A framework of incorporating confidence levels to deal with uncertainty in pairwise comparisons

Georgia Dede[a] Thomas Kamalakis[b] and Dimosthenis Anagnostopoulos[c]

[a,b]Department of Informatics and Telematics, Harokopio University, 9 Omirou Street, Tavros, Athens Greece GR17778

ARTICLE HISTORY
Compiled January 4, 2021

ABSTRACT
Pairwise comparison is a key ingredient in multi-criteria decision analysis. The method is based on a set of comparisons conducted by a group of experts, comparing all possible pairs of alternatives involved in the decision process. The outcome is the estimation of weights determining the ranking of alternatives. In this paper, we introduce a new framework for the incorporation of confidence levels in pairwise comparisons, in order to deal with uncertainty issues related to the individual expert judgments. We discuss how the confidence levels can be related to the probability of rank reversal by introducing a theoretical model based on the multivariate normal cumulative distribution function. A comparison between theoretical and numerical results (Monte Carlo simulations), reveals a very good agreement. The proposed framework may provide a very good basis for pairwise comparison extensions aiming to provide further information regarding the accuracy for the evaluation of the final outcome.

KEYWORDS
Multiple criteria analysis; Decision analysis; Pairwise comparisons; Confidence Levels

1. Introduction

Pairwise comparisons (PWC) are widely used in multi-criteria decision analysis and have successfully been applied in many practical decision-making problems [Saaty, 2008a]. The method provides a structured process for the effective ranking of alternatives and criteria in the context of multi-criteria decision problems, aiming at identifying the most preferable solution [Triantaphyllou, 2000, Yager, 2004]. Within PWC, a group of experts is called to participate in structured surveys and compare the criteria or alternatives in combinations of two rather than ranking them directly, in order to reduce the influence of subjective point of views.

PWCs are applied on a standalone basis or as an integral part of sophisticated decision making methods, such as the analytic hierarchy process (AHP) [Saaty, 2008b] in many areas of everyday life including education, health, technology, energy, etc. [Heldsinger and Humphry, 2010, Anand et al., 2000, Dede et al., 2011,

CONTACT Georgia Dede. Email: gdede@hua.gr
CONTACT Thomas Kamalakis. Email: thkam@hua.gr
CONTACT Dimosthenis Anagnostopoulos. Email: dimosthe@hua.gr

Kok and Lootsma, 1985]. Consequently, the analysis and investigation of its methodological aspects is an issue of paramount importance. Topics such as measurement scales, inconsistency aspects, missing judgment estimation and priority derivation methods have been actively pursued and various approaches have been proposed to address these problems [Barzilai, 1997, Choo and Wedley, 2004, Mamat and Daniel, 2007, Kwiesielewicz and Van Uden, 2004, Boender et al., 1989, Deng, 1999, Mikhailov, 2004, Marimin et al., 2002]. PWCs involve human subjective evaluation and hence uncertainty may undermine the final outcome. Uncertainty and inconsistency issues in PWC have been widely studied and several researchers have attempted to address such crucial aspects in the previous literature. A variety of models have been proposed to represent uncertainty in PWC and address inconsistency issues based usually on probabilities [Durbach et al., 2014, Eskandari and Rabelo, 2007, Mikhailov, 2004, Banuelas and Antony, 2004, Hahn, 2003, Zahir, 1991, Saaty and Vargas, 1987], decision weights, preference weights [Beynon, 2002] risk measures [Millet and Wedley, 2002], interval weights [Dong et al., 2015, Entani and Sugihara, 2012, Sugihara and Tanaka, 2001], fuzzy numbers [Yang et al., 2013, Ramík and Korviny, 2010, Sugihara et al., 2004, Saaty and Vargas, 1987] regression techniques [Alho and Kangas, 1997] and consistency measures [Siraj et al., 2015, Koczkodaj and Szarek, 2010, Limayem and Yannou, 2007, Alonso and Lamata, 2006, Farkas et al., 2003, Barzilai, 1998, Durbach and Stewart, 2012]. In our previous work [Dede et al., 2016, Dede et al., 2015], we have considered the impact of uncertainty in PWC, which originates from the fact that the experts may produce inconsistent pairwise judgments or that their point-of-views can differ. In this context, we have introduced a numerical as well as a theoretical model for the evaluation of the probability of rank reversal P_{RR} in PWC matrices. In many practical circumstances, comparisons may require extensive knowledge and hence a participant may find himself to be uncertain of his judgment. In the standard form of PWCs, he has no means to communicate his uncertainty and hence there is no way in which his certainty level can be taken into account when considering the credibility of the final outcome. In our present paper, we provide a sound methodological framework for extending PWCs in order to reflect the certainty level of the experts.

The basic idea behind our approach is simple: each expert is asked to quantify his confidence level on a particular judgment or a set of judgments, in terms of a number in the range of 0 to 100%. We then apply a methodological framework that translates the individual confidence levels to a measure on the credibility of the outcome and more specifically we estimate the probability P_{RR} that the ranking of at least two alternatives is reversed. The probability of rank reversal is widely considered as an intuitive measure of the credibility of the final outcome [Saaty and Vargas, 1984, Saaty and Vargas, 1987, Saaty, 2003, Dede et al., 2015].

To quantify the credibility of the outcome, we may therefore adopt a probabilistic point-of-view. Each expert m from a group of M $(1 \leq m \leq M)$ fills out specific values $a_{ij}^{(m)}$ in the elements of the pairwise comparison matrices signifying the preference of option i over option j and we consider these values as realizations of the random variables $A_{ij}^{(m)}$ describing the random pairwise comparison matrix $\mathbf{A}^{(m)}$ of the m^{th} expert. The values $a_{ij}^{(m)}$ are usually discrete. For example $a_{ij}^{(m)} \in \left\{\frac{1}{9}, \frac{1}{8}, ..., 1, 2, ...9\right\}$, when the original nine-level scale is applied [Saaty and Vargas, 1987]. The expert also provides the certainty level $c_{ij}^{(m)}$ of his judgment from 0 to 100% which readily translates to the probability that the element of the random matrix $A_{ij}^{(m)}$ will actually be equal to $a_{ij}^{(m)}$, i.e. $P\{A_{ij}^{(m)} = a_{ij}^{(m)}\} = c_{ij}^{(m)}$ where $0 \leq c_{ij}^{(m)} \leq 1$. If $c_{ij}^{(m)} < 1$, we expect that there will be a non-zero probability that $P\{A_{ij}^{(m)} = x\}$ for at least one other value x in the nine-level scale used in the comparisons. The sum of all probabilities $P\{A_{ij}^{(m)} = x\}$ should be equal to 1, $\sum_x P\{A_{ij}^{(m)} = x\} = 1$. We also expect that the expert completed the value $a_{ij}^{(m)}$ which he considers the most probable, hence $P\{A_{ij}^{(m)} = a_{ij}^{(m)}\} \geq P\{A_{ij}^{(m)} = x\}$. Taking into account the unity sum of all probabilities, we readily obtain $1 = \sum_x P\{A_{ij}^{(m)} = x\} \leq \sum_x P\{A_{ij}^{(m)} = a_{ij}^{(m)}\} = N_{\mathrm{el}} c_{ij}^{(m)}$ where N_{el} are the number of possible values for x ($N_{\mathrm{el}} = 17$ for the nine-level scale). We therefore deduce that $c_{ij}^{(m)} \geq 1/N_{\mathrm{el}}$, i.e. $c_{ij}^{(m)}$ is always larger than or equal to $1/N_{\mathrm{el}}$ which is the probability of choosing the value indifferently, without any preference.

