Analytical Group Decision Making in Natural Resources: Methodology and Application

Daniel L. Schmoldt and David L. Peterson

ABSTRACT. Group decision making is becoming increasingly important in natural resource management and associated scientific applications, because multiple values are treated coincidentally in time and space, multiple resource specialists are needed, and multiple stakeholders must be included in the decision process. Decades of social science research on decision making in groups have provided insights into the impediments to effective group processes and on techniques that can be applied in a group context. Nevertheless, little integration and few applications of these results have occurred in resource management decision processes, where formal groups are integral, either directly or indirectly. A group decision-making methodology is introduced as an effective approach for temporary, formal groups (e.g., workshops). It combines the following three components: (1) brainstorming to generate ideas; (2) the analytic hierarchy process to produce judgments, manage conflict, enable consensus, and plan for implementation; and (3) a discussion template (straw document). Resulting numerical assessments of alternative decision priorities can be analyzed statistically to indicate where group member agreement occurs and where priority values are significantly different. An application of this group process to fire research program development in a workshop setting indicates that the process helps focus group deliberations; mitigates groupthink, nondecision, and social loafing pitfalls; encourages individual interaction; identifies irrational judgments; and provides a large amount of useful quantitative information about group preferences. This approach can help facilitate scientific assessments and other decision-making processes in resource management. For. Sci. 46(1):62-75.

Additional Key Words: Analytic hierarchy process, fire research, natural resources planning, priority setting, workshop process.

Fire is an important periodic natural disturbance in most forest, shrubland, and grassland ecosystems of western North America. Extensive, high-intensity fires are infrequent temporally but have substantial impacts spatially and are responsible for rapid changes in vegetation, soils, biogeochemical cycling, and other ecological characteristics. While fire is known to play a critical role in the long-term dynamics of most ecosystems, there are many difficulties associated with scientific assessment and management of large-scale fire phenomena. This problem was brought sharply into focus in 1988 during and following the large fires in the Yellowstone National Park region.

Our ability to understand and manage for the effects of large fires has been limited by a lack of data at large spatial scales. There is a substantial scientific literature on the ecological effects of fire, but the vast majority of scientific data have been collected at scales of 10^{-1} to 10 km^2 (McKenzie et al. 1996). Applying these data to fire phenomena at much larger scales can result in substantial errors in estimating fire effects, a topic that is particularly relevant for modeling fire and ecosystem processes. Extrapolating ecological effects of fire across spatial scales can result in many sources of error (McKenzie et al. 1996), including: (1) extrapolating fire behavior models directly to larger spatial scales, (2) integrat-

Daniel L. Schmoldt (corresponding author), USDA Forest Service, Southern Research Station, Biological Systems Engineering Dept., 460 Henry Mall. University of Wisconsin, Madison WI 53706-1561—Phone: (608) 262-3313; E-mail: dlschmol@facstaff.wisc.edu. David L. Peterson, USGS Forest and Rangeland Ecosystem Science Center, University of Washington Field Station, Box 352100, Seattle, WA 98195-2100—E-mail: wild@u.washington.edu.

Manuscript received March 11, 1998. Accepted May 21, 1999.

ing fire behavior and fire effects models with successional models at the stand level, then extrapolating upward, and (3) aggregating model inputs to the scale of interest. Regardless of the approach used, extreme fire events pose a major problem for modelers due to the problem of propagating and compounding errors across spatial scales. The challenge is to develop or adapt models that are scientifically sound as well as applicable to large-scale resource management issues.

Given the complexity of large-fire phenomena, how do we improve our current scientific assessment and management of natural resources in North America with respect to fire disturbance? In April 1996, a group of scientists and resource managers gathered at the Fire-Disturbance Workshop at the University of Washington to discuss these issues. The workshop objectives were to: (1) identify the current state-of-knowledge with respect to fire effects at large spatial scales; (2) develop priorities for scientific assessment of large-scale fire disturbance and its effects; and (3) develop priorities for assisting scientifically based decision making with respect to fire disturbance in resource management. Addressing this ambitious set of objectives in a timely way required an efficient workshop process and effective group decision making (GDM).

This article introduces the GDM methods that were developed to address this need, and presents some specific examples from the Fire-Disturbance Workshop. First, the following section briefly reviews some of the pertinent social science literature dealing with GDM and provides a basis for formal group processes to aid decision making. This review provides a foundation for the application of our hierarchically based analytical method, which is detailed in the subsequent section. Also included in that section is a specific application of our approach for fire research planning. Finally, we discuss the implementation of our workshop results and the benefits and limitations (theoretical and empirical/practical) of the GDM methodology presented.

Group Decision Making

Few choices in natural resource management are made unilaterally. They vary in importance from which tool should be used for tree planting to how much timber will be cut on a national forest. Decisions are typically made in consultation with others, or by a group. Furthermore, judgmental (value laden) decisions that do not result in group unanimity produce less decision satisfaction for group members (Kaplan 1987), as opposed to informational (intellective) decisions that have a demonstrably "correct" answer. This implies that as current and future land management decisions—both strategic and tactical—are influenced by a wide variety of stakeholders' agendas (notentirely intellective influences), it will become more difficult for a majority to reach a state of agreeable and satisfied acceptance. Therefore, it is increasingly important that differences in preferences be understood and that mechanisms and procedures for describing and handling them be developed and applied.

Many natural resource problems involve selecting among a fixed set of alternatives or treatments or scenarios—a l-of-*N* decision situation. Unfortunately, decision making typically involves a BOGSAT process ("Bunch Of Guys/Gals Sitting Around a Table," Peterson et al. 1994). BOGSAT appears, on the surface, as a very cost-effective decision mechanism as little time or effort is expended. These initial cost savings can become moot, however, when the process produces extensive downstream costs such as time-consuming and expensive litigation and land mismanagement. By expending more organized and systematic effort up front, it may be possible to reduce total costs in terms of time, money, and credibility.

We often assume, however, that the decision produced by a group will always be better than that supplied by an individual. This seems plausible because multiple participants can bring differing expertise and perspectives to bear on difficult or complex problems. In reality, they do not always live up to those expectations. Groups generally perform better than their *average* individual member does but worse than the group's *best* individual (Hall 1970, Hill 1982, Yetton 1982, Bottger 1987, Rogelberg, 1992). Ideally, we should strive to avoid group deficiencies and yet capitalize on inherent group benefits.

McGrath (1984) summarized much of the existing literature on group interaction and performance, and categorized group tasks into four components: (1) generate (identifying alternatives), (2) choose (making value-laden judgments), (3) negotiate (managing conflict), and (4) execute (coordinating detailed implementations). Most social science research on decision making has been limited to examining only one type of group task at a time, which facilitates tractable study designs and straightforward experimental inferences. Multiple-component group tasks have been studied less frequently, even though applied decision processes typically involve most, if not all, of the components McGrath cites. Most resource management decisions and actions, in particular, incorporate aspects of each of these dimensions, which makes analysis and implementation difficult.

In some instances, decision-making groups contain relatively fixed membership and persist for long periods of time, meeting periodically to make strategic, policy, or tactical decisions (e.g., the resource management staff of a national forest—persistent, formal groups). Other groups are assembled for a short period for specific tasks (e.g., technical workshops—temporary, formal groups, q.v. Peterson et al. 1992, Rogelberg et al. 1992, Peterson et al. 1993, Schmoldt et al. 1999). Such task-oriented, temporary groups can be distinguished by high differentiation of members' skills, little synchronization within or across members' organizations, and variable lifespan (Sundstrom et al. 1990). In addition, temporary groups often lack authority or responsibility to execute decisions. While these two types of groups (and specific groups, as well) differ in decision rules, group dynamics, membership, meeting procedures, and organizational support, all types of groups havecommon problems (see Group Liabilities, below), varying only to the extent that each possesses particular problems. All can benefit from group-decision methods that facilitate dialog, mitigate adverse interactions, provide a smooth and efficient process, and produce good collective decisions.

A group-decision context provides several benefits. First, two individuals bring more knowledge to the table than one person does; each new person brings an additional amount. Second, the addition of other people to the decision process also produces an interaction effect, whereby multiple approaches to a problem can eliminate the limited scope that often hinders individual thinking. Third, if more than one person is affected by a decision, it is desirable to have those affected parties involved in the decision process. Participation increases decision acceptance and group members' ability and willingness tochampion the decision when faced with affected parties outside of the group. Because these assets are intrinsic to groups, most research has sought to identify which factors limit GDM.

Group Liabilities

Group communication is problematic due to "process losses" (Steiner 1972) associated with human interaction. On the other hand, when group interaction favors the exchange of relevant decision-making information, favorable decision outcomes occur (Vinokur et al. 1985). Group member shyness, lack of communication skills, and individual dominance affect process losses (Johnson and Johnson 1987), and social pressures to conform can stifle effective discussion (Maier 1967) and lead to group avoidance of viable alternatives (groupthink). Social loafing—relying on others to perform the group's work—is also common (Williams et al. 1981). Additional problems include personality conflicts (Maier 1967), promotion of personal agendas, and uncooperative individuals.

