Multicriteria Evaluation for Ecosystem Services:

A Brief Primer

Lynn A. Maguire

Nicholas School of the Environment, Duke University

February 2014

Acknowledgments

Thanks go to Lydia Olander and Emily Schieffer at the Nicholas Institute for Environmental Policy Solutions at Duke University for proposing the example and executing the figures.

Introduction

The purpose of this primer is to suggest how and where structured methods of multicriteria evaluation might help agencies integrate ecosystem services into their planning processes. I have taken the generalized agency planning process as a starting point, viewing multicriteria evaluation as a supplement to that scheme. Multicriteria evaluation can add clarity and facilitate quantification throughout this process. It is appropriately described as a non-monetary approach for comparing stakeholder preferences and analyzing trade-offs. Although the focus here is on adding value by considering ecosystem services, the methods are also useful generally for decision problems where trade-offs must be made among multiple objectives of resource management.

There are many texts on multicriteria analysis presented at various levels (see Annotated Bibliography), as well as training sessions in structured decision making for federal agency land managers and staff analysts (e.g., http://nctc.fws.gov/courses/SDM/home.html). There are many varieties of multicriteria analysis. Here we follow the principles of multiattribute utility analysis (MAU). The intended audience for this primer is agency planners who have not had formal training in decision analysis. I emphasize concepts and stages of structured decision making, suggest a few simple ways to implement common elements of decision analysis, and advise agency planners when assistance from others with specialized training would be useful, or even necessary. An increasing number of federal agency consultants are being trained in structured decision making and are available to help.

I present these methods using a decision about management of forested wetlands on a national forest (See textbox). Briefly, a national forest wants to enhance forested wetlands, for their own sake and as essential habitat for several at-risk species (two birds, a salamander and a fish). Land managers are considering two alternatives to status quo management: (1) modifying water releases from upstream reservoirs to increase drought-year flows to benefit forested wetlands on the national forest, and (2) restoring a former wetland on the national forest through damming and dredging. The first would affect upstream agricultural interests, and the second would benefit downstream landowners currently subject to flooding, as well as improving conditions on the forest for recreationists, who watch birds, fish, and canoe.

Example - Restoring Eco Forest wetland habitat for species and water flow.

Eco Forest is a federally managed public land that contains habitat for two federally listed endangered migratory bird species and two threatened aquatic species – a salamander and a freshwater fish. A variety of external drivers have led to the loss of riparian wetlands that were important habitat for all of these species. Eco Forest also has high visitation from neighboring communities and states for wildlife viewing, fishing, and canoeing. It is beginning to face crowding at popular sites.

Eco Forest managers have been reviewing information on status and trends for the forest and see a continuing decline in wetland area and reductions in populations of migratory birds and salamanders in their forest. They also have seen a slow but steady increase in regional visitation.

Eco Forest managers assessed the various drivers for these changes and saw reductions in dry-season flows and increased sedimentation as likely drivers of wetland loss. They also noted that climate change may be contributing to this trend. Looking at habitat for the 4 important species, they noted that the river habitat and buffers that remain in the forest and surrounding region seem sufficient for the freshwater fish, but that loss in habitat both in the forest and upstream are impacting the other species. Visitation seems to be increasing in parallel with population growth in the region.

Based on this information and meetings with regional experts, as well as community and stakeholder leaders, Eco Forest managers have set the following objectives and measures of success (metrics):

1. Eco Forest wetland habitat and associated at-risk species
    - Number of high quality wetland habitat acres (on Eco Forest land)
    - Numbers of birds, salamanders and fish
2. Recreational opportunities.
    - Number of people-access days on Eco Forest lands and waters
    - Number of wildlife contact opportunities on Eco Forest lands and waters OR improving satisfaction results in a survey of users
3. Upstream wetland habitat and water quality
    - Number of high quality wetland habitat acres (upstream of Eco Forest)
    - Numbers of birds, salamanders and fish
    - Rate of sedimentation (from upstream) relative to tilled acres of crop land and conversion to urban uses.
    - Water use (upstream), relative to water demand
4. Downstream wetland habitat
    - Number of high quality wetland habitat acres (downstream of Eco Forest)
    - Numbers of birds, salamanders and fish
5. Downstream flood risk
    - Number of water flow events that are above the threshold for flooding
    - # of people affected by flooding

Eco Forest managers have identified two alternative ways to increase wetland habitat area around the river on their lands.

1. Restore a former wetland in a bend in the river by dredging and building a low dam on the river that will slow surface water flow and flood the wetland. This area is in the lower portion of the watershed and will help reduce flooding to the downstream community.
2. Work with upstream water users to increase releases from upstream reservoirs, re-flooding former wetland areas. These areas are in the upper watershed, reducing water supply for some of the agricultural communities that occupy this area. This would only be a problem in drought years.

In addition to the activities on Eco Forest land, the Eco Forest managers would like to develop a program with USDA and the local farm community to increase buffer areas around the river to increase habitat and reduce sedimentation.

Eco Forest managers are working with upstream and downstream communities and other stakeholders to understand their perspectives and concerns.

Eco Forest managers worked with consultants to predict outcomes for their two management alternatives, using the measures identified above.

Measures	Alt 1 – downstream dam	Alt 2 – upstream release
Number of high quality wetland habitat acres (on Eco Forest land; upstream, downstream)	On EF 100 acre (one wetland) Upstream USDA program leads to buffers which provide another 5 acres of habitat. No direct downstream change	On EF 90 acre (multiple wetlands); can also result in around 30 acres on non-EF land upriver. Upstream USDA program leads to 5 acres of habitat
Numbers of birds, salamanders and fish in 10 years	(better habitat) +20 breeding pairs of bird sp 1 ? breeding pairs of bird sp 2 50 more salamanders 20% more of fish sp 1	(better corridor) +5 breeding pairs of bird sp 1 10% increase in breeding pairs of bird sp 2 no more salamanders 10% more of fish sp 1
Number of people access-days on high wildlife opportunity Eco Forest lands and waters (if accompanying walkway constructed)	20% more people can be accommodated in high wildlife viewing and easy access areas	40% more people can be accommodated, but low access, so expect to accommodate only 3% increase in user community
Water use (upstream) relative to water demand	None	5% drop in reservoir and water available for agriculture if releases occur in driest periods. If high water demand crops increase in region, demand will outstrip supply in these periods
Number of water flow events that are above the threshold for flooding	Decrease 5%	No change
Number of people affected by flooding	20% fewer people affected per flooding event	No change

Organizing Management Objectives Hierarchically

Being comprehensive, but not redundant.

One product of the scoping stage of a planning process is a list of objectives that resource managers want to influence by taking action. These are often presented as a list, but it can be useful to organize them hierarchically to show relationships among them (Fig. 1), linking the major categories of objectives (e.g., agricultural production, at-risk species and their habitats) to more specific aspects of those categories that are important (e.g., the types of at-risk species that might be affected by management – two bird species, a fish and a salamander). A hierarchical organization helps detect omissions (e.g., Are there other at-risk species or habitats that should be included?). And, it helps detect redundancy, which can lead to double-counting when evaluating alternatives. (See below for the converse problem of under-counting due to missing objectives.) In this example, we include both at-risk species and their habitats because wetland habitat is considered a valued objective in its own right, even if it does not currently support the at-risk species.

