Ahti Salo and Juuso Liesiö. 2006. A case study in participatory priority setting for a Scandinavian research program. International Journal of Information Technology & Decision Making, volume 5, number 1, pages 65-88.

International Journal of Information Technology & Decision Making
Vol. 5, No. 1 (2006) 65–88

A CASE STUDY IN PARTICIPATORY PRIORITY SETTING FOR A SCANDINAVIAN RESEARCH PROGRAM

AHTI SALO[*,†] and JUUSO LIESIÖ[‡]

Systems Analysis Laboratory, Helsinki University of Technology
P.O. Box 1100, Helsinki 02015 TKK, Finland
[†]*ahti.salo@tkk.fi*
[‡]*juuso.liesio@tkk.fi*

Continuing attempts to align science and technology policies with industrial and societal needs have aroused interest in the determination of research priorities. In this paper, we report a case study where leading experts from industry and public administration were assisted by multicriteria decision analysis in the planning of a collaborative research program for Scandinavian forestry and forest industries. We also address processual and methodological challenges in the deployment of multicriteria methods, and argue that such methods can contribute to the quality of decision support processes in related contexts.

Keywords: Decision analysis; group support systems; research and technology development programs; research evaluation; technology foresight.

1. Introduction

The development of research priorities is a central concern in the preparation of publicly supported research programs (cf. Hartwich and Janssen,[1] Keenan[2] and OECD[3]). This task is complicated by considerations such as major uncertainties, vested interests, lack of unequivocal "alternatives" and asymmetric distribution of relevant information, among others. Arguably, these complications have grown over the past few years, partly due to increasing pressures on the demonstration of the benefits of public science and technology (S&T) expenditure and the recognition of an ever broadening range of concerns in the assessment of these benefits (see, e.g. Smits and Kuhlmann[4]). Because these pressures are likely to persist, there is a demand for decision-aiding processes through which these priorities can be established in a defensible and transparent manner, in view of the concerns that must be brought to bear on them.

In this paper, we describe a case study in which novel methods of multicriteria decision analysis were deployed in the planning of a research program for the

*Corresponding author.

Scandinavian forestry and forest industries. To a considerable extent, this case study was based on a participatory workshop where relevant stakeholders from Finland, Norway and Sweden evaluated complementary research themes with the help of a tailored group support system (GSS). Specifically, this GSS was employed to solicit both quantitative and qualitative inputs in order to inform the shaping of the research program on wood materials science, launched in 2003 with the aim of fostering the long-term industrial competitiveness of forestry and forest industries in Scandinavia. The rationale for this program can be seen in relation to major industrial trends: for example, after numerous mergers and acquisitions, many firms are no longer "national", wherefore the innovation policies of the countries they operate in need to be better coordinated too.

Apart from the novel use of methodological support, there are also other reasons that suggest that our case study is of broader interest. First, the resulting program is truly international in terms of its funding structure, scope of topics and participating organizations, although much of the earlier literature on research priority setting has focused on either national or regional exercises (see, e.g. Grupp and Linstone[5]). Second, the priority setting process was supported by a GSS through which both formal evaluative statements and informal verbal comments were solicited; our concern with the complementary roles of quantitative and qualitative analysis can thus be contrasted with earlier accounts that have tended to emphasize the quantitative aspects of multicriteria analyzes (see, e.g. Henriksen and Traynor[6]). Third, the process in Finland was organized hierarchically so that preliminary results from three national workshops were injected into the Scandinavian workshops for elaboration and validation. The process outline is therefore potentially applicable also in other settings where research priorities need to be synthesized through a hierarchical approach.

The rest of this paper is organized as follows. Section 2 discusses concerns in the development of research priorities and reports experiences from the use of multicriteria methods in the preparation of publicly supported research programs. Section 3 outlines the objectives that were placed on the research program on wood material research and describes the participatory workshops. Drawing upon experiences from this case study, Sec. 4 outlines methodological recommendations and considers the role of systematic methodologies in attaining quality attributes such as transparency, comparability and repeatability.

2. Concerns in Multicriteria-Assisted Priority Setting

Multicriteria methods are extensively employed in many domains of socioeconomic planning, and sometimes they are even endorsed by regulatory requirements (see, e.g. Keefer *et al.*[7] and Hämäläinen[8]). In the development of research priorities, however, they are employed less consistently than in domains such as environmental impact assessment.[9] Indeed, several observations suggest that priority setting in research is an especially challenging problem context:

(i) *Diffuse boundaries*: In the development of research priorities, one can rarely start with an existing list of alternatives consisting of well-defined, mutually exclusive and exhaustive research themes; rather, such a list has to be generated during the earliest phases of the analysis. Moreover, these alternatives may co-evolve as the program objectives are gradually clarified (see, e.g. Salmenkaita and Salo[10]), wherefore it is necessary to address issues of scope, objectives and priorities *in conjunction* with each other.

(ii) *Distributed expertise*: The relevant expertise on scientific, technological and societal questions is scattered among many stakeholders (e.g. researchers, industrialists, civil servants). Thus, there is a need for consultative processes through which the relevant expertise can be synthesized in support of informed decision making.

(iii) *Vested interests*: Proponents of competing research areas may have conflicting interests, if only because the allocation of resources to some specific topics implies that reduced resources remain available for other topics. Hence, it cannot be taken for granted that researchers would not put forth exaggerated future-oriented claims about the beneficial socioeconomic impacts of their work. Although information asymmetries, high uncertainties and long-time horizons make it difficult (if not impossible) to ascertain the truthfulness of these claims (see, e.g. van der Meulen[11]), there is still a need to recognize the possibility of overstatements and to institute mechanisms for controlling them.

(iv) *Systemic objectives*: As an instrument of innovation policy, the functions of research programs range from the mitigation of market failure by means of public funding to the elimination of other deficiencies in the innovation system (e.g. systemic failures, structural rigidities and anticipatory myopia; see Salmenkaita and Salo[10]). Thus, concerns in program planning are not limited to the development of research priorities *per se*, because they extend to the design of purposeful measures for achieving the desired structural and behavioral changes.

In the management of publicly supported research programs, project proposals are typically solicited through the Calls for Proposal mechanism and other procedures through which intended priority areas are communicated to the research community. Because the development of these priority areas differs in many ways from the actual selection of projects, it is useful to distinguish between the following two phases of priority setting:

(i) The *formation of priorities*, which produces a list of research topics (i.e. the "priorities") that are deemed suitable for the program, based on preliminary consultations at a time when the program has not yet been formally started. The resulting list must usually be developed in the absence of (complete) information on the projects through which the priorities might be realized.

(ii) The *implementation of priorities*, which is carried out by soliciting, developing and selecting project proposals that are aligned with the priorities that result from the formation phase.

As we shall argue later on, these two phases differ in terms of their objectives and demand for methodological support. During priority formation, there is typically a need for problem structuring methods to support the clarification of objectives and research topics. The implementation phase, in turn, calls for the ability to evaluate project proposals with regard to the criteria that have been thus elaborated. It is more structured and hence also more readily supported by formal approaches such as multicriteria methods.

This notwithstanding, priority formation is arguably the more important phase, because it exerts a major influence on the kinds of proposals ultimately received. Indeed, improvements in priority formation can thus be motivated by the desire to obtain the best possible proposals for subsequent scrutiny, wherefore these improvements also hold considerable upside potential in terms of enhancing the research quality and industrial relevance of the projects that are finally approved. In contrast, the benefits of methodological support for the implementation phase are inevitably constrained by the quality of available project proposals.