Groupthink (Janis 1971) is a group inertia phenomenon in which strong internal cohesion causes the group to discount and ignore external influences and warnings (Aldag and Riggs-Fuller 1993). For example, technical errors related to the *Challenger* space shuttle may have been affected by groupthink. It is generally less of a problem for temporary groups due to their membership diversity and limited lifespan. However, long-standing groups often feel that they need to justify all actions and consequently isolate themselves in an adversarial mode from the world outside of their group. This may be of particular concern for land management agencies (Vining 1992) that employ persistent groups which might feel assailed and contested at every turn. Nevertheless, groupthink can be reduced by: having decision consequences affect group members directly, removing time constraints, using methodological decision-making procedures, and eliminating a closed-leadership style of facilitation (Neck and Moorhead 1995).

Consensus within a group is important because it: (1) ensures individual ownership in, and commitment to, the group solution, (2) promotes individual satisfaction with the group outcome, (3) provides a unified (even if only majority) group decision that is viewed as more reliable and supportable by outside agents, and (4) produces a group accomplishment and avoids the perception of a lack of consensus. Majority and unanimity are the two basic decision rules used to obtain consensus (conformity in the case of majority rule), and there are several ways of achieving them (Armbruster and Böge 1983). On the other hand, expectations to conform and produce a consensus judgment can often dilute individual, specialized contributions. The failure by groups to adequately consider and accept individual opinion (when correct) often drives suboptimal group performance (Maier and Solem 1952, Janis 1971, Lamm and Trommsdorff 1973). Consequently, groups often choose a middle-ground position that compromises a better alternative for the sake of agreement (cohesion; Callaway and Esser 1984, Leanna 1985) or to merely avoid a less desirable alternative.

Workshops (as temporary formal groups) often contain an abundance of unfocused and rambling discussion, which mixes judgmental and intellective issues (Schmoldt and Peterson 1991, Peterson et al. 1992, Peterson et al, 1993, Schmoldt et al. 1999). Ideas presented in the freeform dialog of some workshops may have merit, but these ideas are not always synchronized with a logical flow of topics. While general discussions of this nature can produce beneficial results due to juxtaposed ideas, there is also a cost in inefficiencies of time and effort and the loss of ideas introduced in the wrong context.

Group Decision Techniques

A number of useful group-interaction techniques have been developed in the fields of behavior science and decision analysis that can mitigate GDM problems. They are intended to add structure to group interaction and provide "problem-centered leadership" (Frankel 1987). Probably the most familiar technique, brainstorming, simply provides for face-to-face discussion between individuals with the intent of generating ideas. In a round-robin fashion, group members offer ideas, which are recorded for later discussion. Ideas that seem to have a nominal amount of group agreement are eventually retained (McGrath 1984). Brainstorming is valuable for generative-type group tasks, such as making lists of things and generating ideas. However, individuals working alone can often generate more ideas than when working in groups, which suggests that group dynamics can have a negative impact (Lamm and Trommsdorff 1973).

Other approaches have sought to minimize or eliminate the impact of social inhibition by removing individual authorship or judgment. The nominal group technique (NGT) (Van de Ven and Delbecq 1971) leans heavily on the brainstorming approach, but augments idea generation with judgmental ranking and group evaluation. Group members provide written and individual judgments of alternative ideas, followed by score aggregation, discussion, and possibly new judgments if consensus is lacking (a round-robin protocol ensures that everyone participates). The Crawford slip method (Crawford and Demidovich 1981) is a variant of brainstorming/NGT in which ideas are written on individual slips of paper, then discussed with each idea having anonymous authorship.

The Delphi technique also maintains anonymity for participants and avoids confrontation by eliminating face-to-face interaction (Dalkey and Helmer 1963). Standard implementation of the Delphi employs questionnaires to which each member of the group anonymously responds.

Questionnaires are repeatedly administered to group members for revision, intermixed with feedback of questionnaire summaries until some convergence of opinion has been reached. The absence of group discussion avoids voice dominance by position or persuasiveness, and reduces group pressure to conform (von Winterfeldt and Edwards 1986), although the benefits of group synergy are sacrificed in the process. A cumulative Delphi procedure expands on the questionnaire approach by circulating "knowledge packets" that cumulatively record group member input and group consensus (Schmoldt and Bradshaw 1989).

NGT can be extended with more detailed quantitative methods, including multidimensional scaling (Franke1 1987), multi-attribute utility (Thomas et al. 1989), and Delphi combined with pairwise comparisons (Gargen and Moore 1984) using interpretive structural modeling (Warfield 1976). By assigning numerical scores more rigorously to alternatives, these methods permit decision makers to evaluate and understand judgments better and to explain them to others more confidently.

The stepladder technique (Rogelberg et al. 1992) is related to the modified Delphi in which face-to-face meetings are reintroduced, and member input is physically and logically cumulative. The process consists of *N-l* steps for a group with *N* members. Initially, a core group is formed by two members who work together on the assigned problem. Then, a third member is added to the group and presents his or her individual ideas to the core group for discussion, evaluation, and incorporation into the group consciousness. This process continues, adding one member at a time, until all group members have been included, all voices have been heard in turn, and the final group product is the cumulative contribution of all members. This is analogous to the cumulative Delphi (Schmoldt and Bradshaw 1989).

Multi-attribute utility theory (MAUT), as its name implies, is a decision method that combines the contributions from several attributes of decision problem. The evaluation task is partitioned into single attributes against which each of the alternatives can be scored (von Winterfeldt and Edwards 1986). These scores are often presented as utility measures (Keeney and Raiffa 1976) where values are subjective. Next, a weighting contribution is determined for each attribute. Finally, these scoring functions and their weights are combined into a mathematical model that can be evaluated for each alternative. The alternative receiving the highest utility score becomes the preferred choice.

The analytic hierarchy process (AHP) (Saaty 1980) is a decision-making framework that uses a hierarchical structure to describe a problem (decomposition), paired comparisons to rank items at each level with respect to importance (or preference or likelihood; i.e., a value judgment), and matrix multiplication to convert level-specific, local priorities into global decision priorities (aggregation). This technique has been applied to a wide variety of decision problems (Zahedi 1986a, Saaty 1990), including resource management and monitoring plans in national parks (Peterson et al. 1994, Schmoldt et al. 1994). A number of other forestry and natural resource applications of the AHP have been reported (Mendoza and Sprouse 1989, Kangas 1992, Kangas and Pukkala 1992, Pesonen 1995, Smith et al. 1995, Pukkala and Kangas 1996), including a framework for participatory decision making (Schmoldt et al. 1995). While not specifically a group-decision technique, the AHP can be used in group settings by obtaining group agreement on each paired comparison (e.g., Peterson et al. 1994, Carlsson and Walden 1995) or by geometric averaging of judgments (Saaty 1980, Schmoldt et al. 1999).

The flexibility and capability of computers make them valuable supplements to traditional tools, such as flip charts, calculators, blackboards, and paper and pencil. Computer-supported "Groupware" was initially designed to enhance office work situations, allowing individuals to jointly author documents and work collaboratively (Engelbart and Lehtman 1988). Further developments have resulted in commercial products, generally referred to as group decision support systems (DeSanctis and Gallupe 1987), that enable video-conferencing and computerized meetings with shared whiteboards and meeting organization/decision tools—including joint development of alternatives and voting (e.g., Quaddus et al. 1992). Collaborative spatial decision making (Luscombe and Poiker 1983), an outgrowth of Groupware that is more applicable to natural resource management, uses a geographic information system (GIS), group display and integration tools, and decision-making methods to provide a spatially-oriented environment for group collaboration. The Active Response GIS system (Faber et al. 1997) is an example of this group-decision support system approach.

Strategic Research Planning

Developing a long-term research program involves strategic planning. Formal studies of strategic decision-making practices have found that logical and sequential steps are rarely used. sophisticated problem formulation methods are lacking, and alternatives are not critically examined (Milliken and Vollrath 1991). The four components of strategic decision making or planning (McGrath 1984) were mentioned previously, and include: generating, choosing, negotiating, and executing. The GDM approach described below is a highly structured process that relies heavily on the AHP for its structure (refining and organizing), and utilizes brainstorming as an idea-generation mechanism. Negotiation (or agreement) is supported within the process but is not absolutely necessary due to the capability to average disparate judgments. When options (or alternatives) are prioritized with respect to both importance and feasibility, an implementation plan emerges naturally (e.g., select alternatives with high importance and high feasibility). However, we have also supplemented the process with a "straw document" that acts as an archetypal template to provide initial content for group discussions. Such a document provides the group with a starting point for deliberations, and removes much of the time-consuming, procedural gymnastics that groups experience while trying to develop an operational protocol for discussion.

