Ranking as parameter estimation

Miroslav Kárný* and Tatiana Valentine Guy

Department of Adaptive Systems,
Institute of Information Theory and Automation,
Academy of Sciences of the Czech Republic
P.O. Box 18, 182 08 Prague 8, Czech Republic
Fax: +420 266052068
E-mail: school@utia.cas.cz
E-mail: guy@utia.cas.cz
*Corresponding author

Abstract: Ranking of alternatives is a common, difficult and repeatedly addressed problem, especially when it requires negotiation of experts. The celebrated Arrow's impossibility theorem expresses formally its difficulty. In spite of the progress made by adopting soft ranking, the problem is far from being generically solved. The paper provides a, probably novel, problem formulation by viewing ranking of alternatives as an estimation of an unknown objective ranking vector. The idea is exposed on a specific task of ranking quality of projects by a large group of experts. The task is important on its own but the proposed methodology is the main message worth of generalisation and use in other application domains.

Keywords: ranking; Bayesian estimation; modelling; negotiation.

Reference to this paper should be made as follows: Kárný, M. and Guy, T.V. (2009) 'Ranking as parameter estimation', *Int. J. Operational Research*, Vol. 4, No. 2, pp.142–158.

Biographical notes: Miroslav Kárný received his CSc (PhD) Degree and DrSc Degree in Theoretical Cybernetics both at the Academy of Sciences of the Czech Republic. In 1976, he joined Institute of Information Theory and Automation, Prague. Since 1990, he has served as Head of the Department of Adaptive Systems of the same Institute. His research experience include: adaptive system, decision making under uncertainty, Bayesian identification, structure estimation, merging prior information into control design, multiple-participant decision making. He is author and co-author of six books and 350 research papers.

Tatiana Valentine Guy received her PhD Degree from the Faculty of Electrical Engineering, Czech Technical University, Prague. She served as a Lecturer at Kiev Polytechnic Institute, Kiev. Since 1994, she is with Department of Adaptive Systems, Institute of Information Theory and Automation, Prague. Her research activities include: modelling, adaptive control, decision making. She is author and co-author of more than 40 research papers and a book.

1 Introduction

Ranking of alternatives is a common, difficult and repeatedly addressed problem, especially when it requires negotiation of experts. This important task is required by many application domains like medicine, environment, engineering, e-democracy, finances, (e.g., Prato, 1999; Rios Insua et al., 2003; Hallerbach and Spronk, 2003). Naturally, the experts have different viewpoints and preferences and systematic approach able to harmonise theirs often contradictory demands is needed. The desired approach should respect personal biases of the experts as well as different scaling presented. A considerable amount of research is devoted to the search for methodology providing the desired solution, see, for instance Jacquet-Lagreze and Siskos (1982), Kim et al. (1999), Butler et al. (2001), Slevin et al. (1998), Rios Insua et al. (2003) and Medaglia et al. (2007). Difficulty of the fair ranking of alternatives is formally expressed by a bit depressing Arrow's impossibility theorem, Arrow (1995). The limitations expressed by it are now attacked by more realistic problem formulations based on a 'soft' description of preferences, Nurmi (2001), but a constructive unambiguous methodology is still missing. The present paper tries to fill this gap by reformulating the ranking problem as estimation of unknown 'objective' rank-determining vector. With this reformulation, it boils down to modelling of relationships of marks assigned by respective experts to the parameter describing the objective ranking and to simple Bayesian parameter estimation, Peterka (1981).

The work has been inspired by the evaluation procedure currently employed by the majority of grant agencies. The procedure relies on marks provided by experts reviewing the submitted project proposals. The complete procedure considers several different tasks including ranking of the submitted proposals according to some predefined criteria. This ranking procedure serves as the problem prototype here.

The layout of the paper is as follows. Section 2 describes the current ranking procedure as it is and outlines its main drawbacks. Formalisation of the ranking as parameter estimation problem is proposed in Section 3. Section 4 describes the key model of the marking process. The solution of the resulting estimation problem is presented in Section 5. The proposed procedure, which modifies the current ranking practice, is summarised in Section 6. Section 7 presents a case study of the developed solution on real data. Section 8 provides the concluding remarks, and outlines the directions of the future work.

The following notation and conventions are used throughout the text. x^* denotes a set of possible values of the variable x. $\mathring{x}$ means cardinality of the finite set x^*. $f(\cdot|\cdot)$ denotes probability density function (pdf). The pdfs are distinguished by the identifiers in their arguments. No formal distinction is made between random variable, its realisation and an argument of a pdf. The correct meaning follows from the context.

2 Standard proposals ranking and its drawbacks

Project proposals, submitted to a granting agency, are numbered by $p \in p^* \equiv \{1, \dots, \mathring{p}\}$. Limited overall budget available for the support motivates the peer review process to select high quality proposals. To respect various grant agency's aims, several criteria, labelled by $c \in c^* \equiv \{1, \dots, \mathring{c}\}$, are established. The importance of the particular criterion is reflected by its weight $w_c > 0, \sum_{c \in c^*} w_c = 1$. To set a lower limit on the proposal's quality, the grant agency a priori determines threshold values $t_c \in m_c^*$ for all criteria $c \in c^*$, where m_c^* defines a set of possible marks for the criterion c. Therefore, only the proposals passed the thresholds for all criteria can be considered for the support.

The review process is performed by $\mathring{e}$ independent experts. Each expert $e \in e^* \equiv \{1, \ldots, \mathring{e}\}$ judges a subset of proposals ${}^e p^* \equiv \{1, \ldots, {}^e \mathring{p}\} \subset p^*$, allocated to him. The allocation is done so that each proposal p is reviewed by several experts forming the set ${}^p e^* \equiv \{1, \ldots, {}^p e^*\} \subset e^*$.

The number of the proposals, judged by the expert e, is $1 < {}^e \mathring{p} \ll \mathring{p}$. The left-hand side restriction allows the expert to compare quality of the competitive proposals and thus decreases the needed number of experts. The right-hand side restriction shortens the reviewing time and reduces inevitable personal biases and inconsistencies.

