Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Convergence properties and practical estimation of the probability of rank reversal in pairwise comparisons for multi-criteria decision making problems

Georgia Dede^{a,b,*}, Thomas Kamalakis^{b,1}, Thomas Sphicopoulos^a

^a Department of Informatics and Telecommunications, University of Athens, Panepistimiopolis, Ilisia, GR15784 Athens, Greece ^b Department of Informatics and Telematics, Harokopio University, Eleftheriou Venizelou 70, Kallithea, GR17671 Athens, Greece

A R T I C L E I N F O

Article history: Received 30 April 2013 Accepted 21 August 2014 Available online 28 September 2014

Keywords: Uncertainty modeling Multiple criteria analysis Decision analysis Pairwise comparisons

ABSTRACT

In this paper, we address the impact of uncertainty introduced when the experts complete pairwise comparison matrices, in the context of multi-criteria decision making. We first discuss how uncertainty can be quantified and modeled and then show how the probability of rank reversal scales with the number of experts. We consider the impact of various aspects which may affect the estimation of probability of rank reversal in the context of pairwise comparisons, such as the uncertainty level, alternative preference scales and different weight estimation methods. We also consider the case where the comparisons are carried out in a fuzzy manner. It is shown that in most circumstances, augmenting the size of the expert group beyond 15 produces a small change in the probability of rank reversal. We next address the issue of how this probability can be estimated in practice, from information gathered simply from the comparison matrices of a single expert group. We propose and validate a scheme which yields an estimate for the probability of rank reversal and test the applicability of this scheme under various conditions. The framework discussed in the paper can allow decision makers to correctly choose the number of experts participating in a pairwise comparison and obtain an estimate of the credibility of the outcome.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Decision making (Bhushan & Rai, 2004; Yager, 2004) consists of choosing a specific course of action between several alternatives and is encountered in countless areas of human activity. In many circumstances, where complex decisions need to be made involving high stakes, it is desirable to proceed in a structured and methodological manner, rather than simply rely on the skills and intuition of a single decision maker. Multi-criteria decision analysis (MCDA) or multi-criteria decision making (MCDM) (Triantaphyllou, 2000) aim at facilitating decision makers in complicated situations where numerous and sometimes conflicting criteria or factors have to be taken into account.

MCDM is further classified into multi-objective decision making (MODM) and multi-attribute decision making (MADM) (Pohekar & Ramachandran, 2004). In MODM, a set of objective functions is

¹ Tel.: + 302109549286.

optimized subject to constraints and hence the efficient solutions in a set of alternatives are sought. MODM typically requires the solution of a series of mathematical programming models in order to reveal implicitly defined efficient solutions. On the other hand, in MADM, a small number of pre-determined alternatives are to be evaluated under a common set of criteria and the best alternative is usually selected by making comparisons between alternatives with respect to each criterion.

A fundamental problem in decision making is to grade the importance of a set of alternatives and assign a weight to each of them. The importance of alternatives usually depends on several criteria which can be evaluated within the decision making framework in which pairwise comparisons (PWC) are an essential ingredient (Saaty & Vargas, 2001). In the context of MADM, PWC enables the ranking of alternatives by allowing the experts to compare the various criteria or alternatives in pairs instead of assigning their priorities in a single step (Saaty, 1977). This reduces the influence of subjective point of views, associated with eliciting weights directly. PWC is usually performed in MADM methods such as the Analytic Hierarchy Process (AHP) (Saaty, 2003), the Weighted Product Method (WPM) (Chang & Yeh, 2001), the preference ranking organization method for enrichment evaluation (PROMETHEE) (Brans, Vincke, & Mareschal, 1986), the Analytic Network Process (ANP) (Saaty, 2004) and so on. The aim





CrossMark

^{*} Corresponding author at: National and Kapodistrian University of Athens, Department of Informatics and Telecommunications, Panepistimiopolis, Ilisia, GR15784 Athens, Greece. Tel.: + 302107275313.

E-mail addresses: gdede@hua.gr, gdede@di.uoa.gr (G. Dede), thkam@hua.gr (T. Ka-malakis), thomas@di.uoa.gr (T. Sphicopoulos).

of this paper is to consider the influence of uncertainty in PWC and how the credibility of the outcome can be assessed. Although the focus is on PWC alone, the results have some implications on the applications of MADM frameworks, particularly in the number of experts that are required and how the credibility of key parts of the framework can be ascertained.

In recent years, PWC has been used either as a stand-alone method or as part of complex MADM frameworks on several areas including government (Huanga, Chub, & Chiang, 2008), business (Lee & Kozar, 2006), industry (Chan, Lau & Ip, 2006), healthcare (Liberatore & Nydick, 2008), technology (Gerdsri & Kocaoglu, 2007), education (Zahedi, 1986), communications (Dede, Kamalakis, & Varoutas, 2011a, 2011b; Dede, Varoutas, Kamalakis, Fuentetaja, & Javaudin, 2010), agriculture (Abildtrup et al., 2006) and energy planning (Kok & Lootsma, 1985). The method itself has also been the focus of extensive research in the field of decision making. Recently in Fan and Liu (2010), a form of uncertain preference information, called ordinal interval numbers, was used in pairwise comparisons in order to rank the alternatives. In Doumpos and Zopounidis (2004), the issue of how pairwise comparisons can be used for the classification of alternatives in different classes of preference is discussed. The problem of deriving the weights from the pairwise comparison matrices using several alternative approaches is studied in Barzilai (1997) and Choo (2004). In Kwiesielewicz and Van Uden (2004) several issues concerning the inconsistency of the pairwise comparison matrix and its impact on the decision making process are highlighted. In addition, fuzzy pairwise comparison for solving the decision making problems has been proposed in Boenderb, Graan, and Lootsma (1989), Deng (1999) and Mikhailov (2005). In Marimin, Umano, Hatono, and Tamura (1998) linguistic labels are used in order to express fuzzy preference relations in pairwise decision problems. Moreover, Shiraishi, Tsuneshi, and Motomasa (1998) dealt with the properties of the principal eigenvector of PWC matrices.

The influence of uncertainty due to the imperfect and subjective expert judgments is of paramount importance when considering the credibility of the outcome of a decision making process. Several studies have attempted to shed some light on this issue in the context of PWC. For example, in Carmone, Karab, and Zanakis (1997), Monte Carlo simulations were performed to study the Incomplete Pairwise Comparisons (IPC) algorithm and investigate the effect of missing information in pairwise comparisons. Furthermore, in Aull-Hyde, Erdogan, and Duke (2006) it has been shown that given a sufficiently large group size, the consistency of the aggregate comparison matrix is guaranteed regardless of the measures used to estimate the consistency of the individual matrices, if the geometric mean method is used to estimate the weights. Moreover, in Hahn (2003) a stochastic characterization of the pairwise comparison judgments is provided, while statistical models for deriving the weights of the alternatives using Markov chain Monte Carlo are also presented. Furthermore, Farkas (2007) theoretically studied the conditions for rank reversal on perturbing the PWC matrices, while Chen and Kocaoglu (2008) also studied the rank reversal problem in this particular context, and came up with an algorithm to analyze the sensitivity of hierarchical decision models.