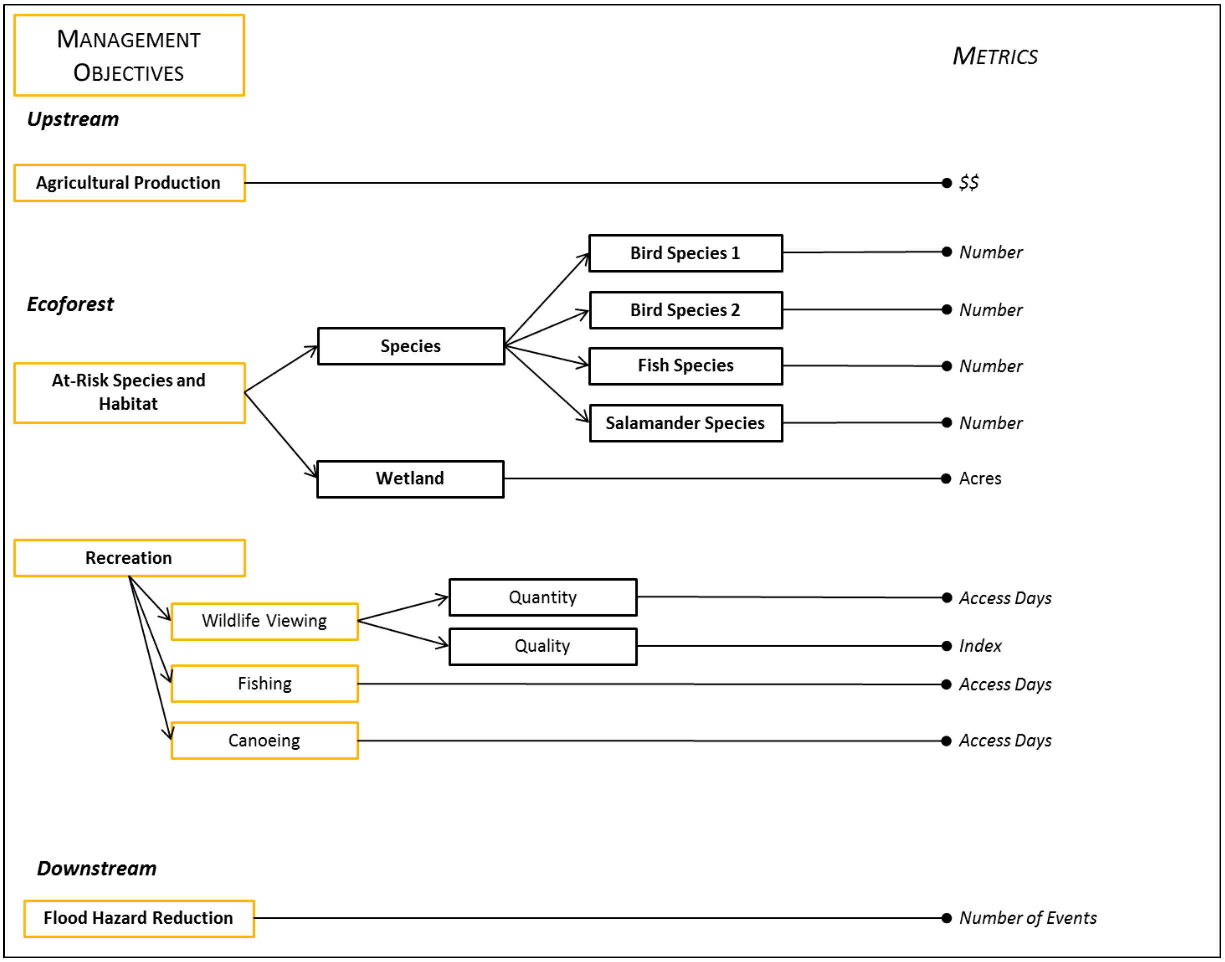


Figure 1. A possible hierarchy of objectives and measures for the Eco Forest wetlands example.

Distinguishing means/actions from ends/objectives.

One principle of constructing objectives hierarchies like the one in Figure 1 is to include only what has value in itself, excluding things that are valuable only because of their contribution to the production of other valued items. In particular, objectives hierarchies should not contain actions, even if those actions create value by positively affecting achievement of objectives. So, we do not include actions like working with upstream stakeholders to modify reservoir releases in the objectives hierarchy. These actions do belong in graphical, mental or mathematical models used to depict relationships between actions taken and achievement of objectives (e.g., state and transition models), as illustrated in the graphical model in Figure 2, where arrows indicate that a factor on the originating end of an arrow influences (positively or negatively) the amount of the item on the terminating end (e.g., reservoir releases influence dry-season flows upstream of forested wetlands within the national forest). Unlike an objectives hierarchy, which has an orderly branching structure (Fig. 1), a means-ends network can become a tangled web of arrows, showing that an action taken (e.g., modification of the schedule of reservoir releases) can affect many ecosystem elements and processes, both proximate and distant in space and time (e.g., reservoir releases affect dry-season flows upstream of national forest wetlands, changing their extent and status, which, in turn, affect health and numbers of wetland-dependent species, on which both passive and active recreation may depend). A means-ends network presents qualitative relationships among management activities and ecosystem responses, all or parts of which can then be translated into quantitative models that predict how management actions will influence achievement of planning objectives.

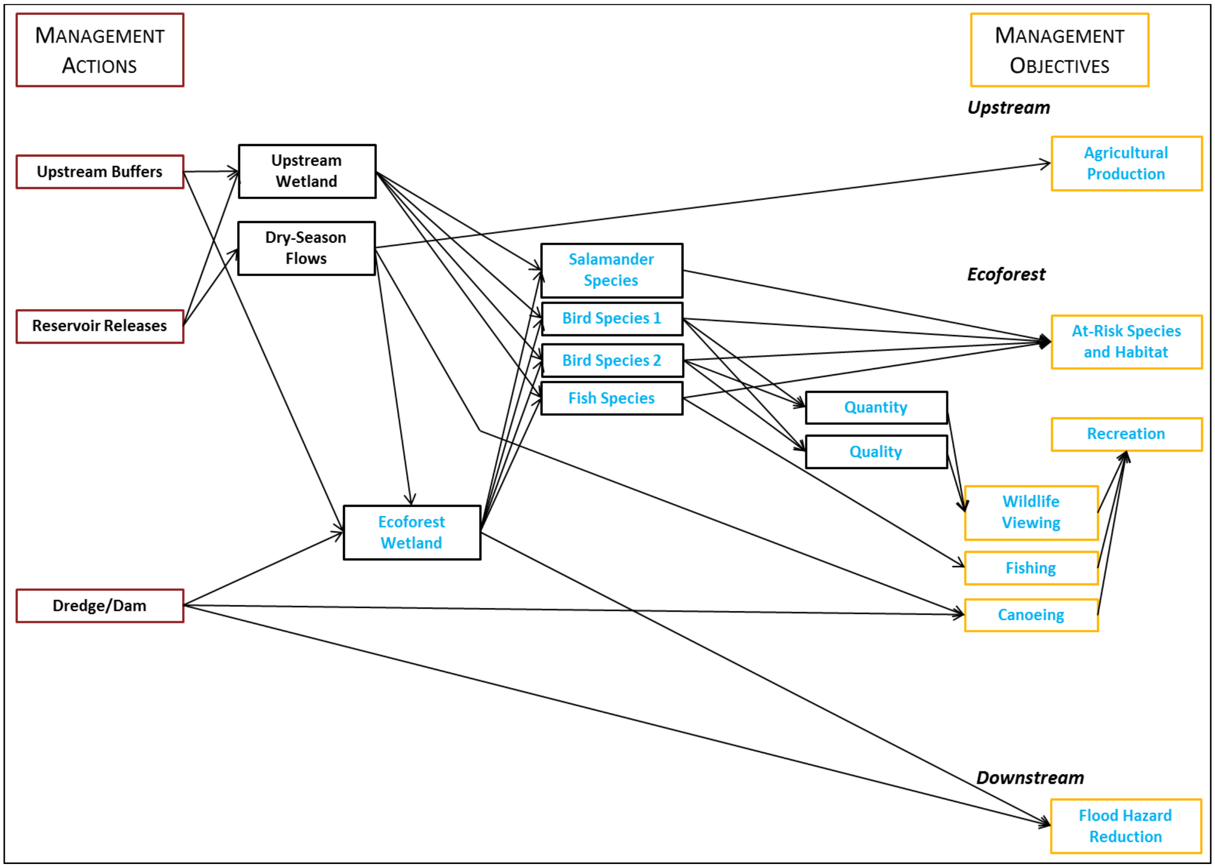


Figure 2. A possible means-ends network for the Eco Forest wetlands example. Arrows indicate that the element on the originating end influences (positively or negatively) the element on the terminating end (e.g., reservoir releases affect dry-season flows, which in turn affect wetland extent and health). Elements on the far left are management actions that might be taken. Elements in blue are part of the objectives hierarchy in Figure 1, making the link between actions taken and achievement of management objectives.