Using PWCs, one can determine the priorities of the alternatives w_i (Section 2.1), corresponding to the pairwise comparison elements $a_{ij}^{(m)}$ completed by the experts. Under the probabilistic framework, w_i are realizations of the random variables W_i which depend on $A_{ij}^{(m)}$. The probability of rank reversal P_{RR} is the probability that the ranking obtained by W_i will be different than that obtained by w_i. If the values of the P_{RR} is high, then the outcome cannot be trusted and the decision maker needs to take measures for increasing the credibility, i.e. by increasing the number of participants. In this paper, we show how this "macroscopic" credibility measure, i.e. the P_{RR} can be related to the "microscopic" measures such as the confidence level $c_{ij}^{(m)}$. In this work, we propose an intuitive probability distribution for the elements of the pairwise comparison matrices $A_{ij}^{(m)}$ that can be readily determined by the input provided by the experts. Based on our previous work on the general weight statistics [Dede et al., 2016], we calculate the weight covariance and show how the multivariate normal cumulative distribution function (MVNCDF) can be used to derive an analytical model for the estimation of the P_{RR}. Although this could in principle be carried out using numerical schemes such as Monte-Carlo (MC) simulation, in practice it is often much better, if possible, to use some analytical technique for evaluating the P_{RR}. This is because, analytical methods are usually much faster than brute-force simulations and at the same time offering a useful insight while constituting a basis for further extending the framework. We compare the values of the P_{RR} against numerical simulations and obtain excellent accuracy. We also discuss another extension of the method where, based on the extra information regarding the confidence levels of each expert judgment contained in $c_{ij}^{(m)}$, one could adopt a "screening" strategy. Throughout our framework we provide analytical formulas for the various parameters involved in the calculations, therefore limiting the application of numerical methods to the calculation of the P_{RR} itself. In other words, by omitting a set of uncertain judgments, we may hope to reduce the probability of rank reversal and hence improve the credibility of the outcome. Our model can be readily used to evaluate the gain by such strategies. We believe that our results are very useful and can be applied to many multi-criteria decision making frameworks in order to estimate the credibility of the intermediate stages of the process. [Stam and Silva, 1997] have previously analyzed stochastic priority vectors in PWC using multivariate normal distributions. In this paper, we extend the concept of this work and propose an additional framework to incorporate the confidence levels in the PWC. More specifically, we introduce confidence levels instead of confidence intervals, we estimate the overall P_{RR} instead of the estimation of P_{RR} between only pairs of criteria/alternatives and we provide a more general framework taking into account multiple experts which is usually adopted in practice, rather than just a single expert. Although in our work we consider the exponential distribution for modelling the uncertainty in the expert judgements, our framework can be easily generalized for other distributions as well as long as the Gaussian distribution can be postulated from the central limit theorem [Rice, 2006]. Moreover, we provide a thorough discussion of the application of the MVNCDF that could take into account possible correlation between the expert judgements, although further analysis is outside the scope of the paper. Finally, we show how thresholds can be applied in order to reduce uncertainty in the pairwise comparison matrices.

The rest of the paper outline is analyzed as follows: In Section 2, we describe the proposed framework based on the confidence levels in the opinion of experts. We also estimate the credibility of the results and evaluating P_{RR} as a metric for uncertainty. In Section 3, we validate our framework by comparing the results obtained by the theoretical model against MC simulations under various circumstances. Finally, Section 4 provides some concluding remarks.

2. Description of the framework

2.1. Estimation of the P_{RR}

In our previous work [Dede et al., 2016], we provided the general theoretical framework for estimating the probability of rank reversal through the MVNCDF. In this subsection, we briefly summarize our earlier findings that will be used in this work. The commonly used geometric mean method [Saaty, 2008a] can be used to elicit the weights $w_k^{(m)}$ for the m^{th} expert as follows:

$$w_k^{(m)} = \left(\prod_{p=1}^{N} a_{kp}^{(m)} \right)^{\frac{1}{N}} \tag{1}$$

where N is the number of alternatives. The overall priorities w_k are calculated through the geometric mean of $w_k^{(m)}$,

$$w_k = \left(\prod_{m=1}^{M} w_k^{(m)} \right)^{\frac{1}{M}} = \left(\prod_{m=1}^{M} \prod_{p=1}^{N} a_{kp}^{(m)} \right)^{\frac{1}{MN}} \tag{2}$$

where M is the number of experts. When prioritizing the alternatives, one can estimate the logarithmic weights,

$$z_k = \ln w_k = \frac{1}{MN} \sum_{m=1}^{M} \sum_{p=1}^{N} \ln a_{kp}^{(m)} \tag{3}$$

The random variables W_k follow from (2) by simply replacing the values $a_{kp}^{(m)}$ by the random variables $A_{kp}^{(m)}$,

$$W_k = \frac{1}{MN}\left(\prod_{m=1}^{M}\prod_{p=1}^{N} A_{kp}^{(m)}\right)^{\frac{1}{MN}} \tag{4}$$

It is reasonable to assume the experts to carry out their judgments independently of each other (for simplicity reasons the violation of independence is not examined) and hence $A_{kp}^{(m)}$ can be assumed independent of each other. The logarithmic random weights are determined by:

$$Z_k = \ln W_k = \frac{1}{MN}\sum_{m=1}^{M}\sum_{p=1}^{N} \ln A_{kp}^{(m)} \tag{5}$$

We also define the difference $\delta Z_k = Z_k - z_k$ between the random weight and the estimated logarithmic weight z_k calculated by the data completed by the experts. According to (5),(3) $\delta Z_k = Z_k - z_k$ are a sum of a large number of nearly statistically independent random variables. In the case where NM is finite, then the Gaussian statistics are still a good approximation, provided that NM is large enough [Dede et al., 2016].

The probability of rank reversal P_{RR} can be estimated as follows: assuming for simplicity that the alternatives C_j are labeled so that C_1 has the largest weight w_1, C_2 the second largest and so on. This implies that $z_{j+1} \leq z_j$ for $1 \leq j \leq N$. The ranking is preserved if $Z_{j+1} \leq Z_j$ for $1 \leq j \leq N$. Denoting $\delta Z_k = Z_k - z_k$, the previous inequality translates to $\delta Z_j - \delta Z_{j+1} \geq z_{j+1} - z_j$. Hence the probability of rank reversal can be estimated as:

$$P_{\mathrm{RR}} = 1 - P\big\{t_2 \geq z_2 - z_1, ..., t_k \geq z_{k+1} - z_k, ...\big\} \tag{6}$$

where

$$t_k = \delta Z_{k+1} - \delta Z_k. \tag{7}$$

Since δZ_k are Gaussian, t_k will also be Gaussian and therefore (6) is given from the MVNCDF of t_k provided that the mean values $\mu_k =< t_k >$ and the covariances $R_{kp} =< (t_k - \mu_k)(t_p - \mu_p) >$ of t_k are determined. Once these statistical parameters are known, a variety of methods can be used to estimate the MVNCDF [Genz and Bretz, 2002]. In the following sections, we will apply this general framework to evaluate the P_{RR} given the pairwise comparison and the confidence level matrices $\mathbf{a}^{(m)}$ and $\mathbf{c}^{(m)}$, respectively.

