

From: [Stephanie DeMay](#)
To: [Becker, Drew N](#); [Reid, Rebekah N](#); [Marshall, Michael E](#)
Subject: [EXTERNAL] SMART analysis HENA thoughts
Date: Monday, June 18, 2018 5:49:31 PM
Attachments: [Smith-Theberge1987 Article EvaluatingNaturalAreasUsingMul.pdf](#)
Importance: High

Hi all,

After talking with Drew today, I did some reading up on SMART analysis and the HENA SSA and did some thinking. Below are my official thoughts on the issue at hand, that you are welcome to use in future discussions about this. Let me know if you have any questions, or if I have gotten wrong how the HENA analysis was done.

Steph

Based on what I have been able to read up about SMART analysis, I agree with Angela that SMART (Simple Multi Attribute Rating Technique) isn't the most appropriate tool to use for the HENA SSA task of identifying EOs that potentially could still harbor extant populations. I do understand the argument that it could be, or could easily be morphed into, a decision context. However, I believe that the analysis that was conducted is valid and justifiable, and that present concerns could be alleviated by changing how the analysis is presented, specifically, by not calling it a SMART analysis.

A large component of SMART has to do with the decision context, such as identifying the decision maker, determining objectives of the resulting decision, coming up with alternatives, etc. This is more than is needed for the task at hand, the simpler task of ranking a pre-defined set of habitat patches on their likelihood of supporting extant HENA based on multiple factors. The ranking method used in the SSA was "simple additive weighting". This is also considered a decision analysis tool, but at its core, it is a just ranking tool that lacks the decision context baggage that comes with SMART. With simple additive weighting, different attributes of an object of comparison are measured, these values are normalized and weighted based on some kind of expert opinion or other justification, and a weighted sum is calculated for each option so that multiple attributes are combined into a single score. This is a basic and intuitive tool that is applicable to a myriad of situations, including the present situation, and is not bound to being used only in decision support settings. It could be used to rank the best burgers in town based on size, texture, toppings, and price. This information could be used in a decision context for a hungry and indecisive biologist (who would also have the alternative of eating pizza or something else not a burger), but a decision context is not necessary. The analysis and results are valuable and relevant even if the only objective is to create a ranked list of burgers. This is a similar objective to the HENA analysis; the objective was to create a ranked list to identify the top most suitable habitat patches from a pre-defined set, and the simple additive weighting analysis that was conducted accomplished that objective.

Smith, P.G. and Theberge, J.B., 1987. Evaluating natural areas using multiple criteria: theory and practice. *Environmental management*, 11(4), pp.447-460.

<https://link.springer.com/content/pdf/10.1007/BF01867653.pdf>

Simple additive weighting specifically discussed on pages 454-455

Stephanie DeMay, PhD
Assistant Research Scientist
Natural Resources Institute, Texas A&M
stephanie.demay@ag.tamu.edu

PROFILE

Evaluating Natural Areas Using Multiple Criteria: Theory and Practice

PAUL G. R. SMITH
JOHN B. THEBERGE

Faculty of Environmental Studies
University of Waterloo Waterloo
Ontario N2L 3G1, Canada

ABSTRACT / Evaluating natural areas involves making measurements for a series of criteria and deciding which areas are most significant based on these measurements. We review the basic theory of measurement and its application to assessing criteria in 20 different conservation evaluation studies. Different types of models exist to assess the overall significance of natural areas based on multiple criteria. A

brief description is presented of these "multicriteria evaluation models" and the underlying theory. The assumptions of these models are outlined as are the types of information that are appropriate to each. We then examine how a final overall ranking is derived in the 20 studies. None explicitly stated the model used and few noted the assumptions involved. Based on theory and current practice, a number of principles are presented to guide the development of evaluation methods. We conclude that evaluation methods should not be allowed to cloud important issues or conceal value judgements. Furthermore, the assumptions of any evaluation method should be clearly stated and rationalized in terms of the data and areas being compared.

The evaluation of natural areas, or conservation evaluation, has become widespread in land use planning, environmental impact assessment, and planning for protected areas (van der Ploeg and Vlijm 1978, Margules and Usher 1981, Goldsmith 1983, Roome 1984, Usher 1986, Wathern and others 1986). Its purpose is to identify the most ecologically valuable natural areas so planning and management practices can be applied so as to maintain the areas' values. Many different criteria form the basis of the evaluations, and we review these criteria elsewhere (Smith and Theberge 1986a). Each of these different criteria must be measured and the measurements used collectively to rank natural areas. This overall assessment is often done by summing subjective scores assigned to different criteria, a practice increasingly criticized (Nilsson 1986, Ratcliffe 1986). Indeed, even the value of composite indices derived from empirical ecological measurements is being questioned (Gotmark and others 1986).

Many composite indices exist, such as those in economics [gross national product and consumer price indices (Eichhorn 1978)], in psychology [IQ and other test scores (Goldstein and Hersen 1984)], and in environmental management [air and water quality indices (Ott 1978)]. The construction of composite indices in all these fields required the development of a strong theoretical basis. No such theoretical basis has been established for the composite indices used in conservation evaluation.

Similarly, decisions based on multiple criteria are

KEY WORDS: Conservation; Evaluation; Criteria; Measurement; Natural areas

common in business and public administration, for example, the hiring of personnel, selection of facility sites, and granting of contracts. Multicriteria decision problems are also common in other fields of environmental management such as waste disposal site location and resource allocation (Keeney and Raiffa 1976, O'Bannion 1980, Bakus and others 1982, Janssen and Nijkamp 1985). A wealth of theory and practice has been used to develop "multicriteria evaluation models" and other methods to facilitate these decisions (Keeney and Raiffa 1976, Hwang and Yoon 1981). Simply put, these models integrate measurements of several separate criteria to arrive at a choice from among a series of alternatives. Indeed, all decisions involving more than one criterion use some sort of "multicriteria evaluation model" whether explicitly or implicitly.

Surprisingly, conservation evaluation has, for the most part, ignored the knowledge gained in these other disciplines in coping with similar issues. The purpose of this article, then, is to draw upon the theories and methods alluded to above for three purposes: (a) to describe how measurement and multicriteria decisions have been treated in the conservation evaluation literature, (b) to critique the methods used and clarify the assumptions and requirements of both the types of measurement and the multicriteria evaluation models used to evaluate natural areas, and (c) to suggest a number of principles to guide the choice and design of the most appropriate evaluation methods. Now long overdue, this synthesis of theory and methods from mathematics, management science, and psychology with the practice of conservation evaluation can only improve the quality of decisions that are made.

The Evaluation Systems Reviewed

To examine how criteria have been measured and what multicriteria models have been used in conservation evaluation, we examined 20 different studies (Table 1). These were selected to reflect the diversity of approaches, applications at different scales, different types of environments, and differences in political organization, as well as changes in methods over a 14-year period (1971–1985). As a result, local, provincial, state, national, and international efforts are represented, as are arctic, temperature, freshwater, terrestrial, wilderness, and human-modified environments (Table 1).

Table 1 briefly describes how each study measures criteria and how an overall ranking is achieved. Other aspects of these studies are described by Smith and Theberge (1986a). Collectively, these studies represent the “practice” of evaluation and will be drawn upon throughout the article.

Methods of Measurement

Theory: Measurement

To understand the nature of measurement, and thus how criteria are assessed, one must grasp three aspects of measurement theory. First, the four scales of measurement, outlined below, have properties that define which types of mathematical operation may be applied to the values. Secondly, measurement of the value of natural areas can involve either measuring actual environmental variables, such as the number of species present, or measuring human perception of an area's value. Finally, any measurements taken are subject to a number of forms of uncertainty, and these affect both how measurement is done and the confidence we can place in the values obtained.

The four scales of measurement. Four basic scales of measurement exist: nominal, ordinal, interval, and ratio (Steele and Torrie 1980, Ghiselli and others 1981, Mason and Swanson 1981). Table 2 summarizes the characteristics of the four scales and gives an example of each from the literature on conservation evaluation.

The nominal scale is qualitative. For example, a human may be male or female and these are mutually exclusive categories. As this scale is qualitative, few mathematical relations or operations can be applied to it (Table 2).

The ordinal scale allows the categories of the scale to be ordered or ranked. A partial ordinal scale consists of a series of categories, for example high, medium, and low, and several items are assigned to each

category. By contrast, in a complete ordinal scale, each item is assigned a unique rank relative to all other items. Arithmetic operations such as addition and multiplication cannot be performed on the categories of an ordinal scale (Table 2). Some consider the ordinal scale qualitative (for example, Voogd 1983:75), while others call it quantitative (Ghiselli and others 1981:12). For the purpose of this article, it will be considered qualitative.

In interval measurement, the numerical distance between the various “categories” of the scale can be specified. Thus, a real number can be assigned to each item. However, an interval scale has no origin or real zero point. Examples are altitude and time measurement, there is no absolute zero altitude or time and it must be arbitrarily assigned to a particular point in the scale. All normal arithmetic operations can be applied to variables measured with an interval scale (Table 2).

A ratio scale has the same characteristics as the interval scale except that it has a true origin or zero point. An example is mass: if an object has zero mass that is its mass under any conditions. Again, all the normal arithmetic operations can be applied to such numbers (Table 2). The last two scales, the interval and ratio are the only completely quantitative or cardinal scales (Mason and Swanson 1981).

Measuring environmental variables or subjective value. Measurement of criteria may consist of simply measuring environmental variables, such as counting the number of rare species present. On the other hand, it may involve assigning scores to measured variables based on subjective preferences. For example, less than five rare species may get 1 point, and 5–10 rare species, 2 points. Rigorous definition of these subjective preferences is the purpose of utility theory (Keeney and Raiffa 1976, Farquhar 1977).

Utility theory is used to measure subjective preferences and probabilities and translate these into quantitative mathematical equations or utility functions. These functions quantitatively express the preferences of particular “decision makers” for different levels of a criterion. Necessarily these functions will vary between people (for example, see Sobral and others 1981) and, indeed, an individual's preferences may vary through time. Figure 1 presents an example of a utility function for salmon catch in a sport fishery that was fitted using utility theory techniques. The function indicates how perceived utility or “satisfaction gained” (vertical axis) changes with increases in catch. The relation is not linear; the rate of increase in perceived utility drops off as catch increases. To use another example, the difference in perceived utility of a natural area with a diversity of 205 species compared with one with

Table 1. A review of 20 conservation evaluation studies.

Reference	Description of measurement techniques and method for overall ranking
International biological program sites (Nicholson 1968, Clapham 1981)	No explicit measurement used. A set of areas judged to represent range of ecological variation, have rare biota or earth science features, have been well studied, or present opportunities for research.
Countryside evaluation, UK (Tubbs and Blackwood 1971)	Overall subjective ranking of sites based on rarity, outstanding features and wildlife reservoir value; subjective scoring of agricultural sites largely based presence of "unsown vegetation."
State natural areas, Wisconsin (Tans 1974)	Scoring of 8 criteria on several different scales (e.g., 1–10 and 1–5). These are summed into four separate indices: biological value, physical characteristics and use value, availability, and threat.
Biosphere reserves, World and Canada (Man and Biosphere Program 1974 and 1976, Francis 1982)	Calls for "best judgement" of representativeness, national importance, effectiveness as a conservation unit and opportunities for research and education. No explicit measurement used.
Local natural areas, Texas (Gehlbach 1975)	Scoring of 5 criteria on a 3-point scale and weighting criteria from 1 to 5. Overall index derived in a number of ways: weighted sum, weighted product, and unweighted sum.
Countryside evaluation, UK (Goldsmith 1975)	Empirical ratio measurements of extent, rarity, plant species richness, and animal species richness are multiplied together for each habitat type and summed across habitat types.
Local natural areas, Vermont (Sargent and Brande 1976)	Both quantitative and qualitative measurements of 12 criteria converted to 5-point scale of value. A natural science rating and planning rating obtained by summing 6 criteria scores for each.
Local natural areas, UK (Wright 1977)	Assigned scores of 1, 2, and 3 to both quantitative and qualitative data on 18 different criteria; 16 of these scores are summed into indices for a scientific appraisal and a management appraisal.
National system of nature reserves, Great Britain (Ratcliffe 1977)	Separate evaluations were undertaken for each habitat type, often using different criteria and almost always formulating criteria differently. Some criteria were quantified, some assessed using a scoring system, and others assessed subjectively based on guidelines. Final grading into 4 ordinal categories based on consensus or adjudication.
Local natural areas, the Netherlands (van der Ploeg and Vlijm 1978)	A review of 10 different methods. These generally involve ratio measurements of biological, soil, geomorphological, and sometimes cultural features. These data were often mapped and in 9 studies were summarized as 3–5 categories (e.g., low, medium, and high). The most common method of deriving these categories was the summation of criterion values, but sometimes the largest single criterion value was emphasized.
State natural aquatic areas, Idaho (Rabe and Savage 1979)	The occurrences of representative and unique aquatic and terrestrial features are scored 0 or 1, presence or absence. These numbers are summed for an overall ranking.
Local natural areas, Southern Ontario (Eagles 1980)	Nine criteria are used separately to assess a site's significance. A particular site must meet one or more of the criteria, which are generally minimum standards. Assessments are qualitative based on guidelines, lists, and classifications.
Ornithological sites, UK (Fuller 1980)	Ratio measurements of population size, diversity, community quality, and rarity. Standards are defined for each of criteria at national, regional, county, and local scales. Each criterion considered separately.
Territorial system of natural areas, Yukon (Theberge and others 1980)	Biophysical rating is a sum of scores of 1 (low), 2 (medium), and 3 (high) for landforms, wildlife, vegetation, and representativeness. Separate subjective ratings are made for threat and recreation potential. Guidelines are given for subjective ratings.
Large roadless areas, western USA (Klopatek and others 1981)	Ratio measurements of indices of the rarity and diversity of vegetation types; the abundance, rarity, and diversity of bird species; the rarity and diversity of mammal species; and the occurrence of endangered species. These are summed for an "overall ecological index."

Table 1. *Continued.*

Reference	Description of measurement techniques and method for overall ranking
State natural areas, North Carolina (Radforth and others 1981)	Five criteria assessed on a subjective five point scale (1–5) and one on a 10-point scale. These scores are summed for an overall score.
Evaluation of potential national parks, Canada (Parks Canada 1979 and 1981)	"Themes" characterizing natural regions are rated 1, 2, or 3 as to commonness on each candidate area and 1 or 2 for significance. These are multiplied and then summed for all "themes." The site with the largest score is deemed the most representative and thus most significant.
Wetlands, southern Ontario (Environment Canada and Ont. Ministry of Nat. Resources 1984)	Complex subjective scoring of many factors which are summed into biological, social, hydrological, and special features indices, and a grand total.
Comprehensive protected area system, northern Canada (Task Force on Northern Conservation 1984)	Subjective yes/no judgement of a site's value with respect to 7 criteria. Overall significance is assigned if site is significant for one or more individual criteria.
Wetlands, southern Ontario (Van Patter and Hilt 1985)	Government agencies, conservation groups, and wetland experts asked to rate each of a list of wetlands as either "more" or "less" important. This produced consensus on half of the sites. Five additional criteria used by study team to select other half.