The main purpose of our work is to provide a suitable characterization of the impact of uncertainty in PWC. A first step in order to characterize the impact of uncertainty in PWCs, is to identify a suitable measure for quantifying its effects. Assume for instance that *N* different alternatives are pairwise compared by *M* experts, each with possibly a different view on the ranking of the alternatives. As discussed further below in Section 2.1, PWC aims at providing an average ranking, encompassing all these diverse opinions of the experts. It is of course natural to expect that the credibility of the overall process will be increased as the size of the expert group increases. Therefore, one possible way of measuring the trustworthiness of the results is to define the probability of rank reversal (P_{RR}) (Saaty & Vargas, 1984) as follows: Let W_1, \ldots, W_N be the weights calculated by the PWC in the case of a very large group of experts $(M \to \infty)$. In a practical situation where *M* is finite, uncertainty may undermine the PWC and the calculated weights w_k may turn out different than W_k . Uncertainty can be due to the difference of opinion among the experts or inconsistent pairwise comparisons. The probability of rank reversal is formally defined as:

$$P_{\rm RR} = P$$
 {the ranking obtained by w_i ,

$$1 \le i \le N$$
, is different than that of W_i (1)

A high P_{RR} implies that the outcome of the PWC in question is not trustworthy and could therefore lead to incorrect decision making. There are two important issues that need to be addressed concerning P_{RR} :

- a) How does P_{RR} relate to the number of experts *M*? An obvious way to reduce the effect of uncertainty is simply to increase *M*, but from a practical point of view, this is not a trivial task. It is usually difficult to locate many experts within a single organization or even in the wider public with sufficient expertise that would be willing to participate in the PWC surveys. On the other hand, there is no clear answer to the question of how many more experts need to participate in order to considerably reduce the uncertainty of the outcome. Say for example, that there are already M = 10 experts participating in the endeavor. How much is there to be gained in terms of P_{RR} by doubling or even quadrupling the size of the group of experts?
- b) How can P_{RR} be estimated from actual expert judgments in practice? When applying PWC in a specific decision making problem, one has access to the elements of a limited number of pairwise comparison matrices $\mathbf{P}^{(m)}$ (where $1 \le m \le M$). So the question becomes whether one can extract any kind of information regarding P_{RR} and hence the credibility of the results based on just the elements of $\mathbf{P}^{(m)}$.

The present paper attempts to deal with both points above. We first discuss a model for incorporating uncertainty in PWC and consider a suitable measure for quantifying the uncertainty level. We then discuss how P_{RR} varies with the group size M depending on the uncertainty level and extract several interesting conclusions from this variation. It is shown that there is not much sense in using more than M = 15 experts in the decision making process because the rate of decrease of P_{RR} is already small for M > 15. We then address the issue of how P_{RR} can be estimated from just the values of the pairwise comparison matrices $\mathbf{P}^{(m)}$ obtained by the experts. Given this information, we discuss a numerical method for estimating P_{RR} based on Monte Carlo simulation. The results indicate that for a sufficiently large group of experts, one can obtain a reasonable approximation to the actual value of P_{RR} .

The rest of the paper is organized as follows: In Section 2, we briefly summarize the PWC method laying the theoretical foundations for incorporating uncertainty and present the model used in our simulations. In Section 3, the results obtained when applying the proposed model are presented and the convergence of the P_{RR} is examined, considering the impact of various aspects which may affect the estimation of probability of rank reversal, such as varying uncertainty level among the experts, alternative preference scales and weight estimation methods. We also consider the case where the judgments are determined in a fuzzy manner. In Section 4 a numerical method is proposed in order to estimate the P_{RR} from the actual user judgments. Finally, some concluding remarks are presented in Section 5.

2. Uncertainty modeling in PWC

In this section, we introduce the model used for incorporating uncertainty in the pairwise comparison matrices. Section 2.1 provides a brief overview of the pairwise comparison framework, while Section 2.2 considers the effects of uncertainty-induced perturbations under general assumptions. Section 2.3 details the uncertainty model adopted in this paper. In Section 2.4 we describe how we can estimate the probability of rank reversal through Monte Carlo simulation.

2.1. The PWC framework

Within the PWC framework, instead of directly assigning priorities to the various alternatives A_i with $1 \le i \le N$, each expert compares all possible combinations of A_i and A_j . The outcome of these judgments for the mth expert are stored in a square $N \times N$ reciprocal matrix $\mathbf{P}^{(m)} = \begin{bmatrix} P_{ij}^{(m)} \end{bmatrix}$, which will henceforth be referred to as a pairwise comparison matrix. The value of the element $P_{ij}^{(m)}$ reflects the importance of alternative A_i over A_j . The expert needs to complete only the upper triangular elements (i < j) of $\mathbf{P}^{(m)}$ since by definition we have $P_{ij}^{(m)} = 1/P_{ii}^{(m)}$ and $P_{ii}^{(m)} = 1$ for a reciprocal matrix. The weights $w_i^{(m)}$ of alternative A_i according to expert *m* can be calculated with various ways. The most widely adopted approach is to solve the eigenvalue problem $\mathbf{P}^{(m)}\mathbf{x}_q^{(m)} = \lambda_q^{(m)}\mathbf{x}_q^{(m)}$, where $\lambda_q^{(m)}$ are the eigenvalues of $\mathbf{P}^{(m)}$ and $\mathbf{x}_q^{(m)} = \begin{bmatrix} x_{pq}^{(m)} \end{bmatrix}$ are the corresponding eigenvectors. Assuming that the eigenvalues are ordered so that $\lambda_1^{(m)}$ is the largest eigenvalue, then the weight of alternative A_i is estimated by normalizing the elements of the principle eigenvector $\mathbf{x}_1^{(m)}$ as follows (Saaty, 2003):

$$w_k^{(m)} = x_{1k}^{(m)} \left[\sum_{l=1}^N x_{1l}^{(m)} \right]^{-1}$$
(2)

In order to further simplify the comparisons, Saaty(2008) introduced the nine-level scale shown in Table 1.

One way of measuring the inconsistency of a pairwise comparison matrix is to calculate the Consistency Ratio (C.R.) defined as:

$$C.R. = \frac{C.I.}{R.I.}$$
(3)

In Eq. (3), C.I. is the Consistency Index which is estimated through:

$$C.I. = \frac{\lambda_1^{(m)} - N}{N - 1}$$
(4)

The denominator in Eq. (3) stands for the Random Index (R.I.) which is an average random consistency index derived from a sample of randomly generated reciprocal matrices with elements scaled according to Saaty (2003). Values of C.R. smaller or equal than 0.1 are generally considered as acceptable and in this case, the matrix is said to be nearly consistent (Saaty, 2003).

After all the comparisons have been completed, the average weight w_k for each alternative A_k is calculated by averaging out the weights $w_k^{(m)}$ obtained by the *M* experts,

$$w_k = \frac{1}{M} \sum_{m=1}^{M} w_k^{(m)}$$
(5)

Table 1

The	nine-	level	fundamental	comparison	scale.

$P_{ij}^{(m)}$	Explanations
1	A _i and A _i are equally important
3	A _i is slightly more important than A _i
5	A _i is strongly more important than A _i
7	A _i is very strongly more important than A _i
9	A _i is absolutely more important than A _i
2, 4, 6, 8	Intermediate values
Reciprocals of above	Used in analogous manner when A _j is more important
	than A _i

The weights w_k determine the priorities of the alternatives and hence the outcome of the PWC process. The method of estimating the weights from the PWC matrices is by no means unique. In the next sections we will also consider what happens when the weights are estimated using the geometric mean method. We will also discuss the situation where the pair wise judgments are carried out in a fuzzy manner and the weights are calculated accordingly.

2.2. First order perturbation analysis

Before proceeding to the description of the uncertainty model used in this paper, it is useful to consider the impact of uncertainty in PWC under very general assumptions. From a theoretical standpoint, we may consider that uncertainty acts as a kind of "noise" that impairs the pairwise comparisons. We may think of $\mathbf{P}^{(m)}$ as a perturbed version of an initial $N \times N$ matrix \mathbf{A} , i.e. $\mathbf{P}^{(m)} = \mathbf{A} + \Delta \mathbf{P}^{(m)}$ where $\Delta \mathbf{P}^{(m)} = [\Delta P_{ij}^{(m)}]$ is a random perturbation matrix, the magnitude of which is related to the level of uncertainty. Since the weights $w_k^{(m)}$ are obtained from the perturbed principal eigenvector $\mathbf{x}_1^{(m)}$ of $\mathbf{P}^{(m)}$, it is interesting to consider how $\mathbf{x}_1^{(m)}$ is related to the principal eigenvector \mathbf{x}_1 of \mathbf{A} . According to Saaty and Vargas (1987), the first order variation $\Delta \mathbf{x}_1^{(m)}$ of the principal eigenvector is determined by:

