Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems ••• (••••) •••–•••
www.elsevier.com/locate/fss

Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms

Faran Ahmed, Kemal Kilic

Sabanci University, Orta Mahalle, Universite Caddesi, No:27, 34956 Tuzla, Istanbul, Turkey

Received 11 November 2016; received in revised form 12 August 2018; accepted 13 August 2018

Abstract

Analytical Hierarchical Process (AHP) along with fuzzy set theory has been used extensively in the Multi-Criteria Decision Making (MCDM) process in which fuzzy numbers are utilized to represent human judgments more realistically. Over the past couple of decades, numerous articles have been published proposing algorithms through which priority vector (or weight vector) can be calculated from fuzzy comparison matrices. The aim of this study is to conduct a comprehensive performance analysis of the most popular algorithms proposed in this domain in terms of accuracy of weights calculated from fuzzy comparison matrices. Such an analysis is much needed by the researchers and practitioners. However none is available. An experimental analysis is conducted and the performance of various algorithms are evaluated with varying three parameters i.e., the size of the comparison matrix, the level of fuzziness and the level of inconsistency. We found that modified Logarithmic Least Squares Method and Fuzzy Inverse of Column Sum Method (FICSM) generally outperformed other algorithms, while Fuzzy Extent Analysis (FEA) which is the most frequently used algorithm in the literature provides the least accurate results. Furthermore, it was observed that a modified version of FEA method significantly improved its performance.

Keywords: Fuzzy Analytic Hierarchy Process (FAHP); Multi-Criteria Decision Making (MCDM); Fuzzy Extent Analysis; Logarithmic Least Square Method

1. Introduction

Multiple Criteria Decision Making (MCDM) methodologies assist decision makers in choosing the best alternative while evaluating various competing and often conflicting criteria. Over the past many years, literature on MCDM has observed a steady growth [1] [2] while Analytical Hierarchical Process (AHP) proposed by Thomas L. Saaty [3] remains the most popular MCDM technique [4]. AHP seeks expert opinions in the form of pairwise comparisons and later derives ratio scales from comparison matrices which indicate the preferences of the decision maker among different alternatives in terms of the criteria as well as the preference i.e., weights of the criteria themselves. The

E-mail addresses: ahmedfaran@sabanciuniv.edu (F. Ahmed), ahmedfaran@sabanciuniv.edu (K. Kilic).

https://doi.org/10.1016/j.fss.2018.08.009
0165-0114/© 2018 Published by Elsevier B.V.

