INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

IFORS International Federation of Operational Research Societies

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Intl. Trans. in Op. Res. 18 (2011) 513–526
DOI: 10.1111/j.1475-3995.2011.00809.x

Composition of probabilistic evaluations of preferences: a case of criteria applied to isolated and clustered options

Annibal Parracho Sant'Anna,[a] Valter de Senna[b] and Hernane Borges de Barros Pereira[b]

[a] *Universidade Federal Fluminense, Rua Passo da Pátria, 156, bl. D. sl. 306, 24210-240, Niterói-RJ, Brazil*
[b] *Programa de Modelagem Computacional SENAI/CIMATEC, Av. Orlando Gomes 1845, 41.650-010, Salvador-BA, Brazil*
E-mail: tppaps@vm.uff.br [Sant'Anna]; vsenna@terra.com.br [de Senna]; hernanebbpereira@gmail.com [Pereira]

Received 15 March 2010; received in revised form 23 January 2011; accepted 27 January 2011

Abstract

This article applies a transformation into probabilities of preference, to rank papers submitted to the XLI Brazilian Symposium of Operational Research. Some 42 papers were divided into sets according to their main topics of interest and measures for 12 attributes were considered, for each paper separately, and also for sets of papers on the same topic, based on the referees' evaluations. Four alternative forms of composition of the probabilistic preferences are illustrated.

Keywords: multiple-criteria decision analysis; probability; performance evaluation; decision analysis

1. Introduction

Multi-criteria composition models have as their main aim to offer the decision maker means of quantifying, and hence making more transparent, the relations between the criteria taken into account in a decision process (Rivett, 1994; Clímaco, 2003; Tsoukiàs, 2008).

Three main obstacles may hinder attaining this objective: errors in the conversion of evaluations given in different scales to a common scale, lack of precision in the evaluations according to each criterion and dependence between the criteria blurring the evaluation of their relative importance.

The importance of the inconvertibility of evaluations increased, historically, as these evaluations extended from the economic sector, where the convertibility effort is, most of the time, limited to setting all values in monetary units (Allais and Hagen, 1979; McFadden, 1999), to other areas where social, environmental and other values are not easily converted to a monetary representation (Tversky and Kahneman, 1981; Ehrgott et al., 2005).

The difficulty in dealing with the criteria evaluated in inconvertible reference settings is clear when the combination has to be made by weighted averages (Aczél and Saaty, 1983; Lootsma,

Published by Blackwell Publishing, 9600 Garsington Road, Oxford, OX4 2DQ, UK and 350 Main St, Malden, MA 02148, USA.

1988; Vincke, 1992; Barron and Barrett, 1996; Hammond et al., 1998; Triantaphyllou, 2000). The need to take into account the scale of measurement when setting the weights has been signalled in the literature long ago (Foster and Sen, 1997; Woodward and Bishop, 1997; Saisana et al., 2005; Zhou et al., 2006).

There is always some amount of subjectivity in deriving preferences for any kind of attribute and the lack of information on the degree of uncertainty in classifications considerably impairs the use of such classifications. The imprecision of the evaluations and the need to take into account uncertainty in order to derive probabilistic conclusions can make it very difficult to combine multiple criteria (Balch et al., 1974; Selvanathan and Prasada Rao, 1994; Stewart, 2005; Sonmez, 2007).

The problem of dependence between the criteria has been generally avoided by enforcing independence assumptions. Nevertheless, there are situations where directly eliminating dependence is unfeasible (Machina, 1982; Govindaluri and Cho, 2005). This is, for instance, the case when comparing the options by combining indicators of preference measured at different levels of aggregation of the evaluated options on an individual basis and in clusters of homogeneous individuals.

By measuring the preferences according to each criterion in terms of the probabilities of being the preferred option (Sant'Anna, 2002), a probabilistic composition of the indicators may be performed in such a way that the correlations between such indicators can be taken into account.

Any vector of measures of preference for each of set of competing options can be transformed into a vector of probabilities of being preferred. The imprecision that is inherent in the initial evaluation may be modelled by considering the measure of preference assigned to each option as a midpoint of a statistical distribution of possible evaluations for that option and these distributions are determined by adding stochastic disturbances to such midpoints.

Finally, the probabilities of being ranked first in a sample drawn from the probability space of such distributions are computed. After determining the probabilities of preference according to the multiple criteria, different probabilistic points of view may be chosen to combine such preferences into global evaluations.

This paper examines the application of this probabilistic composition to the situation where the individuals evaluated interact inside groups. The individual performances must be evaluated taking into account group features. If the same attributes are measured to evaluate the performance of isolated individuals and their performance as a group, the uncertainty affecting the formation of the two evaluations will include common factors. In the probabilistic approach, as the global measurement is a joint probability, the correlation between the individual and the group evaluations can be taken directly into account, as we show in the example considered here, of the selection of the best paper presented at a congress. The papers are divided into groups according to the research area and the criteria are divided into groups of different importance. A total of 12 criteria are combined to rank 42 papers.