Mark ${}^e m_{pc} \in m_c^*$, assigned by the eth expert, expresses expert's subjective opinion about the pth proposal with respect to the degree of satisfaction of the cth criterion. The completely ordered set of all possible marks with respect to the cth criterion m_c^*, $c \in c^*$, is *a priori* defined by a *finite* collection of values on the interval $[\underline{m}_c, \overline{m}_c]$, where $\underline{m}_c$ and $\overline{m}_c$ are the lowest and the highest possible mark, respectively. The overall mark ${}^e m_p$ of the proposal p, assigned by the eth expert, is computed as the weighted sum of marks ${}^e m_{pc}$

$$ {}^e m_p \equiv \sum_{c \in c^*} w_c \, {}^e m_{pc}. \tag{1} $$

As said, each proposal p is judged by a group of experts ${}^p e^* \equiv \{1, \ldots, {}^p \mathring{e}\} \subset e^*$ with a few members. A *full agreement* on the marking ${}^e m_{pc}$, $\forall e \in {}^p e^*$, $\forall c \in c^*$ is required in the specific case, which has motivated this paper. To reach this, a detailed discussion of the group ${}^p e^*$ is supposed. The discussion can, however, modify the experts' original marking with respect to particular criteria ${}^e m_{pc}$ and, consequently, the overall marks ${}^e m_p$.

The overall harmonisation of experts' marking of all proposals is the final stage of the review process. On the final experts' assembly, where all $\mathring{e}$ experts participate:

- the proposals, whose group-harmonised marking of particular criteria fail to cross the given *thresholds*, are taken as unsuccessful and withdrawn from the future consideration
- the remaining proposals are *ranked* according to the overall mark assigned to them
- the final overall harmonisation is performed with the stress on the proposals with the highest overall marks (the most probable candidates for funding).

The proposals with marks the assembly doubts about as well as those with identical marks are ranked by the assembly, which can modify the marks.

The described well-elaborated procedure has the following drawbacks.

- The outcome is probably biased due to naturally different scaling of experts' marking. For instance, some experts take the upper marking bounds as unreachable ideals. Others may interpret them as a basis, which should be left only when some significant flaws are found in the project proposal.

- The reaching of the full agreement is hard even within a small group of experts. Often, the group fixes marks due to the limited negotiation time and sometimes at price of voting. It can be practically reached only when *finite* set of marks m_c^*, $c \in c^*$, is allowed (so-called *discrete-valued ranking*).
- A fair overall ranking via discussion of the experts' assembly is almost impossible as each expert e reviewed only the subset ${}^e p^*$ of all proposals p^*. Moreover, the assembly is too large for an efficient and fair communication.

The procedure proposed below suppresses the mentioned drawbacks:

- personal biases and inconsistencies are counteracted
- need for the full agreement on marking can be relaxed
- a set of possible marks m_c^* is defined as an *infinite* collection of all values within the interval $[\underline{m}_c, \overline{m}_c]$, so-called *real-valued ranking*
- the final ranking is done computationally in a well-justified way, so that the expert assembly (if needed at all) may care only about the exceptional cases requiring complementary peer review
- the proposed procedure offers an additional quality assurance as it points to experts with excessive biases or personal uncertainty.

3 Formalisation of ranking problem as parameter estimation

During the peer review process, each proposal p gets an overall mark $m_p \in [\underline{m}, \overline{m}]$, $0 \leq \underline{m} < \overline{m} < \infty$, see equation (1). The considered *ranking* procedure concerns creation of the *ranking vector*, which ranks the proposals according to their quality: from the worst to the best. A number of high quality proposals, corresponding to the several last elements of the ranking vector, will be considered for funding. The final number of such proposals is determined by the agency's budget that has to cover the sum of budgets of the supported proposals.

The ranking procedure is applied only to the proposals eligible for the support, i.e., those passed corresponding threshold values. Elements of the desired ranking vector are the overall marks reached by the respective proposals, ordered from the smallest to the highest mark. Thus, for example, the last element corresponds to the proposal of the highest quality. Creation of the ranking vector, using the marks provided by experts, is the main aim of the reviewing procedure.

Formulation of the *ranking* problem as *parameter estimation* is the basic idea of the paper. Let us assume that there is an *objective* ranking of proposals (with respect to the a priori known agency aims) expressed by the real-valued *ranking vector* $r \equiv [r_1, \dots, r_{\mathring{p}}]'$,

$$r_p \in [\underline{r}, \overline{r}] \supseteq [\underline{m}, \overline{m}], \quad \underline{m} = \sum_{c \in c^*} w_c \underline{m}_c, \quad \overline{m} = \sum_{c \in c^*} w_c \overline{m}_c. \tag{2}$$

The ranking vector ranks the proposals: the pth is better than $\tilde{p}$th iff $r_p > r_{\tilde{p}}$.

The mark ${}^{e}m_p$ of the pth proposal chosen by the expert e reflects his subjective guess on the unknown objective value r_p. Thus, the expert can be considered as non-ideal observing device: marks he selects represent noisy observations of the unknown ranking vector. The collection of all marks ('observations')

$$D \equiv \left\{ \{{}^{e}m_p\}_{p \in {}^{e}p^*} \right\}_{e \in e^*} \tag{3}$$

represents the data available for estimation of the unknown ranking vector r and some experts' characteristics θ. In accordance with Bayesian view on incomplete knowledge (Peterka, 1981), the unknown r and θ are treated as random variables. The richest information on them is then represented by the conditional pdf $f(r, \theta \,|\, D)$. Its maximiser $(\hat{r}, \hat{\theta})$ is a good point estimate of the unknown 'objective' ranking vector r and experts' characteristics θ.

Bayes rule implies that ($\propto$ means proportionality)

$$\overbrace{f(r,\theta \,|\, D)}^{posterior\ pdf} \propto \underbrace{\overbrace{f(D \,|\, r,\theta)}^{likelihood\ function} \times \overbrace{f(r,\theta).}^{prior\ pdf}}_{\exp(-0.5L(r,\theta)) \equiv \exp(-0.5 \times posterior\ log-likelihood)} \tag{4}$$

A fair evaluation is considered. Consequently, the distribution of the mark ${}^{e}m_p$ of the eth expert to pth proposal is fully determined by the expert characteristics ${}^{e}\theta$, which do not vary with the judged proposal, and by the entry r_p of the ranking vector that corresponds to the proposal p. In other words, observations are conditionally independent and the likelihood function (4) reads

$$f(D \,|\, r, \theta) = \prod_{e \in e^*} \prod_{p \in {}^{e}p^*} f\left({}^{e}m_p \,|\, r_p, {}^{e}\theta\right). \tag{5}$$

Considering the general form of the prior pdf (4), the values r_p, $p \in p^*$, are assumed to be mutually independent as the project proposals are submitted by different proposers. In the considered fair evaluation, they are also independent of evaluating experts, whose characteristics are also a priori independent, i.e.,

$$f(r, \theta) = \prod_{e \in e^*} {}^{e}f({}^{e}\theta) \prod_{p \in {}^{e}p^*} {}^{p}f(r_p). \tag{6}$$

The form of the posterior pdf (4) is thus determined by the specific form of individual factors $f({}^{e}m_p \,|\, r_p, {}^{e}\theta)$, ${}^{e}f({}^{e}\theta)$ and ${}^{p}f(r_p)$. The first factor results from a simple modelling of the expert e. Form of the prior pdf is taken to be similar to the likelihood function on fictitious observations. Its moments are fitted to expected ranges of involved variables. This pragmatic methodology is known to provide self-reproducing (conjugate) prior pdf whenever it exists.