In this paper, we illustrate the application of an AHP-based group decision-making process in the strategic context of formulating aresearch program for assessing the effects of large-scale fire disturbances (Schmoldt et al. 1999). We developed an AHP-based process for workshop settings based on the success of the AHP in similar group settings (Basak and Saaty 1993, Bryson 1996, Choi et al. 1994, Dyer and Forman 1992, Madu and Kuei 1995, Peterson et al. 1994, Reynolds and Holsten 1994) and its ease of application compared to MAUT (Bard 1992). The GDM process described here is potentially applicable to many types of temporary or persistent, formal group tasks. The next section describes our analytical approach in the context of the above literature on GDM, the statistical analyses conducted on AHP-derived priorities, and an illustrative example from the Fire-Disturbance Workshop.

Fire-Disturbance Workshop Process

The structure of this workshop was designed to efficiently elicit expert judgment and generate a research agenda document. The following subsection describes (1) the structure of the workgroups, which were the foci for the accomplishments of the workshop, and (2) a conceptual structure for organizing workgroup discussion. The subsequent section describes estimation and analysis of group priorities developed in this workshop. The last two subsections illustrate: specific process steps used to implement the workshop structure and specific results from one of the workgroups.

Workshop Structure

Workgroups

Workshop discussion centered on four *primary topics:* (1) linkages among fire effects, fuels, and climate, (2) fire as a large-scale disturbance, (3) fire-effects modeling structures, and (4) managerial concerns, applications, and decision support. Because these topics are relatively independent, small workgroups were used rather than one large plenary session. Each of the 25 workshop attendees was assigned to one of the four workgroups, based on their established expertise (several requests for workgroup reassignment were accommodated, however). Both scientists and managers were in attendance—in about a 3 to 1 ratio. The total number of potential experts in the area of fire science and fire management in the Pacific Northwest is relatively small, and there were several critical workgroup topics, so there were some constraints on workgroup organization. The workgroup dealing with *management applications and decision support* naturally contained more managers; this is both reasonably expected and necessary. It would be possible to skew a workgroup based on preferential selection, but the only criteria for selection used here was specialized knowledge, because the knowledge was of paramount importance in this effort.

Added benefits of small groups in the context of GDM is that each participant has more opportunity and greater willingness (less introverted behavior and less social loafing) to contribute, and social inhibitions are less pronounced. Each workgroup consisted of four to six members, dealt with a single topic, and had a discussion leader (not a facilitator, in the sense of being impartial and being uninvolved in discussions) and a recorder. The recorder used a computer to take notes, maintain idea lists, and to record judgments (which were analyzed by AHP software during the sessions). Each workgroup leader used a flip chart as a workgroup memory device. Members of each workgroup were given considerable freedom to move about and participate in other workgroups as appropriate (for informational purposes only). This encouraged wide-ranging contributions by participants (hinders introverted behavior) and between-group interaction (discourages social loafing).

After a half day of general and technical presentations on the first day of the workshop, workgroups met all day on the second day and for 2 hours on the morning of the third day to discuss and synthesize their results. Total time spent in workgroups was 10 hours. Immediately preceding workgroup discussion, all attendees were given a brief introduction to the workshop structure/process (including the use of brainstorming, the AHP, the straw document, and subsequent analyses of priority vectors). After a morning break on the third day, a member from each workgroup made a summary presentation to the entire group. This provided for plenary debriefing and allowed other attendees of the workshop to ask questions or offer suggestions.

Workgroup Discussion

Because the time frame for the workshop was limited, we provided a systematic structure for discussion so that issues were enumerated and reviewed in an organized way (Schmoldt and Peterson 1991). The structure was intended to balance efficient collection of ideas without limiting the content of those ideas. Introducing a small amount of procedural rigor was not expected to hamper timely expression of ideas. We hoped to minimize divergent discussion by maintaining focus on collecting information relevant to developing a research program on large-scale fire disturbance.

A straw document was used as a discussion template that allowed us to pre-assign topics to small workgroups and to help “jump-start” workgroup discussions (Schmoldt and Peterson 1991). All straw document content was open for revision by the workgroups except the four primary topics listed above.

Hierarchical Organization of Topics

Organization of the fire-disturbance assessment problem follows the hierarchical structure of the AHP. We organized this hierarchy using the generic concepts of primary topics, key questions, and responses. The authors and a small group of fire scientists generated the initial hierarchy present in the straw document, although workgroups often modified this structure as they developed their own topics, questions, and responses. The remainder of this description and the subsequent analyses track the hierarchy generated by the workgroup that dealt with *managerial issues, applications, and decision support* (Figure 1). This particular workgroup contained half research scientists and half resource managers.

Because each workgroup discussed a single primary topic, these workgroup subhierarchies could eventually be combined to form a global hierarchy for the workshop—each

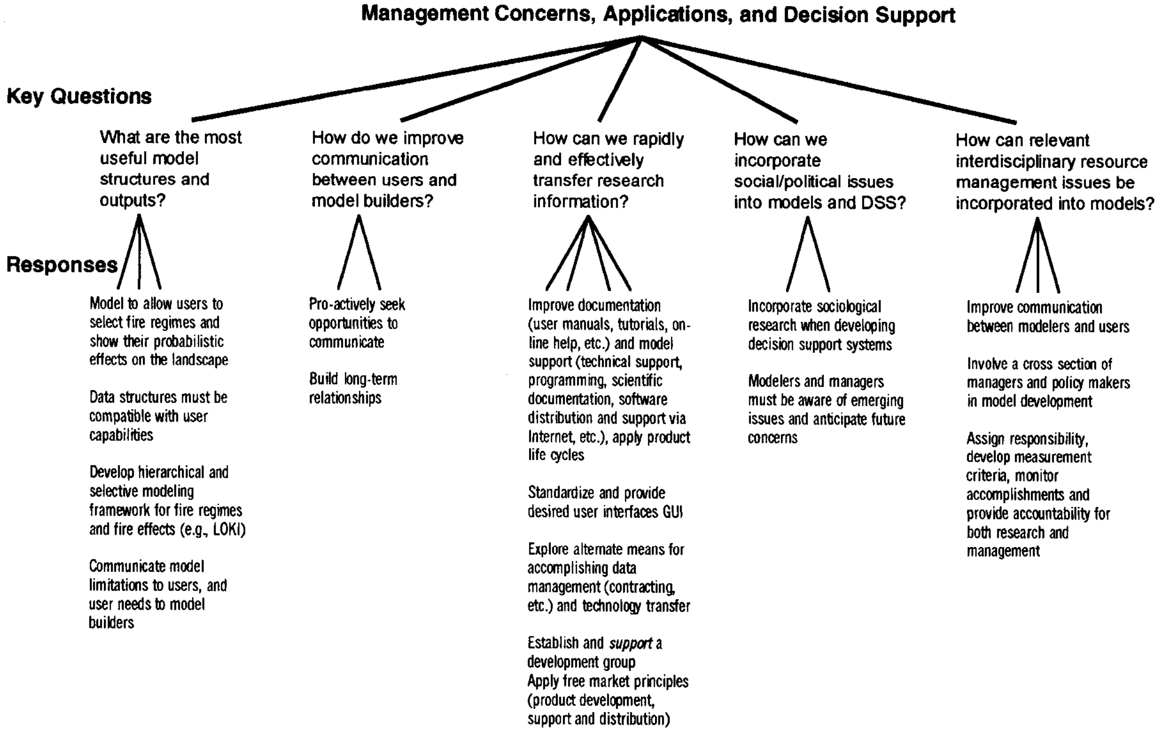

Figure 1. An AHP hierarchy from the Fire-Disturbance Workshop contains only key questions and their responses. Corresponding priority values appear in Table 1.

primary topic would be an element on level one of the global hierarchy. Subsequent levels of each subhierarchy contain *key questions* and *responses* to key questions. These generic terms for hierarchy sublevels were used because they are intuitively understandable and reflect a problem-solving approach to the workshop assignment. Their generic nature also means that the same hierarchical structure and terminology could be used for other technical workshops, or supplanted with more workshop-specific terminology. Workgroups were asked to provide rankings of the key questions with regard to both importance and feasibility. Then discussions were to proceed to *responses* within each key question, which were to be ranked in a similar manner.

The hierarchy presented in Figure 1 is not a traditional AHP hierarchy. Typically, items at each level are compared pairwise with respect to each element in the level above, and priority values are propagated down the hierarchy to alternatives (in this case, responses to key questions) at the lowest level. This produces a fully connected hierarchy, where all items on each level are connected to all items on adjacent levels. For this workshop, however, the hierarchy was singly connected; therefore, each response received only a contribution of importance (or feasibility) from one key question in the preceding level.