2.2. Confidence levels in the opinion of experts

Along with the pairwise comparison matrices $\mathbf{a}^{(m)} = [a_{ij}^{(m)}]$ the experts complete the confidence level matrices $\mathbf{c}^{(m)} = [c_{ij}^{(m)}]$ where $c_{ij}^{(m)}$ designates the confidence level of the m^{th} expert regarding his comparisons for the importance of C_i over C_j. The matrix $\mathbf{c}^{(m)}$ should be symmetrical (Hermitian) since when the level of certainty is the same when comparing C_i with C_j or C_j with C_i. The diagonal elements of the pairwise comparison matrices are always equal to one and hence the diagonal elements $c_{ii}^{(m)}$ should also be one. The confidence level matrices $\mathbf{c}^{(m)}$ therefore have the following form:

$$\mathbf{c}^{(m)} = [c_{ij}^{(m)}] \begin{bmatrix} 1 & c_{12}^{(m)} & \dots & c_{1N}^{(m)} \\ c_{12}^{(m)} & 1 & \dots & c_{2N}^{(m)} \\ \vdots & \vdots & \vdots & \vdots \\ c_{1N}^{(m)} & c_{2N}^{(m)} & \dots & 1 \end{bmatrix} \tag{8}$$

Much like the case of the pairwise comparison matrices, the expert needs only to complete the elements of $\mathbf{c}^{(m)}$ located on the upper triangular part of the matrix. As discussed in the introductory section, we may consider that the matrices $\mathbf{a}^{(m)} = \left[a_{ij}^{(m)}\right]$ are instances of the random matrices $\mathbf{A}^{(m)} = \left[A_{ij}^{(m)}\right]$. In order to describe the statistics of $A_{ij}^{(m)}$ we need the probability distribution $\mathrm{P}\left\{A_{ij}^{(m)} = x\right\} = p_{ij}^{(m)}(x)$. One could have the experts fill out the values of $p_{ij}^{(m)}(x)$ themselves for every x in addition to the pairwise comparison matrices but this will require a lot of additional information and may render the process impractical. It is best if we simply choose an intuitive distribution that can be determined by the the pairwise and the confidence level matrices $\mathbf{a}^{(m)}$ and $\mathbf{c}^{(m)}$, respectively. Alternative types of distributions could also be used, since the estimation of P_{RR} based on MVNCDF is independent of the applied distributions, as shown in Section 2.4. In this context, if the exponential distribution is not adopted, then most of the results of the paper will still apply since the distribution of the logarithmic weights will still be Gaussian as a consequence of the central limit theorem [Rice, 2006]. The proposed distribution should of course be such that $p_{ij}^{(m)}\left(a_{ij}^{(m)}\right) = c_{ij}^{(m)}$. We propose an exponentially decaying distribution $p_{ij}^{(m)}(x) \propto \lambda^{-u}$ with the distance $u = |i_x - i_a|$ between the positions i_x and i_a of x and c_{ij}^{m}, respectively in the nine-level scale. Figure 1 shows an example of this probability distribution when $a_{ij}^{(m)} = 3$ and $c_{ij}^{(m)} = 0.6$. The distribution is determined by the equation,

$$p_{ij}^{m}(x) = c_{ij}^{(m)} \exp\left(-h_{ij}^{(m)} u\right) \tag{9}$$

To calculate $u = |i_x - i_a|$, we number all elements x of the nine-level scale $\left\{\frac{1}{9}, \frac{1}{8}, \dots, 9\right\}$ and calculate the difference between the indices i_x and i_a corresponding to x and $a_{ij}^{(m)}$, respectively. For example, we find that the value $a_{ij}^{(m)} = 3$ is the 11$^{\text{th}}$ element of $\left\{\frac{1}{9}, \frac{1}{8}, \dots, 9\right\}$ and hence $i_a = 11$, while the value $x = \frac{1}{3}$ is the 7$^{\text{th}}$ element, hence $i_x = 7$. The distance between the positions of the elements is therefore $u = |i_x - i_a| = 4$. The parameter $h_{ij}^{(m)}$ in (9) determines the decaying rate of the exponential distribution and is chosen so that the sum of all probabilities is equal to unity, i.e.:

$$S = \sum_x p_{ij}^{(m)}(x) = 1 \tag{10}$$

We will discuss how $h_{ij}^{(m)}$ can be calculated in the following section. The exponential distribution represents an intuitive model where the elements directly adjacent to $a_{ij}^{(m)}$ may occur with much higher probability than the elements situated far from $a_{ij}^{(m)}$.

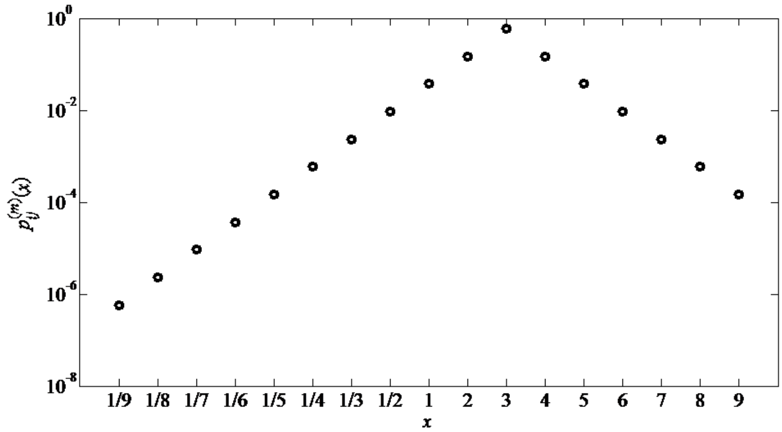


Figure 1.: Example of the proposed probability distribution $p_{ij}^{(m)}(x)$ when $a_{ij}^{(m)} = 2$ and $c_{ij}^{(m)} = 0.6$.

This distribution seems as the most suitable since $p_{ij}^{(m)}(x)$ is expected to be reduced geometrically as x is moving away from $a_{ij}^{(m)}$. For example, we consider the case where $a_{ij}^{(m)} = 3$ and $c_{ij}^{(m)} = 60\%$, then according to Figure 1, the elements $x = 2$ or $x = 4$ directly adjacent to $a_{ij}^{(m)} = 3$ have a probability of 15%, the second nearest elements $x = 1$ and $x = 5$ have probability 3.75% while the outer elements 1/9 or 9 have negligible probability to occur.