The rest of the text follows this line in estimating the overall rank. The same methodology can be used for estimating ranks of the respective partial criteria.

4 Modelling of an expert

The proposed ranking way relies on the modelling of the evaluating experts.

4.1 Probabilistic model of an expert

The allowed set of experts' marks m_c^* (and thus of ranks r^*) can be either

- an interval $m_c^* = [\underline{m}_c, \overline{m}_c]$, or
- a predefined collection of values $m_c^* = \{m_i\}_{i=1}^{\mathring{m}}$, with $m_i \in [\underline{m}_c, \overline{m}_c]$.

Evaluation and ranking corresponding to the first case are called here *real-valued*, while the second case implies *discrete-valued* ranking. We focus on real-valued marks and real-valued ranking vector. The discrete-valued version is just briefly commented.

The expert's mark ${}^e m_p$ always differs from the rank r_p by a 'personal' deviation of the expert e. This deviation can be decomposed into systematic shift ${}^e b$ from the unknown objective value r_p and zero-mean variation ${}^e \varepsilon_p$ of expert's subjective evaluation. Thus,

$$ {}^e m_p = r_p + {}^e b + {}^e \varepsilon_p. \tag{7} $$

Supposed fair evaluation implies that the subjective variations ${}^e \varepsilon_p$ can be assumed mutually independent (over all proposals and over all experts) with expert-dependent variances ${}^e v$. Besides, we treat the variations ${}^e \varepsilon_p$ as normal. This choice is motivated by the desired simplicity of the subsequent treatment. It can also be supported by the well-known fact that the normal pdf has the largest entropy among pdfs having zero mean and a fixed variance ${}^e v > 0$.

Under the normality assumption, parameter ${}^e\theta \equiv ({}^e b, {}^e v)$ characterises the eth expert. The collection $\theta \equiv (b, v) \equiv ({}^1 b, \ldots, {}^{\mathring{e}} b, {}^1 v, \ldots, {}^{\mathring{e}} v)$ parameterises all participating experts and the likelihood function in equation (4) becomes

$$ \begin{aligned} f(D \mid r, b, v) &= \prod_{e \in e^*} \prod_{p \in {}^e p^*} (2\pi {}^e v)^{-0.5} \exp\left[-0.5 {}^e v^{-1} \left({}^e m_p - r_p - {}^e b\right)^2\right] \\ &\propto \prod_{e \in e^*} {}^e v^{-0.5 {}^e \mathring{p}} \exp\left\{-0.5 {}^e v^{-1 e} \mathring{p} \left[({}^e b - {}^e B(r))^2 + \frac{{}^e \mathring{p} - 2}{{}^e \mathring{p}} {}^e V(r)\right]\right\}, \\ {}^e B(r) &\equiv \frac{1}{{}^e \mathring{p}} \sum_{p \in {}^e p^*} ({}^e m_p - r_p), \\ {}^e V(r) &\equiv \frac{1}{{}^e \mathring{p} - 2} \left[\frac{1}{{}^e \mathring{p}} \sum_{p \in {}^e p^*} ({}^e m_p - r_p)^2 - {}^e B^2(r)\right]. \end{aligned} \tag{8} $$

Note that the second line in equation (8) is obtained by completing squares in the exponent with respect to ${}^e b$. Alternatively, the squares can be completed with respect to the ranking vector r. These alternatives motivate the choice of the marginal pdfs forming the prior pdf, see Section 4.2.

Comment on the number of data available

The number $\mathring{D}$ of the data provided by the experts is $\sum_{e \in e^*} {}^e \mathring{p}$. The data D are available for estimating the unknown parameters $\Theta \equiv (r, b, v)$ with $\mathring{\Theta} \equiv \mathring{p} + 2 \times \mathring{e}$ entries. It is desirable to have $\mathring{D} > \mathring{\Theta}$. This situation is reached, iff an average number of proposals judged by an expert is greater than $\frac{\mathring{p}}{\mathring{e}} + 2$.

Comment on discrete-valued ranking

Often, the discrete-valued marking is required. It is simple to handle and makes the groups and the assembly's negotiations easier. However, the quality of proposals usually varies smoothly and unevenly within the proposals' set p^*. Thus, the real-valued ranking vector r is a more realistic model than the discrete-valued one. In spite of the fact that the discrete-valued marking introduces unnecessary rounding errors, the common use of quantised marks calls for the model suitable to this case. Assuming that both the marking and ranking vector are discrete-valued, the following formula provides a possible discrete-valued counterpart of the model (8)

$$f(D\,|\,r,a,w) = \prod_{e\in e^*}\prod_{p\in {}^e p^*} {}^e w^{\delta({}^e m_p - r_p - {}^e a)} \left(\frac{1-{}^e w}{\mathring{r}-1}\right)^{1-\delta\left({}^e m_p - r_p - {}^e a\right)}, \tag{9}$$

where $\delta(0) = 1$ and $\delta(x \neq 0) = 0$; ${}^e a$ is the unknown discrete-valued bias of the expert e; ${}^e w \in (0,1)$ is the probability with which the expert e selects the mark ${}^e m_p = r_p + {}^e a$. The probability ${}^e w$ of this event is expected to be close to one. This makes a finer modelling of events ${}^e m_p \neq r_p + {}^e a$ unnecessary.

The number $\mathring{D}$ of data D provided by experts is again $\sum_{e\in e^*} {}^e \mathring{p}$. The data D serve for estimating the unknown parameters $\Theta \equiv (r,a,w)$ with $\mathring{\Theta} \equiv \mathring{p} + 2 \times \mathring{e}$ entries. Thus, the condition $\mathring{D} > \mathring{\Theta}$ is independent of the valuation version.