Table 2. The four scales of measurement.

Measurement scale	Admissible relations and operations	Example of criteria measurement
Nominal	Equivalence =	"Provides habitat for indigenous species that are rare or endangered at the regional, provincial or national scale" yes or no (Eagles 1980)
Ordinal	Equivalence = Order <, >	Low: "small number of species present" Moderate: "high species diversity" High: "outstanding diversity for particular habitat type" (Wright 1977)
Interval	Equivalence = Order <, > Addition + Subtraction - Multiplication × Division /	"Number of provincially significant plant species," one species, score = 100, two or more species, score = 150 (Environment Canada and Ontario Ministry of Natural Resources 1984)
Ratio	Equivalence = Order <, > Addition + Subtraction - Multiplication × Division /	Number of tree species per half hectare (McKinnon 1982)

200 species may be much less than between two natural areas with five and ten species, respectively.

Uncertainty in measurement. Different sorts of uncertainty affect measurement of criteria. Inaccuracy may exist due to error in measurement. As well, when a parameter is estimated from sample measurements, the laws of probability determine the precision of the

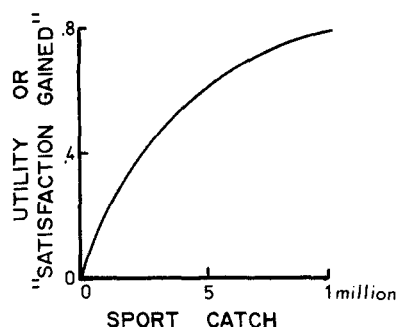


Figure 1. Example of a utility function for fishing sport catch (Holling 1978).

estimate (Steele and Torrie 1980). For example, Oviatt and others (1977) found as much variation in ecological variables within individual marshes as among different marshes and thus questioned the relevance of many evaluation methods. If the definitions of criteria and the values they may take are indistinct, measurements taken are necessarily "fuzzy" and may require the use of fuzzy set theory (Hwang and Yoon 1981, Efstathiou 1984, Zimmerman and others 1984). Finally, variance in the preferences of different people and interest groups have dramatic effect on the subjective scores and weights assigned to particular criteria and thus on the final ranking (for example, Holling 1978, Bakus and others 1982, Sobral and others 1981, Margules 1984).

Practice: Measuring the Value of Natural Areas

The specific approach to measurement used in the 20 evaluation studies is given in Table 1, and Figure 2 shows the number of studies that adopted particular

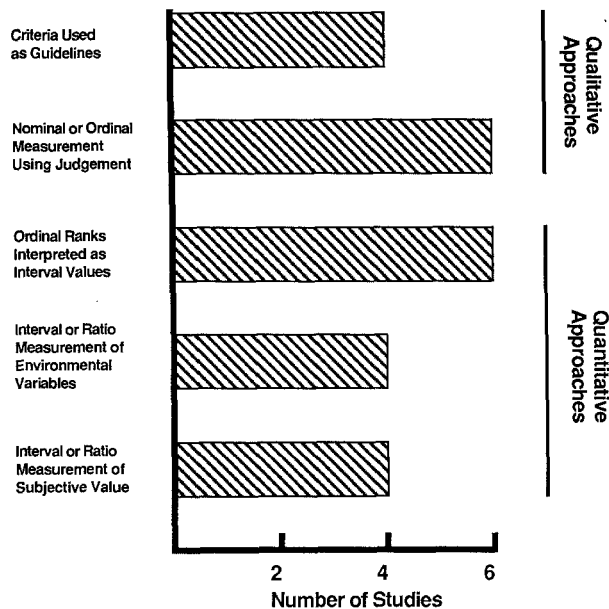


Figure 2. The numbers of studies using different forms of measurement. The sum of the numbers in the Figure exceeds 20 because some studies use more than one form of measurement.

approaches. Those systems that use quantitative (interval or ratio) measurement, either measure actual ecological variables in the field, or subjectively assess perceived "value" or "utility" based on the field measurements. For example, rarity can be expressed as the number of rare biota or scores may be assigned based on the "perceived value" of the rare biota. Several other studies interpret ordinal ranks as interval values (Figure 2). Systems utilizing qualitative "measurements" sometimes use criteria simply as guidelines to be applied based on judgment and professional opinion. In other cases, the criteria are systematically applied using some form of nominal categorization or ordinal ranking of sites as to their overall significance (Table 1 and Figure 2). Some studies incorporate more than one of the foregoing approaches (Table 1).

Quantitative approaches. In those studies where quantitative measurement has been used (Table 1 and Figure 2), methods varied with the quantity and quality of information available. In studies like Fuller's (1980) or Klopatek and others' (1981), the high quality of data allowed the valid use of ratio scales for measurement.

In evaluating ornithological sites, Fuller (1980) used criteria that lend themselves to ratio measurement: population size; diversity or number of species; and rarity. Moreover, extensive documentation existed for the avifauna of the sites being evaluated, such as the

numbers of different bird species occurring in different seasons and years. Thus it was possible to define for example, sites of national significance for population size as those supporting at least 1% of the estimated national population for that species.

Klopatek and others (1981) evaluated 321 tracts of land in the American west for the extent of natural vegetation; the abundance, rarity, and diversity of bird species; the diversity and rarity of mammal species; and the occurrence of endangered and threatened species. The information base used for this evaluation required the creation and integration of large computer data bases to contain all the necessary information. These data were then used to calculate separate indices for vegetation, birds, mammals, and endangered species, which were summed for an "overall ecological index." The vegetation index, for example, was calculated as the area of a particular vegetation type in the tract divided by total area occupied by that vegetation type in all 321 tracts. This was multiplied by a measure of the rarity of that vegetation type at the national, regional, and subregional scales. These values were calculated and summed for all vegetation types occurring in each tract.

Contrasting the foregoing fundamental measurement of environmental variables are systems that have assigned different quantitative "scores" or "points" to levels of each criterion. For example, Tans (1974) assigned points for size and buffer zone as follows: "Greatly exceeds minimum size, excellent buffer, no threat of encroachment," score = 8; "Adequate size and buffer," score = 6; "Adequate size, inadequate buffer," score = 4; "Inadequate size, adequate buffer," score = 2; and "Both inadequate," score = 0. Similarly, Wright (1977) converted low, medium, and high ordinal rankings to scores of 1, 2, and 3 and summed the scores for a number of criteria. This is improper application of addition to ordinal ranks (see Table 2). While such scoring has intuitive appeal, the allocation of points is arbitrary and subjective. Such scores are thought to represent the 'value' of the particular conditions to the decision makers involved. Utility theory offers more rigorous methods for the analysis of such preferences and assigning of scores.

A further problem is that scores assigned to different levels of criteria often have little relation to ecological theory and practice.

Qualitative approaches. All methods by definition involve at least nominal, if not ordinal, measurement by differentiating sites considered significant from those not. Some use explicit qualitative measurement, while others simply use criteria as guidelines and the notion of measurement is not mentioned (Table 1 and Figure 2). Several examples illustrate this.

The terrestrial conservation section of the International Biological Program (IBP) was a pioneering effort in identifying "ecological sites." It operated on a very broad scale in a time when there was little comprehensive survey data and information sources were often not comparable. While specific criteria (Nicholson 1968, Peterson 1975) were used as guidelines and made reference to in site reports (for example, Beckel 1975, Nettleship and Smith, 1975), systematic identification and comparative analysis of sites was, for the most part, lacking (Table 1).

Qualitative measurement of criteria sometimes depends heavily on subjective professional judgement. For example, the criteria advocated by the Task Force on Northern Conservation (1984) are questions as to whether a site is "significant," "necessary," "outstanding," or "essential." One particular criterion is extremely subjective, asking whether areas "contain sites necessary for the preservation of genetic diversity," yes or no. No definitions or guidelines are given. One must conclude that individual interpretation of the question would lead to great variance in the evaluation of sites.

Methods for identifying local "environmentally sensitive areas" in southern Ontario (Eagles 1980) are similar to the foregoing approach (Task Force on Northern Conservation 1984) in requiring judgements as to whether a site is "vital," "distinctive," or "unusual." In contrast, however, specific, detailed guidelines are developed to facilitate the judgements and enhance replicability. For example, one criterion is "the area represents a distinctive and unusual landform within the municipality, Ontario or Canada." The guideline for this criterion is that "the distinctive and unusual landforms are usually defined by the published and verbal comments of authorities in the field of earth science." This approach does not eliminate subjectivity, but rather makes it explicit, spelling out the basis of the judgement and facilitating discussion of that assessment.

The British Nature Conservation Review (Ratcliffe 1977), probably better than any study reviewed, discussed explicitly the issues of quantification and subjectivity and stated the assumptions underlying the method chosen. A scoring system was considered, but it was felt that scoring different habitat types for all the same criteria was inappropriate, that criteria are not independent, and that scoring would only lead to "spurious objectivity." Thus the approach adopted involved a different evaluation method and criteria for different ecosystem or habitat types. For example, grassland and woodland sites were evaluated separately. A final grading into four ordinal categories is the result of a comprehensive analysis of abiotic, biotic,

and cultural factors, some of which could be quantified and others not. Thus grading is "dependent on a series of value judgements" that, although based on consensus between many scientists, are nonetheless "subjective even when collective." A final point not often made in the conservation evaluation literature is that "values are also prone to change in time, and future generations may look at things rather differently" (Ratcliffe 1977).

A method unique among the 20 studies is the Delphi approach used by an environmental group in evaluating wetlands in southern Ontario (Van Patter and Hilt 1985). The Delphi method is a means used in attempts to achieve a collective decision on a particular issue (Linstone and Turloff 1975). Van Patter and Hilt (1985) asked government and nongovernmental organizations as well as a sample of wetland specialists to rate a list of sites as "more important" or "less important" and to suggest areas to be added to the list. Of the sites finally identified as "more important," general consensus among the respondents was reached on about half. The remaining half required an evaluation by the research team using the respondent's information and six additional criteria.

Using Multiple Criteria for an Overall Ranking

One main purpose of conservation evaluation is to reduce the amount of information to be considered in assessing the importance of areas. This need for data reduction has meant that the majority of evaluation systems have attempted to reduce environmental significance to one dimension or value. The extent to which this has been successful, or is possible, is a matter of opinion. Margules and Usher (1981) regard this as necessary, if not desirable, and "predict that ways of aggregating criteria into one index (or a number of indices) will become more sophisticated in the forthcoming decade."

While many view the derivation of a single index of "ecological value" as desirable, or at least a necessary evil, others contest the validity of such a notion (van der Ploeg and Vlijm 1978, Gotmark and others 1986, Nilsson 1986, van der Ploeg 1986). For example, van der Ploeg and Vlijm (1978) state: "If the aim is an ecological evaluation, all aspects should be studied in relation to each other. Any integration which simply adds aspect (criteria) values denies the existence of *ecosystems*. . . . Thus aspects like vegetation, fauna and geomorphology should either be combined from the beginning or they should not be integrated at all." Furthermore, they state that: "None of the studies gives a mathematical specification which is based on the 'ecological' relationships between the criteria used. There-

fore the formulae can be considered as pseudo-objective justifications of *a priori* judgements." It is indeed interesting to note that, while at one time derivation of a single index was common in evaluation in the Netherlands, it is now virtually absent (van der Ploeg 1986). Despite opinions and trends like the foregoing, a majority of the evaluation systems reviewed here used some type of overall summed index (Table 1 and Figure 3) and there is little indication of a decline in this practice in the more recent studies.

To review how a final ranking of sites has been derived in the 20 studies, we take two avenues. First, we introduce the theory of multicriteria evaluation, a discipline within operations research or management science. Secondly, these concepts are used to classify and assess the validity of the methods used in practice to rank sites in the 20 studies.

Theory: Multicriteria Evaluation Models

Multicriteria evaluation models are used to rank a series of alternatives on the basis of a number of criteria. The models vary markedly in the underlying assumptions and are designed to deal with different types of information. Two basic types of models exist: compensatory and noncompensatory (Hwang and Yoon 1981:24–25). Compensatory models generally result in a single number being calculated for each alternative. This generally requires that all criteria be measured on interval or ratio scales and are in comparable units in the sense that a drop in one criterion can be offset by a rise in another. This would imply, for example that the following two sites have equivalent overall value:

	Rarity value	Represent- ativeness value	Productivity value	Recreation value
Site 1	0	4	0	0
Site 2	1	1	1	1

Noncompensatory models, on the other hand, are used when the levels of different criteria cannot be compared; they are like "apples and oranges." Alternatives must then be assessed criterion by criterion.

The simple additive weighting method, the most common compensatory model, can be applied when all criteria are measured on an interval or ratio scale, all values are expressed in comparable units, and weights can be assigned to each criterion. That method involves simply multiplying all criterion values by their respective weights and summing the resulting values for each site. More depth on this method will be provided later.

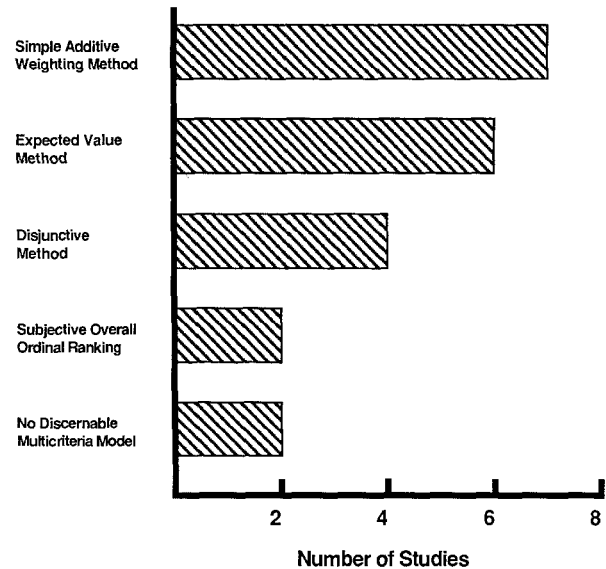


Figure 3. The number of studies using particular multicriteria evaluation methods.