Subsequently, a global AHP hierarchy could be produced by combining each workgroup's subhierarchy, and comparisons could be made among the primary topics according to importance and feasibility. This step could be performed by research program managers, if importance and feasibility have strategic relevance. However, this level of comparison was beyond the scope of the workgroup context, each of which focused on a single primary topic.

The general process for each workgroup was to: (1) identify key questions in the primary topic area assigned, (2) rank those key questions with respect to importance and feasibility, (3) articulate responses to each of those key questions, and (4) rank the responses to each key question with respect to importance and with respect to the feasibility of scientific knowledge, models, and data. Before enumerating specific steps used in the workshop process, the next section describes details of how group judgments were calculated and analyzed.

Estimation and Analysis of Group Judgments

Ratio-scale derivedranking of list items is acritical part of the AHP (Saaty 1980). The AHP uses normalized, principal right eigenvectors to calculate priorities from ratio-scale judgments and matrix multiplication to integrate different levels of criteria and objectives to obtain a prioritized list of decision alternatives. In other words, the mathematical solution to the system of equations that describes the relative weights (or importance) for a set of items (compared pairwise for importance) is equivalent to the principal right eigenvector calculation for the matrix of judgments formed by those pairwise comparisons. Because individual judgments are made in a pairwise fashion, successive judgments can be mutually inconsistent. Saaty's (1980) method also includes a way to measure inconsistency, so that severe deviations from consistency can be identified and, possibly, corrected. The collection and analysis of ratio-scale judgments in the workshop setting is described below.

Ratio-Scale Judgments

During the Fire-Disturbance Workshop, workgroups were asked to make paired comparisons between items under each of their key questions with respect to importance and feasibility of knowledge, models, or data. A matrix can be constructed for each set of items being compared. Entries in this matrix would be ratio-scale values [taken from the set (1,2,..., 9)] that express how much more important (or feasible/ practical, for the other type of ranking) one item (labeling the row) is than another item (labeling the column). Numerical scale interpretations established by Saaty (1980) were used. As always, diagonal entries are 1, 1/x is the entry when the column item is x times more important than the row item, and transposed matrix elements are the multiplicative inverse of the original entry.

There are two ways to obtain pairwise judgments to enter into such a matrix. First, the workgroup could have discussed each comparison and arrived at a consensus. Second, each workgroup member could offer an individual judgment, and then all workgroup judgments would be averaged geometrically (ratio judgments). In effect, this amounts to "arithmetic consensus by contribution," where all judgments become part of the final outcome; there is no loss of minority opinion. The polling technique of the second approach was preferred, because it is much faster than reaching deliberated consensus, it gives each workgroup member equal voice, and the averaging effect of polling mitigates many consistency problems (Smith et al. 1995). Averaging (as in most circumstances) has a tendency to pull extreme values toward the central mass of a sample. In addition, multiple judgments (one for each workgroup member) result in multiple priority vectors for each set of items being compared, which provide a statistical sample of priority vectors that can be used to test for differences in priority vector elements.

A spreadsheet macro was developed to generate matrices and perform calculations during the workshop. The recorder needed only to label the row headings and enter each workgroup member's judgments. The software calculated the priority vector and consistency ratio. Because all judgments are entered into a spreadsheet, it is then possible to modify selected cells (e.g., judgments) and observe how the priorities and consistency change. Statistical analyses of priority vectors were conducted following the workshop.

Analysis of Priority Vectors

Pairwise comparisons by workgroup members allowed us to generate priority vectors for the items compared. Within a workgroup, all corresponding judgments were geometrically averaged to produce a single judgment for each comparison. This produced a group priority vector. There are two critical questions regarding the final priority vectors. One, is there general agreement among workgroup members with respect to their rankings in the priority vectors? Two, are different values within a priority vector really different?

Each judgment provided at this workshop is a sample from the population of experts' judgments on modeling large-scale fire disturbances. Consequently, each priority vector provided by a workgroup member may differ from other workgroup members. Because not all experts agree exactly, there is variation in the results. One way to be more confident in these uncertain results is to perform statistical tests. Individual judgments can be treated as "random" samples' from a population of experts that are independent and identically distributed. The resulting sample of individual priority vectors can then be analyzed statistically to answer the above questions.

Individual judgments are taken from the set (1, 2,..., 9) and their reciprocals. We can assume that this constitutes a truncated log-normal distribution (de Jong 1984, Crawford and Williams 1985, Basak 1990), or some other distribution (e.g., gamma; Vargas 1982, Zahedi 1986b), and then perform the necessary calculations to determine the distribution of the principal right eigenvector, which is the priority vector. However, this locks in assumptions about the distribution of individual judgments and can result in complicated statistical tests. Alternatively, we can assume that final priority vector elements are distributed normally and perform an analysis of variance (ANOVA) with post-hoc tests for mean differences. However, one would not necessarily expect vector elements to be normally distributed and, in fact, with the small sample size, normality tests are not very convincing. The third alternative is to conservatively apply distribution-free tests that are analogous to tests based on the normal distribution of vector elements (Smith et al. 1995), although distribution-free tests use rank information only (no magnitudes) and may fail to detect significant differences.

Three common distribution-free tests that are useful in this context are Friedman's two-way ANOVA, the Kruskal-Wallis one-way ANOVA, and Wilcoxon's signed-ranks test. The Friedman two-way ANOVA test analyzes the rankings by different workgroup members on each set of items compared. The null hypothesis is that there is no systematic variation in the rankings across items by workgroup members. The Kruskal-Wallis one-way ANOVA test indicates whether there are differences between the elements of a priority vector taking into account all workgroup member judgments. The null hypothesis is that there are no differences. While this test can indicate when differences exist, it does not specify which vector elements are different. The Wilcoxon signed-ranks test indicates which pairs of priority vector elements are different. A pairwise table of probability values is created which is equivalent to an ANOVA post-hoc test for mean differences. The combination of these three tests allows us to analyze group, and individual, rankings.

Despite their usefulness, these tests limit our ability to discern true differences because: (1) the Kruskal-Wallis test calculates probability values based on a Chi-square approximation, and the Wilcoxon signed-ranks test uses a normal approximation (for the same data, the first test may yield a significant result while the latter may not), (2) some mathematical precision is lost because ranks are used rather than actual data values, and (3) poor agreement on rankings by workgroup members (a nonsignificant Friedman test) masks

[1] Statistical theory relies on the assumption of random sampling because this ensures that each sample unit has an equal probability of being selected. In reality, however, sampling is always constrained by time and space, which necessarily restricts the choice of sample units and their relative selection probabilities. In our case, we were limited to the judgments of the leading *active* researchers and managers with knowledge of fire modeling in the Pacific Northwest. From that small group, a subgroup was selected and invited without any ideological preference, and those that were able to attend (an additional randomizing component beyond experimenter control) provided judgments. There was no way to know ahead of time what specific issues would require judgments and whatjudgments those attendees mightrender. Therefore, until the actual meeting convened, all possible expert judgments had an equal probability of being sampled.

differences between individual responses in the other two tests. Despite these limitations, we applied the distribution-free tests because they are statistically robust and are the best available tools for analyzing AHP results.

Workshop Process Steps

Each group discussed and analyzed their primary topic with respect to large-scale fire disturbances. The steps enumerated below were recommended to facilitate workgroup conduct, and these steps were followed as closely as topics and time permitted. Ranking of key questions according to feasibility was optional, because this type of comparison could be awkward or unrealistic for some topics (see Step 3 below for greater elaboration on "feasibility" of questions).

Rankings were calculated with the AHP as described below (Step 6). The AHP ranking technique has features that make it easy to implement (pairwise comparisons) and provide reliable results (judgment consistency metrics). To ensure that workgroups did not get bogged down in any particular phase of discussion, periodic deadlines were instituted during the second day. These deadlines were not absolute but were intended to keep discussion moving and to prevent hurried attention to important steps later in the process. All six steps of the workshop process were completed by each workgroup in the time allotted. The level of detail and completion achieved by each workgroup varied by topic (Schmoldt et al. 1999). The following results from the workgroup that dealt with *managerial concerns, applications, and decision support* illustrate the detailed evaluations that the process can produce.

Step 1: Brainstorm the Key Questions

Within each of the four primary topics, important questions needed to be answered regarding large-scale fire disturbances. The straw document (Schmoldt et al. 1999) contained some examples of these, but workshop participants could identify other key questions that might be more appropriate. Key questions were to be simple and concise, and participants were to avoid combining multiple or related ideas within the same question. Because this task involved idea/concept generation, rather than judgment, we felt that workgroup members would be able to reach agreement on these concepts without any formal procedures.

The intent of brainstorming is to generate lots of ideas, or in this case, key questions. In this step, workgroup members were to offer up ideas while someone recorded them. The objective was to generate key questions as quickly as possible. No evaluation of questions was to be made; rather, judgment was deferred until the discussion step below. When the production of additional key questions began to dwindle, further enumeration was to be suspended and discussion commenced. The workgroup settled on five management and application questions, along with their responses, which are listed in Table 1 in order of importance.