normalized weighted sum over the criteria provides an overall score associated with each available alternative and thus help decision maker to choose the best decision. The vast literature on AHP is mainly divided based on two different scales that can be used to record pairwise comparisons namely, scale based on crisp numbers (scale of 1–9) and scale based on fuzzy numbers. The original method as proposed by Saaty [3] uses the crisp scale of 1–9 in which decision maker preferences assessed with natural language labels (e.g. weak, normal, strong etc.,) are represented by one crisp number in the given scale and recorded in comparison matrices. Whereas Fuzzy AHP (FAHP) algorithms use fuzzy numbers to represent the same preferences and they are recorded in corresponding fuzzy comparison matrices. Main motivation behind incorporating fuzzy set theory into original AHP is based on the argument that human judgments and preferences cannot be accurately represented by crisp numbers due to the inherent uncertainty in human perception. Disregarding this fuzziness of the human behavior in the decision making process may lead to wrong decisions [5]. Therefore, in order to address this issue of vagueness and uncertainty, and to accurately transform human judgments into ratio scales, fuzzy set theory introduced by Zadeh [6] has been extensively incorporated into the original AHP in which the weighing scale is composed of fuzzy numbers. In FAHP, weights are calculated from fuzzy comparison matrices which are later used to rank the available alternatives together with the scores attained by the alternatives for each criterion. Therefore, determination of the weights from comparison matrices is one of the key steps of the process. In the conventional AHP these weights are shown to be the eigenvectors of the comparison matrix for a fully consistent decision maker [7]. In addition to the eigenvector approach, other techniques are also proposed in the literature among which arithmetic and geometric mean methods are the most frequently used ones. However in case of FAHP, calculating weights from fuzzy comparison matrices is not straightforward due to complexities associated with the arithmetics of fuzzy numbers and therefore over the past couple of decades, various algorithms have been proposed in the literature with an aim to accurately extract weights from fuzzy comparison matrices. Golany and Kress [8] provide a performance comparison of most commonly used six methods in original AHP, in which they used minimum violation, total deviation, conformity and robustness as criteria for performance analysis. They concluded that Modified Eigenvalue (MEV) [9] is the least effective method, while among the remaining five algorithms, each have their own weaknesses and advantages. Another comparative analysis is performed by Ishizaka and Lusti [10] in which they used Monte Carlo simulations to compare and evaluate four techniques to derive priority vector in original AHP including right eigenvalue method, left eigenvalue method, geometric mean and the mean of normalized values and conclude that number of contradictions increases with increase in the inconsistency as well as the size of the matrix. Some other similar studies that compare crisp AHP approaches in various aspects are also available in the literature [11] [12] [13] [14] [15]. Buyukozkan et al. [16] provide a review of Fuzzy AHP algorithms and list the characteristics and advantages and disadvantages for those methods which are structurally different. However, a *performance analysis* of FAHP algorithms similar to the ones that are available for the original AHP is not conducted so far. More and more papers are being published which apply FAHP as part of the solution process, however the choice of the FAHP technique used in the analysis seems to be arbitrary due to this gap in the literature. Therefore, in this study we attempt to carry out a detailed performance analysis of nine different FAHP methods in terms of accuracy of weights calculated from fuzzy comparison matrices. Such an analysis would address the gap in the literature and guide both the researchers and practitioners while choosing the most appropriate algorithm in their analysis. Generally speaking, most of the AHP literature assumes that the pairwise comparison matrix is complete. Therefore, the incomplete case is left out of the scope of this research and might be a good future research area. Yet, we kindly ask those readers who are interested in how incomplete preference relations can be treated to refer to a good literature review on the related topics [17]. The rest of the paper is organized as follows. We will first provide an overview of the FAHP methods for which performance analysis will be carried out in this study. In section 3, we will describe the methodology used to conduct the performance analysis. Next in section 4 we will discuss the main results and in the final section, conclusions and future research areas will be highlighted.

2. Fuzzy Analytic Hierarchical Process

In the literature various FAHP algorithms are available. A review of citation analysis on Google Scholar is conducted in order to assess the popularity of these algorithms. Five of the FAHP algorithms namely Logarithmic Least Square Method (LLSM, hereinafter referred to as Laarhoven) [18], LLSM with modified normalization (hereinafter referred to as Boender) [19], Fuzzy Extent Analysis (FEA, hereinafter referred to as Chang) [20], FEA with modified normalization (hereinafter referred to as Wang) [21] and Buckley's Geometric Mean method (hereinafter referred to

Please cite this article in press as: F. Ahmed, K. Kilic, Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.08.009

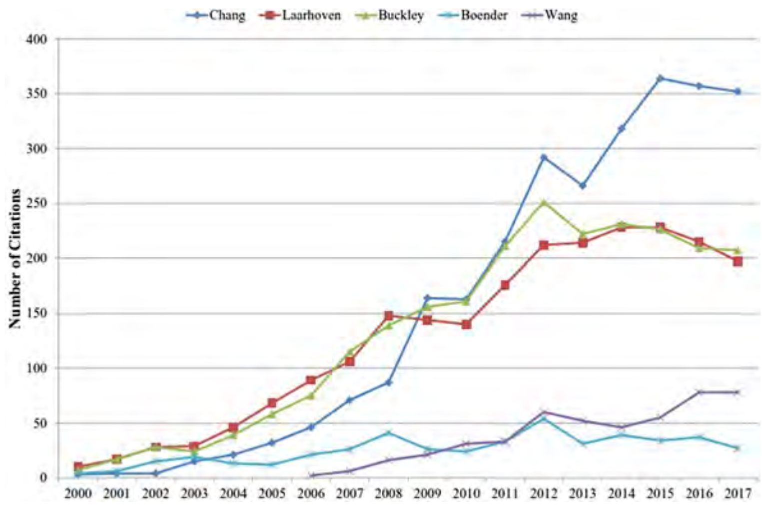

Fig. 1. Number of citations received by the five popular algorithms in Google scholar between years 2000 and 2016.

as Buckley) [22] are among the most popular algorithms in the literature. Fig. 1 illustrates the yearly citation history of these algorithms in Google scholar between 2000 and 2017. Review of this literature shows that some of these techniques (particularly Chang, Laarhoven and Buckley) are still being frequently referred to (and in some cases implemented as well) by the researchers.