2.3. Determination of the exponential distribution

To estimate $h_{ij}^{(m)}$ we can use (9) and fit $h_{ij}^{(m)}$ so that $|S-1|$ is minimized. This procedure can be done numerically with a variety of methods [Hildebrand, 1987]. However in this paper, we are interested in providing analytic formulas so that the application of the framework does not require the application of numerical techniques which require code implementation. We can estimate the sum analytically since for the exponential distribution we obtain the sum of two geometric series to the left and the right of $a_{ij}^{(m)}$. Substituting (9) into (10), we readily obtain,

$$S = c_{ij}^{(m)} \frac{z - z^{18-i_a} - z^{i_a} + 1}{1 - c_{ij}^{(m)}} \tag{11}$$

where:

$$z = \exp(-h_{ij}^{(m)}) \tag{12}$$

In order to have $S = 1$ we therefore need to solve the following polynomial equation:

$$z^{18-i_a} + z^{i_a} - \left(1 + \frac{1}{c_{ij}^{(m)}}\right) z + \left(1 - \frac{1}{c_{ij}^{(m)}}\right) = 0 \tag{13}$$

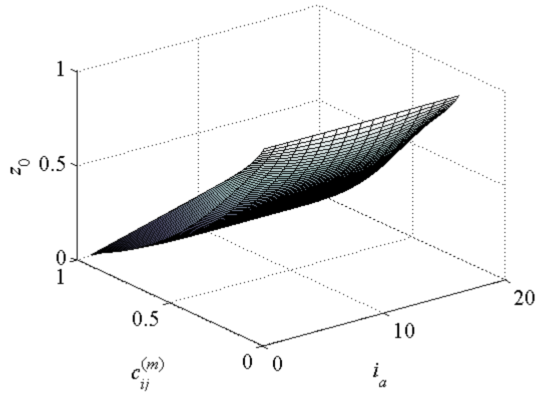


Figure 2.: The values of the single real root z_0 of the polynomial in (13)

For every integer $1 \leq i_a \leq 17$ and every level $\frac{1}{17} \leq c_{ij}^{(m)} \leq 1$, we have numerically confirmed that there is always exactly one positive real root z_0 which is smaller than unity, $0 \leq z_0 \leq 1$. Figure 2 shows the values of z_0 obtained numerically for the intervals in question. The values of $h_{ij}^{(m)}$ can be obtained simply by taking the logarithm of $z_0, h_{ij}^{(m)} = -\ln z_0$. We can also find some approximation for the roots shown in Figure 2 in order to avoid the numerical solution every time. For example, we have found that for the inner scale elements (i.e. $a_{ij}^{(m)} \neq 9$ and $a_{ij}^{(m)} \neq \frac{1}{9}$ or equivalently $i_a \neq 1$ and $i_a \neq 17$) and for $c_{ij}^{(m)} = c$ where $0.5 \leq c \leq 0.95$, i.e. confidence levels ranging from 50% to 95%, the roots can be well approximated by a polynomial function of two variables,

$$z_0 \approx P_2(c, h) \approx 0.22c^2h^2 - 0.4ch^2 + 0.19h^2 + 0.37c^2 - 1.2c + 0.83 \tag{14}$$

where $h = (i_a - 9)/9$ is the normalized index of the scale element. In the case of the outer scale elements (i.e. $a_{ij}^{(m)} = 9$ or $a_{ij}^{(m)} = \frac{1}{9}$ in which case $i_a = 1$ or $i_a = 17$ respectively) we can approximate the root with the following polynomial,

$$z_0 \approx P_1(c) \approx 0.64c^2 - 1.7c + 1.08 \tag{15}$$

The coefficients of the polynomials in (14) and (15) were estimated by fitting the data with second order polynomials in two or one dimensions respectively in the least-square sense. We have found that they approximate the root z_0 with an average error of less than 4% and provide a straightforward means of evaluating $h_{ij}^{(m)}$ since they do not entail any actual root estimation for the polynomial in (13).

2.4. Weight covariance

As stated in Section 2.1, we need to know the mean values $\mu_k =< t_k >$ and the covariances $R_{\kappa\lambda} =< (t_\kappa - \mu_\kappa)(t_\lambda - \mu_\lambda) >$ in order to calculate the probability of rank reversal $P_{\rm RR}$. We evaluate these based on the model presented in Section 2.2. Taking into account (7), we obtain:

$$\mu_k = \langle t_k \rangle = \langle \delta Z_{k+1} \rangle - \langle \delta Z_k \rangle \tag{16}$$

while

$$R_{\kappa\lambda} = \langle t_\kappa t_\lambda \rangle - \langle t_\kappa \rangle \langle t_\lambda \rangle = r_{\kappa\lambda} - r_{\kappa,\lambda-1} - r_{\lambda-1,\mu} + r_{\kappa-1,\lambda-1} \tag{17}$$

where $r_{ij} = \langle \delta Z_i \delta Z_j \rangle - \langle \delta Z_i \rangle \langle \delta Z_j \rangle$ are the elements of the covariance matrix of the variables δZ_k. Using (5) and (3) we readily find that:

$$\delta Z_k = \frac{1}{NM} \sum_{m=1}^{M} \sum_{p=1}^{N} \{ \ln A_{kp}^{(m)} - \ln \alpha_{kp}^{(m)} \} \tag{18}$$

Taking the expected value of both sides in (18), we have:

$$\langle \delta Z_k \rangle = \frac{1}{NM} \sum_{m=1}^{M} \sum_{p=1}^{N} \{ \ln M_{kp}^{(m)} - \ln \alpha_{kp}^{(m)} \} \tag{19}$$

where $M_{kp}^{(m)}$ is the expected value of $\ln A_{kp}^{(m)}$ which is given by:

$$M_{kp}^{(m)} = \langle \ln A_{kp}^{(m)} \rangle = \sum_x p_{kp}^{(m)}(x) \ln x \tag{20}$$

and the sum runs over all elements x of the 9-point scale. We can also infer that:

$$r_{\kappa\lambda} = \frac{1}{(NM)^2} \sum_{mp} \left\{ N_{\kappa p}^{(m)} - [M_{\kappa p}^{(m)}]^2 \right\} - \frac{1}{(NM)^2} \sum_{m} \left\{ N_{\kappa\lambda}^{(m)} - [M_{\kappa\lambda}^{(m)}]^2 \right\} \tag{21}$$

where $N_{kp}^{(m)}$ is the expected value of $(\ln A_{kp}^{(m)})^2$

$$N_{kp}^{(m)} = \langle (\ln A_{kp}^{(m)^2}) \rangle = \sum_x p_{kp}^{(m)}(x) (\ln x)^2 \tag{22}$$

Equations (16)-(22) determine the covariance matrix and hence the MVNCDF that can be used to provide an estimate for the P_{RR}.

2.5. Model application

To summarize the proposed procedure in practical situations, in contrast to traditional decision problems which involve pairwise comparisons and ask the expert to fill in the PWC matrices, crisp and fuzzy intervals or confidence intervals, here the experts have to complete the pairwise comparison matrices $\mathbf{a}^{(m)} = [a_{ij}^{(m)}]$ and the corresponding confidence level matrices $\mathbf{c}^{(m)} = [c_{ij}^{(m)}]$. The PWC matrices can be filled in using a variety of scales that have already been introduced (e.g. nine-level scale [Saaty and Vargas, 1987], logarithmic level scale

[Ishizaka and Labib, 2011], percentage [Gerdsri and Kocaoglu, 2007]). In this paper, the nine-level scale $\left\{\frac{1}{9}, \frac{1}{8}, ..., 1, 2, ...9\right\}$ is used as a simple accepted scale for this kind of decision making problems. As far as the confidence level matrix is concerned, each expert m fills in degrees of their certainty $c_{ij}^{(m)}$ for the corresponding value of their PWC matrix $a_{ij}^{(m)}$. This degree of certainty or confidence level for each possible assignment can be defined in multiple ways. For example, it may be derived using a direct rating approach, which consists of asking the expert to assign a percentage value ranged between [0 100%]. If the confidence level is close to 100% then the expert feels more safe and sure for his PWC judgment. For example, if an expert gives a PWC judgment value 7 and a confidence level 90%, then this means that he is almost sure for the judgment he provides. Otherwise, if he gives a confidence level of 60% then his certainty is not so strong. This is the way followed in this paper. However, since this framework is distribution independent then the confidence levels can also be elicited in terms of intervals as [Stam and Silva, 1997] proposed.