Step 2: Discuss the Key Questions

Key questions identified by brainstorming were further refined, and workgroups were asked to restate key questions to include a clear and unambiguous statement of the question and a thorough explanation of its rationale and its position within the primary topic. Recorders were asked to edit these descriptions as necessary and print out copies for all workgroup members to reference in subsequent discussions.

Step 3: Rank the Key Questions

By ranking only *responses* to each key question (see Step 6), it would be possible to prioritize research within each key question. Nevertheless, priorities that are more global can also be generated if key questions within each primary topic are also ranked with respect to importance *and* feasibility. Therefore, it was decided that ranking key questions would provide valuable additional information for making subsequent research agenda decisions. Feasibility of a key question, in this context, implies an ability/inability to *answer* a question; questions themselves are not infeasible. While workshop continuity would not have been disrupted if this key-question ranking step was delayed until later in the process, it was performed immediately following key question discussion, so that discussion points were reflected in the judgments.

The six workgroup members compared the five key questions appearing in Table 1 with respect to importance and feasibility. In this particular workgroup, members felt comfortable with announcing pairwise judgments verbally, while the recorder entered them into the spreadsheet. Other workgroups entered their values into judgment matrices on data sheets, which were subsequently entered into the spreadsheet model. The former process is faster, and verbal feedback on others' scores helps ensure that individuals are making the proper comparison in each case. Potential pressure to conform does exist using verbal judgments, however.

Step 4: Brainstorm the Responses

As in Step 1 above, brainstorming can also be used effectively to quickly enumerate a list of potential responses to each key question. It was not critical whether the enumeration of responses to *all* the key questions preceded discussing responses to a particular key question or, alternatively, whether enumeration and discussion of all responses to each key question were performed in turn. The only requirement was one of deferred judgment; no evaluation or critique was allowed in this step. The final list of responses to each key question is listed under each question in Table 1.

Step 5: Discuss the Responses

Because responses were intended to resolve an issue or provide a solution to the problem addressed in the key question, supporting rationale for each response was requested of the workgroups. These *justifications* were to include literature references, summarized research results, and other logical or philosophical support. Recorders edited these discussions into electronic summaries that were distributed to workgroup members before ranking.

Step 6: Rank Responses

The same procedure was used to rank responses to each key question in turn. The number of responses varied with each key question (Table 1). Again, however, six workgroup

Table 1. Key questions and responses for the workgroup topic "management concerns, applications, and decision support" are rated according to importance and feasibility. Global priority ratings appear in parentheses for each of the responses.

Importance	Key questions and responses*	Feasibility
0.43	1. What are the most useful *model structures* and outputs, to support issues in planning, operations, monitoring and learning by resource managers, decision makers, policy makers and researchers?	**0.15**
0.53 (0.23)	Model to allow users to select fire regimes and show their probabilistic effects on the landscape	0.14 (0.02)
0.19 (0.08)	Data structures must be compatible with user capabilities	0.32 (0.05)
0.18 (0.08)	Develop hierarchical and selective modeling framework for tire regimes and fire effects (e.g., LOKI)	0.23 (0.03)
0.10 (0.04)	Communicate model limitations to users, and user needs to model builders	0.31 (0.05)
0.28	2. How do we improve *communication* between users and model builders (scientists), relative to the development life cycle?	**0.44**
0.67 (0.19)	Pro-actively seek opportunities to communicate	0.85 (0.37)
0.33 (0.09)	Build long-term relationships	0.15 (0.07)
0.15	3. How can we rapidly and effectively *transfer* research *information?*	**0.17**
0.39 (0.06)	Improve documentation (user manuals, tutorials, on-line help, etc.) and model support (technical support, programming, scientific documentation, software distribution and support via Internet, etc.), apply product life cycles	0.13 (0.02)
0.27 (0.04)	Standardize and provide desired user interfaces GUI	0.31 (0.05)
0.13 (0.02)	Explore alternate means for accomplishing data management (contracting, etc.) and technology transfer	0.33 (0.06)
0.13 (0.02)	Establish and *support* a development group	0.14 (0.02)
0.09 (0.01)	Apply free market principles (product development, support and distribution)	0.10 (0.02)
0.07	4. How can we incorporate *social* and *political* issues into models/ decision support systems?	**0.06**
0.66 (0.05)	Incorporate sociological research when developing decision support systems	0.53 (0.03)
0.34 (0.02)	Modelers and managers must be aware of emerging issues and anticipate future concerns	0.47 (0.03)
0.06	5. How can *relevant* interdisciplinary resource management *issues* be incorporated into models?	**0.18**
0.61 (0.04)	Improve communication between modelers and users	0.40 (0.07)
0.29 (0.02)	Involve a cross section of managers and policy makers in model development	0.38 (0.07)
0.10 (0.01)	Assign responsibility, develop measurement criteria, monitor accomplishment and provide accountability for both research and management	0.22 (0.04)

[a] Terms in italics are used as abbreviations to reference key questions in subsequent tables.

members compared responses for each question with respect to importance. Only five workgroup members were available to compare responses with respect to feasibility. Statistical analyses of those results appear below.

Analysis of Results

Judgment matrices are not shown here for sake of brevity. Complete results for all four workgroups can be found in Schmoldt et al. (1999). For each type of ranking (importance or feasibility), we applied the distribution-free statistical tests described previously to: (1) determine how well workgroup members agreed on their rankings (Friedman test)—i.e., is the ranking of items 1, 2, and 3 for one expert in agreement with the items ranked 1, 2, and 3 by other experts, (2) determine whether there are significant differences between rating scores—i.e., are the values [0.5, 0.3, 0.2] in a group priority vector really different from each other (Kruskal-Wallis test), and (3) identify which priority vector elements are significantly different (Wilcoxon test)—i.e., if the Kruskal-Wallis test indicates that differences exist, which pairs of values are different (0.5 different from 0.3, 0.3 different from 0.2, etc.).

For key question importance, a Friedman two-way ANOVA test rejects the null hypothesis ($P < 0.0005$). This indicates that workgroup members' judgments do vary in a systematic way. That is, there is good agreement on the importance rankings across group members. A Kruskal-Wallis test for differences of mean rating scores for key question importance is also highly significant ($P < 0.0005$), suggesting that real differences exist between the rating scores. Because of this highly significant test result, a Wilcoxon signed-ranks test was used to produce a matrix of pair-wise probabilities (Table 2). These values indicate which of the key question importance scores in Table 1 may actually be different. There does not seem to be any evidence to suggest that the two highest ranked key questions *(model structures* and *communication)* are significantly different from each other, but they do differ significantly from the other three key questions. The third highest ranked key question *(transfer information)* also appears to be significantly different from the two lowest ranked questions *(relevance* and *social/political).* This suggests three different levels of importance for these key questions, with two questions at the top, two at the bottom, and the fifth question lying between the others.

For feasibility comparisons, a Friedman two-way ANOVA test marginally fails to reject the null hypothesis ($P = 0.057$). Still, allowing some latitude in the third decimal place, this test indicates that workgroup members tend to agree on their

Table 2. A Wilcoxon signed-ranks test generates a matrix of probability values for differences across means of the importance rating scores for the key questions.

	Relevant issues	Communication	Transfer information	Model structures	Social/political
Relevant issues	1.000				
Communication	0.028	1.000			
Transfer information	0.028	0.046	1.000		
Model structures	0.028	0.173	0.028	1.000	
Social/political	0.753	0.028	0.075	0.028	1.000

rankings (although weaker agreement than in the importance rankings). A Kruskal-Wallis test for differences of mean rating scores for key question feasibility is significant *(P* = 0.017) suggesting that real differences exist between the rating scores. A Wilcoxon signed-ranks test produces the pairwise probability matrix (Table 3) that indicates which of the feasibility scores in Table 1 may actually be different. The highest ranked key question for feasibility *(communication)* is significantly different from two of the other four questions. The third highest ranked key question *(transfer information)* might be significantly different *(P* = 0.067) from the lowest ranked one *(social/political),* but otherwise there are no discernible differences between key questions with regard to feasibility. Due to the reduced level of agreement on feasibility of the key questions (as compared to importance), it would be difficult to argue that any particular key question (aside from *communication)* has significantly different feasibility than the others.

For the most important and least important key questions there is evidence (based on Friedman tests) to indicate good agreement by workgroup members regarding importance rankings for their respective responses (Table 4). That is, there is good agreement for two key questions on the most important response, second-most important response, and so on. In addition, for all key questions there appear to be significant differences in the actual rating scores for the different key question responses, as indicated by Kruskal-Wallis test probability values.