Conceptual, processual and methodological differences in the formation and implementation of priorities help explain why relatively little research has been done at the *juncture* of these two phases. In effect, much of the applied work in the formation of priorities has appeared under the label "technology foresight" (see, e.g. MacLean *et al.*,[12] Martin[13] and Salo and Cuhls[14]), while case studies in the implementation of priorities have been reported in the literature on decision analysis and technology management, among others (see, e.g. Liberatore[15] and Henriksen and Traynor[16]). In what follows, we cite some examples that extend from methodological accounts of large-scale foresight exercises to the use of multicriteria methods in the implementation of priorities.

In an extensive coverage of the U.K. technology foresight program, Keenan[2] gives a critical appraisal of the methodological approaches with the help of which 15 sectoral expert panels sought to produce recommendations for improving industrial competitiveness and quality of life through S&T policy measures. He also gives an insightful discussion of the difficulties that were encountered in this program, particularly in synthesizing the outputs of the relatively independent panels toward a reduced number of recommendations.

Durand[16] describes the French technology exercise "Key Technologies 2005," the purpose of which was to (i) identify, select and characterize key technologies of national importance and (ii) evaluate the competitive position of France and the European Union (EU) on these technologies. During the early phases of the exercise, more than 600 technologies were evaluated in qualitative terms with regard to some 30 criteria. This resulted in some 120 key technologies that were described with a characterization grid extended from market needs and technological possibilities, in the realization that "technology foresight is bound to be a dual exercise, both technology and demand led." Building on the experiences from this exercise, Durand distills lessons that pertain to the pervasiveness of generic technologies, the recognition of interrelationships among technologies and the management of communication processes in such an exercise, among others.

Hartwich and Janssen[1] present a multicriteria model based on the Analytic Hierarchy Process (AHP)[17] for priority setting in agricultural research. Their model contains three criteria (farm income, food supply, natural resource management) and four alternatives (breeding, agronomy, plant hygiene, post-harvest), which are portrayed as proposals for alternative research projects that might be eligible for funding. Although Hartwich and Janssen discuss the potential of the AHP in this context and suggest that it can be helpful in such contexts, their claims are not vindicated by empirical evidence, as their small example is purely a hypothetical one (i.e. there were no research projects or funding agencies involved).

Braunschweig *et al.*[18] present an extensive case study on the prioritization of projects for a Chilean research program in agricultural biotechnology. Specifically, they outline an iterative framework for the development of evaluation criteria and describe how this framework was employed to support the development of 32 criteria that were applied to the appraisal of research projects. They also describe how the different various tasks of the appraisal necessitated a clear separation of tasks in (i) the elicitation of score information for rivaling project proposals (which were provided by project managers) and (ii) the specification of criteria weights (which were provided by the responsible policy makers in industry and public administration). They also draw attention to possible problems in obtaining truthful inputs from the project managers, and discuss how such problems may be mitigated by obtaining inputs from other experts and by encouraging open-ended mutual critiquing among project managers.

Salo *et al.*[19] survey several multicriteria methods in support of priority setting activities. Among these, value trees,[20,21] the AHP[17,22] and PRIME[23] are based on the use of an additive weighting model (see also Xu[24]), while data envelopment analysis[25] can be used to assess the efficiency of technological options in view of their input–output characteristics. In a related paper, Salo *et al.*[26] report an application of multicriteria methods to the prospective evaluation of a cluster program for the Finnish forestry and forest industries. They also discuss the complementary roles of multicriteria methods in summative and formative evaluations. That is, while conventional uses of multicriteria methods are suitable for *summative* evaluations, which aggregate project-level information for comparative analysis at higher levels of organizational decision making, there is also a need for *formative* evaluations, which seek to interpret results in terms of implications for individual projects and research areas.

3. A Planning Process for a Scandinavian Research Program

After the closing stages of the Wood Wisdom cluster program — which was one of the seven national cluster programs initiated by the Finnish Government in the late 1990s — there was a widespread perception that a continued research program in the field of wood materials science was called for.[26] Moreover, it was felt that such a program should be more international in terms of its funding structure and promotion of collaborative projects. These views were reciprocated in Sweden, whose forestry and forest industries share similar interests in relation to the framework programs of the European Commission, for example. As a result, initial consultations among the representatives of funding agencies and industrial representatives from the two countries led to the decision that a sequel program on wood materials science should be started. This program was launched in early 2003, with a total funding volume of some 20 million euros, of which about 80% consists of public funding. Detailed information on the program, the funding organizations as well as the projects that were approved for funding can be found at http://www.woodwisdom.fi/en/. On this Web site, the documents resulting from this planning process can also be found.

In what follows, the planning process of the research program on wood materials science is described with an emphasis on the participatory activities held in Finland. The international workshop — held at the Helsinki University of Technology to support the formation of Scandinavian research priorities — was organized according to the same principles as the three earlier workshops through which the national priorities in Finland were set. We believe that these workshops are of interest in that they were characterized by (i) a comprehensive *ex ante* characterization of program scope, (ii) "bottom-up" solicitation of suggestions for relevant topics and (iii) a systematic appraisal of research themes, based on several judgmental inputs that were solicited with the help of a multicriteria model.

The methodological design for the participatory process was developed in close collaboration with the first author and the program manager of the Wood Wisdom cluster program (see, e.g. Salo *et al.*[26]). This design phase extended over a 5-month period and consisted of some half a dozen meetings which usually lasted 2 h or less. In the first few meetings, the context and the objectives of the program were clarified. The later meetings focused on the elaboration of complementary research themes and the specification of evaluation criteria. The last meeting was concerned with the scheduling of workshops and the selection of participants, in the realization that the priorities would have to be developed within a relatively tight schedule.

The methodological design described in this paper is not the only possible one, but it is compatible with and partly implied by several overarching principles that were deemed desirable in and of themselves. Thus, as a matter of principle, considerable emphasis was laid on (i) the *a priori* consultation of researchers, in order to enrich workshop discussions and to ensure that the program management would gain a realistic view of the topics the researchers were interested in; (ii) the systematic structuring of program scope and the application of well-defined evaluation criteria so that rivaling research themes could be treated in an equitable manner and (iii) the same "blueprint" of participatory workshops (e.g. agenda, GSS) offered advantages in terms of repeatability (e.g. reuse of instructions and other preparatory material), solicitation of inputs from all workshop participants (e.g. anonymous comments supplied through the GSS) and also comparative analyses across the workshops (e.g. comparisons based on the positioning of research themes with regard to the evaluation criteria).

In effect, a more open-ended and less structured process could have been possible, too, for instance, by catalyzing informal discussions among program managers, senior industrialists and leading researchers in a seminar with a larger number of participants; but then the most vocal and persuasive participants might have dominated the discussions, to the effect that not all the research themes would have received the attention that they deserved. At the other end of the spectrum, more extensive evaluation frameworks (such as those outlined in Sec. 2; see, e.g. Braunschweig *et al.*[18]) could have been employed, with the risk that there would have remained insufficient time for the consideration of verbal arguments and informal discussions in relation to the more formal evaluative statements. Thus, the methodological design presented here can be seen as an attempt to strike a reasonable balance between the advantages and disadvantages of formal and informal analyses in program planning.

3.1. *Development of research themes*

Before the three Finnish workshops, the scope of the research program was modeled as a hierarchical taxonomy that consisted of three *research areas* and 16 *research themes*. The three research areas were defined as relatively broad fields of scientific inquiry: (i) characterization and alteration of wood material properties; (ii) innovative ecoefficient fiber, chemical and bioproducts and (iii) innovative ecoefficient wood products. Under each of these areas, five or six more specific research themes were listed: for example, the theme "measurement methods for wood and fibers" belonged to the first area, "new fiber-based composite structures" was associated with the second area and "intelligent structures and materials" was presented under the third area. Each of the three Finnish workshops sought to assist in the formation of priorities *within* a particular research area, whereby the constituent research themes were regarded as alternatives.