After have filled in the matrices $\mathbf{a}^{(m)}$ and $\mathbf{c}^{(m)}$, we then estimate the probability distributions $p_{ij}^{(m)}(x)$ using the method outlined in Section 2.3. The dampening factors $h_{ij}^{(m)}$ can be obtained by the approximate equations (14) and (15). The next step is to estimate the mean values μ_k and covariance matrix elements $R_{\kappa\lambda}$ using the equations presented in this section. We then apply the MVNCDF in order to obtain an estimate for the probability of rank reversal P_{RR} according to (6).

We also investigate the accuracy of our model in alternative settings such as the acceptance or rejection of pairwise comparison judgments when the confidence level is lower than a given threshold level t_c. This is an attempt to screen out judgments which have low confidence level $c_{ij}^{(m)} < t_c c_{ij}^{(max)}$ where $c_{ij}^{(max)}$ is the maximum certainty level for this particular judgment, $c_{ij}^{(max)} = \max_m\{c_{ij}^{(m)}\}$. In the case where $c_{ij}^{(l)} < t_c c_{ij}^{(max)}$ we replace all $a_{ij}^{(l)}$ with the geometric mean of all the corresponding remaining judgments of the other experts $a_{ij}^{(b)}$, $1 \leq b \leq M$ and consider that the confidence level for this judgment $c_{ij}^{(b)}$ is the mean confidence level from the corresponding remaining judgments. This of course is by no means the only one strategy and other alternative schemes can also be assumed.

Moreover, when applying the model, instead of having the experts complete the pairwise comparison matrices $\mathbf{c}^{(m)}$ we could ask them on their confidence level $c_i^{(m)}$ when judging the alternative C_i. We could then calculate $c_{ij}^{(m)}$ as $c_{ij}^{(m)} = c_i^{(m)} c_j^{(m)}$ and apply the theoretical model described above for the calculation of P_{RR}. This leads to less information being collected by the experts which further enhances the practical application of the method in the case where the number of alternatives is large at the expense of less detailed confidence level designation.

Note that the incorporation of the confidence level is done for each expert separately so that the estimation of the P_{RR} reflects the confidence levels of each of the judgments of the experts. One could think of combining the confidence levels $c_{ij}^{(m)}$ of each expert to calculate an average confidence level c_{ij} for each judgment and then use these averages when estimating the credibility of the final outcome. However, it is not clear how c_{ij} can be consistently calculated from $c_{ij}^{(m)}$ and we plan to investigate this in a future publication. In any case, even if one somehow calculates these average levels c_{ij} from $c_{ij}^{(m)}$ and use them to estimate the P_{RR}, it is still imperative to ask the experts for their individual confidence levels $c_{ij}^{(m)}$ so not much will be gained in the method's practical application.

A repository on github [https://github.com/gdede hua/confidence$_levels_framework$,] has been also created, which contains the main functions used in this framework. This repository also includes an example with M=4 experts and n=4 criteria and their pairwise comparison matrices as well as the corresponding confidence levels matrices. The estimated P_{RR} (∼14%) is also presented.

3. Results and Discussion

3.1. Gaussian statistics

Before establishing the accuracy of calculating the probability of rank reversal using the MVNCDF following the framework outlined in Section 2, it is useful to ascertain whether the logarithmic weights Z_k given by (5) are indeed approximately Gaussian random variables. To do so, we use Monte-Carlo (MC) simulation. We first generate M random matrices $\mathbf{a}^{(m)} = \left[a_{ij}^{(m)}\right]$ along with M uncertainty matrices $\mathbf{c}^{(m)} = \left[c_{ij}^{(m)}\right]$ corresponding to a group of experts and estimate the logarithmic weights z_k according to (3). The elements $c_{ij}^{(m)}$ are randomly generated using a standard uniform distribution in the interval [1/17,1]. We then perform a large number of $N_{\mathrm{MC}}(\geq 10^3)$ MC simulations and in each iteration, we determine an instance of $A_{ij}^{(m)}$, according to the probabilities $p_{ij}^{(m)}(x)$ estimated in (9). These instances are calculated using a commonly known algorithm for obtaining the samples of a discrete random variable [Beaumont, 2005]. In essence, we divide the interval [0 1] into 19 sub-intervals [0 P_1), [P_1 P_2), …, [$P_{\mu-1}$ P_μ),…, [P_{18} 1], where $P_\mu = \mathrm{P}\left\{A_{ij}^{(m)} \leq x_\mu\right\}$ stands for the cumulative probability of $A_{ij}^{(m)}$ and x_μ are the consecutive elements of the nine-point scale. We then generate a uniform random variable V inside the interval [0 1] and we determine in which interval it belongs. If it belongs to the ν^{th} interval [$P_{\nu-1}$ P_ν) then we assign the value x_ν to this instance of $A_{ij}^{(m)}$. The obtained matrices $\mathbf{A}^{(m)}$ are checked for consistency according to the criteria set by [Saaty, 2003]. If a matrix is not consistent, it is discarded and calculated again, until we obtain a consistent matrix [Dede et al., 2015].

We then use (5) to estimate the logarithmic weight Z_k^p for the present iteration p, $1 \leq p \leq N_{\mathrm{MC}}$. If the number of iterations N_{MC} is large enough, we can approximate the probability density function (PDF) of Z_k by the histogram of its realizations Z_k^p. Figure 3 shows the PDF of a typical logarithmic weight obtained by the MC scheme assuming $N = 4$ criteria, $M = 15$ experts and that $c_{ij}^{(m)}$ are initially uniformly chosen inside [1/17 1] with a uniform distribution as discussed in the previous paragraph. We also show the Gaussian fit to the numerically computed PDF which is a very accurate approximation, indicating that the weights are Gaussian-distributed. Similar results are obtained for the other weights as well as for other values of N and M.

3.2. Validation of P_{RR} model

Now that the Gaussian approximation has been validated in Section 3.1, it is time to compare our theoretical model for relating the probability of rank reversal to the confidence levels discussed in Section 2 against a numerical scheme based on MC simulations. We examine the effect of several parameters such as the number of criteria

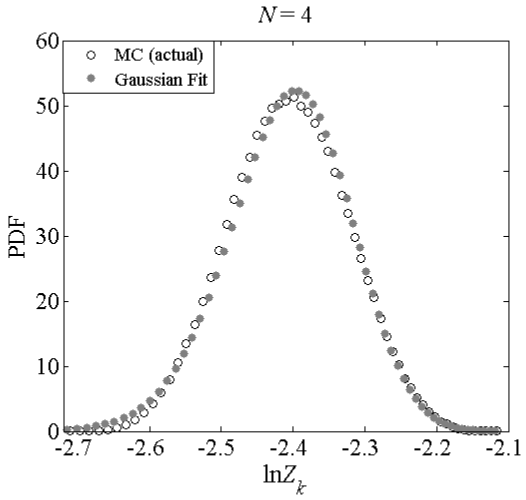


Figure 3.: Comparison of the numerically computed PDF of Z_k and its Gaussian fit.