Another compensatory model, the multiple attribute utility model, is closely related to the simple additive weighting model. Utility models are similar to cost-benefit analyses in attempting to convert all variables to a common scale of measurement. Instead of dollar value, utility models rank alternatives based on the overall expected "utility" of each (Bakus and others 1982). "Utility" is defined as the perceived benefit derived from a particular level of a criterion as discussed earlier. For example, an ecologist or decision maker might decide that the benefit of a site with ten rare plant species was 50% greater than a site with five rare plant species. Utility is a broad concept and can include values other than utilitarian ones, such as perceived intrinsic worth. Assessments of these utility functions are based on a series of systematic techniques for questioning, assessment, and comparison (Farquhar 1977, Zeleny 1981). The overall utility is calculated as the combined utilities of the individual criteria, which may be a simple sum, a weighted sum, a product, or other means of aggregation.

When criteria are measured on different scales and the different criteria are incomparable, the noncompensatory models are useful. To illustrate some of these models we will use a simple, hypothetical example of three sites—A, B, and C—that have been assessed using three criteria—rarity, diversity, and productivity. The values each criterion takes varies from 1 to 4.

	Rarity	Diversity	Productivity
Site A	2	1	4
Site B	2	1	1
Site C	2	3	1

"Dominance methods" (Hwang and Yoon 1981) rank an alternative higher when it is better than other alternatives for one or more criteria and of equal value for the other criteria. In the example, A *dominates* or is preferred over B, because the two have equal values for two criteria and A has a greater value than B for productivity. On the other hand, A does not dominate C because, while A has greater value for productivity, C has more value for diversity.

"Maximin" and "maximax" decision models (Hwang and Yoon 1981) rank alternatives based on their weakest or strongest criterion value, respectively, and ignore other criterion values. Under the maximin approach, sites A, B, and C are all preferred equally because each has a low score of 1 for at least one criterion. Using a maximax rule, site A would be selected as it has the maximum score of 4 for productivity.

Several models are based on minimum standards or thresholds. The "conjunctive method" (Hwang and Yoon 1981) ranks alternatives based on whether an alternative meets acceptable minimum standards for all criteria. The "disjunctive model" (Hwang and Yoon 1981) ranks based on an alternative meeting a minimum standard for at least one criterion regardless of its value for other criteria. If we were to apply minimum standards of 2 to all the criteria in the example, then using the conjunctive model, none of the three sites would be chosen, because none has a value of 2 or more for all the criteria. Under the disjunctive model, all three sites have at least one criterion with a value of 2 or more and thus are acceptable.

Several other models involve direct measurement of preferences for alternatives through comparison of each available alternative with others or with an ideal alternative. An ideal alternative in the example would be a site with a value of 4 for all three criteria. Ranking sites A, B, and C would consist of comparing each with the ideal and assessing which comes closest to that ideal. Full treatment of these and other decision models is available elsewhere (Starr and Zeleny 1977, Nijkamp 1980, Hwang and Yoon 1981, Zeleny 1981, Voogd 1983, Zimmerman and others 1984, Janssen and Nijkamp 1985).

Practice: Multicriteria Models Used to Evaluate Natural Areas

None of the systems reviewed (Table 1) explicitly stated the type of evaluation model used. As a result, we have classified the methods used by comparing the descriptions given with typologies of multicriteria evaluation models in Nijkamp (1980) and Hwang and Yoon (1981). Most evaluation models used in the systems reviewed generally correspond to three types of multicriteria evaluation models (Figure 3): the

simple additive weighting method (Hwang and Yoon 1981); what Nijkamp (1980:38) called the expected value method; and the disjunctive model (Hwang and Yoon 1981). The details of these three multicriteria models are outlined here to allow assessment of the validity of their use in the evaluation systems reviewed here. Figure 4 summarizes the requirements and assumptions of each of the three models. Figure 3 indicates how many of the 20 studies generally correspond to these three multicriteria models. Several methods do not resemble any multicriteria models (Figure 3) including the overall ordinal ranking in Ratcliffe (1977) which was discussed earlier.

Simple additive weighting method. The simple additive weighting method (Figure 4) is the multicriteria model most familiar to many people and that used in seven of 20 studies (Figure 3). Using this model, the alternatives, in our case natural areas, are ranked using the weighted sum of all the criteria measurements. Multiplication and addition are used and "therefore, the criteria values *must* be both *numerical* (i.e. measured using an interval or ratio scale) and *comparable*" (emphasis added; Hwang and Yoon 1981:101). Values for criteria may be either empirical measurements or subjective scores. The values of each criterion are standardized or normalized so that they are comparable between criteria. Standardization is neglected in some studies reviewed (for example, Tans 1974). Standardization often involves a common scoring system for all criteria. The assumption of comparability implies that similar changes in the levels of different criteria are equivalent (Hwang and Yoon 1981:102). For example, it is assumed that a high ranking of a site for rarity is equivalent to a high ranking for diversity. Whether this assumption is valid is probably more of a philosophical issue than a mathematical one. A further assumption is that all criteria are independent of each other, an assumption clearly violated by most ecological data (see Usher 1985, Gotmark and others 1986).

In referring to comparability, it is important to differentiate the "equivalence of units" from standardization or normalization. Criteria are measured using different units that may or may not be equivalent. For example, "we can say that an automobile is wider than it is tall, because both scores are expressed in the same units," whereas "we cannot say that an automobile is wider than it is heavy" (Ghiselli and others 1981). Standardization or normalization involves comparing the scores of a particular alternative with those of all other alternatives or with some real, or theoretical, maximum value.

For the simple additive weighting model, "it is also necessary to find a reasonable basis on which to form the weights reflecting the importance of each of the

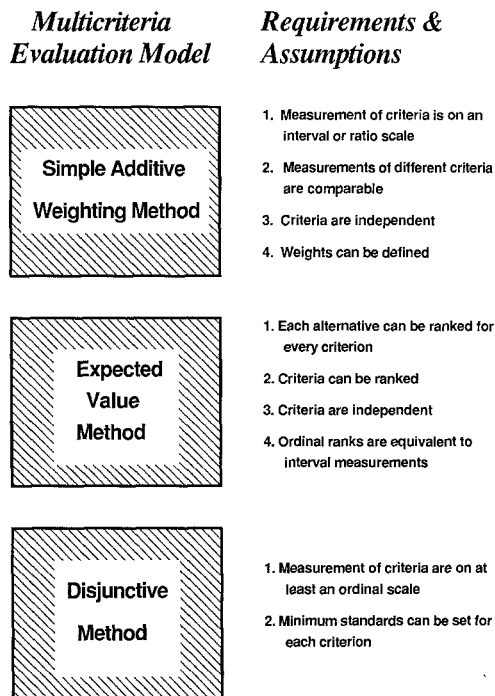


Figure 4. The requirements and assumptions of three multicriteria evaluation models.

criteria" (Hwang and Yoon 1981:101). Weights are necessarily subjective, but there are techniques for measuring subjective preferences precisely and translating these into numerical weights (Hwang and Yoon 1981:41, Zeleny 1981:185). Margules (1984) and Margules and Usher (1984) show that such subjective weights applied by different ecologists and environmental managers can vary widely. When no weights can be assigned subjectively, others methods can be used to assign weights based on an "objective" rationale.

A form of implicit weighting is applied in some systems (for example, Tans 1974) when scores assigned for different criteria have different maximum values and thus are not standardized so as to make the values comparable. So-called equal weights used in some systems (for instance, Wright 1977, Klopatek and others 1981) refer only to the explicit weights. Emphasis on a particular criterion varies with the number of criteria and how they are formulated.

Expected value method. The expected value method (Figure 4) is a crude form of the simple additive weighting method. The decision maker subjectively ranks each alternative for each criterion. For example, three sites may be ranked 1, 2, 3 for rarity and 2, 1, 3 for diversity. The criteria are also ranked; for example, rarity may get 1, diversity 2, and fragility 3. Each site's rank for each individual criterion is then multiplied by the criterion's rank. Thus a site with

rank 2 for diversity would be multiplied by the rank for the diversity criterion, 2, yielding 4. All these values are summed for each site and these sums used to rank the sites. "It is clear that this evaluation method is, however, a rather crude aggregation method based on non-permissible numerical operations on ordinal numbers" (Nijkamp 1980:38). Methods used by six of 20 studies, resemble this method (Figure 3). Two methods correspond closely with the expected value method in using numerical interpretation of both ordinal ranks of sites and of ordinal preferences for different criteria (Gehlbach 1975, Parks Canada 1979 and 1981). Four other studies involve only numerical interpretation of ordinal ranks of sites (Sargent and Brande 1976, Wright 1977, Theberge and others 1980, Radforth and others 1981). The approaches of the latter studies might also be classified as resembling the simple additive weighting method.

In assessing the validity of overall ranking based on any method and particularly the expected value method, an important consideration is the sensitivity of the rank to uncertainty. There may be uncertainty in field measurements, variation between people in assigning scores and in the weights given to different criteria, as well as fuzziness in the definitions of the criteria (see *Uncertainty in measurement* above). The precision of a site's overall rank or score is a function of the amount of uncertainty from all the sources mentioned. One must have a degree of confidence in the derived scores or ranks to defend them and base public policy decisions on their accuracy. Treatment of such uncertainty with mathematical validity necessitates more complex mathematical models (O'Bannon 1980, Sobral and others 1981, Bakus and others 1982, Zimmerman and others 1984).

The disjunctive model. One problem with the simple additive weighting method is that sites with high "overall value" may not be important with respect to any of the individual criteria (van der Ploeg and Vlijm 1978, Klopatek and others 1981). For example, the highest ranking area overall in the study by Klopatek and others (1981) ranked 60, 71, 52, and 19 with respect to birds, mammals, vegetation, and endangered species. Thus that area was not particularly important for any single reason, but rather average for all four, causing it to rank first overall. Emphasis on these "average" conservation values can "hide sites with outstanding qualities" (Gotmark and others 1986). An alternative model for ranking used by several systems reviewed by van der Ploeg and Vlijm (1978) as well as Eagles (1980), Fuller (1980), and the Task Force on Northern Conservation (1984) avoided this problem (Figure 3). In those systems, the overall ranking is

based on the highest rank each area has for any criterion. This is conceptually similar to the disjunctive multicriteria evaluation model (Hwang and Yoon 1981:64; see Figure 4). For example, the area with most rare species would automatically occur in the highest ranking class as would the most diverse area. Furthermore, under this model, if a site is selected because it is quite diverse, it is irrelevant how low it may rank for other criteria.

The point of all this discussion of multicriteria evaluation models is that many other models exist than those often used in conservation evaluation. Different models exist for use with different types of information to avoid inappropriate assumptions and invalid mathematical operations. The principles underlying these models are not difficult. In fact, many methods have been developed to emulate the psychology of human decision making.

Unless a multicriteria ranking scheme is based on valid assumptions, an area's rank may be an artifact with little real meaning. For example, Voogd (1981) found that a multicriteria evaluation model based on assigning scores to ordinal ranks resulted in the selection of different alternatives than a method designed specifically for use with ordinal measurements. Several of the evaluation models not often used may indeed be useful because they involve fewer assumptions than the simple additive weighting and expected value methods. In general, we find the disjunctive model (Figure 4) useful and have adopted it for our own work in Canada's Northwest Territories (Smith and others 1986:22) because the underlying assumptions are valid for the data used in that work and can be communicated easily.

Selecting the "Best" Evaluation Methods

Having examined problems in the ways in which measurement and multicriteria models have been used in conservation evaluation, we can suggest directions for improvement. Little can be said of specific ideal methods in particular situations because this depends on many factors, especially the assumptions that decision makers and the public are willing to accept. Nevertheless, we believe that the principles and assumptions on which methodologies rest must be rationalized ecologically, philosophically, and mathematically. To that end, Table 3 presents a series of principles that we suggest to guide the choice of evaluation methods. These principles have been adapted largely from Jansson and Nijkamp (1985), who provided a guide to the selection of multicriteria methods for a variety of environmental management decisions, including the evaluation of natural areas. Some of the principles

Table 3. Principles for the choice of evaluation method.

An evaluation method should:

- 1) Be based on principles and assumptions that are valid and easily illustrated
 - 2) Yield results understandable to decision makers and the public
 - 3) Make explicit subjective values and judgements
 - 4) Yield results that are repeatable given certain explicit assumptions
 - 5) Allow use of qualitative and quantitative information in a methodologically sound way
 - 6) Stimulate the imagination of decision makers and increase insight into the choice to be made
 - 7) Enable the use of information at different spatial scales
 - 8) Allow consideration of alternatives both separately and in combination
-

may seem to state the obvious, but based on our review of the literature, none can be considered self-evident.

As we have already stated, the underlying principles and assumptions or any method must be ecologically and mathematically valid and furthermore must be acceptable to the decision makers and the public (no. 1). The latter necessitates simple illustration of those underlying principles and assumptions. Grasp of the theory of a method also means that the results can be understood and accepted (no. 2).

Subjective values and judgements must be made explicit in order to assess their effect on the choices made (no. 3). Several authors have recently called for greater separation of the objective component of evaluation, that is, the ecological description of areas, from the subjective component, that of placing value on those descriptions (Roome 1984, Wathern and others 1986).

Methods of assessment must be repeatable given certain assumptions (no. 4). It is a given that subjective preferences for the different levels of criteria and for the criteria themselves vary dramatically among people. As a result, any method employing subjective judgements should provide specific guidelines to structure these judgements so as to enhance replicability.

Evaluation for nature conservation will almost invariably entail use of both qualitative and quantitative information (no. 5). Qualitative data can, in some cases, be scaled into numeric values, but this involves additional assumptions and greater complexity (Keeney and Raiffa 1976). Certain multicriteria models are designed to use both types of data without such transformation.

While one purpose of evaluation is to reduce the amount of information to be digested, another is to stimulate fresh ways of viewing the choice being made

and increase insight into what is being chosen and why (no. 6).

Evaluation of natural areas, almost by definition, entails the use of data at different spatial scales. Often the areas being evaluated vary in size by several orders or magnitude. Some criteria may be more applicable to sites of particular size (Margules and Usher 1984). Moreover, provision for the desirable mix of protected habitats in a region, large and small areas as well as the linkages between them, impinges on an evaluation process that may emphasize specific attributes of sites, like rarity and diversity. These considerations should be built into the method used (no. 7).

Following from the foregoing is the need to compare among combinations of candidate areas as well as evaluating each area on its own merits. For example, an important consideration might be what *combination* of natural areas in a region contains the greatest number of rare species (also see Rudis and Ek 1981). Simple assessment of the number of rare species found on each site neglects this larger perspective.