$$\Delta \mathbf{x}_{1}^{(m)} = \mathbf{x}_{1}^{(m)} - \mathbf{x}_{1} = \sum_{j=2}^{N} \left(\frac{\boldsymbol{\upsilon}_{j}^{\mathrm{T}} \Delta \mathbf{P}^{(m)} \mathbf{x}_{1}}{(l_{1} - l_{j}) \boldsymbol{\upsilon}_{j}^{\mathrm{T}} \mathbf{x}_{j}} \right) \mathbf{x}_{j}$$
(6)

where v_j and \mathbf{x}_j are the left and right eigenvectors of the initial pairwise comparison matrix **A** respectively while l_j are the corresponding eigenvalues. Expanding the matrix products of Eq. (6) in terms of the individual matrix and vector elements, it is revealed that the elements of the perturbation vector $\Delta \mathbf{x}_1^{(m)} = [\Delta \mathbf{x}_{1p}^{(m)}]$ are a linear combination of the elements of the perturbation elements $\Delta \mathbf{P}_{ij}^{(m)}$. This fact has an important bearing on the statistics of the perturbed weights: if one assumes that a sufficient number of the perturbations $\Delta \mathbf{P}_{ij}^{(m)}$ are statistically uncorrelated, then according to the Central Limit Theorem (CLT) (Rice, 1995), $\Delta \mathbf{x}_{1p}^{(m)}$ being a linear combination of statistically uncorrelated variables, will approximately follow a Gaussian distribution (provided that *N* is sufficiently large). The weights $w_k^{(m)}$ are normalized versions of $\mathbf{x}_{1k}^{(m)}$ and we therefore expect that they will also approximately follow a Gaussian distribution. This can be deduced by considering the change $\delta w_k^{(m)}$ in the weights $w_k^{(m)}$ inferred by the perturbations $\Delta \mathbf{x}_{1p}^{(m)}$ in the principle eigenvector. Using the fact that ($\alpha + \Delta \alpha$)/($\beta + \Delta \beta$) $\cong \alpha/\beta + \Delta \alpha/\beta - (\alpha/\beta^2)\Delta\beta$ for small $\Delta \alpha$ and $\Delta \beta$, one obtains from Eq. (2) an estimate for $\delta w_k^{(m)} = w_k^{(m)} - W_k$ as follows:

$$\delta w_k^{(m)} \cong W_k \left\{ \frac{\Delta x_{1k}^{(m)}}{x_{1k}} - \frac{1}{C} \sum_{j=1}^N \Delta x_{1j}^{(m)} \right\}$$
(7)

where $C = \sum_l x_{1l}$. The above equation suggests that since, as mentioned above, $\Delta x_{1p}^{(m)}$ is a linear combination of $\Delta P_{ij}^{(m)}$, the perturbations $\delta w_k^{(m)}$ will also be Gaussian. More importantly, the statistics of the uncertainty perturbations $\Delta P_{ij}^{(m)}$ are expected to mainly influence the variance of the Gaussian-like distributed weights $w_k^{(m)}$ and have smaller impact on their actual statistical behavior (e.g. their PDF). Furthermore the weights w_k in Eq. (5) are the sum of $w_k^{(m)}$ with respect to *m* and the CLT again points out that w_k will also be approximately Gaussian-distributed. We will numerically verify this claim in Section 3.1.

2.3. Description of the uncertainty model

The discussion of Section 2.2 already points out an intuitive way of incorporating uncertainty in the PWC matrices. Instead of assigning

a fixed value for the elements $P_{ij}^{(m)}$ we may assume that $P_{ij}^{(m)}$ varies randomly inside an interval $[D_{ij}, U_{ij}]$ (Saaty & Vargas, 1987). In addition, if the experts carry out their comparisons according to Table 1, then the allowed values for $P_{ij}^{(m)}$ should be discrete. A word of caution is in order here however: the length $U_{ij} - D_{ij}$ of the interval $[D_{ij}, U_{ij}]$ is not a suitable measure for the level of uncertainty. To illustrate this, consider the intervals $I_1 = [1/9, 1]$ and $I_2 = [1, 2]$ with lengths 1 - 1/9 = 8/9 and 2 - 1 = 1 respectively. Although I_1 has a smaller length than I_2 , it is easy to see that it actually corresponds to higher uncertainty level than I_2 since there are nine distinct values of Table 1 (i.e. $1/9, 1/8, \ldots, 1$) that are located inside I_1 while there are only two such values for I_2 . This points out that in order to quantify the uncertainty level it is preferable to consider the number n_{ij} of discrete values contained in the interval rather than its length. In fact, since an interval with $D_{ij} = U_{ij}$ corresponds to no uncertainty a more suitable measure of the uncertainty level is $n_{ij} - 1$.

Based on the above observation we consider the following mapping of the values *V* of Table 1, into a set of integers u = q(V) with $-8 \le u \le 8$ such that:

$$q(V) = \begin{cases} V - 1, & V \ge 1\\ 1 - 1/V, & V < 1 \end{cases}$$
(8)

It is easy to show that the quantity $q(U_{ij}) - q(D_{ij})$ is equal to $n_{ij} - 1$. Also, for the intervals I_1 and I_2 considered in the paragraph above, this quantity is 9 and 1 respectively which indicates that indeed I_1 corresponds to a much larger uncertainty level than I_2 . We also define the inverse mapping q^{-1} of q such that $q^{-1}(q(V)) = V$.

Another interesting question is how the boundaries D_{ij} and U_{ij} are determined in the first place. It is evident that since D_{ij} and U_{ij} can take a variety of values one can think of them as random variables themselves. One way of randomly selecting D_{ij} and U_{ij} is to start with an initial nearly consistent matrix $\mathbf{A} = [A_{ij}]$ and set

$$u_{ij} = \begin{cases} q(A_{ij}) + |\Delta u_{ij}|, & \text{if } q(A_{ij}) + |\Delta u_{ij}| \le 8\\ 8, & \text{otherwise} \end{cases}$$
(9)

$$d_{ij} = \begin{cases} q(A_{ij}) - |\Delta d_{ij}|, & \text{if } q(A_{ij}) - |\Delta d_{ij}| \ge -8\\ -8, & \text{otherwise} \end{cases}$$
(10)

where Δu_{ij} and Δd_{ij} are independent, uniformly distributed variables inside [-s/2, s/2] with variance σ^2 determined by

$$\sigma^2 = \frac{s^2}{12} = \left\langle \Delta u_{ij}^2 \right\rangle = \left\langle \Delta d_{ij}^2 \right\rangle \tag{11}$$

Eqs. (9) and (10) ensure that the integers u_{ij} and d_{ij} remain within the interval [-8, 8] and hence $q^{-1}(u_{ij})$ and $q^{-1}(d_{ij})$ correspond to values contained in Table 1. We can therefore choose the uncertainty interval boundaries according to:

$$D_{ij} = q^{-1}(d_{ij}) \tag{12}$$

$$U_{ij} = q^{-1}(u_{ij}) \tag{13}$$

Given samples of the randomly generated pairwise comparison matrices $\mathbf{P}^{(m)}$, it is interesting to relate the strength of the perturbations *s* to the expected uncertainty level $L = \langle q(U_{ij}) - q(D_{ij}) \rangle = \langle n_{ij} \rangle - 1$ shown in Fig. 1. Note that the mapping in Eq. (9) and (10) forces u_{ij} and d_{ij} to remain bounded inside [-8, 8] even if $|\Delta u_{ij}|$ is large. This implies that *L* is determined by both *s* and the value of $q(A_{ij})$. Given the value of *s*, the corresponding value of *L* is estimated by Monte Carlo simulations. The results are plotted in Fig. 1 which provides a means of translating the value of *L* to the perturbation strength *s* and will be used in the realization of the uncertainty model described below.