The article is developed as follows: in the next section, the transformation into probabilities of being the best option is discussed. Section 3 presents different points of view that may be taken to translate probabilities of being the best according to isolated criteria into global evaluations. Section 4 presents the problem of taking into account collective evaluations and the correlation between the vectors of evaluations that may arise. Section 5 develops the application. Final comments conclude the article.

2. Probabilistic composition of preferences

The key computation in the evaluation of probabilistic preferences is the determination of the probabilities of each option being the preferred one among those in a sample. The probability of considering a particular option as the best one is a natural measure of the decision-maker preference for that option. To be able to compare and compose the different criteria of preference, we transform our measurements into probabilities of being preferred. What cannot be measured cannot be managed, but numbers without a measure of their precision cannot be managed either.

From any kind of measurement, a ranking of the options can be derived. On the other hand, a ranking with possible ties and empty spaces may represent every distance between the different options that one may have in mind. To compute the probabilities of being the best in a set of options after ordering them, all we need is a statistical measure of the uncertainty on each position in that ordering. To take uncertainty into account, the rank of each option, or any other numerical indication of its position, may be seen as a location parameter of a statistical distribution.

Once the observed measurements are seen as estimates of location parameters, estimates of other parameters can be derived from the same set of observations. For instance, the observed range, i.e., the range of observed values for the criterion in the entire set of examined options, may be used as an estimate for a common range for these distributions. Different assumptions on the form of the distribution can therefore be considered for modelling. To compensate the lack of empirical information, simplifying and equalizing assumptions are at the essence of Fuzzy Sets Theory (Zadeh, 1965). By the same token, imposing assumptions of independence between the disturbances affecting the different observations and hypotheses for the form of the distributions will provide the necessary framework. The distributions may be assumed, for instance, normal with identical variance, or uniform with identical range, or triangular distributions with the same maximum and minimum points.

In the present application, triangular continuous distributions with the same minimum and maximum are assumed, as is usual for membership functions of fuzzy set theory (Petrovic et al., 2008; Liang and Cheng, 2009). The use of fixed extremes is a way to emphasize the best and worst evaluations. This reflects our point of view that if, according to a given criterion, no option is classified as "very good", a classification as "good" according to this criterion *should not* make it equivalent to a classification of "very good" according to some other criteria.

Other symmetric distributions may be assumed, with standard deviations derived from the observed range. For instance, uniform distributions are assumed in Sant'Anna (2002) and normal distributions in Sant'Anna (2005).

To facilitate comparisons, the probability of being the preferred option may be computed with respect to a sample of fixed size, drawn at fixed percentiles of the set of observed values. In the application to be discussed, the selection of a best paper in a set of 42, this sample is formed by the nine deciles. In general, with n options being evaluated by the ordered values $x_{(1)}, \ldots, x_{(n)}$, the kth decile will be given by $d_k = x_{(i(k))}$, with $i(k) = k*n/10$, if this is an integer, or $d_k = [x_{(i(k))}+x_{((i(k)+1))}]/2$ for $i(k)$, the largest integer smaller than $k*n/10$, otherwise. To set the dispersion, the maximum and minimum values for each distribution may be derived from the maximum and minimum observed values. Here, estimates given by $x_{(1)}+(d_9-d_1)/8$ and $x_{(n)}-(d_9-d_1)/8$ are used.

Let us consider, for instance, the case of only five options being compared and ranked in a Likert scale of five points by the numbers 1 to 5. The intermediate percentiles will be 1.5, 2.5, 3.5 and 4.5.

The distribution centred at each of these values will be a triangular distribution with extreme values 0.5 and 5.5.

The probability of an option being evaluated as best can be computed by integrating with respect to the joint distribution, as presented in the Appendix.

3. Different approaches to the combination of probabilistic evaluations

After computing the probabilities of being the preferred option according to each criterion, it is easy to combine them into a unique measure of global preference. This can be done, for instance, by treating these probabilities as conditional on the choice of the respective criterion and then computing the total probability of preference by adding the products of these conditional probabilities by the probabilities of choosing each criterion. The real difficulty in this approach is to determine the marginal probabilities of choice for each criterion. This is especially true if the criteria are correlated. If it is possible to rank the criteria and model the correlation between them, the probability of choice for each criterion may be computed in the same way as the probabilities of preference are computed according to each criterion.

We can also combine the probabilistic preferences by computing joint preferences according to different points of view. In this approach, the dependence between the criteria may be directly taken into account. The different points of view considered here are characterized in terms of choice between extreme positions in two basic orientation axes. These extreme positions are, in one axis, an optimistic versus a pessimistic position and, in the other, a progressive versus a conservative position.

In the progressive–conservative axis, the progressive evaluator looks for options that are the first in excellence and the conservative evaluator chooses based on their ability of not minimizing the preference. The term 'conservative' in this terminology is related to the idea of avoiding losses, while the term 'progressive' is related to the idea of constant improvement.