Discussion

Conflicting goals exist in developing any system for evaluating natural areas. Simplicity is required to make the methods accessible to nonspecialists in government, industry, and the public. The need for simplicity no doubt results in the exclusion of complex management science methods. Another, sometimes conflicting, need is for both ecological and mathematical validity. Much of the research on conservation evaluation to date has focused on increasing the ecological validity of methods of evaluation (for example, Margules and Usher 1981). These efforts sometimes have also involved increased mathematical validity (for instance, Dony and Denholm 1985). Nevertheless, we perceive a need for more clarity and rigor in the formulation of criteria measurements and the multicriteria evaluation models used. Furthermore, a more rational basis is needed for the selection of measurement and multicriteria methods. Our review of the pertinent theory and practice may help in that regard, as may our principles to guide method selection (Table 3).

Other multicriteria evaluation models exist than those most often used in evaluating natural areas. Some of these models involve fewer assumptions and assumptions that are more appropriate to ecological data than those underlying the assigning and summing of arbitrary scores. The principles underlying these methods are not difficult and are easily communicated (see *Theory: multi-criteria evaluation models*). Specific reference to the model being used and justifica-

tion of the underlying assumptions seems a reasonable demand (also see Janssen and Nijkamp 1985).

For a number of reasons, the common practice of aggregating criteria assessments into one index masks the complexity of the evaluation process. For example, in the case of evaluating sites for status as protected areas, the problem is much more complex than simply ranking areas. Each area's significance varies depending on how boundaries are drawn, because evaluation is as much an assessment of particular features as of composite geographic areas (see Smith and others 1986). Aggregating criteria also ignores the relationships between criteria (van der Ploeg and Vlijm 1978, Gotmark and others 1986). Inherent linkages exist between criteria such as representativeness and diversity or fragility, size, and rarity (Usher 1980 and 1985, Margules 1984, Margules and Usher 1984). Hence, treatment of each criterion as fully independent is unrealistic. Correlations between criteria can be accounted for in some multicriteria models that aggregate criterion values, but this necessarily involves increased complexity (Hwang and Yoon 1981).

In a similar vein, generally there are natural groupings of criteria that might be applied in separate evaluations, such as ecological, cultural, or human use and management criteria (Smith and Theberge 1986a). Moreover, not all criteria may be applicable to every kind of natural area. For example, rarity and fragility might better apply to small sites and representativeness and naturalness to larger sites (Margules and Usher 1984). Alternatively, different criteria could be applied during different stages of the evaluation process (Margules 1984, Margules and Usher 1984, Nilsson 1986).

The reduction of information to that of most salience is a basic need, if not the purpose of conservation evaluation. However, it is important "to avoid distortions due to aggregation while not presenting the decision maker with too many data to integrate" (Hollick 1981). Evaluation methods should enhance our grasp of the choices being made and not cloud important issues or conceal value judgements. Furthermore, the assumptions of any evaluation method should be clearly stated and rationalized in terms of the data and areas being compared.

Acknowledgments

This research was carried out as background for an evaluation method for environmentally significant areas in the Northwest Territories of Canada (Nelson and others 1985, Smith and others 1986, Smith and Theberge 1986b). It forms part of ongoing research on environmental planning and management in

northern Canada (Theberge and others 1980, Fenge 1982 and 1985, Bastedo and others 1984, Nelson and Jessen 1984). Support for this research was from the Canadian Arctic Resources Committee, the Canadian Department of Indian Affairs and Northern Development, and the Northwest Territories Department of Renewable Resources. The University of Waterloo Faculty of Environmental Studies supplied computing resources. J. G. Nelson, R. Dorney, G. Lee, D. Cole, and C. Price provided critical comment on earlier versions of the manuscript.

Literature Cited

- Bakus, G., W. Stillwell, S. Latter, and M. Wallerstein. 1982. Decision-making with applications for environmental management. *Environmental Management* 6:493–504.
- Bastedo, J. D., J. G. Nelson, and J. B. Theberge. 1984. An ecological approach to resource survey and planning for environmentally significant areas: the ABC method. *Environmental Management* 8:125–134.
- Beckel, D. K. B. (ed.). 1975. IBP ecological sites in subarctic Canada: areas recommended as ecological sites in Region 10, Yukon and Northwest Territories boreal forest to tree-line. University of Lethbridge Production Services, Lethbridge, 163 pp.
- Clapham, A. 1981. The IBP survey of conservation sites: an experimental study. Cambridge University Press, Cambridge, UK.
- Dony, J. G., and Denholm, I. 1985. Some quantitative methods of assessing the conservation value of ecologically similar sites. *Journal of Applied Ecology* 22:229–238.
- Eagles, P. F. J. 1980. Criteria for the designation of environmentally sensitive areas. Pages 68–79 in S. W. Barrett and J. L. Riley (eds.), Protection of natural areas in Ontario. Working Paper no. 3, Faculty of Environmental Studies, York University.
- Eichhorn, W. 1978. What is an economic index? An attempt of an answer. Pages 3–40 in W. Eichhorn, R. Henn, O. Opitz, and R. W. Shephard (eds.), Theory and applications of economic indices. Physica, Wurzburg, Austria.
- Efstathiou, J. 1984. Practical multi-attribute decision-making and fuzzy set theory. Pages 307–322 in H. J. Zimmerman, L. A. Zadeh, and B. R. Gaines (eds.), Fuzzy sets and decision analysis. North-Holland, Amsterdam.
- Environment Canada and Ontario Ministry of Natural Resources. 1984. An evaluation system for wetlands of Ontario south of the precambrian shield. Ottawa.
- Farquhar, P. H. 1977. A survey of multiattribute utility theory and applications. Pages 59–89 in M. K. Starr and M. Zeleny (eds.), Multiple Criteria Decision Making. North-Holland, Amsterdam.
- Fenge, T. 1982. Towards comprehensive conservation of environmentally significant areas in the Northwest Territories, Canada. *Environmental Conservation* 9:305–313.
- Fenge, T. 1985. Conservation of Polar Bear Pass, Bathurst Island, and the emerging comprehensive conservation policy for northern Canada. *Environmental Conservation* 12:231–240.
- Francis, G. 1982. Nomination of Canadian biosphere reserves. Canada/Man and Biosphere Program Report no. 15, Ottawa.
- Fuller, R. J. 1980. A method for assessing the ornithological interest of sites for conservation. *Biological Conservation* 17:229–239.
- Gehlbach, F. R. 1975. Investigation, evaluation and priority ranking of natural areas. *Biological Conservation* 8:79–88.
- Ghiselli, E., J. Campbell, and S. Zedek. 1981. Measurement theory for the behavioral sciences. W. H. Freeman, San Francisco.
- Goldsmith, F. B. 1975. The evaluation of ecological resources in the countryside for conservation purposes. *Biological Conservation* 8:89–95.
- Goldsmith, F. B. 1983. Evaluating nature. Pages 233–246 in A. Warren and F. B. Goldsmith (eds.), Conservation in perspective. Wiley, Chichester, UK.
- Goldstein, G., and M. Hersen (eds.). 1984. Handbook of psychological assessment. Pergamon, New York.
- Gotmark, F., M. Ahlund, and M. Eriksson. 1986. Are indices reliable for assessing conservation value of natural areas? An avian case study. *Biological Conservation* 38:55–73.
- Hollick, M. 1981. The role of quantitative decision-making methods in environmental impact assessment. *Journal of Environmental Management* 12:65–78.
- Holling, C. S. (ed.). 1978. Adaptive environmental assessment and management. John Wiley, New York, 377 pp.
- Hwang, C., and K. Yoon. 1981. Multiple attribute decision making: methods and applications a state-of-the-art survey. Springer-Verlag, New York.
- Janssen, R., and P. Nijkamp. 1985. A multiple criteria evaluation typology of environmental management problems. Pages 495–514 in Y. Haines and V. Chankong (eds.), Decision making with multiple objectives. Springer-Verlag, Berlin.
- Keeney, R. L., and H. Raiffa. 1976. Decisions with multiple objectives: preferences and value tradeoffs. Wiley, New York.
- Klopatek, J. M., J. T. Kitchings, R. J. Olson, K. D. Kumar, and L. K. Mann. 1981. A hierarchical system for evaluating regional ecological resources. *Biological Conservation* 20:271–290.
- Linstone, H., and M. Turoff (eds.). 1975. The delphi method, techniques and applications. Addison-Wesley, Reading, Massachusetts.
- Man and Biosphere Program. 1974. Task force on: criteria and guidelines for the choice and establishment of biosphere reserves. Man and Biosphere Program Report no. 22, Paris.
- Man and Biosphere Program. 1976. Guidelines for the selection of biosphere reserves in Canada. Canada/MAB Report 6, Ottawa.
- Margules, C. R. 1984. Conservation evaluation in practice. II. Enclosed grasslands in the Yorkshire Dales, Great Britain. *Biological Conservation* 18:153–168.
- Margules, C., and M. B. Usher. 1981. Criteria used in assessing wildlife conservation potential: a review. *Biological Conservation* 21:79–109.
- Margules, C. R., and Usher, M. B. 1984. Conservation evaluation.

- ation in practice. I. Sites of different habitats in Northeast Yorkshire, Great Britain. *Biological Conservation* 18:153–168.
- Mason, R. O., and E. B. Swanson. 1981. Measurement for management decision. Addison-Wesley, Reading, Massachusetts, 550 pp.
- McKinnon, J. 1982. National conservation plan for Indonesia. National Parks Development Project of the Food and Agriculture Organization of the United Nations. Bogor. Food and Agriculture Organization of the United Nations.
- Nelson, J. G., and S. Jessen. 1984. Planning and management of environmentally significant areas in the Northwest Territories: issues and alternatives. Canadian Arctic Resources Committee, Ottawa.
- Nelson, J. G., P. G. R. Smith, and J. B. Theberge. 1985. Environmentally significant areas in the Northwest Territories: their role, identification, designation and implementation. *Environments* 17:91–109.
- Nettleship, D. N., and P. A. Smith. 1975. Ecological sites in northern Canada. Canadian Committee for the International Biological Programme, Conservation Terrestrial—Panel 9, Ottawa.
- Nicholson, E. M. 1968. Handbook to the conservation section of the International Biological Program. Blackwell, Oxford, UK.
- Nijkamp, P. 1980. Environmental policy analysis. Operational methods and models. Wiley, New York, 283 pp.
- Nilsson, C. 1986. Methods of selecting lake shorelines as nature reserves. *Biological Conservation* 35:269–291.
- O'Bannon, K. 1980. Use of value functions in environmental decisions. *Environmental Management* 4:3–6.
- Ott, W. 1978. Environment indices: theory and practice. Ann Arbor Science, Ann Arbor, Michigan.
- Oviatt, C., S. Nixon, and J. Garber. 1977. Variation and evaluation of coastal salt marshes. *Environmental Management* 1:201–211.
- Parks Canada. 1979. Parks Canada policy. Department of the Environment. Ottawa.
- Parks Canada. 1981. Terms of reference for a regional analysis of natural region 5, the Rocky Mountains. Ottawa.
- Peterson, E. B. 1975. Development of selection criteria for ecological reserves. Pages 85–90 in R. Franson (ed.), The legal aspects of ecological reserves creation and management in Canada. IUCN Environmental Policy and Law Paper no. 9, Morges, Switzerland.
- Rabe, F. W., and N. L. Savage. 1979. A methodology for the selection of aquatic natural areas. *Biological Conservation* 15:291–300.
- Radford, A. E., D. K. S. Otte, L. J. Otte, J. R. Massey, and P. D. Whitson. 1981. Natural heritage: classification, inventory and information. University of North Carolina Press, Chapel Hill, 485 pp.
- Ratcliffe, D. (ed.). 1977. A nature conservation review, vol. 1. Cambridge University Press, Cambridge, UK.
- Ratcliffe, D. 1986. Selection of important areas for wildlife conservation in Great Britain: the Nature Conservancy Council's approach. Pages 135–159 in M. Usher (ed.), Wildlife conservation evaluation. Chapman and Hall, London.
- Ray, G. C. 1975. Survey of critical marine habitats and requirements for their conservation. Pages 15–59 in Working Paper no. 1 of and International Conference on Marine Parks and Reserves. Tokyo, 12–14 May 1975. IUCN Publication New Series no. 37. Morges, Switzerland.
- Roome, N. J. 1984. Evaluation in nature conservation decision-making. *Environmental Conservation* 11:247–252.
- Rudis, V., and A. Ek. 1981. Optimization of forest island spatial patterns: methodology for analysis of landscape pattern. Pages 241–256 in R. L. Burgess and D. Sharpe (eds.), Forest island dynamics in man-dominated landscapes. Springer-Verlag, New York.
- Sargent, F. C., and J. H. Brande. 1976. Classifying and evaluating unique natural areas for planning purposes. *Journal of Soil and Water Conservation* 31:113–116.
- Smith, P. G. R., and J. B. Theberge. 1986a. A review of criteria used to evaluate natural areas. *Environmental Management* 10:715–734.
- Smith, P. G. R., and J. B. Theberge. 1986b. Evaluating biotic diversity in environmentally significant areas in the Northwest Territories of Canada. *Biological Conservation* 36:1–18.
- Smith, P. G. R., J. G. Nelson, and J. B. Theberge. 1986. Environmentally significant areas, conservation and land use management in the Northwest Territories. Heritage Resources Centre, University of Waterloo, Technical Paper 1, Waterloo, Canada.
- Sobral, M., K. Hipel, and G. Farquhar. 1981. A multi-criteria model for solid waste management. *Journal of Environmental Management* 12:97–110.
- Starr, M. K., and M. Zeleny. 1977. Multiple criteria decision making. North-Holland, Amsterdam.
- Steele, R. G. D., and J. H. Torrie. 1980. Principles and procedures of statistics: a biometrical approach. McGraw-Hill, New York, 633 pp.
- Tans, W. 1974. Priority ranking of biotic natural areas. *Michigan Botanist* 13:31–39.
- Task Force on Northern Conservation. 1984. The report of the task force on northern conservation. Department of Indian Affairs and Northern Development, Ottawa.
- Theberge, J. B., J. G. Nelson, and T. Fenge (eds.). 1980. Environmentally significant areas of the Yukon Territory. Canadian Arctic Resources Committee, Ottawa.
- Tubbs, C. R., and Blackwood, J. W. 1971. Ecological evaluation of land for planning purposes. *Biological Conservation* 3:169–172.
- Usher, M. B. 1980. An assessment of conservation values within a large site of special scientific interest in North Yorkshire. *Field Studies* 5:323–348.
- Usher, M. B. 1985. Implications of species–area relationships for wildlife conservation. *Journal of Environmental Management* 21:181–192.
- Usher, M. (ed.). 1986. Wildlife conservation evaluation. Chapman and Hall, London.
- van der Ploeg, S. W. F. 1986. Wildlife conservation evaluation in the Netherlands: a controversial issue in a small country. Pages 161–180 in M. Usher (ed.), Wildlife conservation evaluation. Chapman and Hall, London.
- van der Ploeg, S. W. F., and L. Vlijm. 1978. Ecological evaluation, nature conservation and land use planning with par-

- ticular reference to methods used in the Netherlands. *Biological Conservation* 14:197–221.
- Van Patter, M., and S. Hilt. 1985. Important wetlands in Ontario south of the precambrian shield. Federation of Naturalists, Toronto.
- Voogd, H. 1981. Qualitative multicriteria evaluation for development planning. *Canadian Journal of Regional Science* 4:73–87.
- Voogd, H. 1983. Multicriteria evaluation for urban and regional planning. Pion, London.
- Wathern, P., S. Young, I. Brown, and D. Roberts. 1986. Ecological evaluation techniques. *Landscape Planning* 12:403–420.
- Wright, D. F. 1977. A site evaluation scheme for use in the assessment of potential nature reserves. *Biological Conservation* 11:293–305.
- Zeleny, M. 1981. Multiple criteria decision making. McGraw-Hill, New York, 563 pp.
- Zimmerman, H. J., L. A. Zadeh, and B. R. Gaines. 1984. Fuzzy sets and decision analysis. North-Holland, Amsterdam, 522 pp.