N and the group of experts M in the accuracy of the model. The MC simulations used in this section are similar to those presented in Section 3.1 except that now we also iterate over the generation of the initial matrices. In each iteration q of the outer loop, we generate groups of M PWC matrices $\mathbf{a}^{(mq)} = \left[a_{ij}^{(mq)}\right]$ and M confidence level matrices $\mathbf{c}^{(mq)} = \left[c_{ij}^{(mq)}\right]$. Using (3) we can obtain the weights z_k^q corresponding to each group of matrices $\mathbf{a}^{(mq)}$. In the inner loop, we then perform the MC iterations described in Section 3.1 and obtain the random weights Z_k^{pq} where as before p denotes the inner iteration. We then count the number N_{RR}^q of times that the weights Z_k^{pq} have different ordering than that obtained by z_k^q. After completing the inner iterations, the numerical estimate of probability of rank reversal P_{NUM}^q at the q^{th} outer iteration is obtained as,

$$P_{\mathrm{NUM}}^q = \frac{N_{\mathrm{RR}}^q}{N_{\mathrm{MC}}} \tag{23}$$

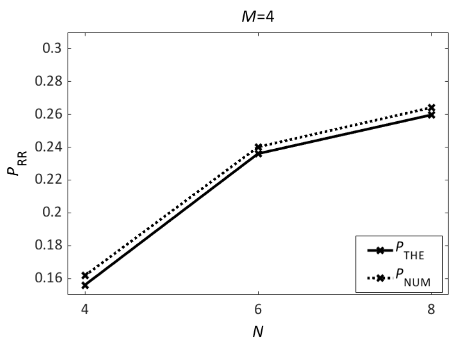
We then calculate the theoretical estimate P_{THE}^q based on the MVNCDF as outlined in Section 2.5. The mean numerical and theoretical probabilities of rank reversals P_{THE} and P_{NUM} are obtained as the averages of P_{THE}^q and P_{NUM}^q respectively,

$$P_{\mathrm{THE}} = \frac{1}{Q}\sum_{q=1}^{Q} P_{\mathrm{THE}}^q \tag{24}$$

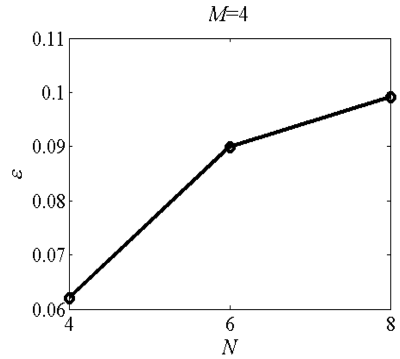
$$P_{\mathrm{NUM}} = \frac{1}{Q}\sum_{q=1}^{Q} P_{\mathrm{NUM}}^q \tag{25}$$

where Q is the total number of outer iterations. We also define the mean relative error ϵ between the two probability estimates,

$$\epsilon = \frac{1}{Q}\sum_{q=1}^{Q}\left|\frac{P^q_{\mathrm{NUM}} - P^q_{\mathrm{THE}}}{P^q_{\mathrm{NUM}}}\right| \tag{26}$$



(a)

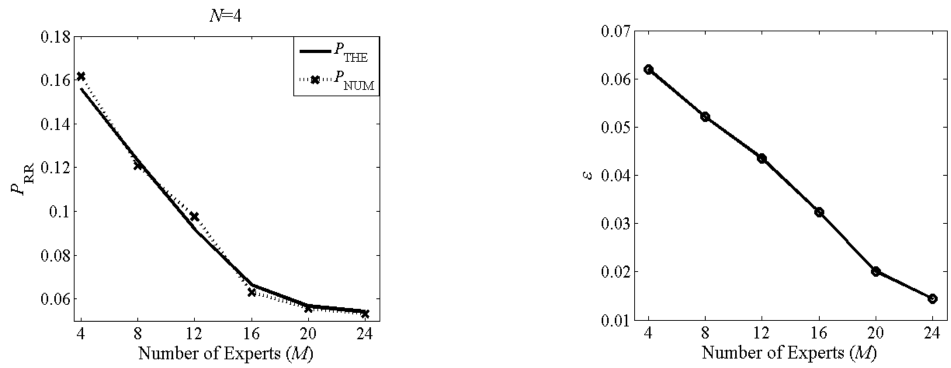


(b)

Figure 4.: (a) P_{RR} and (b) ϵ as a function of the number f criteria for $M = 4$ experts obtained by the MVNCDF and MC simulations.

To ascertain the accuracy of the theoretical estimation of the P_{RR}, one can check the agreement between P_{NUM} and P_{THE} as well as the value of ϵ. Figure 4 shows the values obtained for these quantities when the matrices $\mathbf{a}^{(mq)}$ are randomly generated consistent matrices and $\mathbf{c}^{(mq)}$ are generated as in Section 3.1. In the figure, we consider three different values of $N \leq 8$. The model easily allows to consider cases where $N > 8$, but in practical situations, that many criteria result in a large number of pairwise comparisons required. In Figure 4a, we readily see that the values of P_{RR} obtained by the MVNCDF or the numerical scheme are fairly close indicating that our theoretical model can indeed accurately estimate the probability of rank reversal even for a small number of experts. This is also evidenced by the low values of ϵ obtained in figure 4b. We also note that as the number of criteria N increases, the P_{RR} is also increased, which is to be expected, since rank reversal is typically easier to occur for a larger number of criteria. For $N = 4$, P_{RR} is lower than 5%, while for a larger number say $N = 6$, the P_{RR} is almost 26%. The accuracy of the MVNCDF is very good regardless of the number of criteria, since according to Figure 4b there is no more than 10% relative error in the estimation of the probability estimate, which is certainly acceptable from an application point-of-view.

In Figure 5a, we typically vary the number of experts M from 4 to 24 for $N = 4$ criteria. Inspection of the results reveals a very good agreement between the MVNCDF and the MC simulations for both smaller and larger groups of experts participating in the decision making process. The probabilities of rank reversal initially decrease rapidly at small values of M but gradually, the rate of decrease is reduced. This is consistent with the convergence properties of the probability of rank reversal [Dede et al., 2016]. Figure 5b also presents the mean absolute error, which seems to be lower than 6% even for a small group of experts participating in the PWC process. Interestingly enough the relative error is reduced as more experts are added to the group. This is



(a) (b)

Figure 5.: (a) P_{RR} and (b) ϵ, as a function of the group of experts M obtained by the MVNCDF and MC simulations for $N = 4$ criteria.

a consequence of the fact that for larger M the Gaussian approximation works even better, since more terms are added in the sum of (5).

In order to test the accuracy of the model at even lower confidence levels, we consider the effect of multiplying $c_{ij}^{(mq)}$ of all pairwise judgments of the experts with the same factor $0.5 < F < 1$ and replacing $c_{ij}^{(mq)}$ with $Fc_{ij}^{(mq)}$. In case where $Fc_{ij}^{(mq)}$ is lower than 1/17, this value is discarded and we choose $Fc_{ij}^{(mq)} = 1/17$, in accordance with the description analyzed in Section 2.3. In Figure 6, we show the results obtained in this case. The figure illustrates a very good agreement between the MVNCDF model and the MC simulations, even at small confidence levels ($F \cong 0.5$). It also seems that as the experts feel more certain about their judgment (the factor F is increasing) the probability of rank reversal becomes smaller reflecting the situation that is to be expected in practice. These conclusions hold for both $N = 4$ and $N = 6$ as shown in the figure.