4.2 Choice of the prior pdf

In order to decrease evaluation complexity, the conjugated form, Berger (1985), of prior pdf is desired. However, parameters (b,v) characterising experts and ranking vector r are assumed to be a priori independent. This makes us to select Gauss-inverse-Wishart pdf as the prior pdf but with ${}^e b$ not depending on the ranking vector r, cf. (8). Moreover, experts are a priori independent $f({}^1 b, {}^1 v, \ldots, {}^{\mathring{e}} b, {}^{\mathring{e}} v) = \prod_{e=1}^{\mathring{e}} {}^e f({}^e b, {}^e v)$ and a priori indistinguishable. Thus, ${}^e f(b,v) = f(b,v), \forall e \in e^*$, and statistics determining the single pdf $f(b,v)$ have to be chosen only. This makes us to select

$$f({}^e b, {}^e v) = f({}^e b, {}^e v\,|\,\hat{b}_0, \hat{v}_0, c_0, \nu_0) \equiv f({}^e b, {}^e v\,|\,\mathcal{S}_0), \quad \text{with}$$

$$f(b,v\,|\,\mathcal{S}) \propto v^{-0.5(\nu+3)} \exp\left\{-0.5 v^{-1}\nu\left(\frac{(b-\hat{b})^2}{c} + \frac{\nu-2}{\nu}\hat{v}\right)\right\}. \tag{10}$$

This pdf has the following moments of interest (Kárný et al., 2005), determined by the statistics $\mathcal{S}$ consisting of four entries $\hat{b}$, $\hat{v} > 0$, $c > 0$, $\nu > 0$

$$\mathcal{E}[b|\mathcal{S}, v] = \hat{b},\ \ \mathcal{E}[v\,|\,\mathcal{S}] = \hat{v},\ \ \mathrm{var}(b\,|\,\mathcal{S}, v) = v\frac{c}{\nu},\ \ \mathrm{var}[v\,|\,\mathcal{S}] = \frac{\hat{v}}{\nu-4}. \tag{11}$$

These moments and vague prior knowledge lead to the following specific but universal options (motive for the particular choice is given in brackets)

$\hat{b}_0 = 0$ (experts are a priori believed to be unbiased), (12)

$\hat{v}_0 = [0.25 \times (\overline{m} - \underline{m})]^2$ (expected v is squared half of the half-range of m),

$\nu_0 = 4$ (chosen flat prior implies the first moment of v exists only),

$c_0 = 1$ (expected standard deviation of ${}^e b$ is the half of $\sqrt{v}$).

The entries of the ranking vector r are also a priori independent and indistinguishable. They should be independent of the experts' parameters, too, i.e., ${}^{p}f(r\,|\,b,v) = f(r)$, $\forall p \in p^{*}$. The following normal form of $f(r) = f(r\,|\,\hat{r}_0, \gamma)$, with the expected value $\hat{r}_0$ and variance γ, is assumed

$$f(r\,|\,\hat{r}_0,\gamma) = \prod_{p\in {}^{e}p} (2\pi\gamma)^{-0.5} \exp\left\{-0.5\frac{(r_p-\hat{r}_0)^2}{\gamma}\right\} \text{ with statistics} \tag{13}$$

$$\hat{r}_0 = 0.5 \times (\overline{m} + \underline{m}) \text{ (prior rank expectation is in the middle of interval)} \tag{14}$$
$$\gamma = [0.25 \times (\overline{m} - \underline{m})]^2 = \hat{v}_0 \text{ (standard deviation coincides with that of } v\text{).}$$

The formulae (13) and (14) complete the choice of the prior pdf, i.e., both its form and its statistics.

5 Estimation of unknown parameters

The proposed parameterised model (8) establish a base for estimation of unknown parameters $\Theta = (r, (b, v)) \equiv$ (ranking vector, experts' characteristics). For the considered models, the number $\overset{\circ}{D}$ of data D is larger than the number $\overset{\circ}{\Theta}$ of estimated parameters Θ. Moreover, the maximum likelihood estimate cannot be used as the corresponding normal equations are always singular due to possible compensation of the ranking entries and expert biases. Besides, the excess of data is relatively small. Thus, the exploitation of prior knowledge is inevitable. Due to its regularising effect we can take maximum a posteriori probability estimate as a good point estimate.

Likelihood (8) and the prior pdf (10), (13) with the statistics (12) and (14), provide the following form of the $-2 \times$ logarithm of a posteriori probability density function $L(\Theta)$, (see equation (4), the term not influencing maximiser is omitted)

$$L(\Theta) = \sum_{p\in {}^{e}p^{*}} \frac{(r_p-\hat{r}_0)^2}{\hat{v}_0} + \sum_{e\in e^{*}} \left\{ ({}^{e}\overset{\circ}{p} + \nu_0 + 3)\ln({}^{e}v) + \frac{1}{{}^{e}v}\left[2\hat{v}_0 + \nu_0\,{}^{e}b^2 + \sum_{p\in {}^{e}p^{*}} ({}^{e}m_p - r_p - {}^{e}b)^2\right]\right\}. \tag{15}$$

The maximisation of a posteriori probability is equivalent to the minimisation of the function (15) with respect to (r, b, v). It is quadratic function of r, b with positive-definite kernel. Introducing the weights ${}^{e}\alpha \equiv \frac{1}{{}^{e}\overset{\circ}{p} + \nu_0 + 3}$, the necessary conditions for the extremum read

$${}^{e}\hat{v} = {}^{e}\alpha\left[\hat{v}_0 + \nu_0\,{}^{e}\hat{b}^2 + \sum_{p\in {}^{e}p} ({}^{e}m_p - \hat{r}_p - {}^{e}\hat{b})^2\right], \tag{16}$$

$${}^{e}\hat{b} = \frac{1}{{}^{e}\overset{\circ}{p} + \nu_0} \sum_{p\in {}^{e}p^{*}} ({}^{e}m_p - \hat{r}_p), \tag{17}$$

$$\hat{r}_p = \frac{\hat{r}_0 + \sum_{e\in {}^{p}e^{*}} \frac{\hat{v}_0}{{}^{e}\hat{v}} ({}^{e}m_p - {}^{e}\hat{b})}{\sum_{e\in {}^{p}e^{*}} \frac{\hat{v}_0}{{}^{e}\hat{v}} + 1}. \tag{18}$$

The minimised function can be shown to be strictly convex around the stationary points. Thus, it has single minimiser and the solution of the equations (16)–(18) is a unique maximiser of the a posteriori pdf. This system of equations can be solved by successive approximations as follows.