To broaden the scope of our research, we also introduce four other methods (to the best of authors' knowledge, first time in the literature) in our comparative study which are referred to as Fuzzy Arithmetic Mean (FAM), Fuzzy Geometric Mean (FGM), Fuzzy Row Sum Method (FRSM) and Fuzzy Inverse of Column Sum Method (FICSM). FAM and FGM are the fuzzified version of the two popular conventional AHP methods. Note that FGM is different than the Buckley's Geometric Mean method [22] and should be considered as two separate approaches. FRSM is a modification of FEA method in which comparison of fuzzy numbers and their defuzzification through degree of possibility is replaced with centroid defuzzification. Whereas, FICSM is an intuitive approach in which inverse of column sums of comparison matrices are taken as the corresponding weights. Due to lack of space the details of the five popular FAHP algorithms are not provided here and interested readers are referred to the original papers for the details of the corresponding algorithms [18][22][19][23][21]. However for the sake of clarity, following we provide a brief overview of the latter four methods that are included in the performance analysis.

2.1. Arithmetic and geometric mean

Arithmetic mean and the geometric mean approaches are among the most popular techniques used in crisp AHP. These two techniques originates from the properties of a fully consistent comparison matrix. Note that even though the following discussion will focus only on extracting the weights of criteria, same steps are applicable to assess the score of an alternative for a particular criteria as well. Suppose that there are (n) criteria and one wants to extract the weight vector, i.e., $w = w_1, w_2, \cdots, w_n$ whereas w_i refers to the weight of the ith criteria

Recall that a fully consistent comparison matrix is as follows:

$$W' = \begin{pmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{pmatrix} \tag{1}$$

In the first step, we sum up each column which results in the following:

$$\left(\frac{w_1 + w_2, \cdots, w_n}{w_1}, \frac{w_1 + w_2, \cdots, w_n}{w_2}, \ldots, \frac{w_1 + w_2, \cdots, w_n}{w_n} \right) \tag{2}$$

Note that AHP assumes additive utility, i.e., the overall utility is weighted sum of individual utilities and the $\sum w_i = 1$. Therefore, the column sums provided by Equation (2) are equivalent to

$$\left(\frac{1}{w_1}, \frac{1}{w_2}, \cdots, \frac{1}{w_n}\right) \tag{3}$$

Next if one divides each element of the comparison matrix with its corresponding column sum, we end up with the following matrix which is referred to as W^N.

$$W^N = \begin{pmatrix} w_1 & w_1 & \cdots & w_1 \\ w_2 & w_2 & \cdots & w_2 \\ \vdots & \vdots & \ddots & \vdots \\ w_n & w_n & \cdots & w_n \end{pmatrix} \tag{4}$$

That is to say, for a fully consistent matrix, if one applies the above described normalization process, the resulting matrix W^N will be composed of column vectors which are same with each other and any one of them is the weight vector that we are trying to determine i.e., $(w_1, w_2, \cdots, w_n)$. However, since in practice the comparison matrix obtained from the decision makers are rarely consistent, the resulting normalized matrix would be composed of column vectors that are different from each other. Since each column is a candidate for the weight vector, and the source of the inconsistency cannot be detected, a reasonable thing to do is to average the columns of the normalized matrix W^N. The average can be obtained by either arithmetic means or the geometric means approaches. Equations (5) and (6) represents these two approaches, where w_i^J denotes the candidate weight associated with the ith criteria based on the jth column of the W^N.

$$A.M = \frac{\sum_{J=1}^{n} w_i^J}{n} \quad for\ i = 1, 2, \cdots, n \tag{5}$$

$$G.M = \left[\prod_{J=1}^{n} w_i^J\right]^{1/n} \quad for\ i = 1, 2, \cdots, n \tag{6}$$

These two approaches are commonly used in conventional AHP. In this paper we included the fuzzified version of these approaches in the performance analysis. In the fuzzified versions the starting comparison matrices will be composed of fuzzy numbers and fuzzy algebra will be employed while making calculations for the weight vector.