2.4. Estimation of P_{RR} through Monte Carlo simulation

Taking into account the above discussion, we now proceed to determine how the probability P_{RR} of rank reversal is related to the



Fig. 1. *L* as a function of the perturbation strength *s*.

uncertainty level L and expert group size M. Our approach is based on Monte Carlo (MC) simulations. In each iteration k, we generate a matrix $\mathbf{A}^{(k)}$, the elements of which are randomly chosen from the values of Table 1. If $\mathbf{A}^{(k)}$ is not nearly consistent (i.e. C.R. > 0.1 according to Section 2.1) then it is discarded and we recalculate $\mathbf{A}^{(k)}$ until this criterion is met. Given the value of L and the value of each element $A_{ij}^{(k)}$ of $\mathbf{A}^{(k)}$, we use Eqs. (9) and (10) to estimate $u_{ij}^{(k)}$ and $d_{ij}^{(k)}$ by generating the random perturbations $\Delta d_{ij}^{(k)}$ and $\Delta u_{ij}^{(k)}$. The strength of the perturbations *s* is chosen based on the values of Fig. 1. Once $D_{ij}^{(k)} = q(d_{ij}^{(k)})$ and $U_{ij}^{(k)} = q(u_{ij}^{(k)})$ are determined, we generate a large number of PWC matrices (typically $N_{\text{matrix}} \cong 10^4$) and distribute them in $N_{\rm G} = \lfloor N_{\rm matrix}/M \rfloor$ groups of *M* nearly consistent matrices ${\bf P}^{(\nu)}$ with $1 \le \nu \le N_{\text{matrix}}$ and where $\lfloor x \rfloor$ stands for the largest integer which is smaller than x. In the context of Section 2.2, the matrices $\mathbf{P}^{(\nu)}$ can be assumed as a perturbed version of $\mathbf{A}^{(k)}$ and the perturbation is simply $\Delta \mathbf{P}^{(\nu)} = \mathbf{P}^{(\nu)} - \mathbf{A}^{(k)}$. For each of these $N_{\rm G}$ expert groups we calculate the priorities of the alternatives as discussed in Section 2.1. We then estimate the number F_G of groups for which the order of priorities calculated is different than the order of priorities W'_{μ} calculated by averaging out the weights obtained by all the matrices $\mathbf{P}^{(v)}$. The latter priorities corresponds to the order obtained by a very large ($M = N_{matrix}$) group of experts and since N_{matrix} is very large, one expects that W'_k are approximately equal to the ideal weights W_k corresponding to the limit of infinite group of experts $(M \rightarrow \infty)$. The probability of rank reversal for iteration k is then approximated as $P_{\text{RR}}(\mathbf{A}^{(k)}) \cong F_G/N_G$. In the next iteration k + 1 we estimate a different initial matrix $\mathbf{A}^{(k+1)}$ and estimate the probability of rank reversal $P_{\text{RR}}(\mathbf{A}^{(k+1)})$. The average probability of rank reversal (independent of the choice of the initial matrix) is then calculated as

$$P_{\rm RR} = \frac{1}{N_{\rm MC}} \sum_{k=1}^{N_{\rm MC}} P_{\rm RR}(\mathbf{A}^{(k)}) \tag{14}$$

where N_{MC} is the total number of MC iterations. It is interesting to note that one can estimate P_{RR} for many different values of *M* using the same random samples of the pairwise comparison matrices.

3. Numerical analysis and results

In this section, we present the results obtained when we apply the framework described above in Section 2. Section 3.1 presents some numerical results that further validate the first order analysis of Section 2.2, while Section 3.2 examines the convergence properties of P_{RR} .

3.1. Validation of first order analysis

To validate our first order analysis, as presented in Section 2.2, we numerically calculate the PDF of a representative average weight w_k



Fig. 2. PDF of the weight w_k obtained by averaging the $w_k^{(m)}$, which are estimated through simulation. Also plotted is the Gaussian least squares fit. The uncertainty level is L = 3, the number of experts is M = 10 and the number of criteria is (a) N = 4, (b) N = 6.



Fig. 3. PDF of the weight w_k obtained in the case where M = 40. The rest of the parameters are the same as in Fig. 2.

obtained by a group of experts of size M. The PWC matrices of all experts are calculated as discussed in Sections 2.3 and 2.4 starting from the same nearly consistent $N \times N$ initial matrix **A** with C.R. < 0.1. The results are shown in Figs. 2 and 3, for the cases of M = 10 and M = 40 experts, respectively, both for N = 4 and N = 6 criteria. 10^{6} Monte Carlo (MC) iterations are used and we assume the case where the uncertainty level L is equal to 3. The figure shows the PDFs obtained using either the full weight calculation method outlined in Section 2.1 and the first order approximation given by Eq. (6). Also shown is the Gaussian fit to the former PDF (the Gaussian fit to the latter case is practically the same). The figures illustrate that Eq. (6) is indeed a valid approximation for the individual and hence the average weight perturbations. In addition, the average weights follow a normal distribution in accordance to what has been discussed in the previous paragraph. Similar results are obtained for the rest of the weights as well as for other values of N, L and M. Fig. 3 shows the PDF obtained for M = 40 starting from another initial matrix **A**.

The above discussion in Section 2.2 and the results of this section highlight the fact that to a first order, other than determining the mean value and standard deviation of the priorities w_k , the statistics of $\Delta P_{ij}^{(m)}$ should play a secondary role in the rest of the statistical properties of w_k (e.g. the shape of their PDF).

3.2. Convergence properties of the probability of rank reversal

3.2.1. Implementation of the P_{RR} uncertainty model

Applying the uncertainty model discussed in Sections 2.3 and the Monte Carlo simulation procedure in Section 3.1, we can calculate the probability of rank reversal P_{RR} , in order to estimate its dependence on the expert group size M which is a key issue from a practical



Fig 4. (a) P_{RR} and (b) ΔP_{RR} as a function of the expert group size for N = 3.

point-of-view, as discussed in the introduction. We first examine the convergence properties of $P_{\rm RR}$ with respect to M, in the case of a relative small number of alternatives (N = 3), taking into account a variety of uncertainty level values L. For each value of M, we performed $N_{\rm MC} = 10^3$ iterations and on each iteration we generated $N_{\rm matrix} = 10^4$ random matrices. As indicated in Fig. 4(a), P_{RR} is already low (below 7 percent) even for a group size as small as M = 10 for reasonable values of L. Another important conclusion drawn by the figure is that the convergence speed of P_{RR} is rather slow. For example, for L = 7, $P_{\rm RR}$ changes from 6.8 percent to 3.5 percent when the number of experts is increased from 10 to 20, implying a small practical gain from increasing the expert group size, since P_{RR} is already low for M = 10. The gain is even lower for smaller uncertainty levels. The figure also illustrates that there is not much sense in increasing the number of experts beyond M = 15 since the reduction in P_{RR} is even smaller.



Fig. 5. PDF of the values of P_{RR} obtained from multiple simulations assuming N = 3, M = 10 and L = 7. We also plotted the Gaussian least squares fit.

In order to examine the reliability of the estimated P_{RR} values we can calculate the confidence intervals. We first start with the statistics of the values P_r for the probability of rank reversal P_{RR} obtained by running multiple Monte Carlo iterations for a single value of M. If all these values are similar, we can deduce that our results are credible. Fig. 5 plots the PDF of the calculated P_{RR} values obtained by their histogram over $N_{MC} = 10^3$ simulations for N = 3, M = 10 and L = 7. Since the value of P_{RR} in each simulation is calculated as the average over many independent realizations, the Gaussian PDF fits very well with the estimated PDF. We also notice that there are small variations in the calculated $P_{\rm RR}$ values (the PDF is practically non-zero from 0.067 to 0.070). Since the statistics are Gaussian, we may also readily calculate the confidence interval I_C for a 95 percent confidence level. Using standard confidence interval formulas (Hanke & Wichern, 2008) we readily find that $I_{C} = [0.067, 0.068]$ and $I_{C} = [0.0179, 0.0180]$ in the case where L = 7, N = 3, M = 10 and L = 1, N = 3, M = 10respectively. These tight confidence intervals imply that the number of iterations considered in the calculations of Figure 4 is adequate in order to accurately estimate the probability of rank reversal. Similar results are obtained when estimating the confidence intervals of other points in the graphs. For example for M = 40, L = 7, N = 3 we obtain $I_{\rm C} = [0.0099, 0.01].$

To further understand the convergence of P_{RR} with respect to M, we calculate the differential decrease $\Delta P_{\text{RR}}(M)$ in the probability of rank reversal when the group size is increased by $\Delta M = 1$ expert:

$$\frac{\Delta P_{\rm RR}(M)}{\Delta M} = |P_{\rm RR}(M+1) - P_{\rm RR}(M)| \tag{15}$$

In Fig. 4(b) we have plotted ΔP_{RR} as a function of *M*. The figure indicates that even for L = 7, the changes in the probability of rank reversal when an additional expert is added in the group are very small (below 1 percent per expert) at M = 15. It therefore makes little sense of increasing the expert group size beyond that. This is also shown in Fig. 4(a) where increasing *M* from 15 to 40 has very little impact from a practical point of view, since the additional number of experts only marginally reduces the uncertainty of the outcome.

