B. Csuti, A note on optimal algorithms for reserve site selection, Biological Conservation 78 (1996) 353–355.

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

[14] C.A. Willis, R. Lombard, B. Cowling, B. Heydenrych and C. Burgers, Reserve systems for limestone endemic flora of the Cape Lowland fynbos: iterative versus linear programming techniques, Biological Conservation 77 (1996) 3–62.

[15] J.C. Williams and C.S. ReVelle, Applying mathematical programming to reserve selection, Environmental Modeling and Assessment 2 (1997) 167–175.

[16] S.A. Snyder, L.E. Tyrrell and R.G. Haight, An optimization approach to selecting research natural areas in national forests, Forest Science 45 (1999) 458–469.

[17] 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 (2001) 68–78.

[18] S.A. Malcolm and S.A. Zenios, Robust optimization for power systems capacity expansion under uncertainty, Journal of the Operational Research Society 45 (1994) 1040–1049.

[19] J.M. Mulvey, R.J. Vanderbei and S.A. Zenios, Robust optimization of large-scale systems, Operations Research 43(2) (1995) 264–281.

[20] P. Kouvelis and G. Yu, *Robust Discrete Optimization and Its Applications* (Kluwer Academic Publishers, Netherlands, 1997).

[21] C. Costello and S. Polasky, Dynamic reserve site selection, Energy and Resource Economics (2003, in press).

[22] K.M. Johnson, The changing face of Chicago: demographic trends in the 1990s, Chicago Fed Letter #176, April, Federal Reserve Bank of Chicago, Chicago, IL (2002).

[23] Openlands Project, Under pressure: land consumption in the Chicago region 1998–2028, Openlands Project, Chicago (1999).

[24] J.A. Ruliffson, P.H. Gobster, R.G. Haight and F.R. Homans, Niches in the urban forest: organizations, goals, and cooperation in metropolitan open space acquisition, Journal of Forestry 100(6) (2002) 16–23.

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

[26] GAMS Development Corporation, *General Algebraic Modeling System. Version 2.25.090* (Washington, DC, 1990).