PROFILE

Evaluating Natural Areas Using Multiple Criteria: Theory and Practice

PAUL G. R. SMITH
JOHN B. THEBERGE

Faculty of Environmental Studies
University of Waterloo Waterloo
Ontario N2L 3G1, Canada

ABSTRACT / Evaluating natural areas involves making measurements for a series of criteria and deciding which areas are most significant based on these measurements. We review the basic theory of measurement and its application to assessing criteria in 20 different conservation evaluation studies. Different types of models exist to assess the overall significance of natural areas based on multiple criteria. A

brief description is presented of these "multicriteria evaluation models" and the underlying theory. The assumptions of these models are outlined as are the types of information that are appropriate to each. We then examine how a final overall ranking is derived in the 20 studies. None explicitly stated the model used and few noted the assumptions involved. Based on theory and current practice, a number of principles are presented to guide the development of evaluation methods. We conclude that evaluation methods should not be allowed to cloud important issues or conceal value judgements. Furthermore, the assumptions of any evaluation method should be clearly stated and rationalized in terms of the data and areas being compared.

The evaluation of natural areas, or conservation evaluation, has become widespread in land use planning, environmental impact assessment, and planning for protected areas (van der Ploeg and Vlijm 1978, Margules and Usher 1981, Goldsmith 1983, Roome 1984, Usher 1986, Wathern and others 1986). Its purpose is to identify the most ecologically valuable natural areas so planning and management practices can be applied so as to maintain the areas' values. Many different criteria form the basis of the evaluations, and we review these criteria elsewhere (Smith and Theberge 1986a). Each of these different criteria must be measured and the measurements used collectively to rank natural areas. This overall assessment is often done by summing subjective scores assigned to different criteria, a practice increasingly criticized (Nilsson 1986, Ratcliffe 1986). Indeed, even the value of composite indices derived from empirical ecological measurements is being questioned (Gotmark and others 1986).

Many composite indices exist, such as those in economics [gross national product and consumer price indices (Eichhorn 1978)], in psychology [IQ and other test scores (Goldstein and Hersen 1984)], and in environmental management [air and water quality indices (Ott 1978)]. The construction of composite indices in all these fields required the development of a strong theoretical basis. No such theoretical basis has been established for the composite indices used in conservation evaluation.

Similarly, decisions based on multiple criteria are

common in business and public administration, for example, the hiring of personnel, selection of facility sites, and granting of contracts. Multicriteria decision problems are also common in other fields of environmental management such as waste disposal site location and resource allocation (Keeney and Raiffa 1976, O'Bannion 1980, Bakus and others 1982, Janssen and Nijkamp 1985). A wealth of theory and practice has been used to develop "multicriteria evaluation models" and other methods to facilitate these decisions (Keeney and Raiffa 1976, Hwang and Yoon 1981). Simply put, these models integrate measurements of several separate criteria to arrive at a choice from among a series of alternatives. Indeed, all decisions involving more than one criterion use some sort of "multicriteria evaluation model" whether explicitly or implicitly.

Surprisingly, conservation evaluation has, for the most part, ignored the knowledge gained in these other disciplines in coping with similar issues. The purpose of this article, then, is to draw upon the theories and methods alluded to above for three purposes: (a) to describe how measurement and multicriteria decisions have been treated in the conservation evaluation literature, (b) to critique the methods used and clarify the assumptions and requirements of both the types of measurement and the multicriteria evaluation models used to evaluate natural areas, and (c) to suggest a number of principles to guide the choice and design of the most appropriate evaluation methods. Now long overdue, this synthesis of theory and methods from mathematics, management science, and psychology with the practice of conservation evaluation can only improve the quality of decisions that are made.

KEY WORDS: Conservation; Evaluation; Criteria; Measurement; Natural areas

The Evaluation Systems Reviewed

To examine how criteria have been measured and what multicriteria models have been used in conservation evaluation, we examined 20 different studies (Table 1). These were selected to reflect the diversity of approaches, applications at different scales, different types of environments, and differences in political organization, as well as changes in methods over a 14-year period (1971–1985). As a result, local, provincial, state, national, and international efforts are represented, as are arctic, temperature, freshwater, terrestrial, wilderness, and human-modified environments (Table 1).

Table 1 briefly describes how each study measures criteria and how an overall ranking is achieved. Other aspects of these studies are described by Smith and Theberge (1986a). Collectively, these studies represent the “practice” of evaluation and will be drawn upon throughout the article.

Methods of Measurement

Theory: Measurement

To understand the nature of measurement, and thus how criteria are assessed, one must grasp three aspects of measurement theory. First, the four scales of measurement, outlined below, have properties that define which types of mathematical operation may be applied to the values. Secondly, measurement of the value of natural areas can involve either measuring actual environmental variables, such as the number of species present, or measuring human perception of an area's value. Finally, any measurements taken are subject to a number of forms of uncertainty, and these affect both how measurement is done and the confidence we can place in the values obtained.

The four scales of measurement. Four basic scales of measurement exist: nominal, ordinal, interval, and ratio (Steele and Torrie 1980, Ghiselli and others 1981, Mason and Swanson 1981). Table 2 summarizes the characteristics of the four scales and gives an example of each from the literature on conservation evaluation.

The nominal scale is qualitative. For example, a human may be male or female and these are mutually exclusive categories. As this scale is qualitative, few mathematical relations or operations can be applied to it (Table 2).

The ordinal scale allows the categories of the scale to be ordered or ranked. A partial ordinal scale consists of a series of categories, for example high, medium, and low, and several items are assigned to each

category. By contrast, in a complete ordinal scale, each item is assigned a unique rank relative to all other items. Arithmetic operations such as addition and multiplication cannot be performed on the categories of an ordinal scale (Table 2). Some consider the ordinal scale qualitative (for example, Voogd 1983:75), while others call it quantitative (Ghiselli and others 1981:12). For the purpose of this article, it will be considered qualitative.

In interval measurement, the numerical distance between the various “categories” of the scale can be specified. Thus, a real number can be assigned to each item. However, an interval scale has no origin or real zero point. Examples are altitude and time measurement, there is no absolute zero altitude or time and it must be arbitrarily assigned to a particular point in the scale. All normal arithmetic operations can be applied to variables measured with an interval scale (Table 2).

A ratio scale has the same characteristics as the interval scale except that it has a true origin or zero point. An example is mass: if an object has zero mass that is its mass under any conditions. Again, all the normal arithmetic operations can be applied to such numbers (Table 2). The last two scales, the interval and ratio are the only completely quantitative or cardinal scales (Mason and Swanson 1981).

Measuring environmental variables or subjective value. Measurement of criteria may consist of simply measuring environmental variables, such as counting the number of rare species present. On the other hand, it may involve assigning scores to measured variables based on subjective preferences. For example, less than five rare species may get 1 point, and 5–10 rare species, 2 points. Rigorous definition of these subjective preferences is the purpose of utility theory (Keeney and Raiffa 1976, Farquhar 1977).

Utility theory is used to measure subjective preferences and probabilities and translate these into quantitative mathematical equations or utility functions. These functions quantitatively express the preferences of particular “decision makers” for different levels of a criterion. Necessarily these functions will vary between people (for example, see Sobral and others 1981) and, indeed, an individual's preferences may vary through time. Figure 1 presents an example of a utility function for salmon catch in a sport fishery that was fitted using utility theory techniques. The function indicates how perceived utility or “satisfaction gained” (vertical axis) changes with increases in catch. The relation is not linear; the rate of increase in perceived utility drops off as catch increases. To use another example, the difference in perceived utility of a natural area with a diversity of 205 species compared with one with

Table 1. A review of 20 conservation evaluation studies.

Reference	Description of measurement techniques and method for overall ranking
International biological program sites (Nicholson 1968, Clapham 1981)	No explicit measurement used. A set of areas judged to represent range of ecological variation, have rare biota or earth science features, have been well studied, or present opportunities for research.
Countryside evaluation, UK (Tubbs and Blackwood 1971)	Overall subjective ranking of sites based on rarity, outstanding features and wildlife reservoir value; subjective scoring of agricultural sites largely based presence of "unsown vegetation."
State natural areas, Wisconsin (Tans 1974)	Scoring of 8 criteria on several different scales (e.g., 1–10 and 1–5). These are summed into four separate indices: biological value, physical characteristics and use value, availability, and threat.
Biosphere reserves, World and Canada (Man and Biosphere Program 1974 and 1976, Francis 1982)	Calls for "best judgement" of representativeness, national importance, effectiveness as a conservation unit and opportunities for research and education. No explicit measurement used.
Local natural areas, Texas (Gehlbach 1975)	Scoring of 5 criteria on a 3-point scale and weighting criteria from 1 to 5. Overall index derived in a number of ways: weighted sum, weighted product, and unweighted sum.
Countryside evaluation, UK (Goldsmith 1975)	Empirical ratio measurements of extent, rarity, plant species richness, and animal species richness are multiplied together for each habitat type and summed across habitat types.
Local natural areas, Vermont (Sargent and Brande 1976)	Both quantitative and qualitative measurements of 12 criteria converted to 5-point scale of value. A natural science rating and planning rating obtained by summing 6 criteria scores for each.
Local natural areas, UK (Wright 1977)	Assigned scores of 1, 2, and 3 to both quantitative and qualitative data on 18 different criteria; 16 of these scores are summed into indices for a scientific appraisal and a management appraisal.
National system of nature reserves, Great Britain (Ratcliffe 1977)	Separate evaluations were undertaken for each habitat type, often using different criteria and almost always formulating criteria differently. Some criteria were quantified, some assessed using a scoring system, and others assessed subjectively based on guidelines. Final grading into 4 ordinal categories based on consensus or adjudication.
Local natural areas, the Netherlands (van der Ploeg and Vlijm 1978)	A review of 10 different methods. These generally involve ratio measurements of biological, soil, geomorphological, and sometimes cultural features. These data were often mapped and in 9 studies were summarized as 3–5 categories (e.g., low, medium, and high). The most common method of deriving these categories was the summation of criterion values, but sometimes the largest single criterion value was emphasized.
State natural aquatic areas, Idaho (Rabe and Savage 1979)	The occurrences of representative and unique aquatic and terrestrial features are scored 0 or 1, presence or absence. These numbers are summed for an overall ranking.
Local natural areas, Southern Ontario (Eagles 1980)	Nine criteria are used separately to assess a site's significance. A particular site must meet one or more of the criteria, which are generally minimum standards. Assessments are qualitative based on guidelines, lists, and classifications.
Ornithological sites, UK (Fuller 1980)	Ratio measurements of population size, diversity, community quality, and rarity. Standards are defined for each of criteria at national, regional, county, and local scales. Each criterion considered separately.
Territorial system of natural areas, Yukon (Theberge and others 1980)	Biophysical rating is a sum of scores of 1 (low), 2 (medium), and 3 (high) for landforms, wildlife, vegetation, and representativeness. Separate subjective ratings are made for threat and recreation potential. Guidelines are given for subjective ratings.
Large roadless areas, western USA (Klopatek and others 1981)	Ratio measurements of indices of the rarity and diversity of vegetation types; the abundance, rarity, and diversity of bird species; the rarity and diversity of mammal species; and the occurrence of endangered species. These are summed for an "overall ecological index."

Table 1. *Continued.*

Reference	Description of measurement techniques and method for overall ranking
State natural areas, North Carolina (Radforth and others 1981)	Five criteria assessed on a subjective five point scale (1–5) and one on a 10-point scale. These scores are summed for an overall score.
Evaluation of potential national parks, Canada (Parks Canada 1979 and 1981)	"Themes" characterizing natural regions are rated 1, 2, or 3 as to commonness on each candidate area and 1 or 2 for significance. These are multiplied and then summed for all "themes." The site with the largest score is deemed the most representative and thus most significant.
Wetlands, southern Ontario (Environment Canada and Ont. Ministry of Nat. Resources 1984)	Complex subjective scoring of many factors which are summed into biological, social, hydrological, and special features indices, and a grand total.
Comprehensive protected area system, northern Canada (Task Force on Northern Conservation 1984)	Subjective yes/no judgement of a site's value with respect to 7 criteria. Overall significance is assigned if site is significant for one or more individual criteria.
Wetlands, southern Ontario (Van Patter and Hilt 1985)	Government agencies, conservation groups, and wetland experts asked to rate each of a list of wetlands as either "more" or "less" important. This produced consensus on half of the sites. Five additional criteria used by study team to select other half.

Table 2. The four scales of measurement.

Measurement scale	Admissible relations and operations	Example of criteria measurement
Nominal	Equivalence =	"Provides habitat for indigenous species that are rare or endangered at the regional, provincial or national scale" yes or no (Eagles 1980)
Ordinal	Equivalence = Order <, >	Low: "small number of species present" Moderate: "high species diversity" High: "outstanding diversity for particular habitat type" (Wright 1977)
Interval	Equivalence = Order <, > Addition + Subtraction - Multiplication × Division /	"Number of provincially significant plant species," one species, score = 100, two or more species, score = 150 (Environment Canada and Ontario Ministry of Natural Resources 1984)
Ratio	Equivalence = Order <, > Addition + Subtraction - Multiplication × Division /	Number of tree species per half hectare (McKinnon 1982)

200 species may be much less than between two natural areas with five and ten species, respectively.