In the optimistic–pessimistic axis, the optimistic extreme consists of being happy with the satisfaction of only one criterion. All the criteria are taken into account, but the composition uses the connective 'or'. The joint probability computed is that of maximizing (in a progressive composition or of not minimizing in a conservative one) the preference according to at least one of the criteria. On the opposite end, the pessimistic point of view goes for options that satisfy every criterion. The connective is 'and'. The joint probability computed is that of maximizing (or not minimizing) simultaneously the preference according to all the criteria. The terms optimistic and pessimistic are related to the idea that luck is or is not at one's side.

Let $I = \{i_1, \ldots, i_n\}$ denote a set of n criteria and $K = \{k_1, \ldots, k_n\}$ a set of m options. By combining the positions in the extremes of these two axes, four different measures are generated. In the progressive and optimistic points of view, evaluating each option by its probability of being preferred by at least one criterion, if P_{ik} is the probability of the kth option being preferred according to the ith criterion and the criteria are independent, the value assigned to this option is given by $1 - \prod_{i \in I}(1 - P_{ik})$. The computations for the other points of view are similar and are illustrated in Appendix A.

If the criteria are divided into groups and different points of view are allowed in the computation of the joint probabilities for each group, the number of possibilities increases. A natural division of the criteria into groups is in a subset of criteria to be maximized and another subset for which we want minimization. For instance, criteria of benefits and criteria of losses, criteria related to the production of outputs and criteria related to the use of inputs, criteria related to advantages and criteria related to disadvantages and so on.

Other measures may also be considered. For instance, the criteria may be divided into a variable number of independent subsets and the global evaluation may be given by the probability of presenting the best performance in at least one criterion of each subset.

The choice of the composition algorithm will have to take into account practical considerations. There are cases, as in many segments of public administration, where the situation to avoid is a very bad performance. If such a situation is detected, action must be taken and progress starts. Basing the computation on the probabilities of being the worst would also be advisable in the case where most of the observed values or the more reliable ones are close to the minimum efficiency values. With only a few values on the efficient side, it is safer to evaluate performance by the probability of being away from the inefficiency frontier than by the proximity to the efficiency frontier.

4. Modelling cooperative attributes

Evaluation systems based on the comparison of individual performances may not foster global improvement, as they enhance competitive practices where cooperation would be of more important value. On the other hand, individual efforts should also be prized and evaluating only on the basis of group achievement may leave important drives for improvement out of the performance evaluation.

For instance, rewarding the productivity in scientific research through grants awarded to researchers presenting, comparatively, the best indicators stimulates attitudes that will eventually harm the development of productive research. It induces individual research objectives at the expense of collective projects. Also, a rivalry develops as each researcher has to compete for the same grants reserved for the research field.

Therefore, any evaluation system must take into account variables measuring group effects that affect the individuals evaluated. Otherwise, the evaluator who claims to be judging individual performance may be basically measuring individual results attributable to the context where the work is carried out.

In the present work, the evaluation system used allows for individual and group indicators in such a way that the evaluation of each individual is affected by the group performance but its own individual achievements also have a significant impact on its personal result. As we put together variables measuring individual attributes and variables measuring these attributes in aggregated units, a positive correlation may exist between the stochastic components of these variables.

In Sant'Anna (2009), it was pointed out that the composition of fuzzy logic (Zadeh, 1978) by the necessity and possibility concepts is equivalent to taking, respectively, the minimum and the maximum of the membership probabilities. It corresponds to an extreme of the correlation between indicators of occurrence. This last property follows from the fact that the highest correlation would be that making the probability of the intersection equal to the lowest probability among the events intercepted. Assuming independence has the advantage of allowing the actual numerical values, not only the minimum of them, to be fully taken into account.

To state this feature in a more concrete way, consider the set I of criteria divided into two subsets I_0 and I_1, the criteria in I_1 being applied only to individual options while the criteria in I_0 are applied both to individual options and to clusters. The division in clusters may vary from one criterion to another. For instance, when evaluating research groups, a criterion based on the economic value of projects may place in the same cluster research groups of the same geo-economic region, while for a criterion based on research papers published, the relevant clusters may be composed in terms of research fields.

For any option $k \in K$ and any criterion $i \in I_0$, let $c_i(k)$ denote the cluster the kth option falls into when the ith criterion is applied. For any $k \in K$ and any $i \in I$, let $M_i(k)$ and $m_i(k)$ denote the probabilities of the kth option, respectively, maximizing and minimizing the preference according to the ith criterion. If the criterion $i \in I_0$, let $N_i(k)$ and $n_i(k)$ represent the probabilities of the kth option maximizing and minimizing the preference according to the ith criterion, obtained by substituting for the individual evaluations according to this criterion the evaluations of the clusters they belong to.