Making the hierarchy inclusive of all stakeholders.

The objectives hierarchy in Figure 1 includes objectives important to stakeholders beyond the boundaries of federal land area addressed by this planning process, including (1) farmers upstream whose production may be affected by changes in reservoir management, (2) landowners downstream whose risk of flooding may be affected by reservoir management upstream and by wetlands restoration on the federal forest land, as well as (3) members of the public who may never visit the forest, but who derive satisfaction from its existence. Making the objectives hierarchy comprehensive in this way facilitates communication with diverse stakeholders and creation of management alternatives that can garner widespread support.

Using the hierarchy to prompt brainstorming.

Organizing objectives hierarchically (Fig. 1) and graphically representing the links between actions that might be taken and effects on objectives (Fig. 2) can prompt managers and stakeholders to think creatively about how to accomplish their goals. For example, seeing the influence of upstream flows on Ecoforest wetlands may suggest partnering with upstream water users to modify reservoir operations, an option that might previously have seemed outside the scope of this planning problem.

Including ecosystem services as objectives

Consulting the full range of potential users of forest goods and services, including members of the public who may never visit the forest, but nonetheless value its existence as a whole ecosystem or for any of the elements in it, is essential to making sure the objectives hierarchy is complete. In the ecosystem services literature, these interested parties are often termed "beneficiaries." Different beneficiary groups (e.g., hunters vs. birdwatchers) may emphasize different goods and services and may assign different values to receipt of the same service (e.g., an increase in numbers of an at-risk bird species).

One sometimes controversial point in articulating ecosystem services as objectives of land management is deciding which services have value in themselves (and therefore belong in the objectives hierarchy) and which have value only for their contributions to the production of another ecosystem service. For use-values, such as recreation or agricultural production, it is usually obvious that these are ecosystem services in themselves. For non-use values, such as the value received just because at-risk species or their habitats exist, whether or not such values should be considered management objectives depends on whose perspective is used to frame the decision context. Some users (beneficiaries) may value these "intermediate" services, such as wetland habitat, and others may not. The hierarchy in Figure 1 includes both use values (e.g., fishing, wildlife viewing) and non-use values (e.g., wetlands, at risk-species) among management objectives.

Organizing ecosystem services objectives hierarchically can help ensure that the list of ecosystem goods and services is complete, reflecting the needs and desires of the full suite of stakeholders, but not redundant, alleviating the concern about double-counting in ecosystem services assessments. Constructing both an objectives hierarchy (Fig. 1) and a network of influences on achievement of those objectives (Fig. 2) helps differentiate ecosystem services as end goals from the intermediate ecological elements and functions that help produce those services. This distinction helps clarify the joint production of ecosystem services (where one management action may influence the production of more than one type of ecosystem services).

If the decision problem represented in an objectives hierarchy is going to receive a formal, quantitative evaluation of alternatives where there are trade-offs to be made among conflicting objectives, it would be a good idea to have the hierarchy reviewed by a specialized consultant to make sure that the objectives structure meets the assumptions of independence necessary for such analysis. It is hard for nonspecialists to do this review themselves.

Defining Measures of Success Clearly

Describing what is to be measured, when and how

In order to evaluate different management alternatives that might be used to meet agency goals, each objective in the hierarchy must be represented by a measureable quantity or quality, something that can be observed or predicted for each alternative. Even when these measures seem obvious, such as using the dollar value of agricultural products upstream of the national forest to express the effect of an alternative, care must be taken to define measures clearly: Over what area? What kinds of agricultural products? What period of time? What price-reporting service? To express flood frequency clearly, we must answer these questions: What water level measured where and when constitutes a flood event? Over what time period will we express frequency (monthly, annually)? Over what period might we average the number of flood events (one year, 10 years)? Answers to questions like these depend on the decision context – Whose concerns will be included? What measures are meaningful to them?

Using proxy measures. When performance of alternatives on a particular objective might be difficult or expensive to measure, as might be the case for assessing the size of populations of at-risk species, proxy measures that are easier to observe and demonstrated to be well-correlated with the measure that is really of interest can be used instead. Examples of proxy measures might be using acres of high-quality wetland habitat as a proxy for populations of wetland-dependent species or miles of stream open to fishing and populated by game fish species as a proxy for actual number of anglers. In the case where acres of wetland habitat is used as a proxy for numbers of at-risk species, as well as a direct measure of wetland habitat, care should be taken later in the analysis to make sure that the weight put on acres of wetland relative to other measures reflects the fact that it is doing double (or perhaps triple or quadruple) duty by standing in for other measures. More discussion of how and why to weight measures is included below.

Using Qualitative Measures Correctly

Including what's important, but hard to measure.

Some important features of an ecosystem are hard to express quantitatively. Using proxy measures is one solution, such as using numbers of waterfowl seen in a day to express the quality of a wildlife viewing experience. However, another solution is to define categories of performance verbally, such as characterizing the viewing experience in terms of the rarity of species seen or in terms of seeing species that are iconic for a particular region. It can be tempting to simply omit objectives that are hard to measure, but that is tantamount to declaring those objectives to have no importance at all. A better solution is to make careful use of qualitative measures.

Defining verbal categories transparently.

The trick with verbal categories is to define them clearly enough that they can be used consistently by different evaluators. It is common to use phrases such as "low," "medium" and "high" to define performance on qualitative measures but, without more direction, different evaluators are likely to use different categories to describe the same conditions. To be fully transparent and unambiguous, qualitative measures need to have (a) a category for every condition that is likely to result from the management alternatives being considered (so that evaluators will always be able find a category for any observed or predicted result), and (b) non-overlapping categories (so that evaluators will have only one category for each possible result). For the quality of wildlife viewing, in a region where there are two iconic species of waterfowl, such a measure could include these categories: no iconic species seen in a day's viewing; one but not both iconic species seen in numbers fewer than 5; one but not both iconic species seen in numbers greater than or equal to 5; both iconic species seen in numbers fewer than 5; one iconic species seen in numbers fewer than 5 and the other in numbers greater than or equal to 5; both iconic species seen in numbers greater than or equal to 5. Any number of either iconic species seen in a day will fall into one and only one of these six categories, and any evaluator knowing the number of each species seen will assign a category consistently.

Distinguishing ratings of performance from expressions of relative satisfaction with different levels of performances.

The verbal categories used to define qualitative measures should not embed expressions of relative satisfaction within the verbal description of the categories, either qualitatively or quantitatively. For example, it is common to describe qualitative categories as "worst," "medium" and "best." These categories are not only ambiguous (i.e., Where does "worst" stop and "medium" begin?), but they also assume an order that may not suit all users. In the iconic species measure above, we purposely did not number the verbal categories 1 through 6 because such numbers are likely to be erroneously used as expressions of relative satisfaction later in the analysis (i.e., where the category labeled 6 is assumed to be 6 times as desirable as the category labeled 1), without any scrutiny of whether or not that assumption is warranted and whether it suits all users. Without further investigation we don't even know how to order the six viewing categories: Is it better to see both species in numbers fewer than 5 or only one but in numbers greater than 5? Different users might answer that question differently. Be wary of qualitative scales represented by numbers or ranks. It is better to create a shorthand code for lengthy verbal descriptions of categories by using a word or a letter, rather than a number.

Measuring production of ecosystem services.