More detailed tabular results for the analyses of responses to each key question are not presented here. It is worth noting a few of the significant results, however. The most important response, *fire regimes,* for key question #1 appears to rate significantly different from the other three responses. While Kruskal-Wallis tests for key questions #2 and #4, *communication* and *social/political* (each key question having only two responses), show significant differences between their respective response ratings, Wilcoxon signed-ranks tests do not. This occurs, most likely, because the two tests use different approximations—which can produce different results for small sample sizes. For key question #3, *transfer information,* workgroup members did not agree entirely on rankings in their priority vectors, so although overall means for each response showed significant differences, individual comparisons were less significant because counts of rank differences were mixed. Workgroup members agreed on priority vector ranking for responses to key question #5, *relevant issues* (the least important one). This permitted any rating differences to be easily detected by the other tests; consequently, all three responses for this key question are statistically different from each other.

Only for the most practical key question, *communication,* is there evidence to indicate good agreement by workgroup members regarding feasibility rankings of the responses. When we examine the Wilcoxon signed-rank test for *communication,* the two responses seem to be very different, with "pro-actively seek opportunities to communicate" being a much more practical response than "build long-term relationships" *(P* = 0.039). Few other significant differences are apparent for response priority vectors for any of the key questions.

In general, it was much easier for workgroup members to reach agreement on issues of importance (as opposed to feasibility) and when dealing with more specific concepts, i.e., *responses,* as opposed to more general *key questions.* The most important issues for fire management are "useful model structures" and "output to support decision-making." "Improving communication between users and model builders" also appears to be a critical issue for management, applications, and decision support. There was relatively good agreement that "pro-actively seeking opportunities to communicate" is more important *and* more practical than "building long-term relationships." For the development and application of fire models, proactive communication is an issue that can be readily addressed. It also is the most practical and cost-effective approach to ensuring that models will meet the needs of the fire management community. Combined high scores for importance *and* practicality make *communication* a key factor for the application of large-scale fire-disturbance models to management and decision support.

Table 3. A Wilcoxon signed-ranks test generates a matrix of probability values for differences across means of the feasibility rating scores for the key questions.

	Relevant issues	Communication	Transfer information	Model structures	Social/political
Relevant issues	1.000				
Communication	0.043	1.000			
Transfer information	0.686	0.144	1.000		
Model structures	0.893	0.225	0.893	1.000	
Social/political	0.138	0.043	0.068	0.225	1.000

Table 4. Lists of responses for each key question were statistically tested for workgroup member agreement on importance rankings (Friedman) and for differences in mean rating scores (Kruskal-Wallis).

Key question	Friedman test probability	Kruskal-Wallis probability
Model structures	0.035	0.007
Communication	0.221	0.041
Transfer information	0.119	0.024
Social/political	0.414	0.068
Relevant issues	0.006	0.001

Synthesis of Results

The statistical analyses presented above were conducted following the workshop. Some additional analyses and synthesis reveal several important outcomes. Experts within a workgroup differed significantly in their priority ratings for 33 of 48 comparisons—as determined by Friedman tests that failed to detect a systematic pattern. The workgroups dealing with "linkages between fire effects, fuels, and climate" and "fire as a large-scale disturbance" generally had lower internal agreement on rankings than the other two workgroups. We attribute this effect to both the uncertainty and difficulty associated with those two topics (science questions), as well as the more applied nature of the latter two topics (modeling and decision support).

Because issue importance and feasibility interact to determine research program foci, we can plot priority values with respect to those two dimensions. In Figure 2, we consider key research questions only. Intuitively, one would choose as most pressing those key questions that have both high importance and high feasibility, i.e., high, short-term research priority. In this example, one would choose "communication between model builders and users" based on its relatively high score for both importance and feasibility. Of course, this assumes that equal weight is assigned to both dimensions. Arbitrary lines are drawn in Figure 2 based on an obvious separation between the points in both the importance and feasibility dimensions.

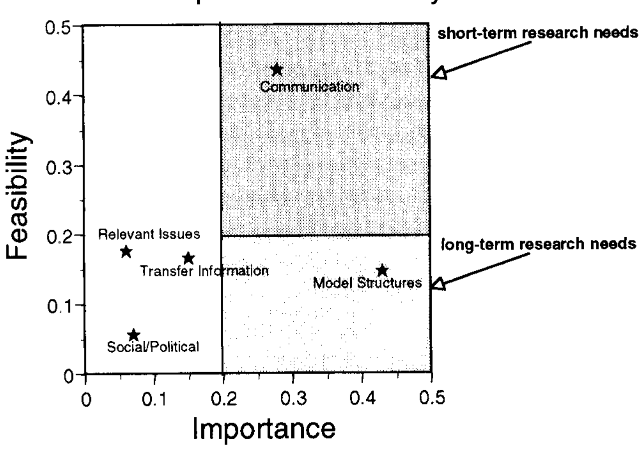

Figure 2. By plotting key question importance and feasibility priorities, we can identify: high importance-high feasibility questions (near-term research needs) and high importance-low feasibility questions (long-term research needs).

There are two immediately obvious ways to select research priorities under this dimensional examination. First, having selected a short-term research key question, we might examine, in a similar manner, responses under the selected key question. Because there are only two responses for this key question, it is obvious from Table 1 that "proactively seek opportunities to communicate between modeler builders and users" would be selected. Other key questions may have high importance but have low feasibility (due to insufficient data, models, or theory). Despite low feasibility, however, they are still important and so may be considered part of a *long-term* research agenda, instead. That is, those key questions are very important, but difficulty in answering them makes them intractable in the short term. Second, we can consider the global priorities of the responses (Figure 3), which are calculated by multiplying local priorities for each response (within each key question) by the priority assigned to the key question. In the example, the results of both approaches are identical, namely, "proactively seek opportunities to communicate between modeler builders and users" is a short-term research priority. As in MAUT, different weights and different mathematical models can be used to make the final scoring evaluation. Other possible options for synthesis are also suggested in the following section.

Discussion

Implementation

How should quantitative data collected at workshops be used in future analyses and implementation? First, one could use the results as is, selecting those items that are most important within each category (key question or response) and then working on the most practical of the important ones, or perhaps developing a combined, importance-feasibility metric to use (as was done above and in Figure 3). Second, one could select specific results from each workgroup, using judgments from only certain members of each workgroup. The members whose judgments are used in each case could vary (i.e., the 3–4 centroid vectors for each matrix could be used), or the judgments from the "most

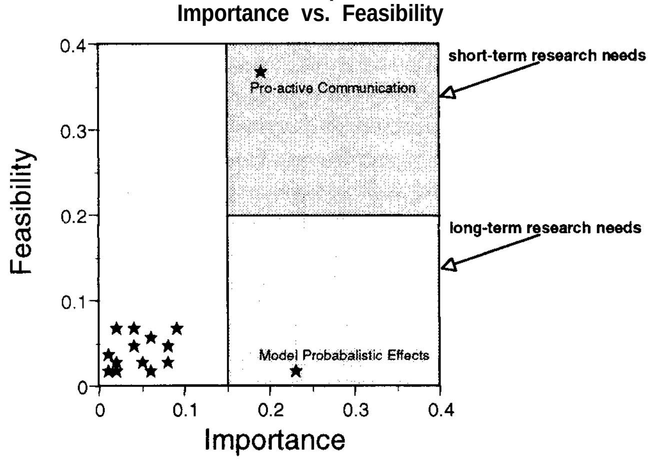

Figure 3. By plotting global importance and feasibility priorities for responses to key questions, we can identify: high importance-high feasibility responses (near-term research needs) and high importance-lowfeasibility responses (long-term research needs).

knowledgeable" or "rational" members of each workgroup could be followed throughout each analysis. In the latter case, "rationality" could be objectively determined by using the consistency ratio provided by the principal right eigenvector calculation. The "most knowledgeable" experts could also be objectively selected by using some criterion, e.g., number of peer-reviewed publications or separate evaluation by a panel of peers or professorial rank/seniority. While none of these criteria for "most knowledgeable" is without failings, the point is that one could select a criterion and use it objectively. A third way to treat the results is to calculate global priorities for averaged workgroup rankings or for each workgroup member separately [i.e., propagating priorities from one level (key questions) down to the next (responses)]. A final approach is to calculate true global priorities that also take into account priorities of the primary topics. It would be appropriate for a program manager or similar administrator to designate these high-level priorities. Including such policy/programmatic input and direction from higher organizational levels is standard for most organizations. Doing so here would add an important missing element to the more detailed decision elements already discussed in the workshop setting. Consequently, the overall decision would be a compromise of perspectives/knowledge resident in multiple organizational levels.

We suggest limiting the number of workgroup-member judgments that are used to develop programs and priorities. This "limiting" could be performed using one of the criteria suggested in the preceding paragraph. It is not necessary to rely on everyone that provides judgments; other, nonvoting workgroup members contribute in other ways (e.g., generating discussion or providing valuable insights). Those same insightful individuals may not necessarily be good at providing judgments or agreeing with others.