Assuming independence between disturbances affecting individual evaluations and the maximum correlation between disturbances affecting individual and cluster evaluations, the pessimistic and progressive approach assigns to the kth option the score

$$\prod_{i_0 \in I_0} \min\{M_{i_0}(k), N_{i_0}(k)\} \prod_{i_1 \in I_1} M_{i_1}(k).$$

Analogously, the pessimistic and conservative point of view assigns to the kth option

$$\prod_{i_0 \in I_0} [1 - \min\{m_{i_0}(k), n_{i_0}(k)\}] \prod_{i_1 \in I_1} [1 - m_{i_1}(k)],$$

the optimistic and conservative point of view assigns the score

$$1 - \prod_{i_0 \in I_0} \min\{m_{i_0}(k), n_{i_0}(k)\} \prod_{i_1 \in I_1} m_{i_1}(k)$$

and, finally, the optimistic and progressive point of view assigns the score

$$1 - \prod_{i_0 \in I_0} [1 - \min\{M_{i_0}(k), N_{i_0}(k)\}] \prod_{i_1 \in I_1} \{1 - M_{i_1}(k)\}.$$

5. Application: evaluation of papers

SOBRAPO – the Brazilian Society of Operational Research awards a prize to the best paper submitted to its symposium. The papers submitted are subjected to a double-blind evaluation by referees who also answer 11 questions. These answers, together with the number of referees who effectively agreed to evaluate the papers, provide an initial basis of 12 attributes for analysis by the Award Committee. A preliminary filter reduces the choice to only those papers written in English. Even so, a total of 42 papers were considered for evaluation in 2009. Combining 12 criteria to evaluate 42 options is a considerable task.

Besides answering the questionnaires, the referees make comments and suggestions. The authors of the articles not immediately rejected then have the opportunity to submit a revised version to the Scientific Committee. At present, the Award Committee uses the initial referees' evaluation to reduce the number of papers to be examined and takes into account attributes of the papers such as the importance of new research areas or the level of interaction between distinct research groups.

SOBRAPO aims, with the prize, to enhance the quality of the papers presented at the Symposium, but also to foster the development of high-quality research in the field. As OR deals with distinct subjects and applications, both from a theoretical and an applied point of view, the papers are spread over a large range of sub-areas. A special concern of the Society is to bring to the Symposium the largest possible variety of themes. In 2009, for instance, almost every sub-area of Operational Research competed for this one prize. Hence, in some sense, it is not only the best paper but also its area that eventually wins the prize.

As already discussed, the model developed provides a way to consider the subject area of the paper. Criteria that take together papers of the same sub-area will tend to place in inner ranks those papers whose sub-areas have more submissions, thereby reducing their chances of being ranked the best. The combination of individual evaluation with group evaluation may improve the rank of good individual papers whose strengths are close to those of its group and harm the chances of those that are good but their sub-area is not so. For these reasons, it is interesting to compare ranks based on individual evaluations alone, with ranks generated by considering group evaluations.

The group evaluation of each paper, for each criterion, is the median of all evaluations according to the criterion, over the papers belonging to the sub-area. The combination of the individual and group probabilities assumes the maximum correlation between these two evaluations and is therefore made by the minimum.

Table 1 shows the 10 areas with their respective number of papers. Table 2 presents the list of criteria considered and Table 3 the probabilities of each paper being the best according to the method discussed in Section 2, using the fixed deciles. We assume triangular distributions with mode in the evaluation attributed to the paper and with maximum and minimum fixed, again as described in Section 2. For the first 11 attributes, as more than one referee examined the paper, the median value was used.

The 12 criteria were divided into three groups, each comprising four criteria. The first group relates to the technical importance of the paper. The second group evaluates the quality of the

Table 1
Areas

Emphases	Areas	No. of papers
Applied	Industrial Applications (I)	8
Applied	Social Applications (S)	6
Theoretical	Optimization (O)	5
Applied	Logistics (L)	4
Theoretical	Data Envelopment Analysis (D)	4
Theoretical	Metaheuristics (M)	4
Theoretical and Applied	Optimization or Metaheuristics/Industrial or Social (OI)	4
Theoretical	Multicriteria (C)	3
Theoretical	Graphs (G)	2
Theoretical and Applied	Optimization or Metaheuristics/Logistics (OL)	2

Table 2
Criteria

Criteria	Questions	Values
Relevance	How relevant is this article?	1–3
Quality	In terms of quality, how do you classify the article?	1–4
Correctness	As for the presence of mistakes, how do you classify the article?	1–4
Relationship	How do you evaluate the references to related papers in the article?	1–5
Organization	How is the organization of the article?	1–5
Language	With respect to the use of language, how do you evaluate the article?	1–5
References	How do you evaluate the references of the article?	1–5
Presentation	How do you evaluate the visual presentation – figures, format, etc of the article?	1–4
Total value	What is your global evaluation of the article?	1–3
Acquaintance	How confident you feel on your evaluation?	1–3
Award value	Would you recommend this article as a candidate for the SOBRAPO Premium?	1–2
Evaluators		1–4

presentation and the last group the referees overall impressions. The three groups are therefore Group 1 (relevance, quality, correctness, relationship); Group 2 (organization, language, references, presentation); Group 3 (total value, acquaintance, award value, number of evaluators).