`Initial` phase

- Specify sets $p^*, e^*, {}^e p^*, {}^p e^*$ and set $\nu_0 = 4$.
- Evaluate the number of elements ${}^e \overset{\circ}{p}$ in ${}^e p^*$ and the weights ${}^e\alpha = \frac{1}{{}^e \overset{\circ}{p} + \nu_0 + 3}$.
- Specify the ranges $\underline{m}$, $\overline{m}$ and define the prior statistics depending on them $\hat{r}_0 = 0.5(\underline{m} + \overline{m})$, $\hat{v}_0 = [0.25(\overline{m} - \underline{m})]^2$.
- Collect the data $D = \{\{{}^e m_p\}_{p \in {}^e p^*}\}_{e \in e^*}$.
- Select the upper bound $\overset{\circ}{n}$ on the possible number of iterations n, and the stopping precision $\varepsilon > 0$, $\varepsilon \approx 0$.
- Set initial values for the counter $n = 0$ and the stopping-rule $norm = 2\varepsilon$.
- Set (at iteration $n = 0$) the initial estimates of variances ${}^e\hat{v}(n) = \hat{v}_0$, biases ${}^e\hat{b}(n) = \hat{b}_0 = 0$ and the negative log-likelihood $L(\widehat{\Theta}(n)) = \infty$.

`Iterative` phase

- Do while $n \leq \overset{\circ}{n}$ and $norm > \varepsilon$.
- Set $\hat{r}_p(n+1) = \frac{\hat{r}_0 + \sum_{e \in {}^p e^*} ({}^e m_p - {}^e\hat{b}(n)) \frac{\hat{v}_0}{{}^e\hat{v}(n)}}{\sum_{e \in {}^p e^*} \frac{\hat{v}_0}{{}^e\hat{v}(n)} + 1}$, $\forall p \in p^*$, cf. (18).
- Set ${}^e\hat{b}(n+1) = \frac{1}{{}^e \overset{\circ}{p} + \nu_0} \sum_{p \in {}^e p} ({}^e m_p - \hat{r}_p(n+1))$, $\forall e \in e^*$, cf. (17).
- Set $n = n + 1$ and ${}^e\hat{v}(n) = {}^e\alpha[\hat{v}_0 + \nu_0 {}^e\hat{b}^2(n) + \sum_{p \in {}^e p^*} ({}^e m_p - \hat{r}_p(n) - {}^e\hat{b}(n))^2]$, $\forall e \in e^*$, cf. (16).
- Evaluate for $\widehat{\Theta}(n) \equiv (\hat{r}(n), b(n), v(n))$ the negative log-likelihood $L(\widehat{\Theta}(n))$, cf. (15), determining the $norm \equiv \text{abs}[L(\widehat{\Theta}(n)) - L(\widehat{\Theta}(n-1))]$.

`Terminal` phase

Take $\hat{r}(n)$ as the best guess of the ranking vector r and use $\hat{v}(n)$, possibly together with $\hat{b}(n)$, for judging reliability of experts.

6 The recommended proposals evaluation procedure

Here, we summarise the ranking procedure that combines the current practice with the derived results in the case of continuous-valued marking and ranking.

- Select $\overset{\circ}{e}$ experts and assign to each of them ${}^{e}\overset{\circ}{p} > 1$ proposals, so that each proposal is judged by ≈3–4 experts.
- Let the experts judge proposals assigned.

  `Continuous − valued marks within pre − specified ranges should be used.`

  This optional change should avoid quantisation errors and increase robustness of the overall evaluation process.
- Estimate ranking vector and quality of experts using their non-harmonised data, i.e., data available before the group discussion, see Section 2.

  This step serves informal judging of the negotiation process and, primarily, evaluation of experts' quality. Experts with high values of personal variance ${}^{e}\hat{v}$ and bias abs$({}^{e}\hat{b})$ are suspected of possible conflict of interests or an insufficient competency.

  `This is novel optional step serving quality assurance.`
- Let the group of experts evaluating the same proposal consult mutually and let them harmonise their marking.

  `Experts should harmonise their comments, but their numerical marking need not be identical!`

  This change will help to avoid necessary voting in controversial cases and will make the harmonisation procedure more fair and robust. This relaxation of the current rules (see Section 2) seems to be highly desirable. Among other, it would speed up the procedure without loosing its quality.
- Estimate ranking vector and quality of experts using the harmonised data.

  `The estimation should replace, or, at least, prepare the last overall harmonisation step more rigorously.`

  The changes of expert-parameter estimates should be mainly in biases: this would reflect improvement in scaling. Again, high values of ${}^{e}\hat{v}$ indicate unreliable expert.

  `The procedure should be also applied to particular criteria and to the probability whether they (do not) crossed thresholds with respect to them.`

Neither the last nor other mechanical steps should violate peer-review principles. The questionable cases should be discussed individually. Note that the proposed evaluations are relatively simple and no new optional parameters are introduced.