Group Decision Making

The structured workshop process has a number of theoretical benefits, which were anecdotally supported during the workshop. First, an explicit decision-making structure—as is afforded by the AHP, brainstorming-elicited idea generation, and the straw document—allows group interaction to focus on exchanging relevant information, with less opportunity for counter-productive group dynamics. A focus on relevant information exchange in this GDM method is afforded by a combination of the *conceptual* structure of the workshop and the *illustrative* template provided by the straw document. There will always be unproductive group interactions, but the straw document and the workshop process cultivate a centering effect that brings workgroups back to the pertinent issues despite occasional digressions. Second, consensus by contribution limits the effects of social pressures to conform (and groupthink) on individual judgments. Final individual priority vectors often variedconsiderably; this would not haveoccurred in a groupthink pitfall. By definition (Janis 1971): IF groupthink has occurred THEN unanimity occurs (group members' judgments agree). This relationship has also been established by much of the groupthink literature (as surveyed by Neck and Moorhead 1995). Of course, there are other symptoms of groupthink and other phenomena that can result in group member agreement (i.e., groupthink is sufficient, but not necessary for group member agreement). Nevertheless, this definition is logically equivalent to: IF group members' judgments disagree THEN groupthink has not occurred. Because we have demonstrated that group members' judgements disagree, this is a very strong test for groupthink nonoccurrence. Third, each workgroup member having equal contribution in the group's results is inconsistent with letting other members perform the work (social loafing). The implied weight of an individual's contribution and the realization that a judgment task is required encourage each workgroup member to participate. Fourth, the requirement that judgment matrices are eventually produced means that workgroups necessarily come to a decision and avoid group nondecision (gridlock). Fifth, by using Saaty's consistency metric, irrational judgments are identified and can be examined and revised (or retained). Sixth, although a middle-ground compromise averaged individual judgments to obtain group agreement in this workshop, those individual judgments have not been discarded and can still be used to improve workshop results. In addition, the structured decision-making process proved to be time effective, allowing workgroups to develop issues, information, and approaches for addressing fire-disturbance effects on ecosystems in a short time frame.

We were concerned about how use of the straw document as a starting point might affect workgroup dialog. That is, workgroups might be inappropriately swayed by the straw document and adopt it prima facie. However, in a post-workshop comparison of the straw document to the results from the workgroups, we found that there is no basis for anchoring concerns, because the straw document and final workgroups' results differed greatly. This may be partly due to the fact—based on prior experiences with scientists in other workshops—that scientists (a large percentage of this workshop's composition) are relatively independent and do not like to be told what or how to think. In fact, two of the four workgroups not only failed to be anchored by the straw document, but also modified their deliberation process to conform to the unique needs of their topic.

Limitations

Brainstorming was used as an idea-generation mechanism for the workshop process because of its simplicity and effectiveness; given the short time frame of the workshop, more detailed or complex idea-generation approaches would have been less practical. As noted earlier, though, individuals can often generate more ideas than brainstorming groups. In addition, brainstorming can result in process losses due to group dynamics. Consequently, more individual-friendly group techniques—e.g., NGT, Crawford slip, Delphi methods, or any mechanism that generates ideas in a group setting—can be substituted into this workshop process.

Statistically significant priority values depend on having relatively good agreement by workgroup members. We did not always get that level of agreement in the Fire-Disturbance Workshop. This does not mean that the results cannot be used, only that interpretation of the results must be done with full knowledge of their limitations.

The group decision process outlined above (or any other similar process, for that matter) will not necessarily result in near-optimal or internally consistent group decisions. In addition, the process will work only if there is a commitment to consensus-building; contentious attitudes, which may be a legitimate component in some group settings, will not lead to efficient and productive decision making in this case. Many factors influence group outcomes, and it is often more convenient to make popular rather than logically consistent decisions (Carlsson and Walden 1995). This is especially problematic when the outcome of the current task or issue is only one small vantage point on the larger political landscape. Nevertheless, the process described here allows a group to examine a problem systematically and rationally, and to generate a quantitative record that details their deliberations and provides for accountability.

Conclusions

Results of the Fire-Disturbance Workshop were submitted to the USDA Forest Service and are now available to administrators and program managers in the research (research stations) and resource management (national forests) branches of the agency. The results are also being utilized by the U.S. Department of the Interior as a component of fire research planning. Two of the general outcomes of the workshop analysis—an emphasis on synthesis and modeling over extensive new field data, and the prioritization of communication and decision support systems at the research-management interface—are components of the interagency Joint Fire Science Program recently implemented in the United States (U.S. Department of the Interior and U.S. Department of Agriculture 1998).

The workshop process facilitated the elicitation of meaningful results at the Fire-Disturbance Workshop (q.v., Schmoldt et al. 1999), although some of the scientific questions discussed were quite general. Temporary, formal groups that deal with clearly defined and applied issues are likely to see the greatest benefit from this process. This was observed in the Fire-Disturbance Workshop, in which the two application-oriented workgroups (modeling; management concerns, applications, and decision support) experienced greater agreement among workgroup members than did other workgroups, and hence, generated more statistically significant results. In fact, we observed that resource managers in the workshop adapted to the structured approach more readily than scientists, a phenomenon observed in other structured workshop settings as well (Peterson et al. 1994). The scientific topics covered in the workshop were likely more complex and less defined than issues faced by other formal groups that deal with more concrete, management-type problems.

Other subject areas, entirely different from large-scale fire disturbances, could easily be cast in the hierarchical workshop framework discussed here. Many of the identical conceptual ideas would translate directly to another subject area, perhaps accompanied by some renaming of hierarchy levels—we just happened to use generic levels, "key questions" and "responses." Therefore, the framework for workshop discussion and idea collection (brainstorming, AHP, and straw document) might also be applied to other subject areas. This GDM method contains all the key components of strategic decision making identified by social scientists (McGrath 1984): generating (ideas are produced in brainstorming sessions); choosing (matrices contain value judgments); negotiating (conflict is handled/mitigated by judgment aggregation, but individual judgments are still retained); and executing (several alternatives are given for implementation plan generation, which emerge naturally from the hierarchy and priority vectors). While this process may not be appropriate for all workshops, it can be most useful where technical discussions are expected to produce concrete and specific recommendations.

Literature Cited

ALDRAG R.J., AND S. RIGGS-FULLER. 1993. Beyond fiasco: A reappraisal of the groupthink phenomenon and a new model of group decision processes. Psych. Bull. 113:533–552.

ARMBRUSTER, W., AND W. BÖGE. 1983. Efficient, anonymous, and neutral group decision procedures. Econometrica 51: 1389–1405.

BARD, J.F. 1992. A comparison of the Analytic Hierarchy Process with multiattribute utility theory: A case study. IIE Transactions 24(5):111–121.

BASAK, I. 1990. Testing for the rank ordering of the priorities of the alternatives in Saaty's ratio-scale method. Europ. J. Oper. Res. 48:148–152.

BASAK, I., AND T.L. SAATY. 1993. Group decision making using the Analytic Hierarchy Process. Math. Comput. Model. 17(4/5): 101–109.

BOTTGER, P.C., AND P.W. YETTON. 1987. Improving group performance by training in individual problem solving. J. Appl. Psych. 72:651–657.

BRYSON, N. 1996. Group decision-making and the Analytic Hierarchy Process: Exploring the consensus-relevant information content. Comput. Oper. Res. 23(1):27–35.

CALLAWAY, M., AND J. ESSER. 1984. Groupthink: Effects of cohesiveness and problem-solving procedures on group decision making. Soc. Behav. Personal. 12:157–164.

CARLSSON, C., AND P. WALDEN. 1995. AHP in political group decisions: A study in the art of possibilities. Interfaces 25:14–29.

CHOI, H., E. SUH, AND C. SUH. 1994. Analytic Hierarchy Process: It can work for group decision support systems. Comput. Indus. Eng. 27(1/4):167–171.

CRAWFORD, C.C., AND J.W. DEMIDOVICH. 1981. Think tank technology for systems management. J. Syst. Manage. Nov.:22–25.

CRAWFORD, G., AND C. WILLIAMS 1985. A note on the analysis of subjective judgment matrices. J. Math. Psych. 29:387–405.

DALKEY, N.C., AND O. HELMER. 1963. An experimental application of the Delphi method to the use of experts. Manage. Sci. 9:458–467.

DE JONG, P. 1984. A statistical approach to Saaty's scaling method for priorities. J. Math. Psych. 28:467–478.

DESANCTIS, G., AND R.B. GALLUPE. 1987. A foundation for the study of group decision support systems. Manage. Sci. 33:589–609.

DYER, R.F., AND E.H. FORMAN. 1992. Group decision support with the Analytic Hierarchy Process. Decision Support Systems 8(2):99–124.

ENGELBART, D., AND LEHTMAN. 1988. Workingsmarter. BYTE 13:245–252.

FABER, B.G., W. WALLACE, K. CROTEAU, V. THOMAS, AND L. SMALL. 1997. Active Response GIS: An architecture for interactive resource modeling. P. 296–301 in Proc. of the GIS '97 Annual Symp. GIS World, Inc., Vancouver.