The development of this hierarchical taxonomy was an exercise "in the abstract" because there were few possibilities of relying on concrete proposals on the topics the researchers were interested in. Thus, in order to probe these interests, and to lend more depth to the research themes, a questionnaire was sent out to some 50 leading researchers who were invited to present research topics that they felt were important. This questionnaire was structured according to the above taxonomy so that for each of the 16 research themes, the respondents were encouraged to (i) describe one or more topics, (ii) state why these topics would be important in view of their expected socioeconomic impacts and (iii) suggest how these topics should be pursued. For each research theme, the respondents were also asked to specify whether they possessed (i) general know-how, (ii) moderate knowledge or (iii) in-depth expertise on this research theme.

This questionnaire study resulted in almost 120 research topics. These topics were enumerated and collated into a summary document that described each topic and its motivation in adjacent columns. This summary document provided a useful source of information, partly because it showed what the different research themes might entail in terms of more concrete research topics.

3.2. *Definition of evaluation criteria*

The evaluation criteria were developed through an iterative process where earlier examples — such as the use of multicriteria models in the U.K. foresight exercise[2] — were first presented to the program manager. In the ensuing discussion, it was felt that it would be beneficial to divide the criteria into two groups under the headings *Attractiveness* and *Feasibility*. Here, *Attractiveness* referred to (i) the novelty of research topics within a particular research theme and (ii) its relevance to industrial competitiveness and the desired long-term environmental and socioeconomic impacts. *Feasibility*, on the other hand, referred to the ability to benefit from research, based on (iii) current research competencies and (iv) the capabilities of potential users and other beneficiaries to exploit results. For each criterion, score information on the research themes was measured through a seven-step Likert scale (see Table 1).

In the Finnish workshops, the evaluation framework also contained two additional criteria on the risks of not achieving research or commercialization objectives. However, it turned out that these two criteria did not offer much additional information that was not linked to the above four criteria. This, together with the quest for an even more parsimonious framework, was the reason why they were not included in the evaluation model used in the Scandinavian workshop.

The development of a rather small evaluation framework consisting of only four criteria (or six, as was the case in the Finnish workshops) was motivated by two concerns. First, priorities for the research themes had to be formed in a single workshop session that lasted no more than 4 h. This placed severe constraints on the amount of time that could be devoted to the elaboration and evaluation of each research theme: if a more extensive evaluation framework had been adopted, there would have been less time for informal discussions and the participants would have been forced to provide their inputs in a rushed manner. Second, although the questionnaire provided information on what the research themes meant in terms of possible research topics, the themes could not be regarded as clear-cut alternatives (such as project proposals) that could have been subjected to a defensible evaluation based on confidently stated expert judgements across a broad range of criteria. From this perspective, too, the elicitation of aggregate evaluative statements with regard to a few broadly defined evaluation criteria was deemed more meaningful.

Table 1. The evaluation criteria.

Dimension	Definition	Scale (1–7)
Attractiveness		
Novelty of research topics	Novelty of proposed research topics in this research theme. • How novel are the topics within this research theme?	1 — Hardly novel at all 3 — Somewhat novel 5 — Very novel 7 — Extremely novel
Relevance of research theme	The relevance of this research theme to industrial competitiveness and desired long-term environmental and socioeconomic impacts. • How relevant is the research theme in view of industrial competitiveness and other desired impacts?	1 — Hardly relevant at all 3 — Somewhat relevant 5 — Very relevant 7 — Extremely relevant
Feasibility		
Research competencies	Current research competencies in this research theme. • How good are the research competencies in relation to the objectives within this research theme?	1 — Poor 3 — Moderately good 5 — Very good 7 — Extremely good
Capabilities for exploitation	Capabilities of potential users and other beneficiaries of research to exploit results from this research theme. • How capable are the beneficiaries of research of exploiting results from this research theme?	1 — Poor 3 — Moderately capable 5 — Very capable 7 — Extremely capable

In the Finnish workshops, the participants were invited to specify to what extent they were able to rate research themes based on their expertise (0 — no judgment; 1 — hardly at all; 3 — to some extent; 5 — rather well; 7 — very well). These expertise ratings were used in forming the aggregate scores through the summation $s_G^j = \sum e_i^j s_i^j / \sum e_i^j$, where s_i^j denotes the ith participant's score for a research theme with regard to the jth criterion and e_i^j is the level of expertise that the participant attached to this score. Thus, in comparative terms, scores with a high degree of expertise counted more in the computation of aggregate scores while the verbal scale helped normalize the ratings so that the experts would understand them in much the same way. This approach is advantageous in that it does capture different levels of expertise and provides a transparent model (for a description of a related approach, see, e.g. Nakayama *et al.*[27]). The disadvantage is that it is difficult to fully calibrate the expertise ratings, even if verbal scales are helpful in this regard.

A major reason for soliciting expertise ratings in the Finnish workshops was that the participants represented a broad range of perspectives and fields of expertise, wherefore they were not consistently "experts" on all research themes in view of all the evaluation criteria. In this setting, the expertise ratings made it possible to ascribe greater significance to the participants with a higher level of relevant expertise while all participants could still be encouraged to contribute to the evaluation task (which was an important process consideration). Moreover, different shades of red color were employed to assist in the visualization of criterion-specific scores, in order to indicate which proportion of aggregate score was contributed by experts with different levels of expertise. A comparison of these visualizations thus conveyed information about which research themes and evaluation perspectives (i.e. sub-criteria under *Attractiveness* and *Feasibility*) the participants had the most expertise on.

In the Scandinavian workshop, on the other hand, the topics were dealt with at a higher level of aggregation, because the five research themes were generated from the national workshops in Finland and Sweden. Furthermore, all participants were either senior decision makers from industrial firms or public funding agencies and could thus be assumed to possess a sufficient degree of expertise. And from the viewpoint of political viability, it was better to avoid uneasy questions that could have offered possibilities for analyzing which countries had the "better" experts, for example.

3.3. *Weighting of criteria*

The planning process was not aimed at the selection of one out of the many research themes but, rather, at the generation of priorities for a *research portfolio* that would reflect existing research competencies and emerging industrial and societal needs. This objective did not call for complete information about criteria weights, because there was no need to determine which research themes would be "better" than others in some absolute sense. Yet, it was felt it would be instructive to present aggregate information about how the research themes could be positioned with regard to the two higher-level criteria, Attractiveness and Feasibility. Toward this end, some preference information about the relative importance of the four sub-criteria in Table 1 was needed.

In discussions on this point, the representatives of program management agreed that under Attractiveness, *Relevance of research theme* was at least as important as *Novelty of research topics*, the argument being that even a novel research topic without socioeconomic or environmental relevance has little value. Under Feasibility, *Capabilities for exploitation* were deemed at least as important as *Research competencies*, because the benefits gained from research results are crucially dependent on the ability to exploit these results. These "policy" statements, however, did not specify whether *Novelty of research topics* should be more or less important than *Research competencies*, or whether *Relevance of research theme* should be more or less important than *Capabilities for exploitation*. Thus, under the two higher level criteria of *Attractiveness* and *Feasibility*, only incomplete information about the relative importance of sub-criteria was available.