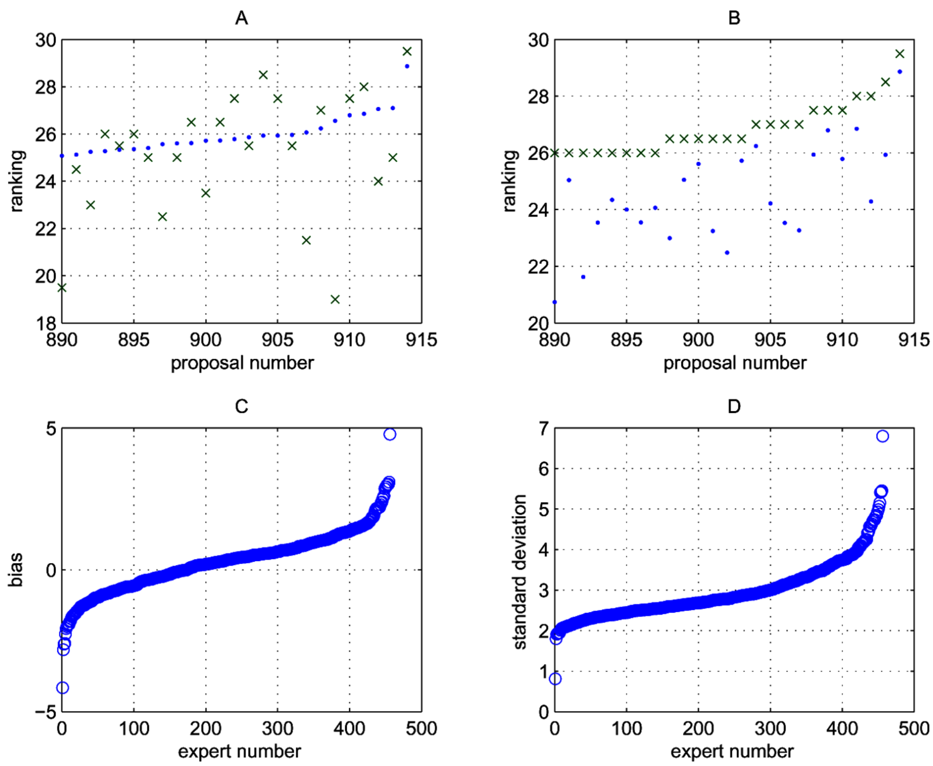
7 Case study with real data

The developed theory was tested on data sets originating from an international granting agency. Before processing, the raw data about proposals and experts were made anonymous. Consequently, no conclusion concerning a specific proposal or expert can be made. The study wanted

- to gain experience with the proposed evaluation technique concerning its behavior (including numerical stability, computational demands)
- to find whether significant personal biases and uncertainties are observable
- to recognise the extent to which the proposed ranking deviates from that made by the procedure used up to now.

The results of processing of the four data sets, labelled by **D1**, **D2**, **D3**, **D4** are presented on figures of a uniform structure (see Figures 1–4).

Figure 1 Results obtained on the data set **D1**: (A) proposed 'o' vs. standard 'x' ranking: proposed ranking is ordered; (B) standard 'x' vs. proposed 'o' ranking: standard ranking is ordered; (C) estimates of experts' biases and (D) estimates of experts' standard deviations (see online version for colours)



The left subplot in the upper part of the figure depicts ranks of proposals ordered according to the proposed ranking procedure (marked by dots). For comparison, the corresponding ranks, assigned by the current procedure, are shown by crosses. The right subplot in the upper part of the figure displays similar comparison, but the proposals ordered by the current ranking procedure. Both subplots display only the proposals with high values of the rank, i.e., near the threshold or greater. The first subplot in the lower part of the figure displays estimates ${}^{e}\hat{b}$ of experts' biases ${}^{e}b$ and the second subplot in the lower part of the figure provides estimates of experts' personal standard deviations.

Figure 2 Results obtained on the data set **D2**: (A) proposed 'o' vs. standard 'x' ranking: proposed ranking is ordered; (B) standard 'x' vs. proposed 'o' ranking: standard ranking is ordered; (C) estimates of experts' biases and (D) estimates of experts' standard deviations (see online version for colours)

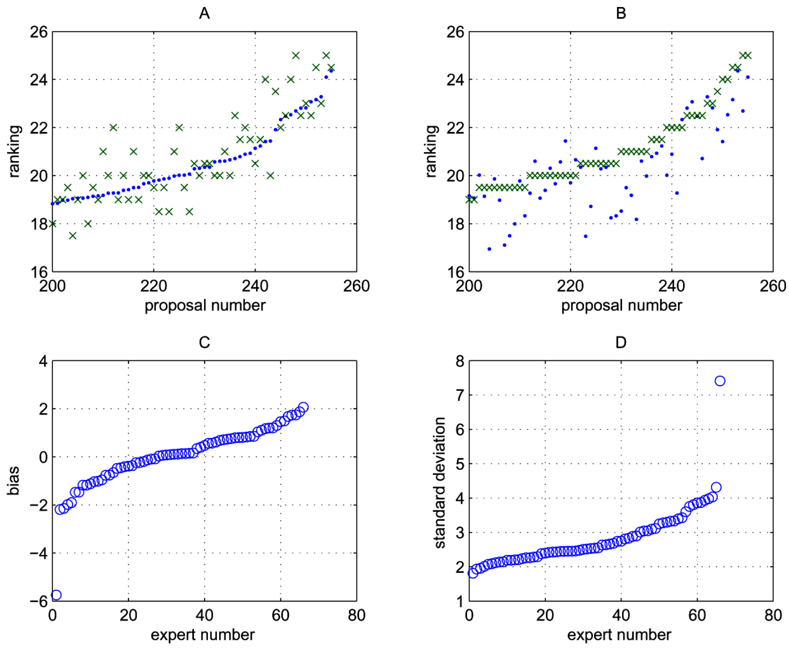


Figure 3 Results obtained on the data set **D3**: (A) proposed 'o' vs. standard 'x' ranking: proposed ranking is ordered; (B) standard 'x' vs. proposed 'o' ranking: standard ranking is ordered; (C) estimates of experts' biases and (D) estimates of experts' standard deviations (see online version for colours)