Fig. 6(a) shows the values of P_{RR} for the case where N = 6 alternatives are pairwise compared, assuming the same uncertainty levels as in Fig. 4. As expected, since the number of alternatives is now larger, one obtains higher values for P_{RR} than in (may even exceed 20 percent in some cases). Indicatively, for L = 7 the confidence intervals are $I_C = [0.22, 0.23]$ and $I_C = [0.0245, 0.0255]$ for M = 10 and M = 40 experts, respectively. Fig. 6(b) shows the corresponding values of ΔP_{RR} and it is deduced that varying the size of the expert group is now more beneficial. For M = 10, P_{RR} is reduced to about 18 percent when the group size is increased by 1 in the case of L = 7. However, for M > 15, the gain becomes progressively smaller as a result of the slow convergence of P_{RR} . Given the practical restrictions of increasing the



Fig. 6. (a) P_{RR} and (b) ΔP_{RR} as a function of the expert group size for N = 6.



Fig. 7. P_{RR} as a function of the expert group size for L = 3.

group size, the figure suggests that using more than 15 experts may not have a strong bearing in the uncertainty of the results.

Fig. 7 illustrates the convergence of PWC for different values of criteria *N* and a constant *L* = 3. We consider the cases where *N* = 2, N = 4 and N = 6. The model easily allows to consider cases where N > 6, but in actual situations, performing PWC with that many alternatives becomes impractical since the number of pairwise comparisons N_c for each expert is given by $N_c = N(N - 1)/2$ and therefore scales as $O(N^2)$. As shown in the figure, P_{RR} does depend on *N*: For a group size of ten experts (M = 10), $P_{RR} \cong 6$ percent when the experts are asked to make pairwise comparisons on six alternatives and this drops slightly to 3.5 percent when the number of criteria is 6. Other than that, the convergence behavior of P_{RR} does not seem dependent on the number of criteria. In all cases, there is no significant gain in increasing the group size beyond 15. This is due to the slow convergence properties of PWC, which seems to hold regardless the number of alternatives used and the uncertainty level.

Up to this point, we estimated P_{RR} taking the same uncertainty level associated with each expert in the group. It is also interesting to carry out this estimation when a different uncertainty level $L^{(m)}$ is assumed for each expert *m*. From a practical point-of-view, this reflects what could happen when the level of expertise varies from participant to participant. Fig. 8(a) shows the values of P_{RR} , as a



Fig. 8. (a) P_{RR} and (b) ΔP_{RR} as a function of the expert group size for different uncertainty levels for each expert.

function of the expert group in the case where the uncertainty level $L^{(m)}$ is randomly selected with equal probability inside the integer interval [1, 7]. The results are shown for N = 3 and N = 6. One obtains similar convergence behavior to the previous case. Similar results were also obtained if we calculate the differential decrease $\Delta P_{\text{RR}}(M)$ as shown in Fig. 8(b). The similar convergence behavior observed in the case of varying uncertainty level can be explained taking into account the theoretical framework of Section 2.2, where we have shown the statistical properties of the uncertainty perturbations other than their mean value and standard deviation, are not expected to play that big of a role in the statistics of the weight and hence of the P_{RR} . We therefore expect a change in the values of P_{RR} but its convergence behavior with respect to M should be more or less the same.

3.2.2. Alternative preference scales

The linear preference scale consisting of integers from one to nine and their reciprocals shown in Table 1 is used quite often in applications of the PWC method. Saaty (1991) advocates this scale as the best one to represent weight ratios. Nevertheless, it is interesting to investigate the convergence properties of P_{RR} considering alternative preference scales such as the inverse linear or the logarithmic scale (Ishizaka & Labib, 2011). Instead of the elements in Table 1, in the former scale we use the elements 9/(10 - x) where $x \in \{1, 2, \dots, 9\}$ as well as their inverses (10 - x)/9. The alternative scale $\{1/9, 2/9, \ldots, n\}$ $1, \ldots, 9/2, 9$ can therefore be indexed in the same way as the scale of Table 1 but with a different mapping q(V). The matrices can again be chosen by randomly selecting the indices as before and applying the alternative mapping. The elements of the logarithmic scale are determined by $\log_{\alpha}(x + (\alpha - 1))$ where $x \in \{1, 2, \dots, 9\}$ and $\alpha > 1$ and their inverses and can be again indexed in a similar way. In any case, we may still use *L* as a measure of uncertainty.

Fig. 9(a) and (b) illustrates ΔP_{RR} for the inverse linear and logarithmic scales, respectively in the case N = 6. For the logarithmic scale we assumed a = 2 (Ishizaka, Balkenborg, & Kaplan, 2011). As shown in Fig. 9, the convergence of ΔP_{RR} is again slow, implying that the convergence of P_{RR} is similar even when an alternative preference scale is used by the experts. From Fig. 9(a) and (b) one may deduce that ΔP_{RR} varies in a similar manner as before, again suggesting that there is no significant gain in increasing the group size beyond 15.

3.2.3. Weight estimation through the geometric mean method

Another important issue to consider is whether the results presented above depend on the weight estimation method. Instead of estimating the weights $w_k^{(m)}$ through the eigenvalue method (EM) discussed in Section 2.1, one could also apply the Geometric Mean Method (GMM) or equivalently the Logarithmic Least Square Method (LLSM) (Saaty, 1990). This method attempts to fit the elements of the pairwise comparison matrix $\mathbf{P}^{(m)}$ with a perfectly consistent matrix, i.e. a matrix whose elements are of the form $w_i^{(m)}/w_i^{(m)}$. The weights

Fig. 9. ΔP_{RR} as a function of the expert group size for different *L* for (a) inverse linear and (b) logarithmic scales.

Fig. 10. (a) P_{RR} and (b) ΔP_{RR} as a function of the expert group size for different *L* for N = 6, for the case of GMM.

are therefore determined by minimizing the following metric through unconstrained nonlinear optimization:

$$Q_{\text{GMM}} = \sum_{i,j=1}^{N} \left(\log P_{ij}^{(m)} - \log \left| \frac{q_i^{(m)}}{q_j^{(m)}} \right| \right)^2$$
(16)

After the minimization is performed, we choose $w_i^{(m)} = |q_i^{(m)}| / \sum_j |q_j^{(m)}|$ to ensure that they are positive and sum up to one. In our implementation we use MATLAB's fminsearch function which is based on the simplex search method (Lagarias, Reeds, Wright, & Wright, 1998).

Fig. 10(a) and (b) shows the values of P_{RR} and ΔP_{RR} for the case of GMM, where N = 6 criteria are pairwise compared assuming several uncertainty levels *L*. It is shown that for both low and high *L*, one obtains a similar convergence behavior for P_{RR} and ΔP_{RR} as in the EM case. Comparing Fig. 10(a) and (b) with Fig. 6(a) and (b) respectively one also deduces that the actual values of ΔP_{RR} and P_{RR} are approximately the same implying that both weight estimation achieve approximately the same credibility.

3.2.4. Impact of fuzzy judgments

In this section we investigate the case where the experts carry out their comparisons in a fuzzy manner (Deng, 1999), i.e. by assigning fuzzy numbers in the elements of the pairwise comparison matrix (Chao Chung, 2011). A fuzzy number \tilde{b} does not refer to one single value but rather to a connected set of possible values *x* centered on *b*, each with its own weight $u_b(x)$ between 0 and 1. This weight distribution is called a membership function. In the special case where the membership function is triangular, then

$$u_{b}(x) = \begin{cases} (x-l)/(b-l), & l \le x \le b \\ (u-x)/(u-b), & b \le x \le u \\ 0, & \text{otherwise} \end{cases}$$
(17)

where [l, u] is the range of values of x, in which the weight $u_b(x)$ is nonzero. The triangular membership function in Eq. (17) is determined by the triplet (l, b, u) which consequently completely determines the fuzzy number \tilde{b} .