FRANKEL, S. 1987. NGT + MDS: An application of the nominal group technique for ill-structured problems. J. Appl. Behav. Sci. 23:543–551.

GARGEN, J.J., AND C.M. MOORE. 1984. Enhancing local government capacity in budget decision making: The use of group process techniques. Publ. Admin. Rev. 44:504–511.

HALL, J., AND W.H. WATSON. 1970. The effects of normative intervention on group decision-making performance. Human Relations 23:299–317

HILL, G.W. 1982. Group versus individual performance: are N + 1 heads better than one? Psych. Bull. 91:517–539.

JANIS, I.L. 1971. Groupthink. Psych. Today 5:43–46, 74–76.

JOHNSON, D., AND F. JOHNSON. 1987. Joining together: Group theory and group skills. Prentice-Hall, Englewood Cliffs, NJ.

KANGAS, J. 1992. Multiple-use planning of forest resources by using the analytic hierarchy process. Scand. J. For. Res. 7:259–268.

KANGAS, J., AND T. PUKKALA. 1992. A decision theoretic approach applied to goal programming of forest management. Silva Fenn. 26:169–179.

KAPLAN, M.F., AND C.E. MILLER. 1987. Group decision making and normative versus informational influence: Effects of type of issue and assigned decision rule. J. Personal. Soc. Psych. 53:306–313.

KEENEY, R.L., AND H. RAIFFA. 1976. Decisions with multiple objectives: Preferences and value trade–offs. Wiley, New York.

LAMM, H., AND G. TROMMSDORFF. 1973. Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review. Europ. J. Soc. Psych. 3:361–388.

LEANNA, C. 1985. A partial test of Janis' groupthink model: Effects of group cohesiveness and leader behavior on defective decision making. J. Manage. 11:5–17.

LUSCOMBE, B.W., AND T.K. POIKER. 1983. Strabo—An alternative GIS approach to decision making for planning applications in data scarce environments. P. 264–269 in Proc. of the Sixth Internat. Symp. on Automated Cartography, vol. 1. Am. Congr. on Surveying and Mapping, Bethesda, MD.

MADU, C.N., AND C. KUEI. 1995. Stability analyses of group decision making. Comput. Indus. Eng. 28(4): 881–892.

MAIER, N.R.F. 1967. Assets and liabilities in group problem solving: The need for an integrative function. Psych. Rev. 74:239–249.

MAIER, N.R.F., AND A.R. SOLEM. 1952. The contribution of adiscussion leader to the quality of group thinking: The effective use of minority opinion. Human Relations 5:277–288.

MCGRATH, J.E. 1984. Groups: Interaction and performance. Prentice-Hall, Englewood Cliffs, NJ.

MCKENZIE, D., D.L. PETERSON, AND E. ALVARADO. 1996. Extrapolation problems in modeling fire effects at largespatial scales: Areview. Int. J. Wildl. Fire 6:165–176.

MENDOZA, G.A., AND W. SPROUSE. 1989. Forest planning and decision making under fuzzy environments: An overview and illustration. For. Sci. 35(2):481–502.

MILLIKEN, F.J., AND D.A. VOLLRATH. 1991. Strategic decision-making tasks and group effectiveness: Insights from theory and research on small group performance. Human Relations 44:1229–1253.

NECK, C.P., AND G. MOORHEAD. 1995. Group think remodeled: The importance of leadership, time pressure, and methodological decision procedures. Human Relations 48:537–557.

PESONEN, M. 1995. Non-industrial private forest landowners' choices of timber management strategies and potential allowable cut: Case of Pohjois-Savo. Acta Forest. Fenn. 247:1–31.

PETERSON, D.L., D.L. SCHMOLDT, J.M. EILERS, R.W. FISHER, AND R. DOTY. 1993. Guidelines for evaluating air pollution impacts on class I wilderness areas in the California. USDA For. Serv. Gen. Tech. Rep. PSW-GTR-136.

PETERSON, D.L., D.L. SCHMOLDT, AND D.G. SILSBEE. 1994. A case study of resource management planning with multiple objectives and projects. Environ. Manage. 18:729–742.

PETERSON, J., ET AL. 1992. Guidelines for evaluating air pollution impacts on class I wilderness areas in the Pacific Northwest. USDA For. Serv. Gen. Tech. Rep. PSW-GTR-299.

PUKKALA, T., AND J. KANGAS. 1996. A method for integrating risk and attitude toward risk into forest planning. For. Sci. 42(2): 198–205.

QUADDUS, M.A., D.J. ATKINSON, AND M. LEVY. 1992. An application of decision conferencing to strategic planning for a voluntary organization. Interfaces 22:61–71.

REYNOLDS, K.M., AND E.H. HOLSTEN. 1994. Relative importance of risk factors for spruce beetle outbreaks. Can. J. For. Res. 24:2089–2095.

ROGELBERG, S.G., J.L. BARNES-FARRELL, AND C.A. LOWE. 1992. The stepladder technique: An alternative group structuring facilitating effective group decision making. J. Appl. Psych. 77:730–737.

SAATY, T.L. 1980. The analytic hierarchy process. McGraw-Hill, New York.

SAATY, T.L. 1990. Multicriteria decision making: The analytic hierarchy process. RWS Publications, Pittsburgh.

SCHMOLDT, D.L., AND W.G. BRADSHAW. 1989. A cumulative Delphi approach to knowledge acquisition. AI Applic. 3:59–67.

SCHMOLDT, D.L., AND D.L. PETERSON. 1991. Applying knowledge-based methods to the design and implementation of an air quality workshop. Environ. Manage. 15:623–634.

SCHMOLDT, D.L., ET AL. 1999. Assessing the effects of tire disturbance on ecosystems: A scientific agenda for research and management. USDA For. Serv. Gen. Tech. Report PNW-GTR-455.

SCHMOLDT, D.L., D.L. PETERSON, AND D.G. SILSBEE. 1994. Developing inventory and monitoring programs based on multiple objectives. Environ. Manage. 18:707–727.

SCHMOLDT, D.L., D.L. PETERSON, AND R.L. SMITH. 1995. The analytic hierarchy process and participatory decision making. P. 129–143 in Proc. of the 4th Internat. Symp. on Advanced Technology in Natural Resources Management, Power, J.M., M. Strome, and T. C. Daniel (eds.). Am. Soc. of Photogramm. and Remote Sens., Bethesda, MD.

SMITH, R.L., R.J. BUSH, AND D.L. SCHMOLDT. 1995. A hierarchical analysis of bridge decision makers. Wood Fiber Sci. 27:225–238.

STEINER, I.D. 1972. Group process and productivity. Academic Press, New York.

SUNDSTROM, E., K.P.D. MEUSE, AND D. FUTRELL. 1990. Work teams: Application and effectiveness. Am. Psych. 45:120–133.

THOMAS, J.B., R.R. MCDANIEL, JR., AND M.J. DOORIS. 1989. Strategic issue analysis: NGT + decision analysis for resolving strategic issues. J. Appl. Behav. Sci. 25:189–200.

U.S. DEPARTMENT OF THE INTERIOR AND U.S. DEPARTMENT OF AGRICULTURE, 1998. Joint fire science plan. U.S. D.I. and U.S. D.A., Washington, DC.

VAN DE VEN, A., AND A. DELBECQ. 1971. Nominal vs. interacting group processes for committee decision making effectiveness. Acad. Manage. J. 14:203–212.

VARGAS, G. L. 1982. Reciprocal matrices with random coefficients. Math. Model. 3:69–81.

VINING, J. 1992. Environmental emotions and decisions: A comparison of responses and expectations of forest managers, an environmental group, and the public. Environ. Behav. 24:3–34.

VINOKUR, A., R. BURNSTEIN, L. SECHREST, AND P.M. WORTMAN. 1985. Group decision making by experts: Field study of panels evaluating medical technologies. J. Personal. Soc. Psych. 49:70–84.

VON WINTERFELDT, D., AND W. EDWARDS. 1986. Decision analysis and behavioral research. Cambridge University Press, Cambridge, MA.

WARFIELD, J. 1976. Societal systems: Planning, policy, and complexity. Wiley, New York.

WILLIAMS, K., S. HARKINS, AND B. LATANE. 1981. Identificability as a deterrent to social loafing: Two cheering experiments. J. Person. Soc. Psych. 40:303–311.

YETTON, P.W., AND P.C. BOTTGER. 1982. Individual versus group problem solving: An empirical test of a best-member strategy. Organiz. Behav. Human Perform. 29:307–321.

ZAHEDI, F. 1986a. The analytic hierarchy process—A survey of the method and its applications. Interfaces 16:96–108.

ZAHEDI, F. 1986b. Group consensus function estimation when preferences are uncertain. Oper. Res. 34:883–894.