2.2. Row sum

The popular FEA Method proposed by Chang [23] consists of two parts. First part is Fuzzy Extent Analysis procedure in which row sums are calculated and afterwards they are normalized. The second part of the methodology is the defuzzification and/or ranking of the weights through principal of comparison of fuzzy numbers based on degree of possibility. In this paper as an alternate to the original ranking technique used in FEA, a centroid based defuzzification approach [24] is adopted and the modified technique, which is referred to as Fuzzy Row Sum Method (FRSM) is included in the performance analysis.

2.3. Inverse of column sum

The fourth algorithm that is developed and included in the performance analysis is referred to as Inverse of Column Sums. This algorithm is very intuitive and require very few arithmetic operations. For the fully consistent comparison matrix, column sum of each column is calculated by Equation (2). Since (due to additive utility assumption) $w_1 + w_2 + \cdots + w_n = 1$, the column sums are equivalent to

$$\frac{1}{w_1}, \frac{1}{w_2}, \cdots, \frac{1}{w_n} \tag{7}$$

Thus, the inverse of column sums will yield the weight vector $w_1, w_2, \cdots, w_n$ for fully consistent comparison matrix.

Adding these four algorithms on top of the five well known FAHP techniques previously discussed, we end up with nine different FAHP algorithms which are listed in Table 1.

Please cite this article in press as: F. Ahmed, K. Kilic, Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.08.009

Table 1
Selected FAHP algorithms.

Abbreviation	Method
Chang	Original Fuzzy Extent Analysis (FEA) proposed by Chang (1996)
Wang	Fuzzy Extent Analysis (FEA) with modified normalization proposed by Wang (2008)
Laarhoven	Logarithmic Least Squares Method (LLSM) proposed by Laarhoven (1983)
Boender	Logarithmic Least Squares Method (LLSM) with modified normalization proposed by Boender (1989)
Buckley	Geometric Mean Method propose by Buckley (1985)
FAM	Fuzzy Arithmetic Mean similar to arithmetic mean method in original AHP
FGM	Fuzzy Geometric Mean similar to geometric mean method in original AHP
FRSM	Fuzzy Row Sum Method which is Fuzzy Extent Analysis (FEA) with centroid defuzzification
FICSM	Fuzzy Inverse of Column Sum Method

Table 2
Fuzzy arithmetics.

Operation	Result
Addition	$(l_1\ m_1\ u_1) \oplus (l_2\ m_2\ u_2) = (l_1{+}l_2\ \ m_1{+}m_2\ \ u_1{+}u_2)$
Multiplication	$(l_1\ m_1\ u_1) \odot (l_2\ m_2\ u_2) = (l_1.l_2\ \ m_1.m_2\ \ u_1.u_2)$
Scalar Multiplication	$\lambda \odot (l\ m\ u) = (\lambda.l\ \ \lambda.m\ \ \lambda.u)$
Inverse (Triangular Fuzzy Number)	$(l\ m\ u)^{-1} = (1/u\ \ 1/m\ \ 1/l)$
Inverse (Trapezoidal Fuzzy Number)	$(l\ m\ n\ u)^{-1} = (1/m\ \ 1/l\ \ 1/u\ \ 1/n)$

2.4. Fuzzification of conventional AHP methods

Conventional AHP algorithms explained in section 2.1, 2.2 and 2.3 are extended in fuzzy case by replacing conventional arithmetic operations to fuzzy arithmetic operations. Let $(l_1\ m_1\ u_1)$ and $(l_2\ m_2\ u_2)$ be two triangular fuzzy numbers and $(l_1\ m_1\ n_1\ u_1)$ be a trapezoidal fuzzy number, than the basic fuzzy arithmetic operations required to extend conventional AHP algorithms to FAHP are tabulated in Table 2.

3. Design of experimental analysis

We conduct our performance analysis under various conditions. That is to say, three major control parameters are employed in the experiments, namely, size of the comparison matrices (n), level of fuzziness (α) and inconsistency of the decision maker (β). Ishizaka and Lusti [10] suggests that inconsistency and size of the matrix are factors that influences the performance of the conventional AHP algorithms. On top of these two parameters, we also included the level of fuzziness as the third control parameter for the performance analysis of the FAHP techniques.