This kind of incomplete ordinal preference information was accommodated by the recently developed RICH method, which admits incomplete ordinal preference information in multicriteria weighting models.[28] More generally, RICH allows the decision maker to characterize the relative importance of criteria by associating sets of attributes with corresponding sets of rankings: for example, he or she may state that a given criterion is among the three most important ones (which assume rankings one, two and three), or that the most important criterion (with ranking one) comes from some given subset of criteria. Such statements correspond to constraints on the criterion weights w_i in the additive model $V(a^j) = \sum_{i=1}^{n} w_i v_i(a^j)$, where the overall value of the jth alternative is expressed as a function of its criterion-specific scores $v_i(a^j)$ and criterion weights w_i. When the weights are allowed to vary subject to these constraints, the overall value of each alternative ranges within a closed interval of real numbers. If the resulting value intervals overlap, decision recommendations can be based on the use of different decision rules, such as maximization of the largest or smallest overall value that an alternative can obtain (see Salo and Hämäläinen[23] and also Danielson[29]).

In our case, the statement that *Relevance of research theme* was at least as important as *Novelty of research topics* could be modeled by associating a ranking of one to *Relevance* under the higher-level attribute Attractiveness. Likewise, the statement that *Capabilities for exploitation* was more important than *Research competencies* with regard to Feasibility was modeled by assigning a ranking of one to *Capabilities*. No further preference statements for the computation of overall values for each research theme (i.e. alternative) were elicited, because it was felt that the workshop discussions would benefit most from informative visualizations of research themes with regard to the higher-level attributes Attractiveness and Feasibility: in consequence, incomplete information was employed to establish upper and lower bounds for the value interval of each research theme with regard to these two higher-level attributes only. Another, related reason for relying on the visualization not emphasizing the comparison of overall values was that the criteria Attractiveness and Feasibility did not fully satisfy conditions of mutual preferential independence: for instance, research themes with very low score on Feasibility would probably not have been of considerable interest, no matter how high their perceived Attractiveness.

3.4. *Process implementation*

The workshops were held in the group decision room in the Systems Analysis Laboratory at the Helsinki University of Technology (see http://www.riihi.hut.fi/). All workshops lasted about 4 h. They were attended by some eight to ten experts who were selected by the funding agencies, with the aim of securing a balanced representation of relevant stakeholders.

The three Finnish workshops were attended by representatives from industry, the research community and public administration. A full summary of the questionnaire study was sent to the participants a week in advance, and it was discussed at the workshops in order to clarify what the different research themes meant. The Scandinavian workshop, in turn, was attended by eight evaluating participants who represented a somewhat higher level of managerial seniority than those of the Finnish workshops. Four of these came from Finland, three from Sweden and one from Norway; moreover, half the participants were senior managers at leading industrial firms, while others held managerial positions at funding agencies with responsibilities in S&T policy development and implementation. No written materials were sent to the participants of the Scandinavian workshop in advance. All workshops were chaired by a process facilitator, the first author of this paper. They were also attended by the program manager of the Finnish Wood Wisdom cluster program (see Salo *et al.*[26]).

A dedicated GSS was developed to support the priority setting process, in recognition of the following requirements: (i) the ability to solicit anonymous inputs (both quantitative and qualitative) from the participating experts through an easy-to-use computer interface, (ii) the presentation of results "on-the-spot," based on the participants' inputs and (iii) the storage of judgmental inputs for later retrieval and analysis.

The GSS hardware consisted of networked laptop computers offering Internet access to all participants, as well as a laptop computer for the facilitator and an attached video projector for the presentation of results. The GSS software was built by enhancing the capabilities of RICH Decisions software (a Web-based decision support software that accommodates incomplete ordinal preference information in hierarchical weighting models[30]) for use in group settings. Toward this end, separate client and server modules were written for the participants and the technical facilitator who coordinated the use of the GSS in the workshops and assisted the participants in technical matters.

In the first phase (see Table 2), the participants introduced themselves to each other, the process facilitator outlined the workshop objectives and the participants were given a brief introduction to the GSS. In the Finnish workshops, the participants identified themselves in the GSS either as researchers or industrialists. In the Scandinavian workshop, the participants identified themselves in the GSS by their nationality, after which responsible persons from research administration gave presentations on the results of the preparatory work in Finland and Sweden.

In the second phase of the Scandinavian workshop, the five research themes — (i) wood- or fiber-based composite structures, (ii) new materials from wood-based polymers or extractives (iii) biotechnical, chemical or physical modification of wood raw material, (iv) innovative applications of traditional wood and fiber products and (v) methods of controlling market-oriented utilization of wood raw material — were addressed one at a time. Here, the participants first engaged in a discussion on the

Table 2. Workshop schedule.

Phase	Tasks	Duration
Introduction	• Presentation of participants, workshop agenda, objectives and tools • Results from the preparatory work in Finland and Sweden	30 min
Analysis of research themes	• Initial comments on presentations (10 min) • Appraisal of research themes (10 min) • Discussion (10 min)	2 h 30 min
Identification of focal research topics	• Comparative results from the appraisal of research themes (10 min) • Proposals for resource allocation (10 min) • Discussion (10 min)	30 min
International collaboration	• Discussion (10 min)	30 min

results of the preparatory work, with the aim of arriving at a shared understanding of these themes. They then evaluated each of the five themes by providing scores and associated verbal comments on them. This information was supplied through a graphical interface (see Fig. 1), where criterion-specific scores were submitted through drop-down boxes and verbal comments were typed into text boxes (in the Finnish workshops the interface also had additional drop-down boxes for indicating the participant's level of expertise on each criterion). The appraisal of research themes often took more than the planned 10 min, even though the number of criteria was relatively small and the GSS interface was easy to use.

The evaluation results were presented on-the-spot through a video projector to catalyze a discussion on the research themes. Each theme was positioned with regard to the four criteria by showing the average values for the three nationalities as well as the average value for the combined group (Fig. 2). Verbal comments, concerning both specific criteria and the research theme as a whole, were also shown (Fig. 3).

In the third phase, support for comparative analyses was offered by illustrating how the research themes performed with regard to Attractiveness and Feasibility, when the scores for their sub-criteria were aggregated through the RICH method. Because the criteria weights were not fully specified, each research theme was associated with a rectangular area on the Attractiveness–Feasibility plane (as opposed to a single point that would have been obtained based on complete weight information).

The interface also made it possible to position research themes from the viewpoint of the three countries. This sparked a constructive debate on the underlying reasons for the similarities and differences that were observed (see Figs. 4 and 5): for example, the Swedish interest in composite structures was partly related to the strong tradition in aerospace research in Sweden. Finally, the participants were asked to make a proposal for a resource allocation by indicating the percentage of funding they would grant to each research theme, in the interest of the

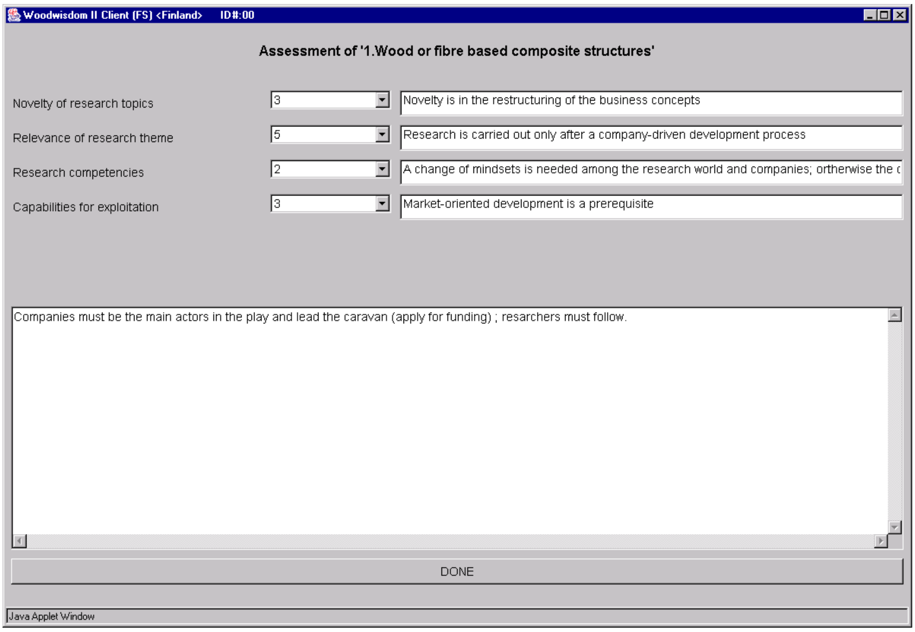

Fig. 1. Participant interface for the evaluation of a research theme.