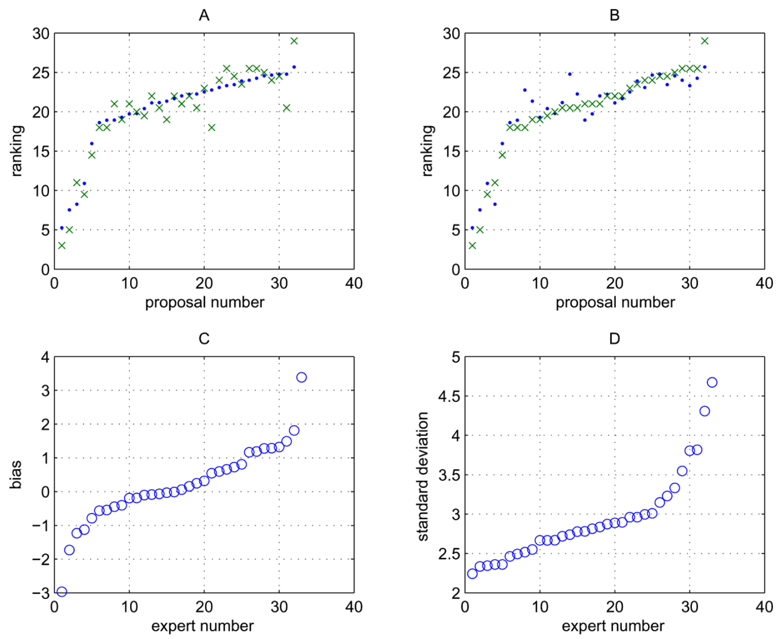
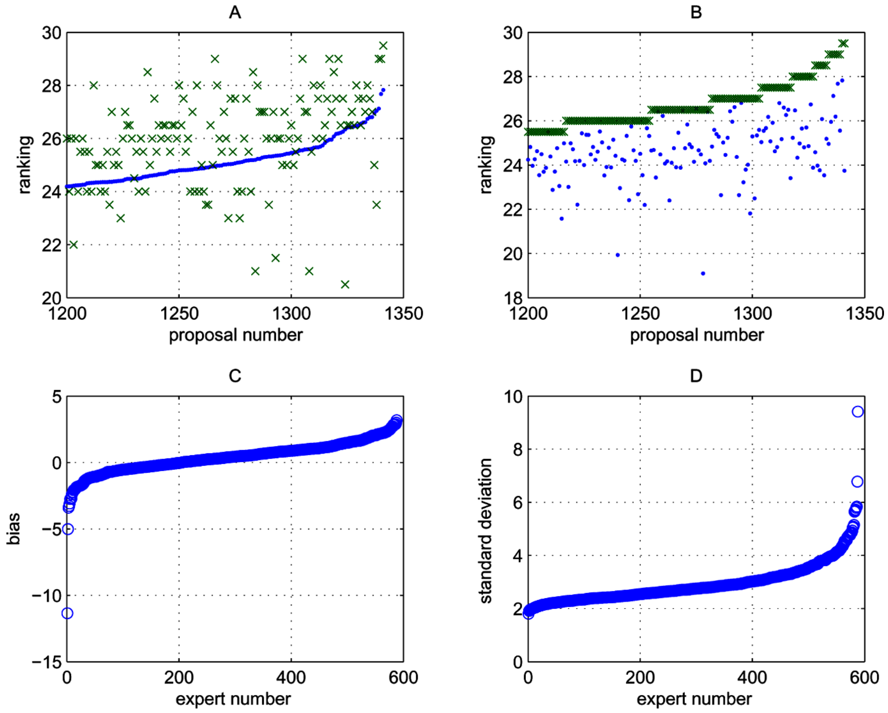


Figure 4 Results obtained on the data set **D4**: (A) proposed 'o' vs. standard 'x' ranking: proposed ranking is ordered; (B) standard 'x' vs. proposed 'o' ranking: standard ranking is ordered; (C) estimates of experts' biases and (D) estimates of experts' standard deviations (see online version for colours)



Descriptive statistics of experiments are given in Table 1. They indicate the degree of coincidence (discrepancy) between the standard and proposed way of ranking. Rank thresholds, determining the border line for the funding, were chosen to illustration purpose only and have no relationships to the thresholds used.

Table 1 Descriptive statistics of experiments

Case	*D1*	*D2*	*D3*	*D4*
Number of proposals	914	255	32	1341
Number of experts	456	66	33	588
Threshold of the overall rank	25	22	22	25
Number of proposals above threshold: proposed	27	11	16	72
Number of proposals above threshold: standard	43	13	11	157
Number of proposals chosen by both ways	16	10	11	57
Degree of coincidence, %	37	77	100	36

Table 2 concerns the experts. The thresholds of the overall rank, given by the Table 1, are used in relative values presented in Table 2. The histograms in Figure 5 put a proper perspective on results shown in Table 2 and indicate sensitivity of the outcomes on estimates of the ranking vector r. It is sufficient to notice that the boxes in histograms correspond roughly 2% change in the values $\hat{r}_p$ assigned to the proposals. At the same time, the number of proposals in boxes adjacent to the selected threshold levels may reach even several tens.

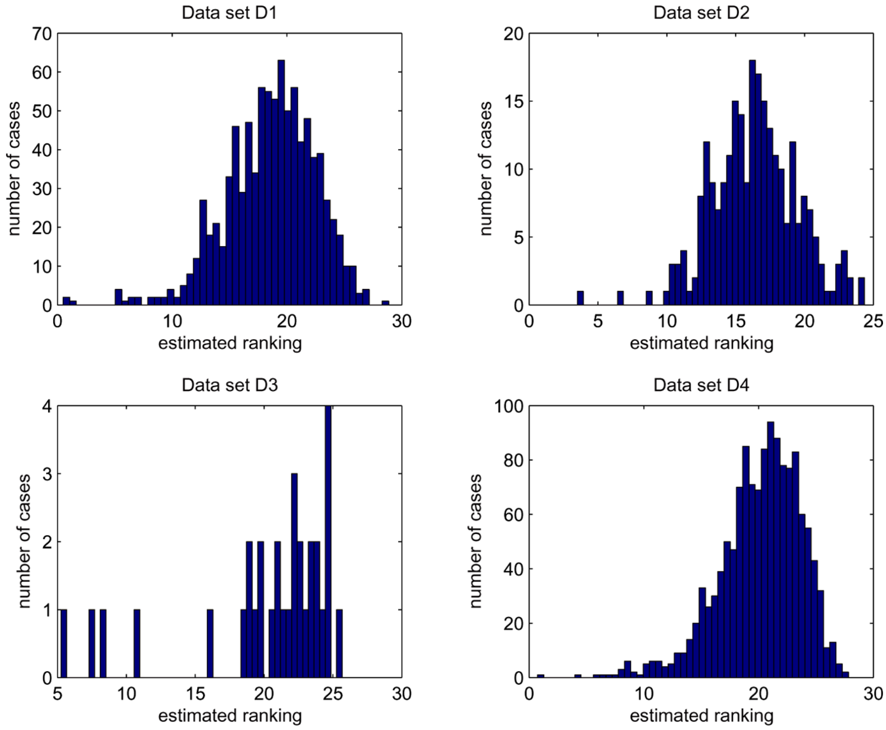
Table 2 Statistics characterising the reviewing quality