By varying the level of these control parameters, we generate a set of matrices on which we apply the nine FAHP methods. As stated earlier, Golany and Kress [8] provide a methodology in order to generate comparison matrices with various levels of inconsistency. However, this technique is limited only for crisp AHP and cannot be replicated for comparison matrices consisting of fuzzy numbers. Therefore, a novel framework is developed through which random fuzzy comparison matrices can be generated for various control parameters as required by the experimental set up.

3.1. Algorithm to generate random fuzzy comparison matrix

Assume that there are n criteria and $w_1, w_2, \cdots, w_n$ are the subjective weights corresponding to these criteria for a decision maker. In practice the decision makers are asked to make pairwise comparisons and it is assumed that they use these weights in order to make the comparisons. In reality, inconsistency as well as fuzziness associated with the natural language are incorporated in the process and the resulting matrix would differ from a *theoretically* fully consistent comparison matrix.

In order to conduct the experimental analysis and determine the performance of the nine FAHP algorithms under various conditions, we developed a methodology that mimics the process used by a human decision maker. We

Please cite this article in press as: F. Ahmed, K. Kilic, Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.08.009

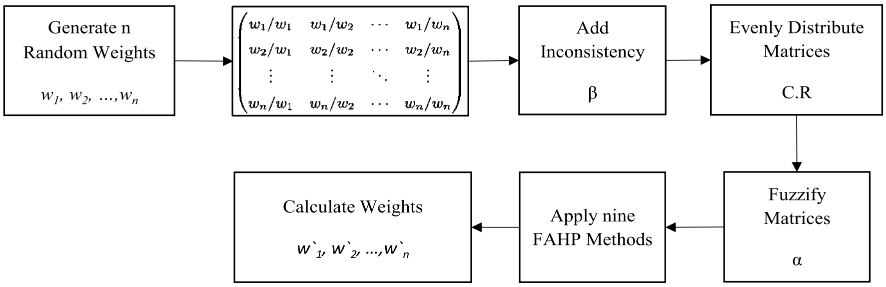

Fig. 2. Process diagram.

generated a random weight vector in each replication and using these subjective weights, we construct a perfectly consistent crisp comparison matrix using Equation (1). Afterwards, we added inconsistency into the matrices and fuzzified it through an approach elaborated below.

3.1.1. *Procedure to add inconsistency*

Various levels of inconsistency are added in the matrices through inconsistency parameter $\beta \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. Inconsistency interval $[a\ b]$ is generated for each element of the comparison matrix such that $a = w_i/w_j - \beta \times w_i/w_j$ and $b = w_i/w_j + \beta \times w_i/w_j$. A number x_{ij} is randomly selected from this interval and replaced with the corresponding element of the comparison matrix. Note that while altering matrices in such a manner, the reciprocal nature of the matrices is preserved. Six different values of β are used to generate inconsistency intervals and add desired level of inconsistency in the matrices. However, due to the randomness inherent in the process employed to incorporate inconsistency, it is possible to end up with a comparison matrix which is less or more inconsistent than aimed with the corresponding β parameter. Therefore, once inconsistency is added to the elements of the comparison matrix, the resulting measure of inconsistency is calculated through $C.R = CI/RI$ where $CI = \frac{\lambda_{max} - n}{n-1}$ as suggested by Satty [3], where λ_{max} is maximum eigenvalue of the comparison matrix and RI is the random index. This C.R measure is employed to classify matrices on different levels of inconsistency i.e., if C.R is between $0 - 0.03$ the corresponding comparison matrix is considered as a low level inconsistent matrix; if C.R is between $0.03 - 0.06$ it is considered as medium level inconsistent matrix and a comparison matrix with C.R between $0.06 - 0.1$ is regarded as highly inconsistent matrix. Note that any comparison matrix with C.R $\geq$0.1 is considered as a matrix not sufficiently consistent and discarded from the data set as suggested by Saaty [3].

To sum up the above discussion, although parameter β is employed to add inconsistency in the matrices, C.R measure is used to classify matrices on different inconsistency levels.