All of these principles for developing good measurement scales to evaluate performance of alternative management actions apply equally to measures of ecosystem services. Properly developed qualitative measurement scales may be especially important when evaluating the production of intangible ecosystem goods and services, such as the existence value of unique species or landscapes.

Evaluating Performance of Alternatives

Creating an alternatives/measures matrix.

It is common to present the results of analyzing the anticipated performance of several management alternatives in a matrix with alternatives on one side and measures of performance on the other. Table 1 shows such a matrix for the wetland example, simplified to include only three alternatives and four measures, including implementation cost over a 10-year period calculated as net present value (NPV).

Table 1. A matrix showing the performance of each of three alternatives for restoring Eco Forest wetlands in terms of four measurement scales, each representing an objective in Figure 1.

	Alternative actions		
Measures	Status quo	Downstream dam	Upstream release
Numbers of bird 1 (breeding pairs on forest)	200	220	205
Wildlife viewing at walkway site (qualitative scale)	One iconic sp < 5	One iconic sp < 5, one >5	Both >5
Flood events (annual average)	0.2	0.15	0.2
Implementation Cost ($MM NPV)	0.1	1.0	0.8

A matrix like this can make it easy to see if there is one alternative that is better (or no worse) on all measures than all the others (and can be chosen without further analysis) or one that is worse (or no better) on all measures (and can be discarded without further analysis). There are no such clear winners or clear losers here.

Using expert opinion.

The entries in the cells of the matrix can be derived from formal or informal models of the effect of management actions on ecological and sociological processes. A diagram relating actions to results, as in Figure 2, can be a first step for both kinds of models. Where such models are not available, it can be tempting to just omit an objective and its measure. But, as explained earlier, leaving out an important but troublesome measure is tantamount to declaring it to have zero importance. A better tactic is to make use of expert opinion to fill in blanks in the matrix. There are established procedures for choosing experts and eliciting their opinions (see references in Gregory et al. 2012), and this is one of the instances where relying on an experienced consultant is smart. In any case, very clearly defined measures and alternatives are necessary precursors to reliable use of expert opinion.

Expressing uncertainty

Imprecise estimates of performance for one or more measures are typical. A common, but undesirable, way of dealing with uncertainty about performance is to create measurement scales that lump quantitative results into "bins," such as 0-10 breeding pairs, 11-20 breeding pairs, and so on. The problem with this tactic is that in order to unambiguously assign a particular result to the correct bin, e.g., to know that it belongs to the 0-10 bin and not to the 11-20 bin, the evaluator must know whether the number of breeding pairs is 10 or 11. A better way to handle uncertainty about the number of breeding pairs is to express performance as a range of values in cells of alternatives/attributes matrix (e.g., 5-8 breeding pairs), or as a probability distribution (e.g., a mean of 6.5 with a standard deviation of 2).Then carry out the rest of analysis using the extremes of the range (or by sampling from the probability distribution) to see if that uncertainty affects the overall rating of alternatives.

Evaluating performance with respect to ecosystem services.

An increasing array of formal models is being developed to predict production of ecosystem services in a variety of landscapes. Where formal models are not available, judicious use of expert opinion will help fill in gaps, thus avoiding the mistake of omitting what is important, but hard to observe or predict.

Expressing Relative Satisfaction with Performance on Individual Measures

Sometimes creating the alternatives/measures matrix is the final step of formal analysis and expressions of relative satisfaction for different levels of performance and any trade-offs among objectives are made intuitively during the decision process. A more formal consideration of relative satisfaction and trade-offs requires a relationship between performance on each measure and a unitless scale (usually 0-1, sometimes 0-100) that describes how relative satisfaction changes over the range of performance levels encountered in an analysis of particular alternatives. (These relationships are often called value functions or utility functions.) These relationships serve the dual purposes of putting unlike measures on a common scale, so that they can be combined, and expressing relative satisfaction with different levels of performance for a single measure.

Linear relationships of relative satisfaction to performance level.

It is common (but not always warranted) to simply assume a linear relationship between relative satisfaction and performance level, as in Figure 3 for number of breeding pairs of bird 1. A linear relationship is more likely to reflect relative satisfaction accurately when the management alternatives change performance relatively little compared to the status quo, as is the case for breeding pairs of bird 1, which varies only 10% from the status quo for any of the new management alternatives.

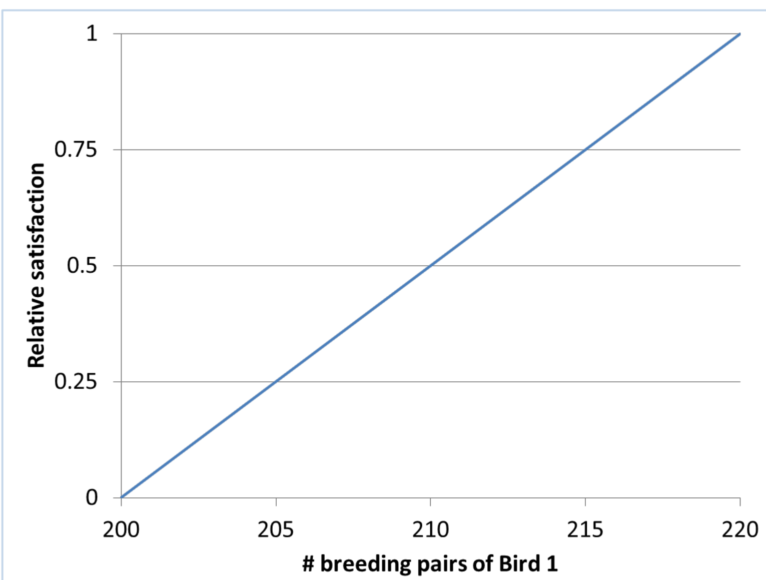


Figure 3. A linear relationship showing that the increase in relative satisfaction for each additional breeding pair of Bird 1 is the same over the range of 200 to 220 breeding pairs.

When performance levels vary more widely, as they do for costs and flood events, assuming a linear relationship may not be adequate. Also, the shape of the relationship between relative satisfaction and performance is tied to the range of performance levels encountered in a particular problem. We should expect a different shape for numbers of breeding pairs ranging from 2 to

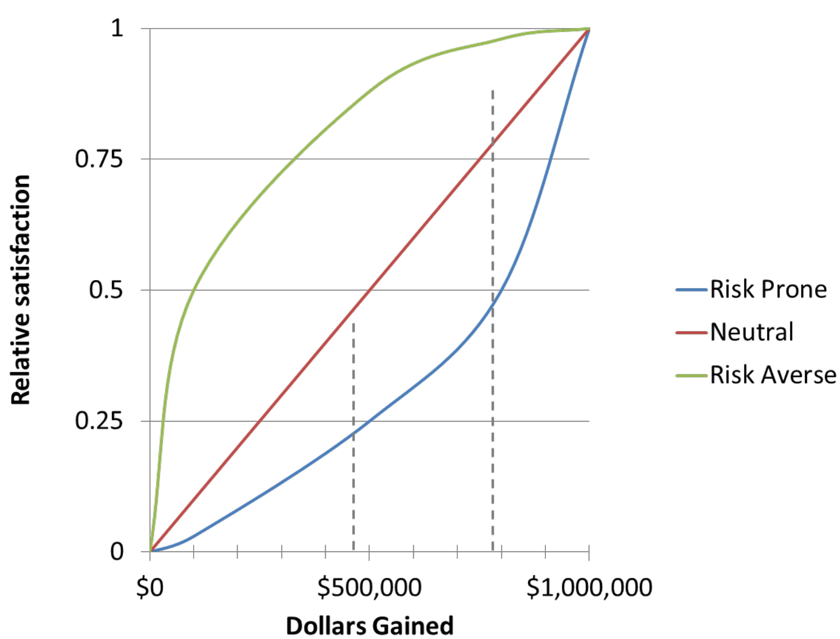
2000 than for numbers of breeding pairs ranging from 200 to 220. This is one reason among many why a relationship derived for one set of alternatives in one context may not be meaningful for another context.