Two different forms of composition were performed. The first used a weighted mean, with a value of 16 for each criterion of the first group, one for each criterion of the second group and eight for each criterion of the last group.

The second form of composition did not use any weights. It used the joint probability of preference. First, a global probability of preference was computed according to each of the three groups of attributes. These three probabilities of being preferred were multiplied together to derive the probability of being preferred jointly.

To combine the criteria of the first group, a pessimistic and progressive approach was used, so that the score of each paper was obtained by multiplying the probabilities of being preferred according to each of the four criteria of the group – this approach reflects the point of view that the paper must be excellent according to all technical criteria. To combine the criteria of the second group, a progressive and optimistic approach was used – this approach requires the paper to be excellent according to at least one of the presentation criteria. We therefore multiply the probability of not being the best according to each criterion in group 2 and subtract from 1. Finally, the criteria for the last group were combined from a conservative and a pessimistic point of view – this approach prefers evaluations that are not the worst, according to every criterion. It is the product of the probabilities of not being the worst for each criterion within the group.

Hence, we give greater importance to the criteria of the first group, treated as essential, and less importance to those of the second group, as it is enough to reach the frontier of excellence according to at least one of them. In the third group, all the criteria are accounted for, but the papers are evaluated by the distance to the worst possible performance, thereby reducing the difference between the papers with the best evaluations. Table 4 presents the ranks according to the two compositions and whether sub-areas are considered in the final evaluation.

It becomes apparent, with a closer look at Table 4, that the inclusion of sub-areas into the analysis increases the stability of the final judgements. In spite of good correlations between the

Table 3
Probabilities of preference for ungrouped criteria

Paper	Relev.	Quality	Correc.	Organ.	Lang.	Relat.	Refer.	Presen.	Total	Acq.	Award	Eval.	Area
1	0.04	0.11	0.07	0.07	0.12	0.11	0.12	0.08	0.17	0.06	0.05	0.2	I
2	0.04	0.11	0.18	0.07	0.12	0.11	0.07	0.08	0.07	0.06	0.05	0.2	I
3	0.04	0.11	0.10	0.12	0.09	0.06	0.12	0.08	0.10	0.08	0.05	0.08	I
4	0.04	0.11	0.10	0.05	0.09	0.11	0.05	0.10	0.10	0.08	0.05	0.08	I
5	0.04	0.11	0.07	0.07	0.07	0.11	0.07	0.08	0.17	0.06	0.05	0.24	I
6	0.04	0.08	0.10	0.09	0.12	0.06	0.12	0.14	0.17	0.12	0.05	0.08	I
7	0.16	0.11	0.07	0.12	0.12	0.06	0.07	0.08	0.17	0.06	0.05	0.03	I
8	0.16	0.11	0.07	0.09	0.12	0.08	0.12	0.14	0.07	0.12	0.05	0.02	I
9	0.07	0.11	0.18	0.12	0.12	0.08	0.09	0.08	0.10	0.08	0.05	0.08	S
10	0.07	0.11	0.18	0.09	0.12	0.11	0.12	0.10	0.10	0.12	0.05	0.02	S
11	0.07	0.11	0.05	0.12	0.07	0.11	0.12	0.14	0.10	0.08	0.05	0.08	S
12	0.16	0.27	0.18	0.12	0.12	0.24	0.12	0.14	0.07	0.06	0.25	0.03	S
13	0.07	0.17	0.18	0.09	0.18	0.11	0.09	0.10	0.10	0.08	0.25	0.08	S
14	0.16	0.11	0.18	0.12	0.12	0.11	0.12	0.08	0.17	0.06	0.05	0.03	S
15	0.16	0.11	0.10	0.09	0.07	0.11	0.07	0.10	0.10	0.32	0.05	0.02	L
16	0.16	0.11	0.07	0.12	0.12	0.11	0.12	0.14	0.07	0.12	0.25	0.02	L
17	0.04	0.06	0.18	0.12	0.07	0.11	0.12	0.08	0.07	0.12	0.05	0.24	L
18	0.16	0.11	0.07	0.12	0.12	0.11	0.07	0.14	0.07	0.12	0.25	0.24	L
19	0.16	0.11	0.07	0.07	0.07	0.11	0.09	0.10	0.10	0.12	0.05	0.08	OI
20	0.16	0.08	0.10	0.12	0.12	0.11	0.12	0.10	0.07	0.08	0.05	0.08	OI
21	0.16	0.11	0.07	0.12	0.12	0.11	0.12	0.08	0.17	0.12	0.05	0.03	OI
22	0.16	0.11	0.07	0.05	0.12	0.11	0.07	0.14	0.17	0.12	0.05	0.03	OI
23	0.07	0.11	0.07	0.18	0.09	0.11	0.18	0.14	0.10	0.19	0.05	0.08	OL
24	0.16	0.11	0.18	0.12	0.12	0.11	0.12	0.14	0.07	0.12	0.25	0.03	Ol
25	0.04	0.27	0.18	0.25	0.25	0.24	0.25	0.14	0.07	0.06	0.05	0.24	G
26	0.16	0.11	0.18	0.25	0.25	0.11	0.25	0.14	0.17	0.06	0.25	0.03	G
27	0.16	0.17	0.18	0.09	0.18	0.11	0.09	0.21	0.07	0.06	0.25	0.08	O
28	0.04	0.27	0.18	0.12	0.25	0.11	0.12	0.14	0.07	0.32	0.25	0.24	O
29	0.07	0.08	0.18	0.09	0.09	0.11	0.09	0.14	0.34	0.08	0.05	0.08	O
30	0.16	0.17	0.07	0.12	0.18	0.11	0.09	0.10	0.10	0.19	0.05	0.08	O
31	0.16	0.11	0.18	0.12	0.12	0.24	0.12	0.31	0.07	0.06	0.25	0.03	O
32	0.04	0.11	0.18	0.12	0.12	0.11	0.12	0.14	0.07	0.12	0.25	0.24	M
33	0.16	0.04	0.04	0.07	0.05	0.04	0.12	0.08	0.17	0.12	0.05	0.03	M
34	0.04	0.06	0.07	0.07	0.07	0.11	0.07	0.08	0.17	0.12	0.05	0.03	M
35	0.07	0.08	0.10	0.07	0.12	0.08	0.07	0.08	0.10	0.06	0.25	0.08	M
36	0.04	0.11	0.10	0.12	0.12	0.06	0.12	0.10	0.07	0.12	0.25	0.08	D
37	0.04	0.11	0.07	0.12	0.12	0.06	0.07	0.08	0.07	0.12	0.25	0.03	D
38	0.04	0.11	0.10	0.09	0.12	0.11	0.12	0.10	0.07	0.08	0.05	0.02	D
39	0.16	0.11	0.07	0.18	0.12	0.24	0.25	0.08	0.17	0.06	0.05	0.02	D
40	0.16	0.11	0.07	0.12	0.12	0.11	0.12	0.14	0.07	0.12	0.25	0.24	C
41	0.04	0.06	0.07	0.12	0.07	0.06	0.12	0.08	0.07	0.06	0.25	0.24	C
42	0.07	0.11	0.10	0.07	0.09	0.06	0.06	0.08	0.10	0.06	0.05	0.08	C