3.1.2. *Fuzzification of matrices*

The final step of synthetic data generation is fuzzification of matrices which is conducted through parameter $\alpha \in [0, 1]$. It converts a crisp number into a triangular fuzzy number $[l\ m\ u]$ such that $l = x_{ij} - \alpha$, $m = x_{ij}$ and $u = x_{ij} + \alpha$. Range of x_{ij} is $(\alpha, 9]$ and during the fuzzification process, the reciprocal nature of matrices is again preserved. That is to say, only one of the entry $x_{ij} \in (\alpha, 9]$ is fuzzified with this process and corresponding reciprocal entry x_{ji} is determined accordingly to preserve reciprocal nature of the resulting matrix (i.e., $x_{ji} = 1/x_{ij}$).

The overall process is illustrated in Fig. 2 and the pseudo code in provided in Algorithm 1. This process generates random matrices for $2 < n \leq 15$, representing comparison matrices assessed from decision makers with inconsistency (β) and fuzziness (α). We intend to make all associated Matlab codes and documentation available for other researchers working in similar domain, therefore, in a pattern similar to the one provided in [25] all necessary Matlab codes and documentation is made available on-line at [26].

3.1.3. *Final data set specifications*

After implementation of Algorithm 1, the *initial data set* is composed of 3840 ($= 4 \times 4 \times 6 \times 40$) matrices as there are four different matrix sizes, four different values of α, six different values of β and for each combination there are 40 replications of matrices. However, as stated earlier regarding the inconsistency levels, matrices are reclassified

Algorithm 1 Generate fuzzy comparison matrices.

n=3
while $n \leq 15$ **do**
  Initial Weights ← *Generate n Random Weight which sums upto 1*
  Perfectly Consistent Matrix ← w_i/w_j
  $\beta = 0$
  while $\beta \leq 1$ **do**
    Add Inconsistency
    $\alpha = 0.25$
    while $\alpha \leq 1$ **do**
      $[l\ m\ u]$ ← *Fuzzify Matrix*
      $\alpha = \alpha + 0.25$
    $\beta = \beta + 0.2$
  $n = n + 4$
Weights ← *Apply FAHP Algorithm*
Perfectly Consistent Matrix (W) ← *Weights*
Compatibility Index Value (CIV)

Table 3
Dataset classification and control parameters.

	Size of Matrix	Number of Matrices	Fuzzification level	Number of Matrices	Inconsistency levels	Number of Matrices
Control Parameters	n = 3	156	α = 0.25	156	CR = Low	208
	n = 7	156	α = 0.50	156	CR = Moderate	208
	n = 11	156	α = 0.75	156	CR = High	208
	n = 15	156	α = 1.00	156		
Total Matrices	624		624		624	

based on the C.R value (instead of β) which results in uneven distribution of matrices where there are large number of low level inconsistent matrices and less number of highly consistent matrices.

Therefore, this initial data set of 3840 matrices is reduced to total of 624 matrices which are *evenly* distributed on all experimental parameters. Note that for each combination there are 13 replication of matrices and the *final data set* is composed of 624 ($= 4 \times 4 \times 3 \times 13$) matrices. Table 3 tabulates final control parameters employed in our comparative study as well as number of matrices used in the analysis.

3.1.4. *Fuzzy inconsistency measure*

When preferences are elicited in the form of crisp numbers, the universally accepted way of measuring inconsistency of comparison matrices is through maximum eigenvalue as explained in section 3.1.1. However, when preferences are represented through fuzzy numbers or membership functions, measuring inconsistency is still a debatable issue [27]. Wang et al. adopts the concept of feasible region to represent inconsistency [28]. Islam et al. proposes a Lexicographic Goal Programming (LGP) method to derive weights from pairwise inconsistent interval judgment matrices as well as an algorithm for identification and modification of inconsistent bounds [29]. Another method uses distance based consistency index is devised in which distance between fuzzy comparison matrix and a specific metric function is used as a measure of inconsistency [30]. Yet another methodology i.e., additive consistency measure is discussed in which a consistency index and threshold of fuzzy preference relations are defined so as to measure whether fuzzy preference relations is sufficiently consistent or not [31]. While another study introduces strong transitivity conditions using route matrices and diagraphs and proposes an adjustment procedure to fix inconsistencies in comparison matrices [32].