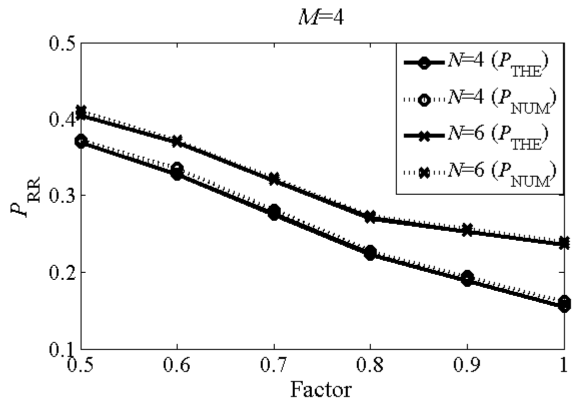
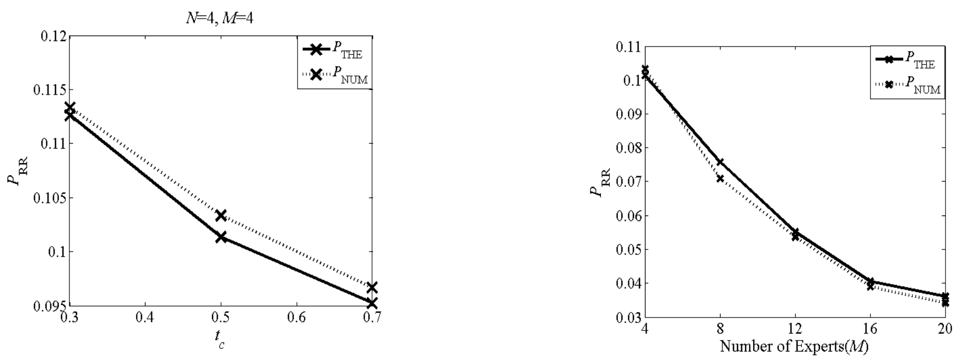


Figure 6.: P_{RR} as a function of a multiplicative factor for the confidence levels for $M = 4$ experts.

It would also be interesting to further investigate its accuracy under different circumstances such as the adoption of screening out judgments by defining threshold levels t_c, as analyzed in Section 2.5. For the screening method in question, we again



(a) (b)

Figure 7.: P_{RR} for $N = 4$ as a function of (a) different t_c for $M = 4$, (b) the number of experts for $t_c = 0.5$.

estimate the P_{RR} by using either the theoretical model or the MC simulations. In Figure 7a, the P_{RR} for several threshold coefficients is depicted for a number of $N = 4$ criteria and $M = 4$ experts. Inspection of the results reveals a very good agreement between the theoretical and the numerical procedure in this case as well. It is also interesting to notice that as the value of the threshold coefficient is increased, the P_{RR} is decreased. This is to be expected, since assuming larger t_c corresponds to accepting fewer judgments which have higher certainty. One should however bear in mind that the outcome is now based on the opinion of fewer experts on the average and it may be imperative to include more experts with high confidence levels in the comparisons. In Figure 7b, we present the P_{RR} by varying the number of experts for a threshold coefficient of 50% ($t_c = 0.5$). It seems once again that the P_{RR} converges after a number of 15 experts and the results comparing MVNCDF and MC are very again very close, proving the validity of the theoretical model.

4. Conclusions

In this paper, we have introduced a new framework against uncertainty in PWC matrices by incorporating the confidence levels in the pairwise comparison process. The method is based on the experts providing the confidence levels for their judgments. In this context, we presented a theoretical model for relating the "microscopic" confidence levels to the probability of rank reversal which is a "macroscopic" measure for the credibility of the outcome. We showed how the probability of rank reversal can be calculated accurately based on the MVNCDF of the successive average logarithmic weight differences, taking into account the confidence levels without resorting to Monte-Carlo iterations. Comparing the results with those estimated using a numerical model, we observed a very good agreement between the two approaches. The proposed approach is therefore an extension of the existing pairwise comparison method providing a way of incorporating the confidence levels in the judgments and assess the reliability of the end result.

As ideas for future studies, it would be interesting to investigate the combination of the confidence levels of each expert towards estimating an average confidence level for each judgement. Another future direction would be to account for some ordinal judgments in the confidence levels instead of numbers, such as certain, uncertain, etc. Finally, it would be interesting enough to investigate the robustness of the final outcomes to possible violations of the independence assumption defined for the experts judgments in this framework.

Acknowledgement

The research leading to these results has received funding from the Project " Productive 4.0: Electronics and ICT as enabler for digital industry and optimized supply chain management covering the entire product lifecycle" (Horizon 2020 ECSEL Joint Undertaking, and National Funding Authorities Grant Agreement number: 737459 — Productive4.0 — H2020-ECSEL-2016-2-IA-two-stage)

References

[Alho and Kangas, 1997] Alho, J. M. and Kangas, J. (1997). Analyzing uncertainties in experts' opinions of forest plan performance. *Forest Science*, 43(4):521–528.

[Alonso and Lamata, 2006] Alonso, J. A. and Lamata, M. T. (2006). Consistency in the analytic hierarchy process: a new approach. *International journal of uncertainty, fuzziness and knowledge-based systems*, 14(04):445–459.

[Anand et al., 2000] Anand, S. S., Yusuf, S., Vuksan, V., Devanesen, S., Teo, K. K., Montague, P. A., Kelemen, L., Yi, C., Lonn, E., Gerstein, H., et al. (2000). Differences in risk factors, atherosclerosis, and cardiovascular disease between ethnic groups in canada: the study of health assessment and risk in ethnic groups (share). *The lancet*, 356(9226):279–284.

[Banuelas and Antony, 2004] Banuelas, R. and Antony, J. (2004). Modified analytic hierarchy process to incorporate uncertainty and managerial aspects. *International Journal of Production Research*, 42(18):3851–3872.

[Barzilai, 1997] Barzilai, J. (1997). Deriving weights from pairwise comparison matrices. *Journal of the operational research society*, 48(12):1226–1232.

[Barzilai, 1998] Barzilai, J. (1998). Consistency measures for pairwise comparison matrices. *Journal of Multi-Criteria Decision Analysis*, 7(3):123–132.

[Beaumont, 2005] Beaumont, G. P. (2005). *Probability and random variables*. Elsevier.

[Beynon, 2002] Beynon, M. (2002). Ds/ahp method: A mathematical analysis, including an understanding of uncertainty. *European Journal of Operational Research*, 140(1):148 – 164.

[Boender et al., 1989] Boender, C., De Graan, J., and Lootsma, F. (1989). Multi-criteria decision analysis with fuzzy pairwise comparisons. *Fuzzy sets and Systems*, 29(2):133–143.

[Choo and Wedley, 2004] Choo, E. U. and Wedley, W. C. (2004). A common framework for deriving preference values from pairwise comparison matrices. *Computers & Operations Research*, 31(6):893–908.

[Dede et al., 2015] Dede, G., Kamalakis, T., and Sphicopoulos, T. (2015). Convergence properties and practical estimation of the probability of rank reversal in pairwise comparisons for multi-criteria decision making problems. *European Journal of Operational Research*, 241(2):458–468.

[Dede et al., 2016] Dede, G., Kamalakis, T., and Sphicopoulos, T. (2016). Theoretical estimation of the probability of weight rank reversal in pairwise comparisons.

European Journal of Operational Research, 252(2):587–600.