To carry out fuzzy PWC, we need to define a fuzzy preference scale where fuzzy numbers are used instead of precise numbers as in a conventional preference scale. The pairwise comparison elements will therefore consist of fuzzy rather than precise numbers. The scale adopted can be a fuzzy extension of a conventional scale such as the one in Table 1. The fuzzy scale adopted in this work consists of the fuzzy numbers \tilde{b} where b is an integer $1 \le b \le 9$ and where $\tilde{b} = (b - 1, b, b + 1)$ if $2 \le b \le 8$, $\tilde{1} = (1, 1, 1)$ and $\tilde{9} = (9, 9, 9)$. The scale also contains the reciprocal elements of \tilde{b} as defined in Chao Chung (2011).

The procedure for the estimation of the fuzzy weights is based on extension principle theory (Gerla & Scarpati, 1998). Once the pairwise comparison elements $(l_{ij}^{(m)}, b_{ij}^{(m)}, u_{ij}^{(m)})$ are determined, we apply a constrained minimization procedure to estimate the fuzzy weights

 $(r_i^{(m)}, w_i^{(m)}, e_i^{(m)})$ for each expert. The function to be minimized is written as:

$$J = \sum_{i=1}^{N} \sum_{\substack{j=1\\i\neq j}}^{N} \left\{ \left[\ln \frac{r_i^{(m)}}{l_{ij}^{(m)} e_j^{(m)}} \right]^2 + \left[\ln \frac{w_i^{(m)}}{w_j^{(m)} b_{ij}^{(m)}} \right]^2 + \left[\ln \frac{e_i^{(m)}}{u_{ij}^{(m)} r_j^{(m)}} \right]^2 \right\}$$
(18)

The minimization constraints are outlined in (Wang, Elhag, & Hua, 2006). In our implementation we use MATLAB's fmincon function which is based on the interior point algorithm (Waltz, Morales, Nocedal, & Orban, 2006). To calculate the average fuzzy weights $\tilde{w}_i = (r_i, w_i, e_i)$, we simply average the points of the individual membership functions, i.e. $r_i = M^{-1} \sum_m r_i^{(m)}$, $w_i = M^{-1} \sum_m w_i^{(m)}$ and $e_i = M^{-1} \sum_m e_i^{(m)}$.

In order to incorporate uncertainty in this fuzzy extension of PWC, we consider the following approach where the pairwise comparison matrices of the experts are obtained again by randomly perturbing an initial consistent pairwise comparison matrix $\mathbf{A}^{(k)}$ in order to obtain the matrices $\mathbf{P}^{(v)}$ as discussed in Section 3.1. Given $\mathbf{P}^{(v)}$ we then consider its fuzzy extension $\mathbf{\tilde{P}}^{(v)}$ with elements $\mathbf{\tilde{P}}^{(v)}_{ij} = (l^{(v)}_{ij}, b^{(v)}_{ij}, u^{(v)}_{ij})$ where $b^{(v)}_{ij} = P^{(v)}_{ij}$ and $I^{(v)}_{ij}$ are determined according to the fuzzy scale as discussed above. From this point forward the simulation procedure is the same as before except that the weights are calculated by the minimization of Eq. (18) instead of the eigenvalue method. In the context of fuzzy judgments, we assume that rank reversal occurs whenever there exist at least two indices *i*, *j* for which the ordering of the center or the upper edge of the corresponding membership function changes compared to the one obtained by a large group of experts.

In Fig. 11(a) and (b) we show ΔP_{RR} in the case of fuzzy judgments as a function of the expert group size, for N = 3, N = 6 criteria respectively. The figures illustrate that even in the case of fuzzy judgments,

Fig. 11. ΔP_{RR} for fuzzy judgments as a function of the expert group size for different *L* for (a) *N* = 3 and (b) *N* = 6.

the convergence of P_{RR} is again slow. For N = 6, ΔP_{RR} is smaller than 0.01 for M > 15 even for large uncertainty levels indicating not much practical gain in further increasing the group size.

4. Estimation of probability of rank reversal from actual user data

The previous analysis dealt with the impact of uncertainty in the evaluation of the priorities of the alternatives, quantified in terms of P_{RR} . We now consider the issue of estimating P_{RR} from actual user data, i.e. a relatively small number of pairwise comparison matrices $\mathbf{P}^{(m)}$ obtained by a single expert group. In this section, we propose a numerical method to achieve this and validate our approach with numerical simulations.

4.1. Numerically evaluating P_{RR}

Our scheme for evaluating the probability of rank reversal is described in the dash line frame of Fig. 12. We first extract some information on the statistical properties of the perturbations (step 2 in Fig. 12). One approach is to calculate the maximum and minimum values of $P_{ij}^{(m)}$ with respect to *m*:

$$U'_{ij} = \max\left\{P^{(m)}_{ij}\right\}, \quad D'_{ij} = \min\left\{P^{(m)}_{ij}\right\}$$
(19)

Another approach would be to map the elements $P_{ij}^{(m)}$ to their corresponding integers according to Eq. (8) and then estimate the mean value μ_{ij} and the standard deviation σ_{ij} of $q(P_{ij}^{(m)})$ with respect to *m*, i.e.

$$u_{ij} = \frac{1}{M} \sum_{m=1}^{M} q(P_{ij}^{(m)})$$
(20)

$$\sigma_{ij}^2 = \frac{1}{M} \sum_{m=1}^{M} \left[q(P_{ij}^{(m)}) - \mu_{ij} \right]^2$$
(21)

Assuming uniform statistics, Eqs.(19) and (20)–(21) are two alternative methods for determining the statistics of the perturbations. Once these statistics are identified, one can perform Monte Carlo simulations (step 3 in Fig. 12) to determine the probability P_{RR} . In the first case, one generates a large number N_{PWC} of random pairwise comparison matrices $\mathbf{R}^{(1)} \ 1 \le l \le N_{\text{PWC}}$, the elements $R_{ij}^{(l)}$ of which are uniformly chosen from $[D'_{ij}, U'_{ij}]$. Note that these matrices do not correspond to pairwise comparison matrices filled out by any actual experts. Rather, they are random matrices numerically generated based on the information extracted by the limited set of user matrices $\mathbf{R}^{(l)}$ assuming that $R_{ij}^{(l)}$ follow a uniform distribution which is determined by the mean value μ_{ij} and standard deviation σ_{ij} . In any

Fig. 12. Illustration of the simulations carried out in Section 4. The dotted frame contains the proposed method for estimating the probability of rank reversal from a limited set of pairwise comparison matrices. The rest of the figure explains how the credibility of the proposed estimation scheme can be verified by numerical simulations.

case the random matrices $\mathbf{R}^{(1)}$ are distributed into $\lfloor N_{PWC}/M \rfloor$ groups containing M matrices each. The probability of rank reversal can be approximated by calculating the number of times for which the priorities obtained by each of these groups is not the same as that obtained by the initial weights w_k (step 4 in Fig. 12). In order to distinguish between the two estimates we designate as $P_{RR}^{(minmax)}(\mathbf{A}^k)$ and $P_{RR}^{(std)}(\mathbf{A}^k)$ the probabilities of rank reversal obtained by Eqs. (19) and (20)–(21) respectively given the initial matrix \mathbf{A}^k . One should note that if the initial group size from which the data are gathered is infinite ($M \rightarrow \infty$), the two alternative methods for calculating $\mathbf{R}^{(1)}$ are equivalent, hence $P_{RR}^{(minmax)}(\mathbf{A}^k) = P_{RR}^{(std)}(\mathbf{A}^k)$. However, in the case of finite M they may provide a different estimate for P_{RR} .