Uncertainty in measurement. Different sorts of uncertainty affect measurement of criteria. Inaccuracy may exist due to error in measurement. As well, when a parameter is estimated from sample measurements, the laws of probability determine the precision of the

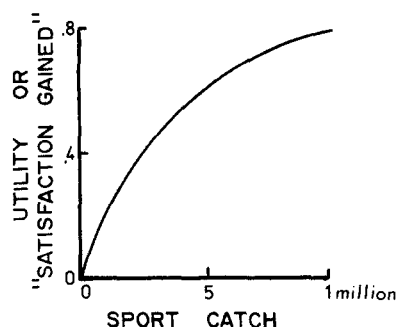


Figure 1. Example of a utility function for fishing sport catch (Holling 1978).

estimate (Steele and Torrie 1980). For example, Oviatt and others (1977) found as much variation in ecological variables within individual marshes as among different marshes and thus questioned the relevance of many evaluation methods. If the definitions of criteria and the values they may take are indistinct, measurements taken are necessarily "fuzzy" and may require the use of fuzzy set theory (Hwang and Yoon 1981, Efstathiou 1984, Zimmerman and others 1984). Finally, variance in the preferences of different people and interest groups have dramatic effect on the subjective scores and weights assigned to particular criteria and thus on the final ranking (for example, Holling 1978, Bakus and others 1982, Sobral and others 1981, Margules 1984).

Practice: Measuring the Value of Natural Areas

The specific approach to measurement used in the 20 evaluation studies is given in Table 1, and Figure 2 shows the number of studies that adopted particular

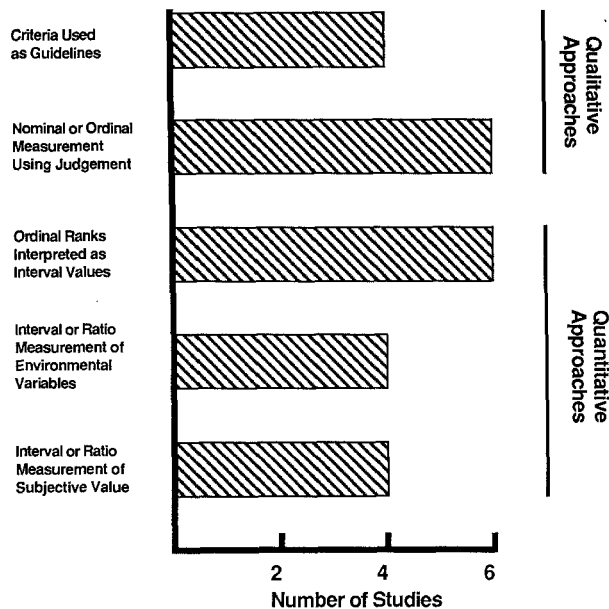


Figure 2. The numbers of studies using different forms of measurement. The sum of the numbers in the Figure exceeds 20 because some studies use more than one form of measurement.

approaches. Those systems that use quantitative (interval or ratio) measurement, either measure actual ecological variables in the field, or subjectively assess perceived "value" or "utility" based on the field measurements. For example, rarity can be expressed as the number of rare biota or scores may be assigned based on the "perceived value" of the rare biota. Several other studies interpret ordinal ranks as interval values (Figure 2). Systems utilizing qualitative "measurements" sometimes use criteria simply as guidelines to be applied based on judgment and professional opinion. In other cases, the criteria are systematically applied using some form of nominal categorization or ordinal ranking of sites as to their overall significance (Table 1 and Figure 2). Some studies incorporate more than one of the foregoing approaches (Table 1).

Quantitative approaches. In those studies where quantitative measurement has been used (Table 1 and Figure 2), methods varied with the quantity and quality of information available. In studies like Fuller's (1980) or Klopatek and others' (1981), the high quality of data allowed the valid use of ratio scales for measurement.

In evaluating ornithological sites, Fuller (1980) used criteria that lend themselves to ratio measurement: population size; diversity or number of species; and rarity. Moreover, extensive documentation existed for the avifauna of the sites being evaluated, such as the

numbers of different bird species occurring in different seasons and years. Thus it was possible to define for example, sites of national significance for population size as those supporting at least 1% of the estimated national population for that species.

Klopatek and others (1981) evaluated 321 tracts of land in the American west for the extent of natural vegetation; the abundance, rarity, and diversity of bird species; the diversity and rarity of mammal species; and the occurrence of endangered and threatened species. The information base used for this evaluation required the creation and integration of large computer data bases to contain all the necessary information. These data were then used to calculate separate indices for vegetation, birds, mammals, and endangered species, which were summed for an "overall ecological index." The vegetation index, for example, was calculated as the area of a particular vegetation type in the tract divided by total area occupied by that vegetation type in all 321 tracts. This was multiplied by a measure of the rarity of that vegetation type at the national, regional, and subregional scales. These values were calculated and summed for all vegetation types occurring in each tract.

Contrasting the foregoing fundamental measurement of environmental variables are systems that have assigned different quantitative "scores" or "points" to levels of each criterion. For example, Tans (1974) assigned points for size and buffer zone as follows: "Greatly exceeds minimum size, excellent buffer, no threat of encroachment," score = 8; "Adequate size and buffer," score = 6; "Adequate size, inadequate buffer," score = 4; "Inadequate size, adequate buffer," score = 2; and "Both inadequate," score = 0. Similarly, Wright (1977) converted low, medium, and high ordinal rankings to scores of 1, 2, and 3 and summed the scores for a number of criteria. This is improper application of addition to ordinal ranks (see Table 2). While such scoring has intuitive appeal, the allocation of points is arbitrary and subjective. Such scores are thought to represent the 'value' of the particular conditions to the decision makers involved. Utility theory offers more rigorous methods for the analysis of such preferences and assigning of scores.

A further problem is that scores assigned to different levels of criteria often have little relation to ecological theory and practice.

Qualitative approaches. All methods by definition involve at least nominal, if not ordinal, measurement by differentiating sites considered significant from those not. Some use explicit qualitative measurement, while others simply use criteria as guidelines and the notion of measurement is not mentioned (Table 1 and Figure 2). Several examples illustrate this.

The terrestrial conservation section of the International Biological Program (IBP) was a pioneering effort in identifying "ecological sites." It operated on a very broad scale in a time when there was little comprehensive survey data and information sources were often not comparable. While specific criteria (Nicholson 1968, Peterson 1975) were used as guidelines and made reference to in site reports (for example, Beckel 1975, Nettleship and Smith, 1975), systematic identification and comparative analysis of sites was, for the most part, lacking (Table 1).

Qualitative measurement of criteria sometimes depends heavily on subjective professional judgement. For example, the criteria advocated by the Task Force on Northern Conservation (1984) are questions as to whether a site is "significant," "necessary," "outstanding," or "essential." One particular criterion is extremely subjective, asking whether areas "contain sites necessary for the preservation of genetic diversity," yes or no. No definitions or guidelines are given. One must conclude that individual interpretation of the question would lead to great variance in the evaluation of sites.

Methods for identifying local "environmentally sensitive areas" in southern Ontario (Eagles 1980) are similar to the foregoing approach (Task Force on Northern Conservation 1984) in requiring judgements as to whether a site is "vital," "distinctive," or "unusual." In contrast, however, specific, detailed guidelines are developed to facilitate the judgements and enhance replicability. For example, one criterion is "the area represents a distinctive and unusual landform within the municipality, Ontario or Canada." The guideline for this criterion is that "the distinctive and unusual landforms are usually defined by the published and verbal comments of authorities in the field of earth science." This approach does not eliminate subjectivity, but rather makes it explicit, spelling out the basis of the judgement and facilitating discussion of that assessment.

The British Nature Conservation Review (Ratcliffe 1977), probably better than any study reviewed, discussed explicitly the issues of quantification and subjectivity and stated the assumptions underlying the method chosen. A scoring system was considered, but it was felt that scoring different habitat types for all the same criteria was inappropriate, that criteria are not independent, and that scoring would only lead to "spurious objectivity." Thus the approach adopted involved a different evaluation method and criteria for different ecosystem or habitat types. For example, grassland and woodland sites were evaluated separately. A final grading into four ordinal categories is the result of a comprehensive analysis of abiotic, biotic,

and cultural factors, some of which could be quantified and others not. Thus grading is "dependent on a series of value judgements" that, although based on consensus between many scientists, are nonetheless "subjective even when collective." A final point not often made in the conservation evaluation literature is that "values are also prone to change in time, and future generations may look at things rather differently" (Ratcliffe 1977).

A method unique among the 20 studies is the Delphi approach used by an environmental group in evaluating wetlands in southern Ontario (Van Patter and Hilt 1985). The Delphi method is a means used in attempts to achieve a collective decision on a particular issue (Linstone and Turloff 1975). Van Patter and Hilt (1985) asked government and nongovernmental organizations as well as a sample of wetland specialists to rate a list of sites as "more important" or "less important" and to suggest areas to be added to the list. Of the sites finally identified as "more important," general consensus among the respondents was reached on about half. The remaining half required an evaluation by the research team using the respondent's information and six additional criteria.

Using Multiple Criteria for an Overall Ranking

One main purpose of conservation evaluation is to reduce the amount of information to be considered in assessing the importance of areas. This need for data reduction has meant that the majority of evaluation systems have attempted to reduce environmental significance to one dimension or value. The extent to which this has been successful, or is possible, is a matter of opinion. Margules and Usher (1981) regard this as necessary, if not desirable, and "predict that ways of aggregating criteria into one index (or a number of indices) will become more sophisticated in the forthcoming decade."

While many view the derivation of a single index of "ecological value" as desirable, or at least a necessary evil, others contest the validity of such a notion (van der Ploeg and Vlijm 1978, Gotmark and others 1986, Nilsson 1986, van der Ploeg 1986). For example, van der Ploeg and Vlijm (1978) state: "If the aim is an ecological evaluation, all aspects should be studied in relation to each other. Any integration which simply adds aspect (criteria) values denies the existence of *ecosystems*. . . . Thus aspects like vegetation, fauna and geomorphology should either be combined from the beginning or they should not be integrated at all." Furthermore, they state that: "None of the studies gives a mathematical specification which is based on the 'ecological' relationships between the criteria used. There-

fore the formulae can be considered as pseudo-objective justifications of *a priori* judgements." It is indeed interesting to note that, while at one time derivation of a single index was common in evaluation in the Netherlands, it is now virtually absent (van der Ploeg 1986). Despite opinions and trends like the foregoing, a majority of the evaluation systems reviewed here used some type of overall summed index (Table 1 and Figure 3) and there is little indication of a decline in this practice in the more recent studies.

To review how a final ranking of sites has been derived in the 20 studies, we take two avenues. First, we introduce the theory of multicriteria evaluation, a discipline within operations research or management science. Secondly, these concepts are used to classify and assess the validity of the methods used in practice to rank sites in the 20 studies.

Theory: Multicriteria Evaluation Models

Multicriteria evaluation models are used to rank a series of alternatives on the basis of a number of criteria. The models vary markedly in the underlying assumptions and are designed to deal with different types of information. Two basic types of models exist: compensatory and noncompensatory (Hwang and Yoon 1981:24–25). Compensatory models generally result in a single number being calculated for each alternative. This generally requires that all criteria be measured on interval or ratio scales and are in comparable units in the sense that a drop in one criterion can be offset by a rise in another. This would imply, for example that the following two sites have equivalent overall value:

	Rarity value	Represent- ativeness value	Productivity value	Recreation value
Site 1	0	4	0	0
Site 2	1	1	1	1

Noncompensatory models, on the other hand, are used when the levels of different criteria cannot be compared; they are like "apples and oranges." Alternatives must then be assessed criterion by criterion.

The simple additive weighting method, the most common compensatory model, can be applied when all criteria are measured on an interval or ratio scale, all values are expressed in comparable units, and weights can be assigned to each criterion. That method involves simply multiplying all criterion values by their respective weights and summing the resulting values for each site. More depth on this method will be provided later.

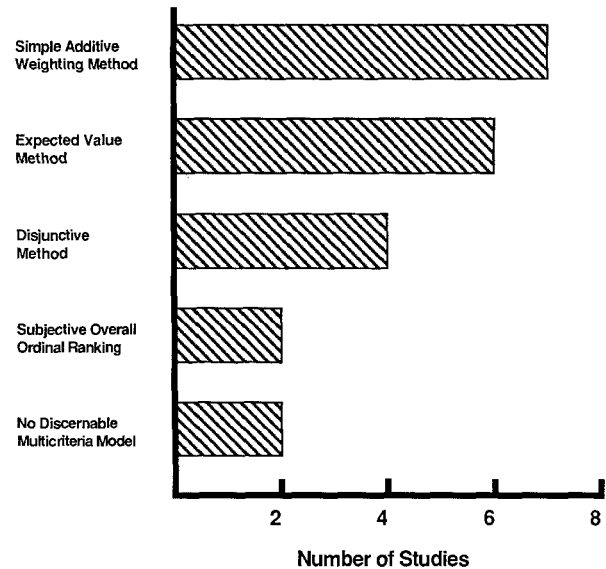


Figure 3. The number of studies using particular multicriteria evaluation methods.

Another compensatory model, the multiple attribute utility model, is closely related to the simple additive weighting model. Utility models are similar to cost-benefit analyses in attempting to convert all variables to a common scale of measurement. Instead of dollar value, utility models rank alternatives based on the overall expected "utility" of each (Bakus and others 1982). "Utility" is defined as the perceived benefit derived from a particular level of a criterion as discussed earlier. For example, an ecologist or decision maker might decide that the benefit of a site with ten rare plant species was 50% greater than a site with five rare plant species. Utility is a broad concept and can include values other than utilitarian ones, such as perceived intrinsic worth. Assessments of these utility functions are based on a series of systematic techniques for questioning, assessment, and comparison (Farquhar 1977, Zeleny 1981). The overall utility is calculated as the combined utilities of the individual criteria, which may be a simple sum, a weighted sum, a product, or other means of aggregation.

When criteria are measured on different scales and the different criteria are incomparable, the noncompensatory models are useful. To illustrate some of these models we will use a simple, hypothetical example of three sites—A, B, and C—that have been assessed using three criteria—rarity, diversity, and productivity. The values each criterion takes varies from 1 to 4.

	Rarity	Diversity	Productivity
Site A	2	1	4
Site B	2	1	1
Site C	2	3	1

"Dominance methods" (Hwang and Yoon 1981) rank an alternative higher when it is better than other alternatives for one or more criteria and of equal value for the other criteria. In the example, A *dominates* or is preferred over B, because the two have equal values for two criteria and A has a greater value than B for productivity. On the other hand, A does not dominate C because, while A has greater value for productivity, C has more value for diversity.