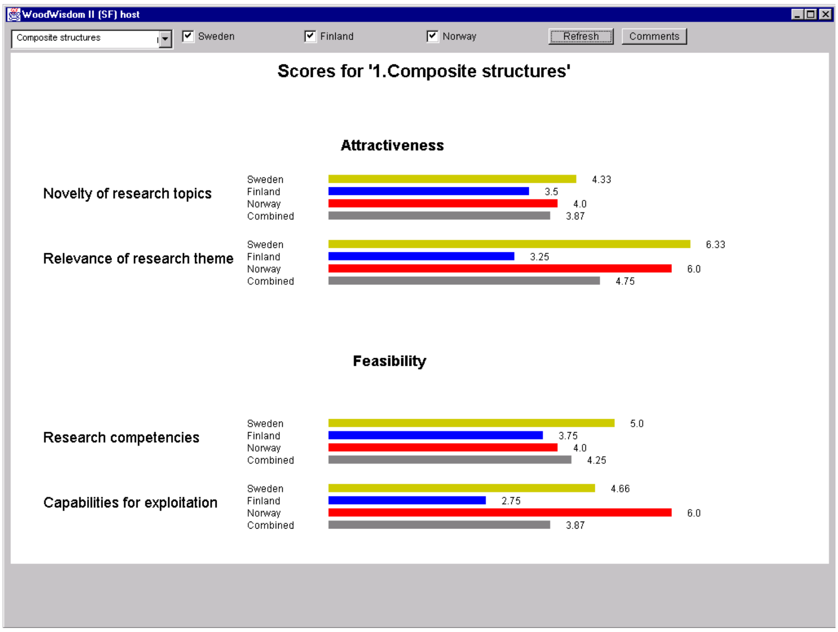

Fig. 2. Quantitative results of a research theme evaluation.

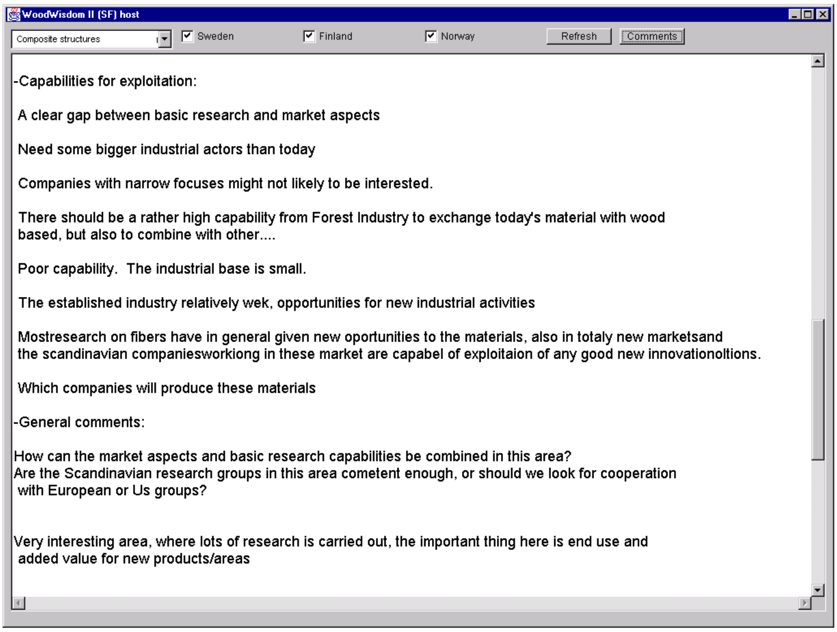

Fig. 3. Verbal comments on a research theme.

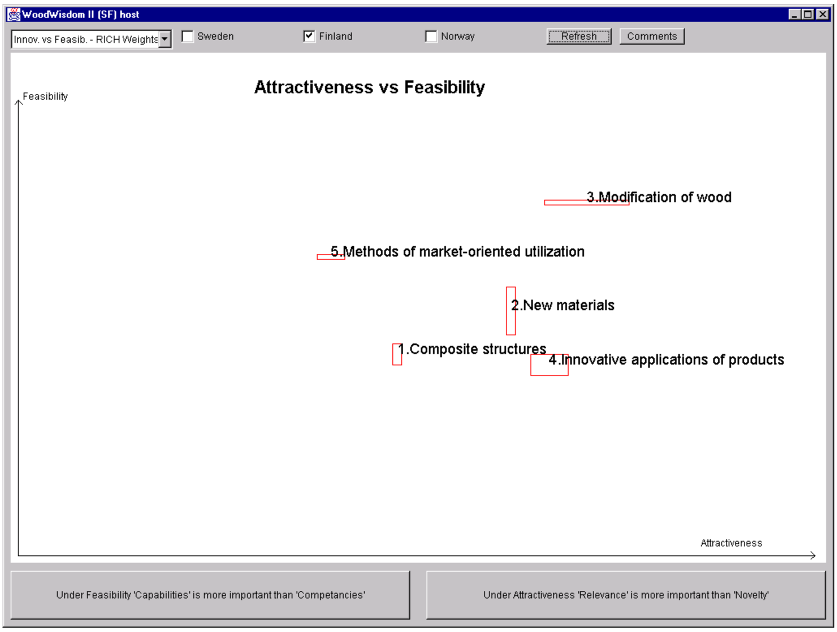

Fig. 4. Finnish positioning of research themes in the *Attractiveness–Feasibility* plane.

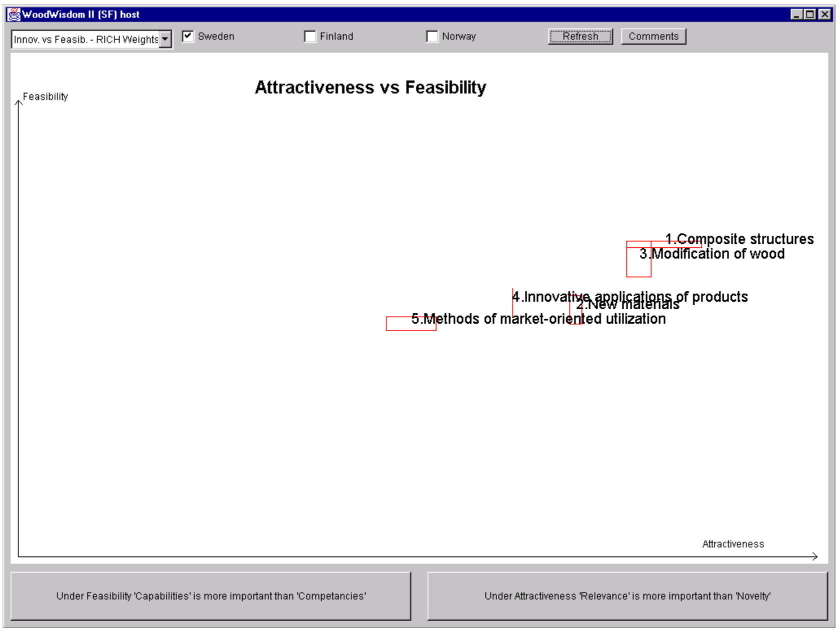

Fig. 5. Swedish positioning of research themes on the *Attractiveness–Feasibility* plane.

