USING THE AHP IN A WORKSHOP SETTING TO ELICIT AND PRIORITIZE FIRE RESEARCH NEEDS

Daniel L. Schmoldt
Research Forest Products Technologist
USDA Forest Service
Blacksburg VA USA

David L. Peterson
Professor of Forest Ecology
USGS, Biological Resources Division
Seattle WA USA

ABSTRACT

The benefits of convening a group of knowledgeable specialists together in a workshop setting to tackle a difficult problem can often be offset by an over-abundance of unfocused and rambling discussion and by counterproductive group dynamics. In light of this workshop paradox, we have created a generic workshop framework based on the analytic hierarchy process, that efficiently elicits both workshop content and products. The 2 key components of this workshop structure are: (1) a straw document that is both generic and hierarchical and (2) pair-wise comparisons of elements at each level of the hierarchy. Top-down levels of the generic hierarchy include: primary topics, key questions, and responses to key questions. We applied this workshop structure to elicit and prioritize research needs for large-scale fire disturbances from a large (25 attendees) group of scientist and managers. Advantages of this approach are: (1) it allows for *conceptual* workshop facilitation and eliminates most traditional facilitation chores, which aim to mitigate undesirable group dynamics, (2) it does not hinder or burden technical discussions, but rather focuses and enhances them, (3) it allows workshop organizers to efficiently collect and structure ideas generated in technical discussions, and (4) it enables statistical comparisons between competing ideas.

INTRODUCTION

Issues of regional or national scope can often benefit from recommendations produced by technical workshops convened with many experts. It is always more advantageous to enlist the accumulated knowledge and expertise of many specialists, rather than one individual or small group. Each group member brings a different background and set of past experiences that can work together synergistically and complement those of other group members. This intermixing of ideas can produce innovative and effective solutions to difficult problems.

Nevertheless, our past experiences with technical workshops (Peterson et al. 1992, Peterson et al. 1993, Schmoldt and Peterson 1991) indicate that workshops often contain an abundance of unfocused and rambling discussion. The ideas presented in the, often, freeform debate during workshops are not without merit; however, they are just not always properly synchronized with any logical flow of topics. While general discussions of this nature can produce beneficial and unique results due to juxtaposed issues and ideas, the cost is often considerable waste. This surfaces as inefficiencies in time and effort and can result in ideas being lost because they are introduced at the wrong time or in the wrong context. This characterizes a workshop paradox, of sorts, that seems to arise from group dynamics.

The analytic hierarchy process (AHP) is a decision-making framework (Saaty (1980) that uses a hierarchical structure to describe a problem and paired comparisons to rank decision alternatives with respect to importance (or preference or likelihood). This technique has been applied to a wide variety of decision problems (Saaty 1990, Zahedi 1986). Schmoldt et al. (1994) describe its use for inventory and monitoring program planning and give an example (Peterson et al. 1994). Despite the AHP's effectiveness for structuring decision problems, we know of no prior attempt to apply it in a workshop setting. Its ability to systematically analyze technical issues and to provide numerical assessments of competing alternatives makes it a potentially useful format for conducting workshop discussions. Nevertheless, it remained unclear as to how well this approach would be accepted by workshop attendees and whether its structuring and quantitative capabilities could be successfully incorporated into a workshop environment where time can be constraining.

In April 1996, a group of scientists and resource managers gathered at the Fire-Disturbance Workshop on the University of Washington campus (USA) to discuss research priorities for large-scale fire disturbance modeling. The strategic objectives of the workshop were to: (1) identify the current state-of-knowledge with respect to fire effects at large spatial scales, (2) develop priorities for a scientific approach to model 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. While the focus was on the Pacific Northwest region of the USA, issues of broader national and global concern were also addressed in the workshop.

To test the AHP in a workshop setting, discussions in the fire-disturbance workshop were formulated into a hierarchical framework, and the ranking of key questions and responses was used to provide concrete recommendations for prioritizing future research studies. The structured workshop process described below was used to conduct workshop discussions, to compile information, and to elicit knowledge from participants. This process consists of: (1) a hierarchical structure to organize topics regarding large-scale fire disturbance, (2) a "straw document" to provide attendees with a starting point and an example of hierarchical topics, and (3) pair-wise comparisons among elements at each level in the hierarchy. In addition to describing the workshop process, we also describe how workgroup responses were analyzed statistically and provide abbreviated results from one of the four workgroups at the meeting.