"Maximin" and "maximax" decision models (Hwang and Yoon 1981) rank alternatives based on their weakest or strongest criterion value, respectively, and ignore other criterion values. Under the maximin approach, sites A, B, and C are all preferred equally because each has a low score of 1 for at least one criterion. Using a maximax rule, site A would be selected as it has the maximum score of 4 for productivity.

Several models are based on minimum standards or thresholds. The "conjunctive method" (Hwang and Yoon 1981) ranks alternatives based on whether an alternative meets acceptable minimum standards for all criteria. The "disjunctive model" (Hwang and Yoon 1981) ranks based on an alternative meeting a minimum standard for at least one criterion regardless of its value for other criteria. If we were to apply minimum standards of 2 to all the criteria in the example, then using the conjunctive model, none of the three sites would be chosen, because none has a value of 2 or more for all the criteria. Under the disjunctive model, all three sites have at least one criterion with a value of 2 or more and thus are acceptable.

Several other models involve direct measurement of preferences for alternatives through comparison of each available alternative with others or with an ideal alternative. An ideal alternative in the example would be a site with a value of 4 for all three criteria. Ranking sites A, B, and C would consist of comparing each with the ideal and assessing which comes closest to that ideal. Full treatment of these and other decision models is available elsewhere (Starr and Zeleny 1977, Nijkamp 1980, Hwang and Yoon 1981, Zeleny 1981, Voogd 1983, Zimmerman and others 1984, Janssen and Nijkamp 1985).

Practice: Multicriteria Models Used to Evaluate Natural Areas

None of the systems reviewed (Table 1) explicitly stated the type of evaluation model used. As a result, we have classified the methods used by comparing the descriptions given with typologies of multicriteria evaluation models in Nijkamp (1980) and Hwang and Yoon (1981). Most evaluation models used in the systems reviewed generally correspond to three types of multicriteria evaluation models (Figure 3): the

simple additive weighting method (Hwang and Yoon 1981); what Nijkamp (1980:38) called the expected value method; and the disjunctive model (Hwang and Yoon 1981). The details of these three multicriteria models are outlined here to allow assessment of the validity of their use in the evaluation systems reviewed here. Figure 4 summarizes the requirements and assumptions of each of the three models. Figure 3 indicates how many of the 20 studies generally correspond to these three multicriteria models. Several methods do not resemble any multicriteria models (Figure 3) including the overall ordinal ranking in Ratcliffe (1977) which was discussed earlier.

Simple additive weighting method. The simple additive weighting method (Figure 4) is the multicriteria model most familiar to many people and that used in seven of 20 studies (Figure 3). Using this model, the alternatives, in our case natural areas, are ranked using the weighted sum of all the criteria measurements. Multiplication and addition are used and "therefore, the criteria values *must* be both *numerical* (i.e. measured using an interval or ratio scale) and *comparable*" (emphasis added; Hwang and Yoon 1981:101). Values for criteria may be either empirical measurements or subjective scores. The values of each criterion are standardized or normalized so that they are comparable between criteria. Standardization is neglected in some studies reviewed (for example, Tans 1974). Standardization often involves a common scoring system for all criteria. The assumption of comparability implies that similar changes in the levels of different criteria are equivalent (Hwang and Yoon 1981:102). For example, it is assumed that a high ranking of a site for rarity is equivalent to a high ranking for diversity. Whether this assumption is valid is probably more of a philosophical issue than a mathematical one. A further assumption is that all criteria are independent of each other, an assumption clearly violated by most ecological data (see Usher 1985, Gotmark and others 1986).

In referring to comparability, it is important to differentiate the "equivalence of units" from standardization or normalization. Criteria are measured using different units that may or may not be equivalent. For example, "we can say that an automobile is wider than it is tall, because both scores are expressed in the same units," whereas "we cannot say that an automobile is wider than it is heavy" (Ghiselli and others 1981). Standardization or normalization involves comparing the scores of a particular alternative with those of all other alternatives or with some real, or theoretical, maximum value.

For the simple additive weighting model, "it is also necessary to find a reasonable basis on which to form the weights reflecting the importance of each of the

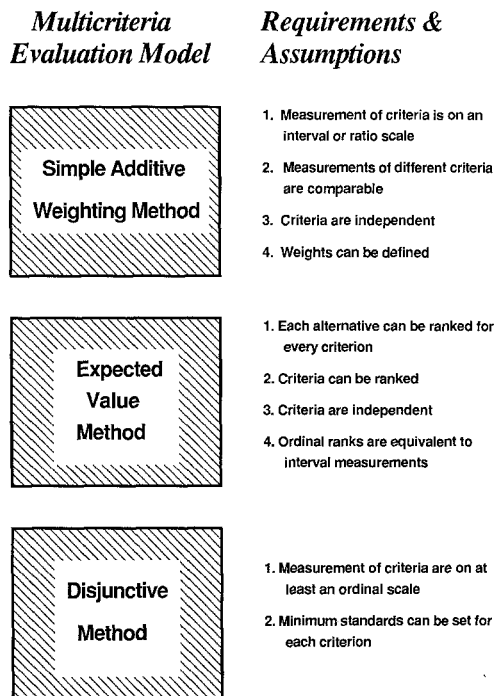


Figure 4. The requirements and assumptions of three multicriteria evaluation models.

criteria" (Hwang and Yoon 1981:101). Weights are necessarily subjective, but there are techniques for measuring subjective preferences precisely and translating these into numerical weights (Hwang and Yoon 1981:41, Zeleny 1981:185). Margules (1984) and Margules and Usher (1984) show that such subjective weights applied by different ecologists and environmental managers can vary widely. When no weights can be assigned subjectively, others methods can be used to assign weights based on an "objective" rationale.

A form of implicit weighting is applied in some systems (for example, Tans 1974) when scores assigned for different criteria have different maximum values and thus are not standardized so as to make the values comparable. So-called equal weights used in some systems (for instance, Wright 1977, Klopatek and others 1981) refer only to the explicit weights. Emphasis on a particular criterion varies with the number of criteria and how they are formulated.

Expected value method. The expected value method (Figure 4) is a crude form of the simple additive weighting method. The decision maker subjectively ranks each alternative for each criterion. For example, three sites may be ranked 1, 2, 3 for rarity and 2, 1, 3 for diversity. The criteria are also ranked; for example, rarity may get 1, diversity 2, and fragility 3. Each site's rank for each individual criterion is then multiplied by the criterion's rank. Thus a site with

rank 2 for diversity would be multiplied by the rank for the diversity criterion, 2, yielding 4. All these values are summed for each site and these sums used to rank the sites. "It is clear that this evaluation method is, however, a rather crude aggregation method based on non-permissible numerical operations on ordinal numbers" (Nijkamp 1980:38). Methods used by six of 20 studies, resemble this method (Figure 3). Two methods correspond closely with the expected value method in using numerical interpretation of both ordinal ranks of sites and of ordinal preferences for different criteria (Gehlbach 1975, Parks Canada 1979 and 1981). Four other studies involve only numerical interpretation of ordinal ranks of sites (Sargent and Brande 1976, Wright 1977, Theberge and others 1980, Radforth and others 1981). The approaches of the latter studies might also be classified as resembling the simple additive weighting method.

In assessing the validity of overall ranking based on any method and particularly the expected value method, an important consideration is the sensitivity of the rank to uncertainty. There may be uncertainty in field measurements, variation between people in assigning scores and in the weights given to different criteria, as well as fuzziness in the definitions of the criteria (see *Uncertainty in measurement* above). The precision of a site's overall rank or score is a function of the amount of uncertainty from all the sources mentioned. One must have a degree of confidence in the derived scores or ranks to defend them and base public policy decisions on their accuracy. Treatment of such uncertainty with mathematical validity necessitates more complex mathematical models (O'Bannon 1980, Sobral and others 1981, Bakus and others 1982, Zimmerman and others 1984).

The disjunctive model. One problem with the simple additive weighting method is that sites with high "overall value" may not be important with respect to any of the individual criteria (van der Ploeg and Vlijm 1978, Klopatek and others 1981). For example, the highest ranking area overall in the study by Klopatek and others (1981) ranked 60, 71, 52, and 19 with respect to birds, mammals, vegetation, and endangered species. Thus that area was not particularly important for any single reason, but rather average for all four, causing it to rank first overall. Emphasis on these "average" conservation values can "hide sites with outstanding qualities" (Gotmark and others 1986). An alternative model for ranking used by several systems reviewed by van der Ploeg and Vlijm (1978) as well as Eagles (1980), Fuller (1980), and the Task Force on Northern Conservation (1984) avoided this problem (Figure 3). In those systems, the overall ranking is

based on the highest rank each area has for any criterion. This is conceptually similar to the disjunctive multicriteria evaluation model (Hwang and Yoon 1981:64; see Figure 4). For example, the area with most rare species would automatically occur in the highest ranking class as would the most diverse area. Furthermore, under this model, if a site is selected because it is quite diverse, it is irrelevant how low it may rank for other criteria.

The point of all this discussion of multicriteria evaluation models is that many other models exist than those often used in conservation evaluation. Different models exist for use with different types of information to avoid inappropriate assumptions and invalid mathematical operations. The principles underlying these models are not difficult. In fact, many methods have been developed to emulate the psychology of human decision making.

Unless a multicriteria ranking scheme is based on valid assumptions, an area's rank may be an artifact with little real meaning. For example, Voogd (1981) found that a multicriteria evaluation model based on assigning scores to ordinal ranks resulted in the selection of different alternatives than a method designed specifically for use with ordinal measurements. Several of the evaluation models not often used may indeed be useful because they involve fewer assumptions than the simple additive weighting and expected value methods. In general, we find the disjunctive model (Figure 4) useful and have adopted it for our own work in Canada's Northwest Territories (Smith and others 1986:22) because the underlying assumptions are valid for the data used in that work and can be communicated easily.

Selecting the "Best" Evaluation Methods

Having examined problems in the ways in which measurement and multicriteria models have been used in conservation evaluation, we can suggest directions for improvement. Little can be said of specific ideal methods in particular situations because this depends on many factors, especially the assumptions that decision makers and the public are willing to accept. Nevertheless, we believe that the principles and assumptions on which methodologies rest must be rationalized ecologically, philosophically, and mathematically. To that end, Table 3 presents a series of principles that we suggest to guide the choice of evaluation methods. These principles have been adapted largely from Jansson and Nijkamp (1985), who provided a guide to the selection of multicriteria methods for a variety of environmental management decisions, including the evaluation of natural areas. Some of the principles

Table 3. Principles for the choice of evaluation method.

An evaluation method should:

- 1) Be based on principles and assumptions that are valid and easily illustrated
 - 2) Yield results understandable to decision makers and the public
 - 3) Make explicit subjective values and judgements
 - 4) Yield results that are repeatable given certain explicit assumptions
 - 5) Allow use of qualitative and quantitative information in a methodologically sound way
 - 6) Stimulate the imagination of decision makers and increase insight into the choice to be made
 - 7) Enable the use of information at different spatial scales
 - 8) Allow consideration of alternatives both separately and in combination
-

may seem to state the obvious, but based on our review of the literature, none can be considered self-evident.

As we have already stated, the underlying principles and assumptions or any method must be ecologically and mathematically valid and furthermore must be acceptable to the decision makers and the public (no. 1). The latter necessitates simple illustration of those underlying principles and assumptions. Grasp of the theory of a method also means that the results can be understood and accepted (no. 2).

Subjective values and judgements must be made explicit in order to assess their effect on the choices made (no. 3). Several authors have recently called for greater separation of the objective component of evaluation, that is, the ecological description of areas, from the subjective component, that of placing value on those descriptions (Roome 1984, Wathern and others 1986).

Methods of assessment must be repeatable given certain assumptions (no. 4). It is a given that subjective preferences for the different levels of criteria and for the criteria themselves vary dramatically among people. As a result, any method employing subjective judgements should provide specific guidelines to structure these judgements so as to enhance replicability.

Evaluation for nature conservation will almost invariably entail use of both qualitative and quantitative information (no. 5). Qualitative data can, in some cases, be scaled into numeric values, but this involves additional assumptions and greater complexity (Keeney and Raiffa 1976). Certain multicriteria models are designed to use both types of data without such transformation.

While one purpose of evaluation is to reduce the amount of information to be digested, another is to stimulate fresh ways of viewing the choice being made

and increase insight into what is being chosen and why (no. 6).

Evaluation of natural areas, almost by definition, entails the use of data at different spatial scales. Often the areas being evaluated vary in size by several orders or magnitude. Some criteria may be more applicable to sites of particular size (Margules and Usher 1984). Moreover, provision for the desirable mix of protected habitats in a region, large and small areas as well as the linkages between them, impinges on an evaluation process that may emphasize specific attributes of sites, like rarity and diversity. These considerations should be built into the method used (no. 7).

Following from the foregoing is the need to compare among combinations of candidate areas as well as evaluating each area on its own merits. For example, an important consideration might be what *combination* of natural areas in a region contains the greatest number of rare species (also see Rudis and Ek 1981). Simple assessment of the number of rare species found on each site neglects this larger perspective.

Discussion

Conflicting goals exist in developing any system for evaluating natural areas. Simplicity is required to make the methods accessible to nonspecialists in government, industry, and the public. The need for simplicity no doubt results in the exclusion of complex management science methods. Another, sometimes conflicting, need is for both ecological and mathematical validity. Much of the research on conservation evaluation to date has focused on increasing the ecological validity of methods of evaluation (for example, Margules and Usher 1981). These efforts sometimes have also involved increased mathematical validity (for instance, Dony and Denholm 1985). Nevertheless, we perceive a need for more clarity and rigor in the formulation of criteria measurements and the multicriteria evaluation models used. Furthermore, a more rational basis is needed for the selection of measurement and multicriteria methods. Our review of the pertinent theory and practice may help in that regard, as may our principles to guide method selection (Table 3).

Other multicriteria evaluation models exist than those most often used in evaluating natural areas. Some of these models involve fewer assumptions and assumptions that are more appropriate to ecological data than those underlying the assigning and summing of arbitrary scores. The principles underlying these methods are not difficult and are easily communicated (see *Theory: multi-criteria evaluation models*). Specific reference to the model being used and justifica-

tion of the underlying assumptions seems a reasonable demand (also see Janssen and Nijkamp 1985).