Scandinavian forestry and forest industries. By and large, this proposal (Fig. 6) was compatible with the positioning of research themes in the Attractiveness–Feasibility plane, in the sense that Pareto-efficient themes (i.e., themes that performed best on both dimensions) tended to receive a high share of resources.

The workshop was concluded by asking the participants to evaluate how important it would be to pursue international collaboration beyond Scandinavia within the five research themes. Here, a 5 min discussion was carried out before the participants gave their quantitative and qualitative evaluation through the GSS. Both combined and national averages were presented before concluding observations were made.

3.5. *Workshop feedback*

At the end of the Scandinavian workshop, the participants were requested to provide feedback on the workshop. When asked to agree or disagree with the evaluative statements, the large majority of participants agreed, for instance, that similar kinds of workshops should be organized in the context of other research and technology programs, and that GSS was easy to use and also useful in the commenting and generation of research topics. At the same time, some participants were concerned with the extent to which the different research themes and stakeholders were really represented by the workshop participants. This points to challenges in securing a balanced representation of relevant interests and expertise in a single workshop

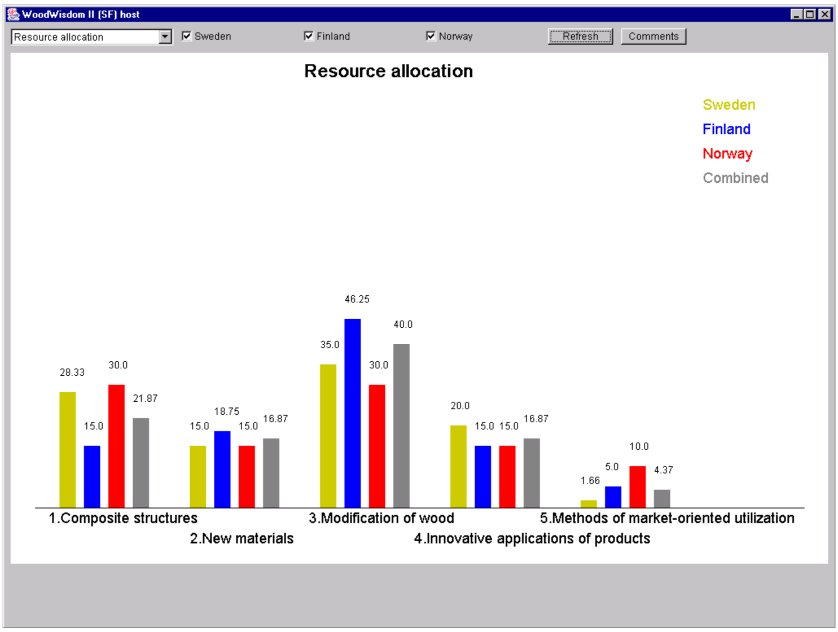

Fig. 6. Proposals for a resource allocation.

that seeks to cover an exceptionally broad range of topics. There is, in effect, a fundamental trade-off between (i) comprehensiveness in terms of interests and fields of expertise and (ii) possibilities for engaging all the participants in a constructive discussion, which is difficult in larger groups.

Later on, the Finnish program manager reported that hardly any complaints were made after the projects for the program had been approved. This can be contrasted with the not too uncommon situation where at least those researchers whose project proposals have been rejected tend to complain. It therefore appears that a systematic and methodologically structured priority-setting process helps establish these priorities in a manner that is perceived as just and impartial.

4. Implications for Program Planning

Drawing upon experiences from the above case study, we next discuss challenges in the formation of research priorities and associated implications for methodological choices. We also argue that the use of systematic methodologies may be warranted in view of quality considerations, too.

4.1. *Challenges in priority formation*

A major challenge in the formation of research priorities is that these priorities must be established based on limited information about the prospective projects through which these priorities can be possibly implemented: in effect, the very purpose of priorities is to guide the development of project proposals that reflect the perceived importance of different research themes. This setting entails an inherent dilemma in that the formation of priorities is partly guided by an implicit understanding of what projects might be carried out within these themes; yet a confirmation of the correctness of this understanding can be obtained only *after* the priorities have been defined and communicated. Although one might wish to form priorities iteratively through several such rounds, this is rarely possible within the available time frames, nor is it economical due to the high costs of developing project proposals. But interestingly enough, a common praxis in the management of many programs is that researchers are encouraged to submit preliminary "light-weight" plans as an input to later priority-setting activities.[31]

A further challenge is that systematic methodological support for the formation of priorities requires that a taxonomy of alternatives is constructed for the full scope of the program. Especially in large programs, these alternatives (cf. research themes in our case study) must be defined in broad terms that subsume several research topics, or else the sheer number of rivaling themes becomes prohibitive. But when the themes contain several topics, they may not be easy to interpret in concrete terms. It is therefore pertinent to clarify what these themes mean, for instance, by developing mutually agreed definitions for them at the aggregate level, or by offering concrete examples to illustrate what specific topics might be pursued within each theme.

In effect, some workshop participants noted that the evaluation of research themes with regard to the four criteria was challenging, partly because the themes contained both highly attractive and less appealing research topics. The participants thus had to "average" over several potential topics through an informal yet cognitively demanding task when appraising research themes. In this regard, the presentation of concrete examples in the Finnish workshops seemed beneficial: for had the themes not been linked to such examples, the experts would have more likely understood these themes differently or too superficially.

4.2. *Implications for the deployment of methodologies*

In view of the above discussion, it is useful to distinguish between two intertwined processes in the formation of research priorities: (i) determination of how important different research themes are in relation to each other (e.g. in terms of the amount of funding that they merit) and (ii) characterization of what topics seem particularly promising within these themes. In what follows, we refer to these two processes as *weighting* and *shaping*, respectively.

More generally, Fig. 7 illustrates interrelationships among several activities in program management. Here, the early consultations during priority formation are crucial for the development of adequate organizational structures and the

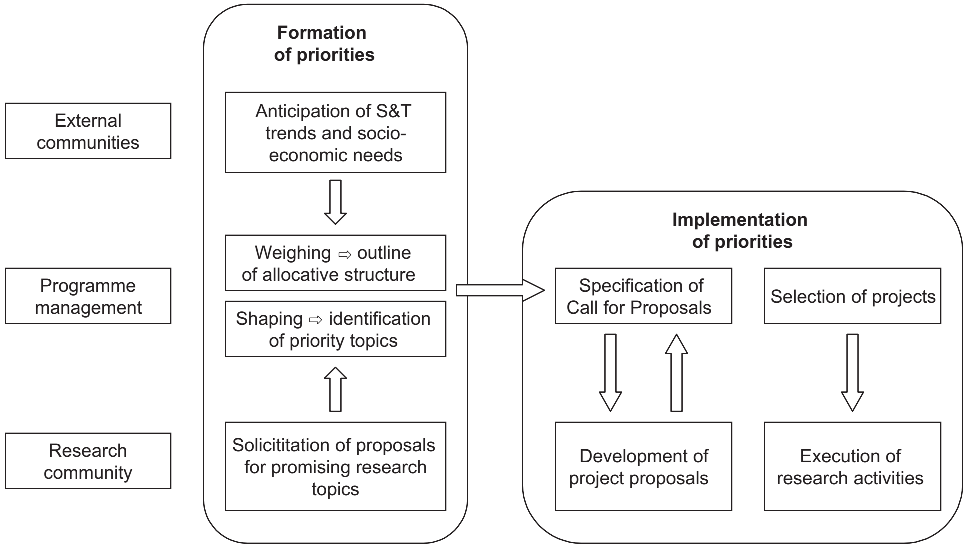

Fig. 7. Phases in priority setting.

earliest decisions of program implementation (e.g. Calls for Proposals). A natural borderline between the phases of priority formation and priority implementation can be drawn when the program as a decision making entity is formally established.