THE WORKSHOP PROCESS

Workgroups

Workshop discussion centered around four, broad content areas, or *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 disjoint and workshop attendees possessed very specialized knowledge regarding these topics, we opted for small workgroups rather than one large plenary session. Each workgroup consisted of four to six members, dealt with a single fire topic, and had a discussion leader and a recorder. Each recorder took notes using a desktop computer, which also facilitated recording judgments and calculating priorities.

An initial four-hour period was used on the first day of the workshop to introduce important fire disturbance issues and to familiarize attendees with the workshop structure that was to be used. Workgroups met all day on the second day of the workshop, and for about two hours on the morning of the third day to tie up loose ends and discuss final results. The total amount of time actually spent in workgroup discussion during the workshop was approximately 10 hours.

A straw document was one aspect of this structured workshop process. It was intended to provide an outline that allowed us to pre-assign topics to small workgroups of participants and to provide content to help jump-start workgroup discussions (Schmoldt and Peterson 1991). The straw document (for brevity, we do not include it here) contained a preliminary list of key questions and responses for each workgroup topic, All straw document content was open for revision by the workgroups as discussions commenced.

Hierarchical Organization of Topics

Each of the primary topic areas was arranged hierarchically and organized into primary topics, key questions, responses, and response rankings. A simple graphical presentation of this organization appears in Figure 1. At successive levels, more and more details were requested. Initially, workgroups were to enumerate key questions for their primary topic. Depending on workgroup preferences, they could then rank key questions with respect to importance and/or feasibility. Then discussions were to proceed to responses to key questions, and finally to ranking responses within each key question. Specific details that would eventually be expressed in the combination of responses and response rankings were expected to provide (1) research issue priorities, (2) feasibility or practicality of those issues, and (3) recommendations about which issues should be targeted for research study.