When relationships between satisfaction and performance have been determined for the range of performance levels for a particular set of alternatives, analyzing an additional alternative has the potential to produce a level of performance outside the original range, necessitating a new relationship. Sometimes it is useful to anticipate the need to analyze additional alternatives and extend the performance range used to establish the relationship between satisfaction and performance a bit beyond that encountered in the initial set of alternatives.

Nonlinear relationships of relative satisfaction to performance level.

Relationships between relative satisfaction and performance levels can easily be nonlinear. A common nonlinear relationship is diminishing marginal increases in relative satisfaction as the level of performance increases, as is often the case for monetary gains (Fig. 4, risk-averse curve; the meanings of risk-averse and risk-prone are explained below in the section on risk attitudes).

Figure 4. Two nonlinear relationships between relative satisfaction and dollars gained compared to a linear relationship (red). The green line shows decreasing marginal satisfaction as more dollars are gained, a commonly encountered relationship for gains from the status quo. The blue line shows increasing marginal satisfaction as more dollars are gained; this is uncommon for dollars gained, but common for dollars lost. When curves like these are used to express attitudes toward uncertain levels of performance (e.g., returns from a lottery), the green line is termed "risk-averse" and the blue line "risk-prone," as explained further below in the section on risk attitudes.



These nonlinear relationships can be different for different stakeholders. It is easy to imagine that downstream landowners, who most immediately feel the pain of flood events, might experience a larger boost in relative satisfaction from decreasing flood event frequency from its status quo level of 0.2 than upstream landowners (Fig. 5, red line vs. Fig. 5, blue line). (Note that this relationship represents a wider range of flood frequency than the 0.15 to 0.2/yr in Table 1, in order to show how nonlinear relationships might differ for different user groups; the relationship between flood frequency and relative satisfaction might well be more nearly linear for the narrower 0.15 to 0.2 range of performance.)

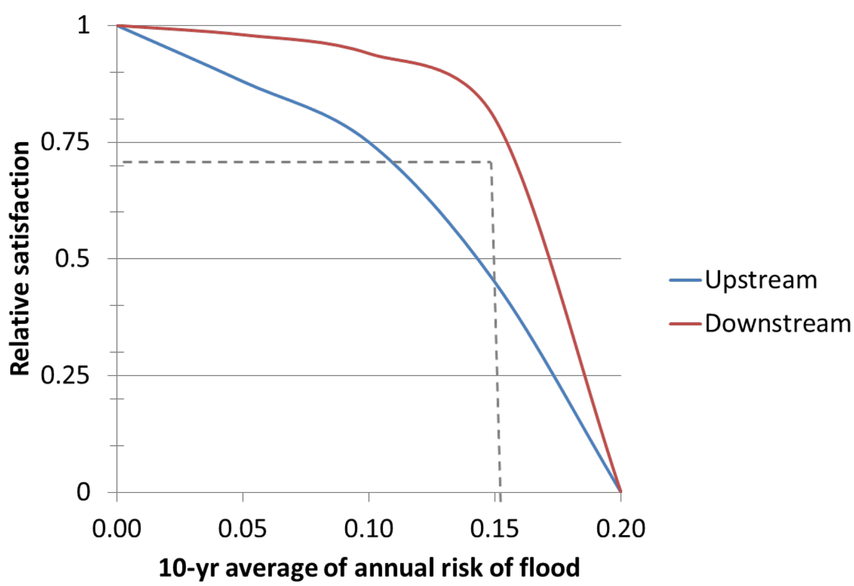


Figure 5. Curves representing a steeper increase in relative satisfaction as the probability of flooding decreases from the status quo of 0.2 for downstream landowners (red line) compared to upstream landowners (blue line). Both curves can also be said to represent risk-averse attitudes for uncertain probabilities of flooding (see section on risk attitudes below).

Finding these nonlinear relationships can seem daunting. There are structured methods for eliciting such relationships (see Clemen 2001), and it is probably best to use the services of specialized consultants to implement them. Usually these elicitations are done face to face with decision makers or stakeholder representatives, but there are some ways of gathering information remotely via surveys or social media.

A stopgap approach is to simply draw shapes that seem to capture the way relative satisfaction increases or decreases with performance level on a graph (such as Figs. 4 and 5) and read off the relative satisfaction that corresponds to a particular level of performance. It will often be the case that the choice of a particular alternative is not highly sensitive to the exact relationship between satisfaction and performance.

Expressing risk attitudes.

A third purpose served by relationships between relative satisfaction and performance is to express attitudes about uncertainty. The shape in Figure 4 (green line) describes diminishing marginal increases in satisfaction with increasing increments of money gained when the amount of money gained is certain. It can also express a risk-averse attitude toward uncertain amounts of money gained. When faced with an uncertain amount of money gained (e.g., a 50-50 gamble between $0 and $1 million), it is common to prefer an amount of money for certain that is considerably less than the expected value of the gamble (0.5 x $0 + 0.5 x $1 million = $500K). The upward-bulging curve in Figure 4 (green line) expresses that willingness to accept less (about $100K). A downward-bulging curve expresses a risk-prone attitude to uncertainty, i.e., requiring more money for certain than the expected value (about $800K) (Fig. 4, blue line). It is common to be risk-averse with respect to gains from the status quo and risk-prone with respect to losses from the status quo (e.g., additional costs).

We can also interpret the curves in Figure 5 as expressions of risk attitudes regarding uncertain annual probabilities of flooding. Both curves are upwardly bulging as satisfaction rises with declining probability of flooding, indicating a risk-averse attitude for improvements in the risk of flooding from the status quo level of 0.2 (i.e., lower probabilities of flooding). To see this more clearly, reverse the direction of the x-axis, producing curves that have the same shape as the green line in Figure 4.

The linear relationship, which expresses a risk-neutral attitude (Fig. 4, red line), is probably adequate for modest changes from the status quo, even when those changes are uncertain. If such a relationship doesn't seem adequate to express attitudes toward uncertain outcomes, it is probably best to engage a specialized consultant to develop nonlinear risk-attitude functions.

Expressing relative preference for qualitative measures.

I purposely did not label the qualitative scale for quality of wildlife viewing above with numbers because numerical labels are often misinterpreted as indicators of relative satisfaction. If we need to express relative satisfaction for use in a more formal analysis of alternatives, we must assign numerical values between 0 and 1 to the categories of the qualitative scale. The first step is to order the verbal categories from worst to best. This order might differ for different users, although the worst and best categories are likely to be obvious. For this scale, seeing none of either iconic species is worst; it will receive a numerical value of 0; seeing both in numbers greater than 5 is best and it will receive a numerical value of 1. As mentioned above, the order of the four intermediate categories isn't entirely obvious since one stakeholder group might prefer seeing both species in smaller numbers to seeing one species, but in larger numbers, and another stakeholder group might prefer the opposite. The order shown in Table 2 represents the preferences of the first group.

Table 2. An implementation of the ratio method for eliciting relative satisfaction values for intermediate categories of a qualitative scale for viewing of iconic wetland wildlife species.

Category	Ratio	Points	Relative Satisfaction
Neither		10	0
One, <5	2x	20	0.14
One, ≥5	2.5x	25	0.21
Both <5	5x	50	0.57
One <5, other ≥5	7x	70	0.86
Both ≥ 5	8x	80	1

There are a number of structured techniques for capturing the numerical values for the four intermediate categories (see Clemen 2001). It would be best to enlist the help of a specialized consultant to implement one of these, but one fairly simple method is to ask users to express relative satisfaction with each level of the scale as a ratio compared to the worst level of the scale (seeing neither iconic species). In Table 2, the second column lists the ratios given by one user; 2x in the second row of that column means that seeing one of the iconic species in numbers less than 5 is twice as good as seeing neither. The last entry in the second column, 8x, means that seeing both in numbers greater than or equal to 5 is eight times as good as seeing neither. The next column assigns an arbitrary number of points to the lowest level of the scale (10 is a handy number to use) and then assigns points to the remaining levels of the scale using the ratios in the second column. The last column re-calibrates this list of points to the 0-1 scale we need for expressing relative satisfaction (e.g., (20-10)/(80-10) = 0.14).