four rankings in Table 4 – the correlation coefficients between the vectors of ranks are all above 0.6 – they increase noticeably when group evaluations are added to the model. Considering only individual evaluations, the coefficient of correlation between the vectors of evaluations derived

Table 4
Final Rankings

Paper	Area	Weighted average		Joint probability	
		Simple	Grouped	Simple	Grouped
1	I	31	28	34	33
2	I	25	35	30	32
3	I	35	39	33	34
4	I	41	40	38	37
5	I	34	29	39	36
6	I	36	30	31	38
7	I	23	33	26	35
8	I	33	42	17	30
9	S	21	10	21	11
10	S	27	22	16	12
11	S	37	24	27	21
12	S	4	14	4	8
13	S	11	16	11	13
14	S	13	15	9	9
15	L	14	18	23	17
16	L	15	12	12	14
17	L	18	26	32	26
18	L	8	8	20	15
19	OI	28	17	28	20
20	OI	24	23	13	19
21	OI	19	11	18	18
22	OI	30	21	24	22
23	OL	22	20	15	16
24	OL	7	7	7	7
25	G	3	1	1	1
26	G	2	5	2	2
27	O	5	2	6	3
28	O	1	3	8	6
29	O	16	13	22	10
30	O	12	6	10	5
31	O	10	4	3	4
32	M	9	38	19	39
33	M	39	37	41	42
34	M	42	34	42	40
35	M	32	41	35	41
36	D	20	25	29	25
37	D	29	27	36	28
38	D	40	36	25	27
39	D	17	31	5	24
40	C	6	9	14	23
41	C	26	19	40	31
42	C	38	32	37	29

from the weighted average and from the computation of joint probabilities is 0.79 (the Spearman rank correlation coefficient is 0.82). When group evaluations are taken into account, the coefficient of correlation becomes 0.93 (the Spearman coefficient is 0.90).

The effect of the size of the sub-area could be detected for extreme cases. The sub-area with the best average rank was 'Graphs', with only two papers presented, while the sub-area with the worst average rank was 'Metaheuristics', with four papers. The rank of two of these four papers changed considerably as sub-areas were included in the analysis.