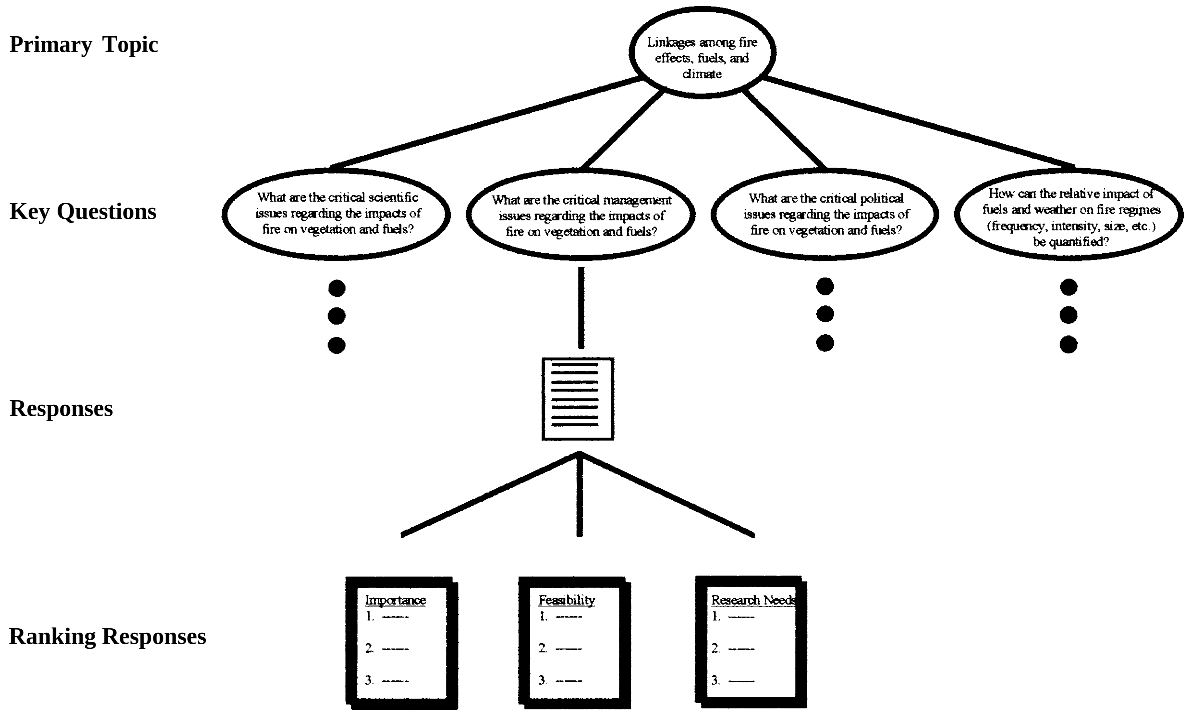

Figure 1. An example from one of the primary topics in the straw document illustrates key questions, responses, and response rankings. Rankings for the key questions are not depicted here.

To maintain proper perspective as issues were identified and discussed, all items at each level were to be considered before moving down to the next level. That is, all key questions were to be listed prior to listing any responses to key questions, and all responses to a key question were to be listed before discussing or ranking them.

The overall task for each workgroup was to: (1) identify *key questions* — of narrow and broad scope—in the *primary topic* area assigned, (2) if appropriate, rank key questions both with regard 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 for answering the question and also with respect to the feasibility/practicality of scientific knowledge, models, or data.

Conduct of the Discussion

Using the straw document content and the workshop structure outlined above as guides, each group discussed and analyzed their primary topic with respect to large-scale fire disturbances. The steps enumerated below were strongly recommended to facilitate workgroup conduct, and these steps were followed as closely as possible.

It was acceptable for workgroups to apply techniques other than brainstorming (see Steps 1 and 4, below) to obtain lists of key questions and responses. The brainstorming method, however, is easy, familiar, and fast, which is why it was recommended to the groups. Workgroups could potentially consume valuable time reinventing their own list-generation protocol, and it was strongly discouraged given the limited timeframe. Also, it was recommended that rankings be performed using the approach described below (Step 6). The ranking technique (using ratio-scale judgment matrices) has features that make it easy to implement and ensure reliable results. As noted above, this process was designed to facilitate the collection and organization of ideas and was not proposed to predetermine the content of the ideas themselves.

STEP 1: Brainstorm the Key Questions The intent of brainstorming is to generate lots of ideas—'in this case, key questions. In this step, workgroup members were to offer up ideas while someone recorded them on a flip chart. No evaluation of questions was to be made; rather, judgment was to be deferred until the discussion step below. The concept of deferred judgment is a critical aspect of the brainstorming method. Mixing enumeration and evaluation obfuscates idea generation, so it was discouraged. When the production of additional key questions began to dwindle, further enumeration was to be suspended and discussion commenced.

STEP 2: Discuss the Key Questions Because the initial enumeration phase of brainstorming should not have fully described key questions, they were to be further refined here. Workgroups were asked to restate key questions carefully and unambiguously. They were also asked to include additional prose to elaborate on the key question, for example, why it is important, what various terms mean, whether it addresses a single question or multiple issues, etc. After these discussions were completed, each key question should have included: (1) a clear and unambiguous statement of the question and (2) a thorough explanation of its rationale and its position within the primary topic. At this point, the recorders were asked to add some minimal polish to these descriptions and print out copies for all workgroup members to reference in the remainder of the day's discussions.

STEP 3: Rank the Key Questions As research is conducted on large-scale fire disturbance as part of Forest Service research, it will be necessary to make decisions about which studies receive higher priorities than others. By ranking responses to each key question later (see Step 6), it would only be possible to prioritize research within each key question individually. A more global priority can be generated if key questions within each primary topic are also ranked with respect to importance and feasibility y/practicality. Then, by multiplying the priority of each response with the priority of its associated key question, a global ranking for all issues within a primary topic can be produced. Later, the research program managers could, if they desired, also prioritize the primary topics for each key question to further define global priorities.

STEP 4: Brainstorm the Responses As in Step #l, above, brainstorming could 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 was performed in turn. The only requirement, as before, was one of deferred judgment, that is, the activities of enumerating responses and discussing them remained separate.