4.2. Results and validation

In order to validate the P_{RR} estimation procedure we again resort to Monte Carlo simulations as suggested by the rest of Fig. 12. In each Monte Carlo iteration k, we calculate an initial matrix \mathbf{A}^k and estimate the intervals $[D_{ij}, U_{ij}]$ as before (Section 3.1). We then generate the matrices $\mathbf{P}^{(m)}$ where $1 \le m \le N_{\text{matrix}}$. We choose the first *M* matrices $\mathbf{P}^{(m)}$ where $1 \le m \le M$ to correspond to the user matrices, while the rest of the matrices are to estimate the actual probability of rank reversal $P_{RR}(\mathbf{A}^k)$ which is obtained in the fashion explained in Section 3.1. In accordance to what was discussed in Section 4.1, the matrices $\mathbf{P}^{(m)}$ where $1 \le m \le M$ are used to estimate either D'_{ii} and U'_{ii} according to Eq. (19) or μ_{ij} and σ_{ij} according to Eqs. (20) and (21) respectively. Afterward we generate the matrices $\mathbf{R}^{(l)}$ and use them to estimate $P_{RR}^{(minmax)}(\mathbf{A}^k)$ and $P_{RR}^{(std)}(\mathbf{A}^k)$ (steps 1 to 4 in the dotted portion of Fig. 12 explained in Section 4.1). To measure the error introduced by the aforementioned estimation methods, we calculate the error function defined by

$$e^{(u)}(\mathbf{A}^k) = \left| P_{\mathrm{RR}}^{(u)}(\mathbf{A}^k) - P_{\mathrm{RR}}(\mathbf{A}^k) \right|$$
(22)

where *u* stands for either "minmax" or "std". Averaging over many possible of the initial matrices obtained in the different Monte Carlo iterations, we can therefore obtain an estimate for the average error:

$$e^{(u)} = \frac{1}{N_{\rm MC}} \sum_{l=1}^{N_{\rm MC}} e^{(u)}(\mathbf{A}^k)$$
(23)

Fig. 13(a) and (b) shows the values of $e^{(\text{std})}$ and $e^{(\text{minmax})}$ obtained by the MC simulations as well as the probability of rank reversal P_{RR} obtained by the procedure outlined in Section 3, for L = 5 and (a) N = 3and (b) N = 6. We see that the average error depends on the method used for the estimation of the perturbation statistics. The values of $e^{(\text{minmax})}$ are higher than $e^{(\text{std})}$ demonstrating that determining the perturbation statistics through the mean value and standard deviation is much better. This is in accordance with what was discussed in Section 2.2 where the CLT was used to discuss the perturbation

Fig. 13. P_{RR} and $e^{(u)}$ as function of the expert group size for L = 5 and (a) N = 3, (b) N = 6.

Fig. 14. P_{RR} and $e^{(u)}$ as function of the expert group size for L = 7 and (a) N = 3, (b) N = 6.

statistics. The statistical properties of the perturbed weights resemble those of a Gaussian distribution. In addition, according to the CLT, it is the mean value and the standard deviation that primarily determine this distribution rather than other measures like D'_{ii} and U'_{ii} used in the "minmax" method. We see that for N = 3, the "std" method results in an average error of $e^{(\text{std})} = 2$ percent for M = 10which corresponds to less than half the value of P_{RR} implying that the method correctly estimates the order of magnitude of $P_{\rm RR}$ which is important in practical applications of the method. Similar observations are drawn in the case where N = 6, shown in Fig. 13(b) indicating that the method provides reasonable accuracy for increasing number of alternatives. On the other hand, the "minmax" method produces larger errors in the estimation of P_{RR} and especially in the case where N = 6 produces errors which are larger than the value of P_{RR} itself. It is also interesting to investigate the accuracy of both methods under an even greater uncertainty level. In Fig. 14(a) and (b) we show the values of $e^{(\text{std})}$ and $e^{(\text{minmax})}$ in the case where L = 7 again for N = 3 and N = 6 respectively. It is again deduced that the "std" method provides sufficient accuracy when estimating the probability of rank reversal while the "minmax" approach leads to significant larger errors. Both methods however seem to provide a correct estimate regarding the order of magnitude of $P_{\rm RR}$.

Fig. 15 discusses the accuracy of the "std" approach for a group size equal to M = 15 which according to Section 3, is a sufficient group size to estimate the priorities of the alternatives, given the limitations of practical PWC applications. It is verified that $e^{(\text{std})}$ is significantly smaller than P_{RR} implying a tolerable error in the estimation of the probability of rank reversal.

In the aforementioned results, the eigenvalue method was used for the estimation of PWC weights. Nevertheless, it would be very interesting to investigate the estimation of P_{RR} from actual user data applying the Geometric Mean Method, as analyzed previously in Section 3.2.3. Fig. 16 shows the values of P_{RR} and $e^{(u)}$ for the case where N = 3 criteria are pairwise compared assuming an uncertainty level L = 5. The results reveal that the values of P_{RR} , estimated by actual user data, increasing the expert group size is more or less the same as mentioned before in the context of EM in Fig. 13(a). It is again deduced that the "std" method provides sufficient accuracy while the "minmax" approach leads to larger errors.

Fig. 15. (a) $e^{(\text{std})}$ and (b) P_{RR} as a function of the uncertainty level for M = 15.

Fig. 16. P_{RR} and $e^{(u)}$ as function of the expert group size for L = 5 and N = 3, for GMM.

Fig. 17. P_{RR} and $e^{(u)}$ as function of the expert group size in the case of fuzzy judgments.

4.3. Impact of fuzzy judgments

It is also interesting to investigate the estimation of $P_{\rm RR}$ from actual user data considering the representation of judgments by fuzzy numbers, as discussed in Section 3.2.4. Toward this end, we apply the scheme shown in Fig. 12 except that the matrices are now replaced by their fuzzy extensions and the weights are calculated in a fuzzy manner as discussed previously. Again the probability of rank reversal from actual user data can be estimated using both the "minmax" and "std" methods. The uncertainty model is modified in order to incorporate the fuzzy judgments in the opinion of experts when generating the matrices $\mathbf{R}^{(l)}$, where the elements $R^{(l)}_{ij}$ either are uniformly chosen from $[D'_{ij}U'_{ij}]$ for the case of "minmax", or follow a uniform distribution determined by the mean value μ_{ii} and standard deviation σ_{ii} for "std" method. In Fig. 17 we show the values of $e^{(\text{std})}$ and $e^{(\text{minmax})}$ in the case where L = 5 and N = 6. Inspection of the results depicted in the figure reveals that similar conclusions are drawn for the estimation of P_{RR} from actual user data in the case of fuzzy judgments. It is obvious that applying either the traditional PWC or the fuzzy triangular representation of PWC judgments, the "std" method prevails against the "minmax" for the estimation of P_{RR} from actual user data.

5. Conclusions

In this paper, we applied an uncertainty model to address two important issues concerning the probability of rank reversal in pairwise comparisons. The first issue concerned how this probability is reduced by augmenting the number of experts participating in the surveys. The results dictate that there is not much to be gained by increasing the number of experts beyond 15, even if uncertainty level is large. Using perturbation theory we have argued that the convergence of the probability of rank reversal with the number of experts is not crucially dependent on the uncertainty statistics. We have also shown numerically that the choice of comparison scale and weight selection method does not significantly affect the convergence. The second issue concerned the problem of how the probability of rank reversal can be estimated in practice, from the elements of the pairwise comparison matrices of a single expert group. Two alternative methods have been discussed for extracting information on the statistical behavior of the uncertainty-induced perturbations. It was shown that one of them provides reasonable good accuracy and can therefore be used in practical applications of the method for estimating the credibility of the outcome. Interestingly enough, this conclusion holds under other situations such as fuzzy pairwise judgments, alternative preference scales, weight estimation methods and accounting for a different uncertainty level for each expert in the group.