The observation that the composition based on weighting the criteria seems to be reasonably insensitive to the values of the weights is also relevant. Compositions were tried with weights given by observed standard deviations. These weights, which go from 0.44 to 0.99 for the vectors of individual evaluations and from 0.32 to 0.56 for the vectors generated by substituting for the evaluations the medians of the clusters, vary less than the weights 16, 8 and 1. Still, both led to practically the same scores of global preference. The correlation between the scores derived from the two weightings was 0.95 (the Spearman coefficient is 0.92), for the analyses that did not take into account sub-areas, and increased to 0.98 (Spearman coefficient of 0.95) when sub-areas were considered.

6. Conclusion

In this paper, a new proposal for combining evaluations by means of probabilistic preferences was explored. The application given shows that the probabilistic composition allows us to combine evaluations with different levels of aggregation and to achieve results that are easy to interpret. It allowed for comparing papers presented at a scientific meeting, combining very different criteria according to well-defined points of view.

The application involved a set of 42 options with a reasonably large set of criteria. The probabilistic composition generated ranks that summarize in a comprehensive way this complex information. The composition that uses the joint probability of being preferred by the criteria of different groups and applies different points of view to combine the criteria within each group provides a way to take into account the importance of the different sub-groups without the need to decide on quantitative weights for them.

The only dependence relations assumed were those relating the evaluations according to the same criterion of the papers looked at in isolation and of the papers of the same area taken together. In that case, the maximum possible positive correlation was assumed. The considerable agreement between the rankings both taking and not taking into account group evaluation makes us consider that other dependence structures might not be usefully explored.

Acknowledgements

We are grateful to the three referees who evaluated the paper for their comments. Undoubtedly, any mistakes still remaining are our full responsibility.

We are grateful to CNPq and to FAPESB for grants that supported the development of this work.

References

Aczél, J., Saaty, T.L., 1983. Procedures for synthesizing ratio judgments. *Journal of Mathematical Psychology* 27, 93–102.

Allais, M., Hagen, O., (eds) 1979. *Expected Utility Hypotheses and the Allais Paradox*. D. Reidel, Dordrecht.

Balch, M., McFadden, D., Wu, S., 1974. *Essays on Economic Behavior under Uncertainty*. North Holland, Amsterdam.

Barron, F.H., Barrett, B.E., 1996. The efficacy of SMARTER. Simple multi-attribute rating technique extended to ranking. *Acta Psychologica* 93, 23–36.

Clímaco, J.C.N., 2003. A critical reflection on optimal decision. *European Journal of Operational Research* 153, 506–516.

Ehrgott, M., Greco, S., Figueira, J., 2005. *State of the Art in Multiple Criteria Decision Analysis*. Springer, Berlin.

Foster, J., Sen, A., 1997. *On Economic Inequality*. Clarendon Press, Oxford.

Govindaluri, S.M., Cho, B.R., 2005. Robust design modeling with correlated quality characteristics using a multicriteria decision framework. *International Journal of Advanced Manufaturing Technology* 32, 423–433.

Hammond, J.S. III, Keeney, R.L., Raiffa, H., 1998. *Smart Choices: A Practical Guide to Making Better Decisions*. Harvard Business School Press, Boston, MA.

Liang, T.F., Cheng, H.W., 2009. Application of fuzzy sets to manufacturing/distribution planning decisions with multi-product and multi-time period in supply chains. *Expert Systems with Applications* 36, 3367–3377.

Lootsma, F.A., 1988. Numerical scaling of human judgment in pairwise-comparison methods for fuzzy multicriteria decision analysis. *NATO ASI F, Computer and System Sciences* 48, 57–88.

Machina, M.J., 1982. Expected utility without the independence axiom. *Econometrica* 50, 277–323.

McFadden, D., 1999. Rationality for economists. *Journal of Risk and Uncertainty* 19, 73–105.

Petrovic, S., Fayad, C., Petrovic, D., 2008. Sensitivity analysis of a fuzzy multiobjective scheduling problem. *International Journal of Production Research* 46, 3327–3344.

Rivett, P., 1994. *The Craft of Decision Modelling*. John Wiley, New York.

Saisana, M., Saltelli, A., Tarantola, S., (2005). Uncertainty and sensitivity analysis as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society A* 168, 1–17.

Sant'Anna, A.P., 2002. Data envelopment analysis of randomized ranks. *Pesquisa Operacional* 22, 203–215.

Sant'Anna, A.P., 2005. Composição Probabilística de Critérios na Avaliação de Cursos. *Revista Brasileira de Pós-Graduação* 2, 40–54.

Sant'Anna, A.P., 2009. Probabilistic majority rules: A new approach to the composition of social preferences. *International Journal of Industrial and Systems Engineering* 4, 32–45.

Selvanathan, E.A., Prasada Rao, D.S., 1994. *Index Numbers: A Stochastic Approach*. University of Michigan Press, Ann Arbor, MI.

Sonmez, M., 2007. Data transformation in the evidential reasoning-based decision making process. *International Transactions in Operations Research* 14, 411–427.

Stewart, T.J., 2005. Dealing with uncertainties in MCDA. In Figueira, J., Greco, S., Ehrgott, M. (eds) *Multiple Criteria Decision Analysis – State of the Art Annotated Surveys*. Springer, New York, pp. 445–470.