STEP 5: Discuss the Responses Responses to key questions were expected to be much shorter and specific than the questions themselves. Therefore, the statement of each response would be easier to formulate and would require less explanation. However, because responses were intended to resolve an issue—i.e., provide a solution to the problem addressed in the key question—there needed to be supporting rationale for each response. These justifications were to include literature references, summarized research results, and other logical or philosophical support. Again, as in Step 2, recorders were asked to record and polish these discussions into an electronic summary that could be printed and distributed to workgroup members for the final step of ranking.

STEP 6: Rank the Responses Each workgroup could have been asked to ordinally rank all responses with respect to each key question, but this would have provided only an ordered list (1 .—, 2.—, etc.). An ordinal ranking conveys no information about the relative distance between list items—the relative importance of #1 versus #2, for example. By generating a cardinal list, instead, where each list item has an associated numerical value on a meaningful scale, an explicit measure is obtained for the relative differences between list items. Such a list would provide much more information about the relative merit of issues. One good way to do this is to apply the paired comparison method used in the AHP. Here, items are compared pair-wise using a ratio scale; that is, list items are compared two at a time, and one is judged as x times more important (or preferred or likely) than the other. An aggregate set of comparisons can then be mathematically treated to produce a prioritized numerical list.

CALCULATING AND ANALYZING PRIORITY VECTORS

Because each workgroup contained 4-6 members, we needed some way to deal with multiple (and most likely different) judgments. There are two ways to obtain pairwise judgments from a group to enter into a matrix. First, the workgroup could have discussed each comparison and arrived at a consensus judgment. Second, each workgroup member could offer an individual judgment, and then all workgroup judgments would be averaged geometrically. The polling technique of the second approach was preferred in this workshop for several reasons. First, it is much faster than reaching consensus. Second, it gives each workgroup member equal voice. Third, the averaging effect of polling usually mitigates inconsistency problems. And fourth, multiple judgments for each comparison, in effect, provides us with a statistical sample of priority vectors, which we can then use to test for differences in priority vector elements.

To avoid having this method for prioritizing become excessively cumbersome, an Excel* Add-in was developed to generate initially empty matrices and subsequently to perform all the calculations from filled-in matrices. The recorder needed only to label the matrix row headings and enter each workgroup member's judgments. The software calculates the priority vector and consistency ratio. Because all judgments were entered into a spreadsheet, it would then be possible to modify selected cells (e.g., judgments) and observe how the priorities and consistency change. There was not sufficient time for this sensitivity analysis during the workshop, but subsequent application of these analyses to prioritize future research studies could effectively use this capability.

Pairwise comparisons by workgroup members allowed us to generate priority vectors for the items being compared. These priorities were for either "importance" or for "practicality/feasibility". Within a workgroup, all corresponding judgments were geometrically averaged to produce a single judgment for each comparison. This produced a *group* priority vector. But, there were two questions that we could ask about the final priority vectors. One, Is there general agreement among workgroup members with respect to the rankings in the priority vector? Two, Are different priority values in a priority vector really different? Answers to these 2 questions would have a significant bearing on how the final rankings would be used to select research priorities for large-scale fire disturbance modeling.

---

* Tradenames are used for informational purposes only, No endorsement by the U.S. Department of Agriculture is implied.

The aggregate judgments used to create a group priority vector can be treated as samples from a population of experts that are independent and identically distributed. Then, individual priority vectors can be generated from the judgments of each workgroup member, separately. The resulting *sample* of 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 could assume that this constitutes a truncated log-normal distribution (Basak 1990, Crawford and Williams 1985, de Jong 1984), or some other distribution, e.g. gamma (Vargas 1982, Zahedi 1986), and then perform the necessary calculations to determine the distribution of the principal right eigenvector, which is the priority vector calculated by the AHP. This, however, locks in assumptions about the distribution of individual judgments, and can result in unnecessarily complicated statistical tests. Alternatively, we could assume that final priority vector elements are distributed normally and perform an analysis of variance, 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, a normality assumption cannot be justified, The third alternative, and the one chosen here and used by Smith et al. (1995), is to conservatively apply distribution-free tests that are analogous to tests which are based on the normal distribution of vector elements. The drawback, however, is that distribution-free tests are conservative and so they may fail to detect significant differences when they exist.