One caution here is to avoid jumping to the conclusion that the levels of a qualitative scale are equally spaced (e.g., 0, 0.2, 0.4, 0.6, 0.8, and 1 for a six-level scale) without checking carefully to see that that assumption accurately reflects the ratios of relative satisfaction between levels.

More generally, don't assume that all users have the same shapes for the relationships expressing relative satisfaction with different levels of performance on a particular measure without checking.

Table 3 inserts numerical values for relative satisfaction in parentheses below the evaluations of performance for each alternative, using the relative satisfaction relationships shown in Figure 3 for number of breeding pairs, Table 2 for wildlife viewing, Figure 5 (downstream), and a slightly risk-prone relationship for cost (i.e., not as curved as the blue line in Fig. 4).

Table 3. Performance evaluations for the three alternatives for restoring forested wetlands with performance levels for each of the four alternatives translated to a 0-1 scale expressing relative satisfaction (in parentheses) using Figure 3 for number of breeding pairs, Table 2 for wildlife viewing, Figure 5 (downstream), and a slightly risk-prone relationship for cost (i.e., not as curved as the blue line in Fig. 4).

	Status quo	Downstream dam	Upstream release
Numbers of bird 1 (breeding pairs on forest)	200 (0)	220 (1)	205 (0.25)
Wildlife viewing at walkway site (qualitative scale)	One iconic sp < 5 (0.14)	One iconic sp < 5, one >5 (0.86)	Both >5 (1)
Flood events (annual average)	0.2 (0)	0.15 (0.8)	0.2 (0)
Cost ($MM NPV)	0.1 (1)	1.0 (0)	0.8 (0.6)

Monetization as an expression of relative satisfaction.

Monetization of non-market goods or services, such as wildlife viewing quality, is an alternative way of expressing relative satisfaction with different levels of performance. Monetization allows market and non-market goods and services to be compared on a common scale, but it is sometimes difficult, or perceived as inappropriate, to come up with monetary values.

Transferring expressions of relative satisfaction.

When using a monetary valuation approach, benefits-transfer methods can be used to transfer dollar values (or functional relationships between dollars and levels of a performance measure) to different user groups and different decision contexts, but there are significant limitations to doing this. Many of the same caveats regarding sensitivity to decision context and performance ranges pointed out here for nonmonetary expressions of relative satisfaction apply equally to monetization as an expression of relative satisfaction.

Relative satisfaction for production of ecosystem services.

All of the principles discussed above apply to expressing relative satisfaction with different levels of provision of ecosystem services. Of particular importance is recognizing that different user groups may well exhibit different patterns of increasing or decreasing satisfaction as the level of production of a particular ecosystem service changes, as suggested above for upstream versus downstream landowners and moderation of flooding. Capturing these differences is necessary for analyzing potential distributional and equity effects of managing for ecosystem services.

Using Weights to Express Trade-offs Among Objectives

In addition to numerical expressions of relative satisfaction for the levels of individual measures, the other ingredient needed to support a formal evaluation of trade-offs among multiple measures is some expression of the priority to put on each measure. A common way to express those priorities is as a set of fractional weights that add up to 1. These weights express willingness to accept worse performance on one measure in order to secure better performance on another, i.e., willingness to make trade-offs among conflicting objectives.

Weights depend on the ranges of performance.

As with the relationships that express relative satisfaction, willingness to make trade-offs is tied to the range of possible performance levels that might be encountered when evaluating a particular set of alternatives. It is easy to see that this is so by imagining that the range of costs in Table 3 is $500 thousand to $550 thousand instead of $100 thousand to $1 million. If the ranges for the other three performance measures (breeding pairs, wildlife viewing, and flood events) stay the same, the impact that different levels of cost have in determining overall satisfaction with each alternative will be far lower when the range of costs is small than when cost levels vary more widely. The weight on cost will be lower in the former case than in the latter. (And, because the weights must add up to one, the weights on the other three measures will be correspondingly larger in the former case.)

Weights may not be transferrable between decision contexts.

Unlike the relationships that express relative satisfaction, where the range of performance levels for a single measure affects only the pattern of relative satisfaction for that measure independent of the performance levels on other measures, the weights for a set of measures have meaning only in relation to each other and only in relation to the ranges of performance on all measures for a particular problem. It is not credible to elicit weights for one set of performance levels in one decision context and then apply those weights in another context without first verifying that the performance levels and the particulars of the decision context are similar enough to justify such a transfer.

Weights differ among user groups.

For most contentious decisions, the fundamental disagreements among stakeholder or user groups are about the priorities placed on different objectives, as expressed by weights. Capturing these differences by eliciting separate sets of weights for different users is very helpful to both decision makers and the user groups themselves, showing why different groups prefer different management alternatives and, sometimes, suggesting where compromises that satisfy some of the needs of each group can be found. Attempting to gloss over differences in priorities by eliciting weights from only one or a few perspectives, or by averaging weights across user groups, is a recipe for continued contention.

Eliciting weights.

There are a variety of structured ways to assess weights (see Clemen 2001) and using a specialized consultant to implement one of these is a good idea. If that is not possible, a stopgap approach is to use a visual representation of weights, such as a bar with segment lengths proportional to the weight on each measure (e.g., Fig. 6). The fact that the length of the whole bar is fixed at 1 requires that any increase in weight on one measure be compensated by decreases in the weight on one or more of the other measures.

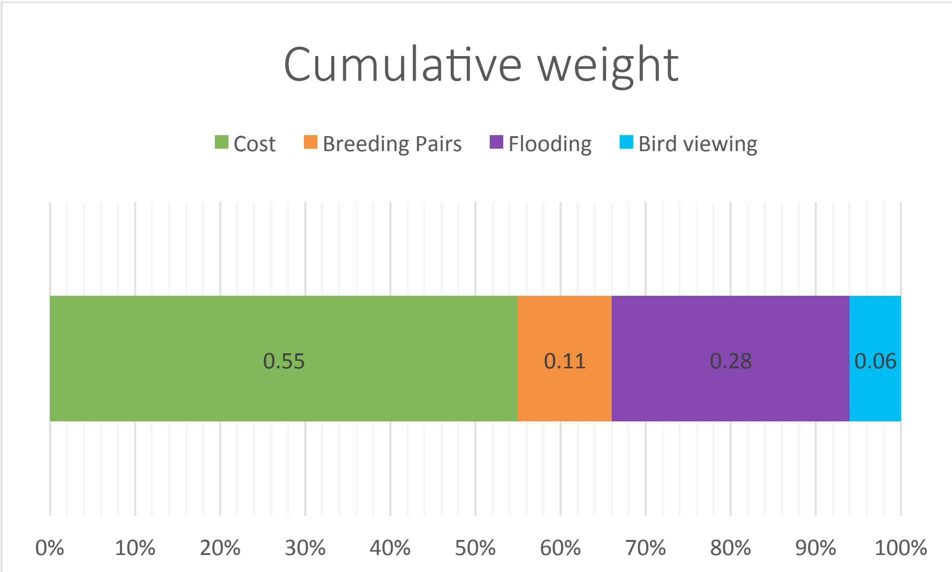


Figure 6. A visualization of the relative weights assigned to the four measures of performance in the wetland restoration example. The length of each colored bar represents the fractional weight on that measure. The total weight is fixed at 1.