At the outset, there is a need to anticipate S&T trends and emerging socioeconomic needs on a broad front, or else the research program may not be effective in responding to new opportunities. At the same time, there is a need to understand to what extent the research community is interested in and capable of pursuing these opportunities. Arguably, the processes of examining both external and internal considerations must be closely linked, in order to mitigate potential pitfalls in priority setting.

For example, if priorities are derived in a top-down manner whereby allocative proposals for broadly characterized research themes are derived primarily from the analysis of external developments, there may be a mismatch between (i) the tacit assumptions that are made about the feasibility of specific topics that might be pursued within these themes and (ii) the interest and capabilities of the research community in pursuing such topics. Thus, at worst, strongly prioritized themes may attract only few (if any) high-quality proposals.

Conversely, if top-down processes of priority-setting processes are weak, the program may not respond to new opportunities or it may not fully succeed in mitigating structural deficiencies in the innovation system (e.g. structural rigidities, anticipatory myopia).[10] That is, if priorities are formed in a bottom-up manner, they are likely to reflect well-established research agendas as well as organizational structures and interests.

Further to the above challenges, the formation of priorities can be enhanced through several methodological recommendations:

Consultation of multiple stakeholders: The need to rely on complementary sources of expertise implies that the formation of priorities should be organized as a group activity, because no experts possess expertise across all the relevant dimensions. This has several methodological implications. First, the participating experts must be carefully chosen to secure a balanced representation of relevant fields of expertise and stakeholder interests. Second, if multicriteria methods or other formal evaluative approaches are employed, the participants may be invited to rate their level of expertise depending on the particular theme and criterion (as was done in the Finnish workshops), to ensure that the judgments of the more informed experts are not diluted by those of the less informed. Third, it may be instructive to contrast the viewpoints of different experts and stakeholder groups (cf. Figs. 4 and 5), because this may reveal hidden assumptions. Fourth, due to information asymmetries and vested interests, it may be beneficial to devote explicit attention to the de-biasing of possibly exaggerated claims about the beneficial impacts of specific topics (e.g. mutual critiquing of results in the broader community; see, e.g. Braunschweig *et al.*[18]). Fifth, from an organizational perspective, it is vital that some experts (e.g. representatives of funding agencies) are active in priority formation *and* implementation, to eliminate the possibility that the priorities are interpreted differently in the implementation phase from how they were seen in the formation phase.

Combination of top-down and bottom-up approaches: To some extent, pitfalls in priority-setting processes can be reduced by combining top-down and bottom-up approaches. In our case study, the development of the initial taxonomy of research themes was a top-down process where the entire program scope was structured in terms of comparable research themes; on the other hand, the solicitation of suggestions for research topics was a bottom-up process, the purpose of which was to probe the interests of the research community in these themes. Furthermore, the evaluation criteria comprised both outward-looking (cf. Attractiveness) and inward-looking concerns (cf. Feasibility).

Deployment of complementary methodological approaches: The formation of priorities is essentially an explorative exercise, rather than that of finding an "optimal" solution to a well-defined problem. It is therefore likely to benefit from the deployment of multiple approaches that bring in complementary perspectives through different modes of activity (e.g. formal multicriteria analyses vs. informal discussions) and help juxtapose the results of one approach with those of another (e.g. suggestions for direct resources allocation vs. positioning of research themes with regard to Attractiveness and Feasibility). The use of multiple approaches can be consequently motivated as a means of validating results.

Recognition of uncertainties: The program objectives (when measured in terms of the relative emphasis on evaluation criteria) are often vague until the envisaged content of each research theme has been clarified. Moreover, the topics that are pursued within the themes remain uncertain until the projects have been selected. Both these observations suggest that it may not be meaningful to solicit complete information about the weights of the evaluation criteria. In our case study, we therefore elicited only incomplete ordinal preference information with the RICH method[28] while the results were presented with the aim of highlighting similarities and differences in the positioning of research themes. This differs from conventional multicriteria analyses, where an attempt is made to determine the best alternative in terms of its overall value.

In the context of multicriteria project portfolio selection, the problem of dealing with uncertainties has been addressed through the Robust Portfolio Modeling (RPM) methodology.[32] This methodology captures uncertainties through incomplete preference information and generates all the non-dominated portfolios in view of this information (i.e. project portfolios for which another feasible portfolio with a certainly better overall value cannot be found). In this case, projects included in *all* non-dominated portfolios can be surely recommended, because they would belong to all non-dominated portfolios even if additional information were to be acquired. Likewise, projects that are not included in *any* non-dominated portfolio are candidates for rejection, because it would be possible to construct a better portfolio from the other projects. In between these extremes, RPM thus defines a negotiation zone of projects on which the elicitation of further information can be focused. To date, the RPM approach has been successfully applied to the development of a product strategy at a telecommunication company, among others.[33]

Treatment of horizontal considerations: The "vertical" priorities that are set through the weighing of research themes provide a foundation for the development of Calls for Proposals. Yet, since the eventual socioeconomic impacts of research come about through many-faceted collaborative interactions, it is necessary to consider several cross-cutting "horizontal" considerations, too (e.g. networking, international collaboration, socioeconomic impacts). These can be addressed, for instance, by means of informal discussions where the workshop participants put forth suggestions about how such considerations can be accounted for in the context of different research themes. Alternatively, it is possible to develop tailored horizontal evaluation frameworks for this purpose: Salo *et al.*,[19] for example, describe a multicriteria model that accounts for several modes of collaboration.

In general, vertical frameworks seem suitable for the *weighting* of research themes, while horizontal frameworks can be helpful in catalyzing discussions on the *shaping* of such themes. Moreover, because the "broad contours" of priority formation must be set before any specifications for the implementation of themes are given, it follows that vertical frameworks are needed earlier in the overall process of program planning. Once these overall priorities have been set, horizontal frameworks (e.g. focusing on alternative modes of collaboration) can be utilized to develop recommendations on specific themes (e.g. to what extent international collaboration should be pursued within some given theme, and among which groups).

4.3. *Quality dimensions*

Apart from the generation of well-founded results, systematic evaluation frameworks seem warranted in the light of several quality dimensions, too. Because priority setting is one of the preparatory phases in resource allocation, it is plausible to require that all research areas and their constituent research themes are treated fairly and equally. This requirement — which we call *equitability* — implies that research themes should be approached through the same methodology, to ensure that rivaling themes receive the same amount of attention and that conclusions about them are arrived at by adhering to the same process. This helps eliminate the possibility that one theme stands out favorably in comparison with others, merely because it has been defended at greater length by the more vocal and persuasive proponents.

Fundamentally, equitability builds on *comparability* and *repeatability*. Here, comparability refers to the ability to contrast results from the analysis of different research themes (or some other relevant units of analysis). Repeatability, in turn, refers to the ability to apply the methodology in the same manner to all themes. Apart from its contribution to equitability, repeatability is advantageous also in that it entails "economies of scale." For example, documentation and other supporting materials can be re-used across workshops so that new activities can be organized at a lower marginal cost (see, e.g. Salo *et al.*[19]).