Each of the following distribution-free tests ranks the data prior to calculating statistics, so relative magnitude information is lost (SYSTAT 1992). This constitutes the conservative nature of these tests. The Friedman two-way analysis of variance test analyzes the rankings of each set of items being compared for the different workgroup members. The null hypothesis is that there is no systematic variation in the rankings across items by workgroup members. The Kruskal-Wallis one-way analysis of variance test indicates whether there are differences between the priority vector elements taking into account all workgroup member's judgments. The null hypothesis is that there are no differences. While this test identifies that differences exist, it does not indicate which vector elements are different. The Wilcoxon signed ranks test determines which pairs of priority vector elements are different by creating a pair-wise table of probability values that is equivalent to an ANOVA post-hoc test for mean differences.

RESULTS

As an example of workgroup output, we provide some results from one of the workgroups that dealt with issues related to managerial concerns and decision support for fire disturbance models. After some initial discussion covering a broad range of topics, the workgroup settled on a short list of key questions. These 5 management and application questions are listed below, in order of importance (Table 1). The following analyses examine rankings of both importance and practicality for the key questions only (in the interest of brevity). For each type of ranking (importance or practicality), we applied the distribution-free statistical tests described previously to: (1) determine how well workgroup members agreed on their rankings of key questions, (2) determine whether there are significant differences between rating scores for the key questions, and (3) identify which key questions are significantly different.

Table 1. Management concerns, applications, and decision support key questions are rated according to importance and practicality.

Importance	Key Questions	Practicality
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
0.28	2. How do we improve communications between users and model builders (scientists), relative to the development life cycle?	0.44
0.15	3. How can we rapidly and effectively transfer research information?	0.17
0.07	4. How can we incorporate social and political issues into models/ decision support systems?	0.06
0.06	5. How can relevant interdisciplinary resource management issues be incorporated into models?	0.18

Six workgroup members compared the 5 key questions appearing in Table 1. Actual judgments are not shown here for brevity reasons. A Friedman two-way analysis of variance test rejects the null hypothesis ($p < 0.001$), indicating that workgroup members judgments do vary in a systematic way. That is, there is good agreement on the rankings across group members. A Kruskal-Wallis test for differences of mean rating scores for the key questions is also highly significant ($p < 0.001$) suggesting that real differences exist between the rating scores. A Wilcoxon signed-ranks test produces a matrix of pair-wise probabilities (Table 2) that indicates 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 2 highest ranked key questions *(model structures* and *communication) are* significantly different. These 2 key questions do differ significantly, however, from the other 3 key questions. The third highest ranked key question *(information transfer) also* appears to be significantly different from the 2 lowest ranked questions *(relevance* and *social/political).* So, there seems to be 3 significant levels of importance for these key questions—with 2 questions at the top, 2 at the bottom, and the 5th question lying between the others.

For practicality comparisons, a Friedman two-way analysis of variance test only marginally rejects the null hypothesis ($p = 0.057$), indicating that workgroup members tend to agree on their rankings. A Kruskal-Wallis test for differences of mean rating scores for key question practicality is significant ($p= 0.017$) suggesting that real difference exist between the rating scores. A Wilcoxon signed-ranks test produces a matrix of pair-wise probabilities (Table 3) that indicates which of the practicality scores in Table 1 may actually be different. The highest ranked key question for practicality *(communication)* is significantly different from 2 of the other 4 questions. The second highest ranked key question *(information transfer)* 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 practicality.

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.

	Relevance	Communi-cation	Info transfer	Model structures	Social/ Political
Relevance	1.000				
Communication	0.028	1.000			
Info transfer	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

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

	Relevance	Communi-cation	Info transfer	Model structures	Social/ Political
Relevance	1.000				
Communication	0.043	1.000			
Info transfer	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

CONCLUSIONS AND DISCUSSION

Improving communication between users and model builders appears to be a very critical key question for the primary topic, “management, applications, and decision support.” *Communication is* the second highest ranked key questions with respect to importance. However, its rating score is not significantly different from the highest ranked key question, "useful model structures and output to support decision-making”, but it is significantly different from the other 3 lower scores. This gives this key question high importance overall. Also, it is the highest ranked key question with respect to practicality, and its practicality score is significantly different from 2 of the other key questions. Although, better agreement among workgroup members for the practicality ranking (Friedman test, p = 0.057) would have made communication significantly different from all other key questions. Nevertheless, communication seems to be as highly practical as any other key question. Combined high scores for importance and practicality make “communication between users and model builders” a key managerial concern for the application of large-scale fire disturbance models.