Another stopgap approach to eliciting weights is to use ratios in a manner somewhat similar to what was described above for quantifying relative satisfaction for qualitative measurement scales. To assess weights using ratios, first call attention to the range of performance levels for each measure (because that information is essential to assessing weights). Then, put the measures in order of priority, showing which objective is the most important in determining the overall success of a management alternative, next most important, and so on (Table 4). Assign a ratio of 1 to the least important, and then ask how much more important the next most important is compared to the least, and so on, bearing in mind how widely performance of each of these might vary for the particular alternatives being considered. Do the same for the remaining measures, asking how much more important the second-ranked measure is compared to the fourth, and so on. It may also help to ask how important the second most important is compared to the third, and so on, in order to make sure the relative importance has been captured accurately. Then, add up the ratios that have been assigned and calculate the weight for each measure by dividing its ratio by the total (e.g., 10/18 = 0.55). Adjust the rounded fractions so that the weights sum to 1. (Note that this set of calculations differs from that for obtaining relative satisfaction.)

Table 4. Illustration of a method of obtaining fractional weights to represent the relative importance of the four measures of performance in determining the overall value of a wetland restoration alternative.

Measure (units)	Range	Rank	Ratio	Weight
Cost ($MM)	0.1-1.0	1	10	0.55
Bird 1 (pairs)	200-220	3	2	0.11
Flood events (ave #/yr)	0.15-0.2	2	5	0.28
Viewing (index)	One<5 - both >5	4	1	0.06

Weights express priorities and trade-offs among ecosystem services.

For objectives hierarchies that include ecosystem services, the weights placed on measure of an ecosystem service express willingness to trade off gains or losses in that service in relation to changes in performance on the other objectives included in the analysis. The potential to for either "leverage," where two or more ecosystem services can be improved by a course of action, or "trade-offs," where gains in one ecosystem service are accompanied by losses in another service (or in another valued objective) has been a concern in evaluation of ecosystem services production. Weights in a multicriteria analysis can help address that concern.

Other important issues in ecosystem services evaluation that assessment of weights can help address are distributional and equity effects of the pursuit of ecosystem services. It is very likely that different user groups will put different priorities on different ecosystem services (e.g., downstream landowners affected by flooding versus upstream farmers affected by availability of irrigation water) or on ecosystem services relative to other desired objectives. Assessing weights from different user groups, and using them to understand why different groups prefer different combinations of gains and losses on different objectives, can help adjudicate conflicts and address equity concerns.

Aggregating Value for Multiple Objectives

Calculating overall value of an alternative.

Sometimes decision makers do not want to create aggregated numerical values that express the relative merits of each alternative being analyzed, preferring to integrate information about performance, relative satisfaction and/or weights intuitively in a decision process. In situations where an aggregated score for each alternative is wanted, a commonly used method is to calculate an overall value on a 0-1 scale by adding up, for all measures, the weight on each measure multiplied by the relative satisfaction associated with performance on that measure. Table 5 includes the weights derived in Table 4 as well as the performance and corresponding relative satisfaction values in Table 3 and reports overall value for each of the three alternatives being analyzed (e.g., (0.11)(0) + (0.06)(0.14) + (0.28)(0) + (0.55)(1) = 0.56 for the status quo).

Table 5. A full representation of a multicriteria analysis of trade-offs in performance for three wetland restoration alternatives evaluated using four performance measures. The weights from Table 4 have been added to Table 3 and overall values for each alternative calculated by summing the weight times the relative satisfaction value associated with performance across the four measures.

	Alternatives		
Measures (Weights)	Status quo	Downstream dam	Upstream release
Numbers of bird 1 (breeding pairs on forest) (w = 0.11)	200 (0)	220 (1)	205 (0.25)
Wildlife viewing at walkway site (qualitative scale) (w = 0.06)	One iconic sp < 5 (0.14)	One iconic sp < 5, one >5 (0.86)	Both >5 (1)
Flood events (annual average) (w = 0.28)	0.2 (0)	0.15 (0.8)	0.2 (0)
Cost ($MM NPV) (w = 0.55)	0.1 (1)	1.0 (0)	0.8 (0.6)
Overall value	0.56	0.39	0.42

Sensitivity analysis.

The resulting overall values of 0.56, 0.39 and 0.42 have meaning only in relation to each other and for the decision context, set of alternatives and users whose relative satisfaction relationships and weights were used in the calculations. Because these overall values depend on so many ingredients, it is a good idea to vary some of the ingredients systematically to see how the overall values respond in a sensitivity analysis. All the ingredients are candidates to be changed: (1) The predicted performance of each alternative can be raised or lowered, either in concert (e.g., flood frequency higher for all three alternatives) or separately. Particularly where there is uncertainty about predicted performance, perhaps indicated by recording a range of values rather than a single number in one or more cells of the alternatives/measures matrix in Table 1, re-calculating overall value of each alternative at the high and low ends of that range can help reveal whether resolving that uncertainty is critical to understanding the relative merits of the three alternatives. (2) The relationships between relative satisfaction and performance can be changed for one or more of the measures, indicating either a different pattern for a particular user group or indicating different patterns for different user groups, to see if changing the type of relationship (e.g., linear vs. nonlinear) changes the overall values of the alternatives enough to re-order them. (3) The set of weights used to calculate overall value can be changed (note that all the weights must be changed at once to keep their sum at 1 to reflect either errors in elicitation of weights or to reflect differing priorities for different user groups. In situations where it may be impractical to elicit weights from the full variety of user groups, hypothesizing sets of weights that capture the stated priorities of different groups can help reveal where differing priorities may be critical to disagreement about the best course of action.

Aggregating value for multiple ecosystem services.

Articulating the combined value of a suite of ecosystem services has been the target of much research and the expressed desire of federal regulatory and budgetary organizations. The type of multicriteria analysis described here offers one way of meeting that need, producing numerical values that describe the relative abilities of a set of management alternatives (usually including the status quo) to produce desired ecosystem services. As emphasized throughout, these numerical expressions of relative merit are tied to a particular decision context, a particular set of alternatives, and particular characterizations of relative satisfaction with performance and priorities among conflicting objectives. This type of analysis easily blends measures that are typically monetized (e.g., financial costs of implementing management actions) with those that are not easily monetized (e.g., the experience of viewing iconic wildlife species). Distaste for monetization of non-market ecosystem goods and services is avoided with this type of analysis, although some user groups are wary of any quantification of intangible services, whether via monetization or otherwise. Some of those demanding summary values of ecosystem services may believe that aggregated monetary values will be more widely understood and more readily compared among decision contexts than the unitless relative values produced by multicriteria analysis but, for reasons discussed below, many of these apparent advantages of monetization may fail to be realized in practice.

Comparison with other methods for aggregating multiple services.

Most of the other methods of expressing relative satisfaction and priorities for multiple objectives and for aggregating those to produce an overall value for a management alternative involve some degree of monetization of costs incurred and/or benefits produced. Monetizing non-market goods and services to express relative satisfaction at different performance levels has been controversial, desired by some and criticized by others, in part because methods for monetization have often been misused. Different user groups in different places will often exhibit different monetizations for a particular ecological metric or ecosystem service. Sensitivity to the distributional effects of managing for ecosystem services may require separate monetization functions (or separate value/utility functions) for different stakeholders.

It might appear simple to express trade-offs among competing objectives and to aggregate gains and losses for multiple objectives by simply adding the dollar values obtained for each measure used to evaluate management alternatives in a cost-benefit analysis. However, monetization is sensitive to the decision context, the suite of measures represented in valuation experiments, the ranges of performance levels for all measures for the management alternatives being evaluated, and the user groups whose views are being captured numerically. These constraints limit the use of dollar values developed in one context for decisions made in another.

Where costs of management are monetized and other measures are not, cost-effectiveness analysis (CEA) reports how much it costs to increase performance of a particular measure by one unit (e.g., acres, number of animals). Only one non-monetized measure can be evaluated at a time, so trade-offs among non-monetized measures cannot be expressed.

Cost-utility analysis (CUA) does much the same calculation as CEA, but uses a multi-attribute function (weight times value or utility of each measure) to aggregate all measures except those for costs of implementing management actions into a unitless metric on a 0 to 1 scale, as in the multicriteria analysis presented here.