Triantaphyllou, E., 2000. *Multi-Criteria Decision Making Methods: A Comparative Study*. Kluwer Academic Publishers, Dordrecht.

Tsoukiàs, A., 2008. From decision theory to decision aiding methodology. *European Journal of Operational Research* 187, 138–161.

Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. *Science* 211, 453–458.

Vincke, P., 1992. *Multi-Criteria Decision-Aid*. John Wiley, Chichester.

Woodward, R.T., Bishop, R.C., 1997. How to decide when experts disagree: Uncertainty-based choice rules in environmental policy. *Land Economics* 73, 492–507.

Zadeh, L.A., 1965. Fuzzy sets. *Information and Control* 8, 338–353.

Zadeh, L.A., 1978. Fuzzy sets as the basis for a theory of possibility. *Fuzzy Sets and Systems* 1, 3–28.

Zhou, P., Ang, B.W., Poh, K.L., 2006. Comparing aggregating methods for constructing the composite environmental index: An objective measure. *Ecological Economics* 59, 305–311.

Appendix A

A.1. Converting preferences into probabilities

This Appendix develops the computational aspects of the conversion of initial numerical preferences into probabilities. Any indication of preferences among a set of options, according to a given criterion, can be transformed into a vector of probabilities of preference for each option. We can look at each individual value assigned to the options, for that criterion, as some location measure, e.g. mean, mode or median, for an uncertainty distribution. Hence, the imprecision that is inherent in every human evaluation process may be taken into account properly. These distributions are determined by allowing stochastic disturbances in such central measures. Therefore, the probabilities of being ranked first or last in a sample withdrawn from the space of such distributions can be derived to replace the point values that once represented the initial preferences.

Given a vector of preferences' values, we need to transform these values into probabilities of being the best (or worst) choice for each criterion, according to the prescribed uncertainty model. The appropriate transformation can be derived as follows:

Consider m options evaluated by a given criterion i. Let X_k, for $k=1$ to m, be the random variables associated with the k-th option for that criterion. We need to evaluate, for each option $k \in K$, $M_i(k) = \text{Prob}[X_k = \max\{X_1, \ldots, X_m\}]$ and $m_i(k) = \text{Prob}[X_k = \min\{X_1, \ldots, X_m\}]$.

Assuming the X_ks to be independent,

$$M_i(k) = \int_{Dk} \left\{ \prod_{j \neq k} [F_k(x_k)][f_k(x_k)] \right\} dx_k,$$

where F_k, f_k and Dk are, respectively, the cumulative distribution, the density function and the support of the random variable X_k.

Analogously, we find that

$$m_i(k) = \int_{DK} \left\{ \prod_{j \neq k} [1 - F_k(x_k)][f_k(x_k)] \right\} dx_k.$$

A.2. Points of view

Different points of view may be characterized in terms of choices between extreme positions in two basic orientation axes. These extreme positions are an optimistic versus a pessimistic position, in one axis, and a progressive versus a conservative one, in the other axis.

In the progressive–conservative axis, the progressive evaluator looks for options that are the first in excellence; the conservative evaluator chooses options for their ability of not being the worst. In the optimistic–pessimistic axis, the optimist considers enough to excel (or not to be the worst) in just one criterion. On the opposite end, the pessimist tries to be as good as possible (or not to be the worst) in every criterion. By combining the positions in the extremes of these two axes, four different measures are generated.

Table A1
Example

Option	Criterion 1	Criterion 2	Criterion 3	Criterion 4
A	0.34	0.26	0.36	0.05
B	0.09	0.17	0.12	0.28
C	0.29	0.16	0.32	0.25
D	0.12	0.18	0.12	0.23
E	0.16	0.23	0.08	0.19

For instance, for P_{ik} as defined in Section 3, if the probabilities of the options in a set K of size 5 being the best according to a set I of 4 criteria are as given in Table A1, we have:

(i) For the progressive and pessimistic evaluator, the choice would be option C, which has 0.4% as the product of the probabilities of being the best option according to all four criteria, that is, $\max_{k\in \mathrm{K}}\{\Pi_{i\in I}P_{ik}\}$.
(ii) For the progressive and optimistic evaluator, the choice would now be option A, which maximizes the probability of being the best according to at least one of the criteria, given as $\max_{k\in K}\{1-\Pi_{i\in I}(1-P_{ik}\}$. The value for option A is 70.3%.

If we let Table A1 now represent the probabilities of being the worst according to each criterion, we have:

(iii) For the conservative and pessimistic evaluator, the choice would be option D, which maximizes the product as given by the expression $\max_{k\in K}\Pi_{i\in I}(1-P_{ik})$. The value for option D is 48.9%.
(iv) Finally, for the conservative and optimistic evaluator, the best choice is option B, with the value 99.9%, given as $\max_{k\in K}\{1-\Pi_{i\in I}P_{ik}\}$.