Leaders from each of the workgroups have recast their workshop results into a separate chapter of a workshop report. This report will be published by the USDA Forest Service. It is expected that these results will not only help guide the current research program on large-scale fire disturbance modeling, but will also help other researchers focus on high profile research needs.

From the length of the current draft of this document, it is apparent that each workgroup accomplished a considerable amount of work in a 10-hour period. We attribute this productivity to the workshop structure and process described above. In fact, several workshop participants acknowledged anecdotally that they could not have analyzed their topic area as effectively in the time allotted without this amount of organization. Yet, no one expressed dissatisfaction with the process or thought that it hindered them in any way. Thus, we feel that the workshop structure was successful and accomplished these four objectives:

- The workshop was facilitated *conceptually* by formulating a rigorous, yet flexible, a priori structure and process.
- The workshop organization did not get in the way of, or burden, technical discussions.
- Workshop participants were readily convinced that this structure would help them and simplify their efforts.
- Ideas generated in technical discussions were efficiently collected and organized.

Other subject areas, entirely different from large-scale fire disturbance modeling, could easily be cast into this framework. Many of the same conceptual ideas—i.e., key questions and responses to them—should carry over directly to another subject area, perhaps accompanied by some renaming of hierarchy levels. Therefore, while we initially set out to provide a flexible framework for workshop discussion and idea collection dealing with large-scale fire disturbances, we have instead produced a workshop process that can be applied broadly to many other subject areas. 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.

REFERENCES

Basak, I. (1990). Testing for the rank ordering of the priorities of the alternatives in Saaty’s ratio-scale method. *European Journal of Operational Research,* 48:148-152.

Crawford, G., and C. Williams. (1985). A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology,* 29:387-405.

de Jong, P. (1984). A statistical approach to Saaty's scaling method for priorities. *Journal of Mathematical Psychology,* 28:467-478.

Peterson, J., D. L. Schmoldt, D. L. Peterson, J. M. Eilers, R. W. Fisher, and R. Bachman. (1992). Guidelines for evaluating air pollution impacts on class I wilderness areas in the Pacific Northwest. Gen. Tech. Rep. PNW-GTR-299. Portland OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 83 p.

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. Gen. Tech. Rep. PSW-GTR-136. Albany CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station. 34 p.

Peterson, D. L., D. L. Schmoldt, and D. G. Silsbee. (1994). A case study of resource management planning with multiple objectives and projects. *Environmental Management,* 18(5): 729-742.

Schmoldt, D. L., and D. L. Peterson. (1991). Applying knowledge-based methods to design and implement an air quality workshop. *Environmental Management,* 15:623-634.

Schmoldt, D. L., D. L. Peterson, and D. G. Silsbee. (1994). Developing inventory and monitoring programs based on multiple objectives. *Environmental Management,* 18(5): 707-727.

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 PA.

Smith, R. L., R. J. Bush, and D. L. Schmoldt. (1995). A hierarchical analysis of bridge decision makers. *Wood and Fiber Science,* 27(3): 225-238.

SYSTAT. (1992). *Statistics, Version 5.2 Edition.* SYSTAT Inc.: Evanston IL

Vargas, G. L. (1982). Reciprocal matrices with random coefficients. *Mathematical Modelling,* 3(1); 69-81.

Zahedi, F. (1986). A simulation study of estimation methods in the analytic hierarchy process. *Socio-Economic Planning Science,* 20(6): 347-354.

1997
ACSM/ASPRS

Annual Convention & Exposition Technical Papers

ACSM 57th Annual Convention
ASPRS 63rd Annual Convention

Seattle, Washington
April 7-10, 1997

Resource Technology

Volume 4
Resource Technology Institute