A well-structured methodological framework is also useful in that those in charge of designing the priority-setting process can rely on it in the management of stakeholder interactions. For instance, such a framework helps communicate the specific objectives of the process and the role of the chosen methodology in achieving these objectives. Methodology can thus be regarded as a management tool that, at best, stimulates stakeholders' interest in the process and organizes their interactions in a purposeful manner. It also helps ensure that the workshop discussions remain on the intended agenda, as the facilitator can appeal to it, should the discussions tend to drift to irrelevant topics.

5. Conclusion

In this paper, we have reported a case study of the application of novel multicriteria methods to the planning of a Scandinavian research program on wood material science. Apart from describing these methods and the corresponding decision support tools, we have discussed challenges in priority formation and discussed methodological considerations that may improve the quality of these processes: for example, due to the presence of high uncertainties and the need to account for multiple perspectives, these processes may benefit from the deployment of multiple approaches, a combination of both top-down and bottom-up approaches, explicit modeling of the participants' degree of expertise and the use of mechanisms for de-biasing possibly exaggerated statements about the beneficial impacts of proposed research themes.

Based on the positive feedback from this case study, we conjecture that a similar approach may be suitable for the preparation of research programs in other fields, or even for the design of other policy instruments where priorities must be developed as a precursor to the solicitation of project proposals. In such contexts, the benefits from the deployment of systematic multicriteria methods relate to (i) the transparency of decision support activities; (ii) the repeatability of the planning process, whereby additional research themes can be addressed through the same format and — perhaps most importantly — (iii) the requirement for equitability in addressing alternative research themes, which helps ensure that these themes are treated equally and receive the same amount of attention.

Acknowledgements

We are grateful to Leena Paavilainen and Christine Hagström-Näsi for their major contribution to the design of the priority-setting process. We also wish to thank the referees for their instructive comments and suggestions. This work has benefited from the financial support of the National Technology Agency (Tekes).

References

1. F. Hartwich and W. Janssen, Setting research priorities: an example from agriculture using the analytic hierarchy process, *Res. Eval.* **9**(3) (2000) 201–210.
2. M. Keenan, Identifying emerging generic technologies at the national level: the UK experience, *J. Forecasting* **22**(2–3) (2003) 129–160.
3. OECD, *Choosing Priorities in Science and Technology* (OECD, Paris, 1991).
4. R. Smits and S. Kuhlmann, The rise of systemic instruments in innovation policy, *Int. J. Foresight Innovation Policy* **1**(1–2) (2004) 4–32.
5. H. Grupp and H. H. Linstone, National technology foresight activities around the globe, *Technol. Forecasting Social Change* **60** (1999) 85–94.
6. A. D. Henriksen and A. J. Traynor, A practical R&D project-selection scoring tool, *IEEE Trans. Eng. Mgmt.* **46**(2) (1999) 158–170.
7. D. L. Keefer, C. W. Kirkwood and J. L. Corner, Perspective on decision analysis applications, 1990–2001, *Decision Anal.* **1**(1) (2004) 5–24.
8. R. P. Hämäläinen, Reversing the perspective on the applications of decision analysis, *Decision Anal.* **1**(1) (2004) 26–31.
9. D. Loveridge, Technology and environmental impact assessment: methods and synthesis, *Int. J. Technol. Mgmt.* **11**(5–6) (1996) 539–553.
10. J.-P. Salmenkaita and A. Salo, Rationales for government intervention in the commercialization of new technologies, *Technol. Anal. Strategic Mgmt.* **14**(2) (2002) 183–200.
11. B. van der Meulen, Science policies as principal-agent games: institutionalization and path dependency in the relation between government and science, *Res. Policy* **27**(4) (1998) 397–414.
12. M. MacLean, J. Anderson and B. R. Martin, Identifying research priorities in public sector funding agencies: mapping science outputs to user needs, *Technol. Anal. Strategic Mgmt.* **10**(2) (1998) 139–155.
13. B. R. Martin, Foresight in science and technology, *Technol. Anal. Strategic Mgmt.* **7**(2) (1995) 139–168.

14. A. Salo and K. Cuhls, Technology foresight — past and future, *J. Forecasting* **22**(2–3) (2003) 79–82.
15. M. J. Liberatore, An extension of the analytical hierarchy process for industrial R&D project selection and resource allocation, *IEEE Trans. Eng. Mgmt.* **34**(1) (1987) 12–18.
16. T. Durand, Twelve lessons drawn from "Key Technologies 2005", the french technology foresight exercise, *J. Forecasting* **22**(2–3) (2003) 161–177.
17. T. L. Saaty, *The Analytic Hierarchy Process* (McGraw-Hill, New York, 1980).
18. T. Braunschweig, W. Janssen and P. Rieder, Identifying criteria for public agricultural research decisions, *Res. Policy* **30**(5) (2001) 725–734.
19. A. Salo, T. Gustafsson and R. Ramanathan, Multicriteria methods for technology foresight, *J. Forecasting* **22**(2–3) (2003) 235–256.
20. R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs* (John Wiley & Sons, New York, 1976).
21. R. L. Keeney, *Value-Focused Thinking: A Path to Creative Decisionmaking* (Harvard University Press, Cambridge, MA, 1992).
22. S. Zahir, Making public policy decisions using a web-based multi-criteria electoral system (MCES), *Int. J. Inf. Technol. Decision Making* **1**(2) (2002) 293–309.
23. A. Salo and R. P. Hämäläinen, Preference ratios in multiattribute evaluation (PRIME) — Elicitation and decision procedures under incomplete information, *IEEE Trans. Syst. Man Cybern.* **31**(6) (2001) 533–545.
24. Z. Xu, An approach to group decision making based on incomplete linguistic preference relations, *Int. J. Inf. Technol. Decision Making* **4**(1) (2005) 153–160.
25. A. Charnes, W. W. Cooper and E. Rhodes, Measuring the efficiency of decision making units, *Eur. J. Oper. Res.* **2**(6) (1978) 429–444.
26. A. Salo, T. Gustafsson and P. Mild, Prospective evaluation of a cluster programme for Finnish forestry and forest industries, *Int. Trans. Oper. Res.* **11**(2) (2004) 139–154.
27. H. Nakayama, T. Tanino, K. Matsumoto, H. Matsuo, K. Inoue and Y. Sawaragi, Methodology for group decision support with and application to assessment of residential environment, *IEEE Trans. Syst. Man Cybern.* **9**(9) (1979) 477–485.
28. A. Salo and A. Punkka, Rank inclusion in criteria hierarchies, *Eur. J. Oper. Res.* **163**(2) (2005) 338–356.
29. M. Danielson, Handling imperfect user statements in real-life decision analysis, *Int. J. Inf. Technol. Decision Making* **3**(3) (2004) 513–534.
30. J. Liesiö, *RICH Decisions — A Decision Support Software* (Helsinki University of Technology, Finland, 2002), http:\\www.rich.hut.fi\.
31. A. Salo and J.-P. Salmenkaita, Embedded foresight in RTD programs, *Int. J. Technol. Policy Manage.* **2**(2) (2002) 167–193.
32. J. Liesiö, P. Mild and A. Salo, Preference programming for robust portfolio modeling and project selection, to appear in *Eur. J. Oper. Res.*
33. M. Lindstedt, J. Liesiö and A. Salo, Participatory development of a strategic product portfolio in a telecommunication company, to appear in *Int. J. Technol. Mgmt.*
34. W. H. Cooper and R. B. Gallupe, Some liberating effects of anonymous electronic brainstorming, *Small Group Res.* **29**(2) (1978) 147–178.
35. J.-P. Salmenkaita and A. Salo, Emergent foresight processes: industrial activities in wireless communications, *Technol. Forecasting Social Change* **71**(9) (2004) 897–912.