[Dede et al., 2011] Dede, G., Kamalakis, T., and Varoutas, D. (2011). Evaluation of optical wireless technologies in home networking: An analytical hierarchy process approach. *Journal of Optical Communications and Networking*, 3(11):850–859.

[Deng, 1999] Deng, H. (1999). Multicriteria analysis with fuzzy pairwise comparison. *International journal of approximate reasoning*, 21(3):215–231.

[Dong et al., 2015] Dong, Y., Chen, X., Li, C.-C., Hong, W.-C., and Xu, Y. (2015). Consistency issues of interval pairwise comparison matrices. *Soft Computing*, 19(8):2321–2335.

[Durbach et al., 2014] Durbach, I., Lahdelma, R., and Salminen, P. (2014). The analytic hierarchy process with stochastic judgements. *European Journal of Operational Research*, 238(2):552 – 559.

[Durbach and Stewart, 2012] Durbach, I. N. and Stewart, T. J. (2012). Modeling uncertainty in multi-criteria decision analysis. *European Journal of Operational Research*, 223(1):1 – 14.

[Entani and Sugihara, 2012] Entani, T. and Sugihara, K. (2012). Uncertainty index based interval assignment by interval ahp. *European Journal of Operational Research*, 219(2):379 – 385.

[Eskandari and Rabelo, 2007] Eskandari, H. and Rabelo, L. (2007). Handling uncertainty in the analytic hierarchy process: A stochastic approach. *International Journal of Information Technology & Decision Making*, 6(01):177–189.

[Farkas et al., 2003] Farkas, A., Lancaster, P., and Rózsa, P. (2003). Consistency adjustments for pairwise comparison matrices. *Numerical linear algebra with applications*, 10(8):689–700.

[Genz and Bretz, 2002] Genz, A. and Bretz, F. (2002). Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, 11(4):950–971.

[Gerdsri and Kocaoglu, 2007] Gerdsri, N. and Kocaoglu, D. F. (2007). Applying the analytic hierarchy process (ahp) to build a strategic framework for technology roadmapping. *Mathematical and Computer Modelling*, 46(7-8):1071–1080.

[Hahn, 2003] Hahn, E. D. (2003). Decision making with uncertain judgments: A stochastic formulation of the analytic hierarchy process. *Decision Sciences*, 34(3):443–466.

[Heldsinger and Humphry, 2010] Heldsinger, S. and Humphry, S. (2010). Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher*, 37(2):1–19.

[Hildebrand, 1987] Hildebrand, F. B. (1987). *Introduction to numerical analysis*. Courier Corporation.

[https://github.com/gdede hua/confidence$_levels_framework$,] https://github.com/gdede hua/confidence$_levels_framework$.

[Ishizaka and Labib, 2011] Ishizaka, A. and Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert systems with applications*, 38(11):14336–14345.

[Koczkodaj and Szarek, 2010] Koczkodaj, W. W. and Szarek, S. J. (2010). On distance-based inconsistency reduction algorithms for pairwise comparisons. *Logic Journal of IGPL*, 18(6):859–869.

[Kok and Lootsma, 1985] Kok, M. and Lootsma, F. (1985). Pairwise-comparison methods in multiple objective programming, with applications in a long-term energy-planning model. *European Journal of Operational Research*, 22(1):44–55.

[Kwiesielewicz and Van Uden, 2004] Kwiesielewicz, M. and Van Uden, E. (2004). Inconsistent and contradictory judgements in pairwise comparison method in the ahp. *Computers & Operations Research*, 31(5):713–719.

[Limayem and Yannou, 2007] Limayem, F. and Yannou, B. (2007). Selective assessment of judgmental inconsistencies in pairwise comparisons for group decision rating. *Computers & Operations Research*, 34(6):1824–1841.

[Mamat and Daniel, 2007] Mamat, N. J. Z. and Daniel, J. K. (2007). Statistical analyses on time complexity and rank consistency between singular value decomposition and the duality approach in ahp: A case study of faculty member selection. *Mathematical and Computer Modelling*, 46(7):1099–1106.

[Marimin et al., 2002] Marimin, M., Umano, M., Hatono, I., and Tamura, H. (2002). Hierarchical semi-numeric method for pairwise fuzzy group decision making. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 32(5):691–700.

[Mikhailov, 2004] Mikhailov, L. (2004). A fuzzy approach to deriving priorities from interval pairwise comparison judgements. *European Journal of Operational Research*, 159(3):687 – 704.

[Millet and Wedley, 2002] Millet, I. and Wedley, W. C. (2002). Modelling risk and uncertainty with the analytic hierarchy process. *Journal of Multi-Criteria Decision Analysis*, 11(2):97–107.

[Ramík and Korviny, 2010] Ramík, J. and Korviny, P. (2010). Inconsistency of pairwise comparison matrix with fuzzy elements based on geometric mean. *Fuzzy Sets and Systems*, 161(11):1604 – 1613. Theme: Decision Systems.

[Rice, 2006] Rice, J. (2006). *Mathematical statistics and data analysis*. Nelson Education.

[Saaty, 2003] Saaty, T. L. (2003). Decision-making with the ahp: Why is the principal eigenvector necessary. *European journal of operational research*, 145(1):85–91.

[Saaty, 2008a] Saaty, T. L. (2008a). *Decision making for leaders: the analytic hierarchy process for decisions in a complex world*. RWS publications.

[Saaty, 2008b] Saaty, T. L. (2008b). 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. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 102(2):251–318.

[Saaty and Vargas, 1984] Saaty, T. L. and Vargas, L. G. (1984). Inconsistency and rank preservation. *Journal of Mathematical Psychology*, 28(2):205–214.

[Saaty and Vargas, 1987] Saaty, T. L. and Vargas, L. G. (1987). Uncertainty and rank order in the analytic hierarchy process. *European Journal of Operational Research*, 32(1):107–117.

[Siraj et al., 2015] Siraj, S., Mikhailov, L., and Keane, J. A. (2015). Contribution of individual judgments toward inconsistency in pairwise comparisons. *European Journal of Operational Research*, 242(2):557 – 567.

[Stam and Silva, 1997] Stam, A. and Silva, A. P. D. (1997). Stochastic judgments in the ahp: The measurement of rank reversal probabilities. *Decision Sciences*, 28(3):655–688.

[Sugihara et al., 2004] Sugihara, K., Ishii, H., and Tanaka, H. (2004). Interval priorities in ahp by interval regression analysis. *European Journal of Operational Research*, 158(3):745 – 754.

[Sugihara and Tanaka, 2001] Sugihara, K. and Tanaka, H. (2001). Interval evaluations in the analytic hierarchy process by possibility analysis. *Computational intelligence*, 17(3):567–579.

[Triantaphyllou, 2000] Triantaphyllou, E. (2000). Multi-criteria decision making methods. In *Multi-criteria decision making methods: A comparative study*, pages 5–21. Springer.

[Yager, 2004] Yager, R. R. (2004). Modeling prioritized multicriteria decision making. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(6):2396–2404.

[Yang et al., 2013] Yang, X., Yan, L., and Zeng, L. (2013). How to handle uncertainties in ahp: The cloud delphi hierarchical analysis. *Information Sciences*, 222:384 – 404. Including Special Section on New Trends in Ambient Intelligence and Bio-inspired Systems.

[Zahir, 1991] Zahir, M. S. (1991). Incorporating the uncertainty of decision judgements in the analytic hierarchy process. *European Journal of Operational Research*, 53(2):206–216.