For a number of reasons, the common practice of aggregating criteria assessments into one index masks the complexity of the evaluation process. For example, in the case of evaluating sites for status as protected areas, the problem is much more complex than simply ranking areas. Each area's significance varies depending on how boundaries are drawn, because evaluation is as much an assessment of particular features as of composite geographic areas (see Smith and others 1986). Aggregating criteria also ignores the relationships between criteria (van der Ploeg and Vlijm 1978, Gotmark and others 1986). Inherent linkages exist between criteria such as representativeness and diversity or fragility, size, and rarity (Usher 1980 and 1985, Margules 1984, Margules and Usher 1984). Hence, treatment of each criterion as fully independent is unrealistic. Correlations between criteria can be accounted for in some multicriteria models that aggregate criterion values, but this necessarily involves increased complexity (Hwang and Yoon 1981).

In a similar vein, generally there are natural groupings of criteria that might be applied in separate evaluations, such as ecological, cultural, or human use and management criteria (Smith and Theberge 1986a). Moreover, not all criteria may be applicable to every kind of natural area. For example, rarity and fragility might better apply to small sites and representativeness and naturalness to larger sites (Margules and Usher 1984). Alternatively, different criteria could be applied during different stages of the evaluation process (Margules 1984, Margules and Usher 1984, Nilsson 1986).

The reduction of information to that of most salience is a basic need, if not the purpose of conservation evaluation. However, it is important "to avoid distortions due to aggregation while not presenting the decision maker with too many data to integrate" (Hollick 1981). Evaluation methods should enhance our grasp of the choices being made and not cloud important issues or conceal value judgements. Furthermore, the assumptions of any evaluation method should be clearly stated and rationalized in terms of the data and areas being compared.

Acknowledgments

This research was carried out as background for an evaluation method for environmentally significant areas in the Northwest Territories of Canada (Nelson and others 1985, Smith and others 1986, Smith and Theberge 1986b). It forms part of ongoing research on environmental planning and management in

northern Canada (Theberge and others 1980, Fenge 1982 and 1985, Bastedo and others 1984, Nelson and Jessen 1984). Support for this research was from the Canadian Arctic Resources Committee, the Canadian Department of Indian Affairs and Northern Development, and the Northwest Territories Department of Renewable Resources. The University of Waterloo Faculty of Environmental Studies supplied computing resources. J. G. Nelson, R. Dorney, G. Lee, D. Cole, and C. Price provided critical comment on earlier versions of the manuscript.

Literature Cited

- Bakus, G., W. Stillwell, S. Latter, and M. Wallerstein. 1982. Decision-making with applications for environmental management. *Environmental Management* 6:493–504.
- Bastedo, J. D., J. G. Nelson, and J. B. Theberge. 1984. An ecological approach to resource survey and planning for environmentally significant areas: the ABC method. *Environmental Management* 8:125–134.
- Beckel, D. K. B. (ed.). 1975. IBP ecological sites in subarctic Canada: areas recommended as ecological sites in Region 10, Yukon and Northwest Territories boreal forest to tree-line. University of Lethbridge Production Services, Lethbridge, 163 pp.
- Clapham, A. 1981. The IBP survey of conservation sites: an experimental study. Cambridge University Press, Cambridge, UK.
- Dony, J. G., and Denholm, I. 1985. Some quantitative methods of assessing the conservation value of ecologically similar sites. *Journal of Applied Ecology* 22:229–238.
- Eagles, P. F. J. 1980. Criteria for the designation of environmentally sensitive areas. Pages 68–79 in S. W. Barrett and J. L. Riley (eds.), Protection of natural areas in Ontario. Working Paper no. 3, Faculty of Environmental Studies, York University.
- Eichhorn, W. 1978. What is an economic index? An attempt of an answer. Pages 3–40 in W. Eichhorn, R. Henn, O. Opitz, and R. W. Shephard (eds.), Theory and applications of economic indices. Physica, Wurzburg, Austria.
- Efstathiou, J. 1984. Practical multi-attribute decision-making and fuzzy set theory. Pages 307–322 in H. J. Zimmerman, L. A. Zadeh, and B. R. Gaines (eds.), Fuzzy sets and decision analysis. North-Holland, Amsterdam.
- Environment Canada and Ontario Ministry of Natural Resources. 1984. An evaluation system for wetlands of Ontario south of the precambrian shield. Ottawa.
- Farquhar, P. H. 1977. A survey of multiattribute utility theory and applications. Pages 59–89 in M. K. Starr and M. Zeleny (eds.), Multiple Criteria Decision Making. North-Holland, Amsterdam.
- Fenge, T. 1982. Towards comprehensive conservation of environmentally significant areas in the Northwest Territories, Canada. *Environmental Conservation* 9:305–313.
- Fenge, T. 1985. Conservation of Polar Bear Pass, Bathurst Island, and the emerging comprehensive conservation policy for northern Canada. *Environmental Conservation* 12:231–240.
- Francis, G. 1982. Nomination of Canadian biosphere reserves. Canada/Man and Biosphere Program Report no. 15, Ottawa.
- Fuller, R. J. 1980. A method for assessing the ornithological interest of sites for conservation. *Biological Conservation* 17:229–239.
- Gehlbach, F. R. 1975. Investigation, evaluation and priority ranking of natural areas. *Biological Conservation* 8:79–88.
- Ghiselli, E., J. Campbell, and S. Zedeck. 1981. Measurement theory for the behavioral sciences. W. H. Freeman, San Francisco.
- Goldsmith, F. B. 1975. The evaluation of ecological resources in the countryside for conservation purposes. *Biological Conservation* 8:89–95.
- Goldsmith, F. B. 1983. Evaluating nature. Pages 233–246 in A. Warren and F. B. Goldsmith (eds.), Conservation in perspective. Wiley, Chichester, UK.
- Goldstein, G., and M. Hersen (eds.). 1984. Handbook of psychological assessment. Pergamon, New York.
- Gotmark, F., M. Ahlund, and M. Eriksson. 1986. Are indices reliable for assessing conservation value of natural areas? An avian case study. *Biological Conservation* 38:55–73.
- Hollick, M. 1981. The role of quantitative decision-making methods in environmental impact assessment. *Journal of Environmental Management* 12:65–78.
- Holling, C. S. (ed.). 1978. Adaptive environmental assessment and management. John Wiley, New York, 377 pp.
- Hwang, C., and K. Yoon. 1981. Multiple attribute decision making: methods and applications a state-of-the-art survey. Springer-Verlag, New York.
- Janssen, R., and P. Nijkamp. 1985. A multiple criteria evaluation typology of environmental management problems. Pages 495–514 in Y. Haines and V. Chankong (eds.), Decision making with multiple objectives. Springer-Verlag, Berlin.
- Keeney, R. L., and H. Raiffa. 1976. Decisions with multiple objectives: preferences and value tradeoffs. Wiley, New York.
- Klopatek, J. M., J. T. Kitchings, R. J. Olson, K. D. Kumar, and L. K. Mann. 1981. A hierarchical system for evaluating regional ecological resources. *Biological Conservation* 20:271–290.
- Linstone, H., and M. Turoff (eds.). 1975. The delphi method, techniques and applications. Addison-Wesley, Reading, Massachusetts.
- Man and Biosphere Program. 1974. Task force on: criteria and guidelines for the choice and establishment of biosphere reserves. Man and Biosphere Program Report no. 22, Paris.
- Man and Biosphere Program. 1976. Guidelines for the selection of biosphere reserves in Canada. Canada/MAB Report 6, Ottawa.
- Margules, C. R. 1984. Conservation evaluation in practice. II. Enclosed grasslands in the Yorkshire Dales, Great Britain. *Biological Conservation* 18:153–168.
- Margules, C., and M. B. Usher. 1981. Criteria used in assessing wildlife conservation potential: a review. *Biological Conservation* 21:79–109.
- Margules, C. R., and Usher, M. B. 1984. Conservation evaluation.

- ation in practice. I. Sites of different habitats in Northeast Yorkshire, Great Britain. *Biological Conservation* 18:153–168.
- Mason, R. O., and E. B. Swanson. 1981. Measurement for management decision. Addison-Wesley, Reading, Massachusetts, 550 pp.
- McKinnon, J. 1982. National conservation plan for Indonesia. National Parks Development Project of the Food and Agriculture Organization of the United Nations. Bogor. Food and Agriculture Organization of the United Nations.
- Nelson, J. G., and S. Jessen. 1984. Planning and management of environmentally significant areas in the Northwest Territories: issues and alternatives. Canadian Arctic Resources Committee, Ottawa.
- Nelson, J. G., P. G. R. Smith, and J. B. Theberge. 1985. Environmentally significant areas in the Northwest Territories: their role, identification, designation and implementation. *Environments* 17:91–109.
- Nettleship, D. N., and P. A. Smith. 1975. Ecological sites in northern Canada. Canadian Committee for the International Biological Programme, Conservation Terrestrial—Panel 9, Ottawa.
- Nicholson, E. M. 1968. Handbook to the conservation section of the International Biological Program. Blackwell, Oxford, UK.
- Nijkamp, P. 1980. Environmental policy analysis. Operational methods and models. Wiley, New York, 283 pp.
- Nilsson, C. 1986. Methods of selecting lake shorelines as nature reserves. *Biological Conservation* 35:269–291.
- O'Bannon, K. 1980. Use of value functions in environmental decisions. *Environmental Management* 4:3–6.
- Ott, W. 1978. Environment indices: theory and practice. Ann Arbor Science, Ann Arbor, Michigan.
- Oviatt, C., S. Nixon, and J. Garber. 1977. Variation and evaluation of coastal salt marshes. *Environmental Management* 1:201–211.
- Parks Canada. 1979. Parks Canada policy. Department of the Environment. Ottawa.
- Parks Canada. 1981. Terms of reference for a regional analysis of natural region 5, the Rocky Mountains. Ottawa.
- Peterson, E. B. 1975. Development of selection criteria for ecological reserves. Pages 85–90 in R. Franson (ed.), The legal aspects of ecological reserves creation and management in Canada. IUCN Environmental Policy and Law Paper no. 9, Morges, Switzerland.
- Rabe, F. W., and N. L. Savage. 1979. A methodology for the selection of aquatic natural areas. *Biological Conservation* 15:291–300.
- Radford, A. E., D. K. S. Otte, L. J. Otte, J. R. Massey, and P. D. Whitson. 1981. Natural heritage: classification, inventory and information. University of North Carolina Press, Chapel Hill, 485 pp.
- Ratcliffe, D. (ed.). 1977. A nature conservation review, vol. 1. Cambridge University Press, Cambridge, UK.
- Ratcliffe, D. 1986. Selection of important areas for wildlife conservation in Great Britain: the Nature Conservancy Council's approach. Pages 135–159 in M. Usher (ed.), Wildlife conservation evaluation. Chapman and Hall, London.
- Ray, G. C. 1975. Survey of critical marine habitats and requirements for their conservation. Pages 15–59 in Working Paper no. 1 of and International Conference on Marine Parks and Reserves. Tokyo, 12–14 May 1975. IUCN Publication New Series no. 37. Morges, Switzerland.
- Roome, N. J. 1984. Evaluation in nature conservation decision-making. *Environmental Conservation* 11:247–252.
- Rudis, V., and A. Ek. 1981. Optimization of forest island spatial patterns: methodology for analysis of landscape pattern. Pages 241–256 in R. L. Burgess and D. Sharpe (eds.), Forest island dynamics in man-dominated landscapes. Springer-Verlag, New York.
- Sargent, F. C., and J. H. Brande. 1976. Classifying and evaluating unique natural areas for planning purposes. *Journal of Soil and Water Conservation* 31:113–116.
- Smith, P. G. R., and J. B. Theberge. 1986a. A review of criteria used to evaluate natural areas. *Environmental Management* 10:715–734.
- Smith, P. G. R., and J. B. Theberge. 1986b. Evaluating biotic diversity in environmentally significant areas in the Northwest Territories of Canada. *Biological Conservation* 36:1–18.
- Smith, P. G. R., J. G. Nelson, and J. B. Theberge. 1986. Environmentally significant areas, conservation and land use management in the Northwest Territories. Heritage Resources Centre, University of Waterloo, Technical Paper 1, Waterloo, Canada.
- Sobral, M., K. Hipel, and G. Farquhar. 1981. A multi-criteria model for solid waste management. *Journal of Environmental Management* 12:97–110.
- Starr, M. K., and M. Zeleny. 1977. Multiple criteria decision making. North-Holland, Amsterdam.
- Steele, R. G. D., and J. H. Torrie. 1980. Principles and procedures of statistics: a biometrical approach. McGraw-Hill, New York, 633 pp.
- Tans, W. 1974. Priority ranking of biotic natural areas. *Michigan Botanist* 13:31–39.
- Task Force on Northern Conservation. 1984. The report of the task force on northern conservation. Department of Indian Affairs and Northern Development, Ottawa.
- Theberge, J. B., J. G. Nelson, and T. Fenge (eds.). 1980. Environmentally significant areas of the Yukon Territory. Canadian Arctic Resources Committee, Ottawa.
- Tubbs, C. R., and Blackwood, J. W. 1971. Ecological evaluation of land for planning purposes. *Biological Conservation* 3:169–172.
- Usher, M. B. 1980. An assessment of conservation values within a large site of special scientific interest in North Yorkshire. *Field Studies* 5:323–348.
- Usher, M. B. 1985. Implications of species–area relationships for wildlife conservation. *Journal of Environmental Management* 21:181–192.
- Usher, M. (ed.). 1986. Wildlife conservation evaluation. Chapman and Hall, London.
- van der Ploeg, S. W. F. 1986. Wildlife conservation evaluation in the Netherlands: a controversial issue in a small country. Pages 161–180 in M. Usher (ed.), Wildlife conservation evaluation. Chapman and Hall, London.
- van der Ploeg, S. W. F., and L. Vlijm. 1978. Ecological evaluation, nature conservation and land use planning with par-

- ticular reference to methods used in the Netherlands. *Biological Conservation* 14:197–221.
- Van Patter, M., and S. Hilt. 1985. Important wetlands in Ontario south of the precambrian shield. Federation of Naturalists, Toronto.
- Voogd, H. 1981. Qualitative multicriteria evaluation for development planning. *Canadian Journal of Regional Science* 4:73–87.
- Voogd, H. 1983. Multicriteria evaluation for urban and regional planning. Pion, London.
- Wathern, P., S. Young, I. Brown, and D. Roberts. 1986. Ecological evaluation techniques. *Landscape Planning* 12:403–420.
- Wright, D. F. 1977. A site evaluation scheme for use in the assessment of potential nature reserves. *Biological Conservation* 11:293–305.
- Zeleny, M. 1981. Multiple criteria decision making. McGraw-Hill, New York, 563 pp.
- Zimmerman, H. J., L. A. Zadeh, and B. R. Gaines. 1984. Fuzzy sets and decision analysis. North-Holland, Amsterdam, 522 pp.