References

- Abildtrup, J., Audsley, E., Fekete-Farkas, M., Giupponi, C., Gylling, M., Rosato, P., et al. (2006). Socio-economic scenario development for the assessment of climate change impacts on agricultural land use: A pairwise comparison approach. *Environmental Science & Policy*, 9, 101–115.
- Aull-Hyde, R., Erdogan, S., & Duke, J. (2006). An experiment on the consistency of aggregated comparison matrices in AHP. *European Journal of Operational Research*, 171, 290–295.
- Barzilai, J. (1997). Deriving weights from pairwise comparison matrices. Journal of the Operational Research Society, 48, 1226–1232.
- Bhushan, N., & Rai, K. (2004). Strategic decision making: Applying the analytic hierarchy process. London, New York: Springer.
- Boenderb, C. G. E., Graan, J. G., & Lootsma, F. A. (1989). Multi-criteria decision analysis with fuzzy pairwise comparisons. *Fuzzy Sets and Systems*, *29*, 133–143.
- Brans, J. P., Vincke, Ph., & Mareschal, B. (1986). How to select and how to rank projects: The Promethee method. *European Journal of Operational Research*, 24, 228–238.
- Carmone, F. J., Karab, A., & Zanakis, S. H. (1997). A Monte Carlo investigation of incomplete pairwise comparison matrices in AHP. *European Journal of Operational Research*, 102, 538–553.
- Chan, F. T. S., Chan, H. K., Lau, H. C. W., & Ip, R. W. L. (2006). An AHP approach in benchmarking logistics performance of the postal industry. *International Journal of Benchmarking*, 13, 636–666.
- Chang, Y. -H., & Yeh, C. -H. (2001). Evaluating airline competitiveness using multiattribute decision making. Omega, 29, 405–415.
- Chao Chung, H. (2011). Optimal evaluation of infectious medical waste disposal companies using the fuzzy analytic hierarchy process. *Waste Management*, 31, 1553–1559.
- Chen, H., & Kocaoglu, D. F. (2008). A sensitivity analysis algorithm for hierarchical decision models. European Journal of Operational Research, 185, 266–288.
- Choo, E. U., & Wedley, W. C. (2004). A common framework for deriving preference values from pairwise comparison matrices. *Computers & Operations Research*, 31, 893–908.
- Dede, G., Kamalakis, T., & Varoutas, D. (2011). Towards a roadmap for future home networking systems: An analytical hierarchy process approach. *IEEE Systems Journal*, 5, 374–384.
- Dede, G., Kamalakis, T., & Varoutas, D. (2011). Evaluation of optical wireless technologies in home networking: An analytical hierarchy process approach. *Journal of Optical Communications and Networking*, 3, 850–859.
- Dede, G., Varoutas, D., Kamalakis, T., Fuentetaja, R. G., & Javaudin, J. P. (2010). Evaluation of technological and socio-economic issues affecting the deployment of home networks: Evidence from the ICT-OMEGA project. Springer Netnomics: Economic Research and Electronic Networking, 11, 181–200.
- Deng, H. (1999). Multicriteria analysis with fuzzy pairwise comparison. Fuzzy Systems Conference Proceedings, FUZZ-IEEE '99, 2, 726–731.
- Doumpos, M., & Zopounidis, C. (2004). A multicriteria classification approach based on pairwise comparisons. European Journal of Operational Research, 158, 378–389.
- Fan, Z. P., & Liu, Y. (2010). An approach to solve group-decision-making problems with ordinal interval numbers. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40, 1413–1423.
- Farkas, A. (2007). The analysis of the principal eigenvector of pairwise comparison matrices. Acta Polytechnica Hungarica, 4, 99–115.
- Gerdsri, N., & Kocaoglu, D. F. (2007). Applying the Analytic Hierarchy Process (AHP) to build a strategic framework for technology roadmapping. *Mathematical and Computer Modelling*, 46, 1071–1080.
- Gerla, G., & Scarpati, L. (1998). Extension principles of fuzzy set theory. Journal of Information Sciences, 106, 49–69.
- Hahn, E. D. (2003). Decision making with uncertain judgments: A stochastic formulation of the analytic hierarchy process. *Decision Sciences*, 34, 443–466.
- Hanke, J. E., & Wichern, D. (2008). Business forecasting (9th ed.). Prentice Hall.
- Huanga, Ch. -Ch., Chub, P. Y., & Chiang, Y. -H. (2008). A fuzzy AHP application in government-sponsored R&D project selection. Omega, 36, 1038–1052.
- Ishizaka, A., Balkenborg, D., & Kaplan, T. (2011). Influence of aggregation and measurement scale on ranking a compromise alternative in AHP. *Journal of the Operational Research Society*, 62, 700–710.
- Ishizaka, A., & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. Expert Systems with Applications, 38, 14336–14345.
- Kok, M., & Lootsma, F. A. (1985). Pairwise-comparison methods in multiple objective programming, with applications in a long-term energy-planning model. *European Journal of Operational Research*, 22, 44–55.
- Kwiesielewicz, M., & Van Uden, M. (2004). Inconsistent and contradictory judgments in pairwise comparison method in the AHP. Computers & Operations Research, 31, 713–719.

- Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM Journal of Optimization, 9, 112–147.
- Lee, Y., & Kozar, K. A. (2006). Investigating the effect of website quality on e-business success: An analytic hierarchy process (AHP) approach. *Decision Support Systems*, 42, 1383–1401.
- Liberatore, M. J., & Nydick, M. J. (2008). The analytic hierarchy process in medical and health care decision making: A literature review. *European Journal of Operational Research*, 189, 194–207.
- Marimin, M., Umano, M., Hatono, I., & Tamura, H. (1998). Linguistic labels for expressing fuzzy preference relations in fuzzy group decision making. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 28*, 205–218.
- Mikhailov, L. (2005). A fuzzy approach to deriving priorities from interval pairwise comparison judgments. *European Journal of Operational Research*, 159, 687–704.
- Pohekar, S. D., & Ramachandran, M. (2004). Application of multi-criteria decision making to sustainable energy planning—A review. *Renewable and Sustainable Energy Reviews*, 8, 365–381.
- Rice, J. A. (1995). Mathematical statistics and data analysis (2nd ed.). Belmont, C.A.: Duxbury Press.
- Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15, 234–281.
- Saaty, T. L. (1990). Eigenvector and logarithmic least squares. European Journal of Operational Research, 48, 156–160.
- Saaty, T. L. (1991). Response to Holder's comments on the analytic hierarchy process. Journal Operational Research Society, 42, 909–929.
- Saaty, T. L. (2003). Decision-making with the AHP: Why is the principal eigenvector necessary?. European Journal of Operational Research, 145, 85–91.

- Saaty, T. L. (2004). Fundamentals of the analytic network process Dependence and feedback in decision-making with a single network. *Journal of Systems Science and Systems Engineering*, 13, 129–157.
- Saaty, T. L. (2008). Relative measurement and its generalization in decision making: Why pairwise comparisons are central in mathematics for the measurement of intangible factors – The analytic hierarchy/network process. RACSAM, Review of the Royal Spanish Academy of Sciences, Series A, Mathematics, 251–318.
- Saaty, T. L., & Vargas, L. G. (1984). Inconsistency and rank preservation. Journal of Mathematical Psychology, 28, 205–214.
- Saty, T. L., & Vargas, L. G. (1987). Uncertainty and rank order in the analytic hierarchy process. European Journal of Operational Research, 32, 107–117.
- Saaty, T. L., & Vargas, L. G. (2001). Models, methods, concepts & applications of the analytic hierarchy process. Boston: Kluwer Academic Publishers.
- Shiraishi, S., Tsuneshi, O., & Motomasa, D. (1998). Properties of a positive reciprocal matrix and their application to AHP. Journal of the Operations Research Society of Japan-Keiei Kagaku, 4(1), 404–414.
- Triantaphyllou, E. (2000). Multi-criteria decision making methods: A comparative study.. Kluwer Academic Publishers.
- Waltz, R. A., Morales, J. L., Nocedal, J., & Orban, D. (2006). An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming*, 107, 391–408.
- Wang, Y. M., Elhag, T. M. S., & Hua, Z. (2006). A modified fuzzy logarithmic least squares method of fuzzy analytic hierarchy process. *Fuzzy Sets and Systems*, 157, 3055–3071.
- Yager, R. R. (2004). Modeling prioritized multicriteria decision making. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 34, 2396–2404.
- Zahedi, F. (1986). The analytic hierarchy process, a survey of the method and its applications. *Interfaces*, *16*, 96–108.