Similarly, various other methods have been proposed in the literature, which employ different inconsistency measures depending on specific needs of that particular study. Through this study, we intend to conduct a comprehensive comparative analysis involving various techniques based on experiment parameters which are consistent across all FAHP algorithms. Therefore, we decided to adopt a uniform inconsistency measure as proposed by Saaty [3] to measure inconsistencies of matrices before fuzzifying them. This approach helps us find true measure of inconsistencies added in our synthetic data and helps in classifying matrices more effectively based on the selected inconsistency measure.

Please cite this article in press as: F. Ahmed, K. Kilic, Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.08.009

3.2. Metrics of performance analysis

Fuzzy comparison matrices were constructed by assuming normalized crisp weights. Afterwards, the nine FAHP methods included in our comparative analysis is applied to these random fuzzy comparison matrices and priority vector or weight vectors are calculated. Among all the FAHP methods, only FEA methods (Chang and Wang) computes the defuzzified weights through principle of comparison of fuzzy numbers based on degree of possibility, while rest of the FAHP methods calculate weights in the form of fuzzy numbers. In order to effectively conduct performance analysis, these fuzzy weights are defuzzified using centroid defuzzification technique.

For our study, the metrics of performance analysis will be Compatibility Index Value (CIV) [33] which provides a measure of the deviation between pairwise comparison matrix provided by the decision maker (which in our case will be randomly generated pairwise inconsistent matrix) and matrix $W = (w_i/w_j)$ constructed from the derived priority vector from fuzzy comparison matrix. Let $A = (a_{ij})$ be randomly generated inconsistent matrix and $W = (w_i/w_j)$ be perfectly consistent matrix constructed from derived priority vector, then CIV is defined as

$$CIV = n^{-2}.e^T A \circ W^T e \tag{8}$$

where n is the size of the matrix and $e^T A \circ W^T e$ is the Hadamard product of matrix A and W^T. Note that if A is perfectly consistent matrix than both matrices A and W will be similar and CIV becomes one, else it will have value greater than one.

In order to carry out statistical testing, Analysis of Variance (ANOVA) tests are conducted under various experimental conditions. These results are summarized in the next section.

4. Computational results and discussions

In this section, the results of the performance analysis conducted on the nine FAHP methods will be discussed. In order to determine the effect of the three experimental parameters, we employed one-way ANOVA (also referred to as a between-subjects ANOVA or one-factor ANOVA) which will help in determining any statistical differences between the mean differences of CIV while we change the experimental parameters. Note that one-way ANOVA is an omnibus test statistic and therefore cannot tell which specific groups of data are significantly different from each other; rather it just provide information that at least two of the groups are significantly different from each other. Therefore, for detailed analysis, we also conducted a post hoc test in order to analyze the results in more detail.

However, in order to conduct one-way ANOVA the following six assumptions must be satisfied:

1. Dependent variables are continuous.
2. Independent variable consists of two or more categorical, independent groups.
3. There is no relationship between the observations in each group or between the groups themselves i.e., independence of observations must hold.
4. There should be no significant outliers, which might have a negative effect on the one-way ANOVA, thus reducing the validity of the results.
5. Dependent variable should be approximately normally distributed for each category of the independent variable. However, one-way ANOVA only requires approximately normal data because it is quite “robust” to violations of normality, meaning that assumption can be a little violated and still provide valid results.
6. Homogeneity of variances must hold.

As explained in the previous section, the structure of the data obtained from the experimental analysis ensures that the first three assumptions are satisfied. On the other hand, we employed the methodology provided in [34] in order to check the latter three assumptions. In order to check the validity of the fourth assumption box-plots are utilized and some outliers are determined. These outliers were neither result of data entry errors nor due to measurement errors but determined to be genuinely unusual values. There are various ways through which these outliers can be treated [35]. We resolve this problem by conducting the analysis with and without these outliers and no significant difference in the results were observed. So we decided to keep these values in our analysis. In order to check the validity of the fifth assumption we conducted both Kolmogorov–Smirnov test as well as Shapiro–Wilk test for each subgroup and determined that our data is normally distributed for each sub group. To test the final assumption, we conducted

Table 4
Selected nine FAHP methods.