Case	*D1*	*D2*	*D3*	*D4*
Mean of expert biases	0.27	0.03	0.20	0.36
$\frac{\text{mean of expert biases}}{\text{threshold of the overall rank}}$, %	1.0	1.2	6.0	4.1
Standard deviation (std) of expert biases	1.04	1.25	1.15	1.20
$\frac{\text{std of expert biases}}{\text{threshold of the overall rank}}$, %	4.2	5.7	5.2	4.8
Minimum of expert biases	−4.15	−5.75	−3.00	−11.35
$\frac{\text{minimum of expert biases}}{\text{threshold of the overall rank}}$, %	−16.6	−26.1	−13.6	−45.4
Maximum of expert biases	4.78	2.06	3.39	3.18
$\frac{\text{maximum of expert biases}}{\text{threshold of the overall rank}}$, %	19.1	9.4	15.4	12.7
Mean of expert stds	2.94	2.82	2.94	2.92
$\frac{\text{mean of expert stds}}{\text{threshold of the overall rank}}$, %	11.8	12.8	13.4	11.7
Std of expert stds	0.69	0.83	0.57	0.74
$\frac{\text{std of expert stds}}{\text{threshold of the overall rank}}$, %	2.8	3.8	2.6	3.0
Minimum of expert stds	0.81	1.81	2.24	1.81
$\frac{\text{minimum of expert stds}}{\text{threshold of the overall rank}}$, %	3.2	8.2	10.2	7.2
Maximum of expert stds	6.80	7.41	4.67	9.42
$\frac{\text{maximum of expert stds}}{\text{threshold of the overall rank}}$, %	27.2	33.7	21.2	37.7

A deeper discussion would require individual inspection of the outlying cases (both with respect to proposals and experts). For example, the comparative study of the estimated overall rank and scientific results achieved by the project supported would be desirable. Such a study will be published in the future. The current results indicate that:

- discrepancy in proposals selected by the standard and proposed procedure may be significant
- the majority of experts has a small biases but there is a non-negligible portion of those having significant biases
- the majority of experts has similarly small personal variations but there is a non-negligible portion of experts having significant variations
- use of prior pdf allows to apply the proposed methodology even in cases when each expert deals with a single proposal (the case D3); the introduced prior information does not seem to influence the results in an adverse way.

Figure 5 Histograms of ranking results (see online version for colours)



8 Concluding remarks

The paper addresses the so-called ranking problem. The task arises every time, when relative ranking based upon peer-review results should be assigned. Relative ranking of student performance and evaluation of project proposals are the typical application examples. The last one served as prototype for the research. The efficiency of the proposed approach has been verified on the real-data application. The proposed method:

- respects the review procedure already used, no extra steps are required
- simplifies the evaluation process, as no agreement of the experts on marks of individual proposals is required
- provides an effective evaluation of the quality of experts: the suspicious results of this evaluation (performed *before* experts' discussion) should warn about expert's incompetency and/or conflict of interests.

The presentation does not cover particular criteria, but it is obvious that:

- the particular criteria characterising proposals can be evaluated in the exactly same way
- the evaluation of the particular criteria can make the judgement about thresholds-crossing more robust

- the discrete-valued marking with the discrete-valued ranking vector can be treated in the same way as the continuous-valued one: the model (8) offers the necessary basis for it.

The research has opened several topics to be studied. In the considered application area, the solution of the following problems will lead to more fair judging of proposals. In particular:

- a comparative study of the ranking of a proposal supported by the granting agency, and quality of the final results obtained by this project (this needs inclusion of the same marking, ideally by the same reviewers, into the evaluation of the completed project)
- study of the influence of the quantisation
- extended inspection of influence the scoring weights allocated to the particular criteria improving fairness of the judgement of the future project proposals.

The described procedure tries to meet needs of granting agencies. We however believe that it offers more.

- It is applicable in other domains that need a sort of ranking. For instance, methods solving the same, for instance diagnostic, problem are often compared in order to select the best of them.
- It provides a, probably novel, view on an important version of the hard negotiation problem appearing in multi-participant decision making.

Acknowledgements

This research was supported by: MŠMT 2C06001, AV ČR 1ET100750401 and MŠMT 1M0572. The authors also would like to acknowledge helpful comments of anonymous reviewers.

References

Arrow, K. (1995) *Social Choice and Individual Values*, 2nd. ed., Yale University Press, New Haven.

Berger, J. (1985) *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York.

Butler, J., Morrice, D. and Mullarkey, P. (2001) 'A multiple attribute utility theory approach to ranking and selection', *Management Science*, Vol. 47, pp.800–816.

Hallerbach, W. and Spronk, J. (2003) 'The relevance of MCDM for financial decisions', *Multi-Criteria Decision Analysis*, Vol. 11, pp.187–195.

Jacquet-Lagreze, E. and Siskos, Y. (1982) 'Assessing a set of additive utility functions for multicriteria decision making: the UTA method', *European Journal of Operational Research*, Vol. 10, No. 2, pp.151–164.

Kárný, M., Böhm, J., Guy, T. V., Jirsa, L., Nagy, I., Nedoma, P. and Tesař, L. (2005) *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer-Verlag, London.

Kim, S., Choi, S. and Kim, J. (1999) 'An iterative procedure for multiple attribute group decision making with incomplete information: Range-based approach', *European Journal of Operational Research*, Vol. 118, No. 1, pp.139–152.

Medaglia, A., Graves, S. and Ringuest, J. (2007) 'A multiobjective evolutionary approach for linearlky constrained project selection under uncertanty', *European Journal of Operational Research*, Vol. 179, pp.869–894.

Nurmi, H. (2001) 'Resolving group choice paradoxes using probabilistic and fuzzy concepts', *Group Decision and Negotiation*, Vol. 10, pp.177–198.

Peterka, V. (1981) 'Bayesian system identification', *Trends and Progress in System Identification*, Pergamon Press, Oxford, pp.239–304.

Prato, T. (1999) 'Multiple attribute decision analysis for ecosystem management', *Ecological Economics*, Vol. 30, pp.207–222.

Rios Insua, D., Holgado, J. and Moreno, R. (2003) 'Multicriteria e-negotiation systems for e-democracy', *Multi-Criteria Decision Analysis*, Vol. 12, Nos. 2–3, pp.213–218.

Rosen, S., Harmonosky, C. and Traband, M. (2007) 'A simulation optimisation method that considers uncertanty and multiple performance measures', *European Journal of Operational Research*, Vol. 178, pp.315–330.

Slevin, D., Boone, L., Russo, E. and Allen, R. (1998) 'CONFIDE: A collective decision-making procedure using confidence estimates of individual judgements', *Group Decision and Negotiation*, Vol. 7, No. 2, pp.179–194.