Choosing a Course of Action

Types of decision processes for public lands management are probably best viewed as a continuum from fully shared decision-making among agency leaders and stakeholders (generally precluded by agency rules) to a decision by a single agency leader (also rare). In between are various levels of input from and consultation with stakeholders and higher-level agency authorities.

Agency decision makers (and stakeholders engaged in advisory roles) might receive information from many of the stages above and use it to recommend a management alternative. For example, the analysis might stop with the prediction of performance for alternative management actions, with the remaining steps of relative satisfaction, weighting and aggregation carried out intuitively. Or, information on relative satisfaction and weights might be developed for various user groups, but not aggregated across all objectives, so that user preference information can influence the decision process, but perhaps more "flexibly" than when a numerical value of the relative merit has been calculated for each alternative.

Even when all of the stages above have been carried out and a fully aggregated numerical value produced for each alternative, those values are likely to be treated as advisory to the human decision maker, rather than decisive in themselves. Variation among stakeholders, information not incorporated in previous stages, and other factors are likely to influence the final decision.

A Caveat about Going It Alone vs. Engaging Specialized Consultants

Many of the elements of structured decision making can be employed to good effect by anyone with a head for orderly and rigorous thinking. These include: articulating objectives, scrutinizing objectives for completeness and redundancy, defining measures clearly, scrutinizing proposed measures for implicit inclusion of relative satisfaction (where it does not belong), and recognizing where different decision makers or user groups may have different beliefs or different preferences that impinge on the decision structure (e.g., differing assessments of performance, differing relationships of relative satisfaction to performance, differing priorities among objectives).

Other elements of structured decision making benefit greatly from the experience and judgment of specialized consultants in decision making. These include (1) scrutinizing objectives hierarchies to make sure that they meet the assumptions for independence of different parts of the hierarchy, which are necessary to support subsequent stages of the analysis (i.e., eliciting relative satisfaction with performance levels, eliciting weights to express priorities, and aggregating overall value of alternatives by summing the products of relative satisfaction and weight for each measure) and (2) identifying where and how to use sensitivity analysis to illuminate essential trade-offs and establish how robust the results of an analysis are to changes in the ingredients used to compose it. Other elements use rather finicky procedures to elicit numerical representations of beliefs and preferences (e.g., eliciting expert opinion to fill in gaps in performance predictions, eliciting relationships for relative satisfaction with different levels of performance, eliciting weights to express priorities among objectives). Although we have offered some simplified methods of performing these tasks here, there are lots of potentially misleading pitfalls for inexperienced users and we urge use of the services of specialized consultants for these portions of an analysis. Fortunately, the number of agency consultants trained in structured decision making is increasing.

Anyone at any level can clarify complex decision problems by taking a structured approach such as that recommended here. But, where the stakes are high, it would be just as foolish to rely on multicriteria analysis without engaging the help of structured decision making consultants as it would be to attempt an economic analysis of benefits and costs without engaging the help of economists.

Comparison with Other Types of Analysis

There are many kinds of multicriteria analysis that operate on different assumptions and different calculations (e.g., Department of Communities and Local Government 2009, Appendices 1 through 8). The merits of these different approaches have been debated vigorously in the academic literature, but the characteristics that are likely to mean the most to agency practitioners are ease of use, transparency and minimizing the potential for misleading results. This document describes the use of a particular type of multicriteria analysis, multiattribute utility analysis (MAUA) (DCLG, Appendices 3 and 4). Probably the charges most often leveled at MAUA, from a practitioner's point of view, are (1) that it is too hard and/or too much work to implement, (2) that the results are not meaningful beyond the decision context for which the analysis was made, and, a related complaint, (3) that the resulting valuations are useful only for comparison and have no absolute meaning. Charges 2 and 3 apply equally to other kinds of multicriteria analysis and, therefore, don't constitute a strong argument against MAUA in particular. Monetary valuation methods, to be used in benefit-cost analysis or some other type of economic analysis, are sometimes promoted as solutions to charges 2 and 3. However, for the reasons described above in the section on Other Methods for Aggregating Value, monetary valuations may be neither as transferable to other contexts nor as absolute as is commonly believed.

The remaining charge, that MAUA is too much work to implement or is too hard for nonspecialists to implement competently, has some merit, as described in the preceding section on Going It Alone. Many of the alternative types of multicriteria analysis described in the DCLG Appendices cited above attempt to reduce the data required to implement an analysis and to simplify, or even automate, the judgments that must be elicited from decision makers or stakeholders (e.g., by requiring only pairwise comparisons of alternatives instead of numerically expressed evaluations, as in the Analytic Hierarchy Process, DCLG 2009, Appendix 5, or by applying a set of rules for reducing the dimensions of a decision problem, as in outranking methods, DCLG 2009, Appendices 6 and 7). Indeed, the availability of user-friendly software to implement some of these methods (especially Analytic Hierarchy Process) has undoubtedly enhanced their use in environmental applications. However, some of these alternate methods have structural flaws that can lead to results that don't accord with common sense (see DCLG 2009, Appendix 5, re Analytic Hierarchy Process), and many of them incorporate assumptions that are not wholly transparent to users (e.g., ratios of preference used in Analytic Hierarchy Process). Thus these methods can fall short in terms of producing potentially misleading results and lacking transparency.

Several arguments can be made in favor of MAUA: (1) It may be difficult and a lot of work to implement, but that is because the problems being addressed are genuinely difficult for a host of reasons (e.g., disputes among parties, technical disagreements, limited information, conflicting goals and mandates). MAUA helps to identify and articulate these difficulties. (2) Going through the steps of MAUA (i.e., stating objectives, developing measurement criteria, evaluating performance, assessing relative satisfaction and weights), obliges decision makers and stakeholders to address all these sources of potential difficulty explicitly, even if done only qualitatively. (3) Addressing all these stages of analysis explicitly enhances transparency, an especially important characteristic for public decision making (a conclusion also endorsed by the DCLG 2009 report). (4) The dependence of analytical results on decision context, and the inherently relative nature of those results, is real. That MAUA makes these limitations more obvious than some other types of analysis is to its credit rather than to its detriment.

Annotated Bibliography on Structured Decision Making

Clemen, R.T. (with T. Reilly). 2001. Making Hard Decisions. 2nd ed. Rev. Duxbury Press, Pacific Grove CA. This is a quite comprehensive, reference for the more technical aspects of decision analysis, including elicitation of expert opinion, assessment of value and utility functions and assessment of weights.

Department of Communities and Local Government. 2009. Multi-criteria analysis: a manual. ISBN 978-1-4098-1023-0. Online via the Communities and Local Government website: www.communities.gov.uk. A good overview of the steps in multicriteria analysis aimed at nonspecialist readers. Ch. 5 discusses decision context, objectives and measures, and evaluations of performance. Ch. 6 adds expression of relative satisfaction, determination of weights and calculation of overall value using purchase of a toaster as an example. Ch. 7 presents some more complex examples that will resonate with environmental managers.

Gregory, R., L. Failing, M. Harstone, G. Long, T. McDaniels, and D. Ohlson. 2012. Structured decision making: a practical guide to environmental management choices. ISBN 978-1-4443-3342-8 (paper). Wiley-Blackwell, Oxford, UK. A book-length description of the use of multicriteria analysis for real-world environmental decision making, especially where multiple stakeholders are involved. Not a "how-to" manual.

Hammond, J.S., Keeney, R.L. and H. Raiffa. 1999. Smart Choices. (ISBN 0-87584-857-5, cloth, or 0-7679-0866-4, paper). Harvard Business School Press, Cambridge MA. A non-mathematical presentation of the

concepts of decision analysis aimed at the general public, using everyday decisions (such as buying a house). It presents some simplified tools for expressing relative satisfaction and weights and for determining overall value.