Model	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Boender	624	1.06203	0.04618	0.00185	1.05840	1.06566	1.00003	1.37271
Buckley	624	1.06667	0.04772	0.00191	1.06291	1.07042	1.00003	1.20900
Chang	624	1.23835	0.38388	0.01537	1.20817	1.26853	1.00146	6.48521
FAM	624	1.14855	0.14130	0.00566	1.13744	1.15966	1.00003	2.10831
FGM	624	1.06707	0.04764	0.00191	1.06333	1.07082	1.00003	1.21053
FICSM	624	1.05968	0.04665	0.00187	1.05601	1.06335	1.00003	1.17599
FRSM	624	1.10803	0.09460	0.00379	1.10060	1.11547	1.00006	1.67369
Laarhoven	624	1.06675	0.04763	0.00191	1.06300	1.07049	1.00005	1.32394
Wang	624	1.30748	0.41495	0.01661	1.27486	1.34010	1.00111	6.03032
Total	5616	1.12496	0.21717	0.00290	1.11927	1.13064	1.00003	6.48521

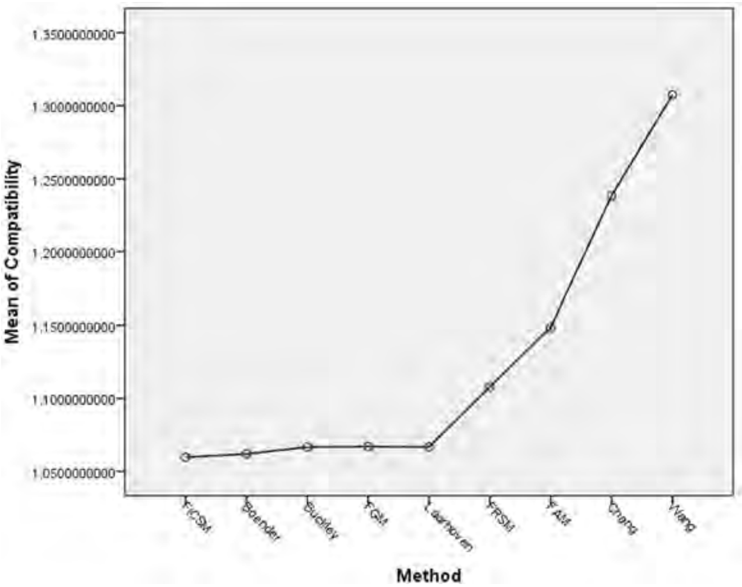

Fig. 3. Estimated marginal means of compatibility index.

Table 5
Welch ANOVA analysis between nine different methods.

Test	Statistic[a]	df1	df2	Sig.
Welch	84.146	8	2315.467	.000

[a] *Asymptotically F distributed.*

Levene's test for equality of variances. The results suggested that the assumption of homogeneity of variances was violated. Hence, as suggested in [36] we decided to conduct a modified version of one-way ANOVA which is Welch one-way ANOVA in the analysis.

4.1. Comparison of selected nine FAHP methods

We first tabulate overall descriptive statistics which shows that mean CIV for FICSM (1.05968, ±0.04665) is lowest and for Wang (1.30748, ±0.41495) is highest. Same is illustrated in Table 4 and Fig. 3.

In order to conclude that these differences are significant, one-way Welch ANOVA test is conducted for which results are tabulated in Table 5.

Note that Welch ANOVA results only imply that group means differ. It does not show in which particular way these group means differ among various subgroups. Therefore, post hoc test is conducted in order to investigate the significant differences more effectively.

Table 6
Games Howell post hoc test – Comparison of FAHP algorithms.

Method (I)	Method (J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
FICSM	Boender	-0.00235	0.00263	0.99327	-0.01052	0.00581
FICSM	Buckley	-0.00699	0.00267	0.18104	-0.01529	0.00131
FICSM	Chang	**-0.17867**	0.01548	0.00000	-0.22685	-0.13049
FICSM	FAM	**-0.08887**	0.00596	0.00000	-0.10740	-0.07034
FICSM	FGM	-0.00740	0.00267	0.12509	-0.01569	0.00090
FICSM	FRSM	**-0.04835**	0.00422	0.00000	-0.06148	-0.03523
FICSM	Laarhoven	-0.00707	0.00267	0.16753	-0.01536	0.00122
FICSM	Wang	**-0.24780**	0.01672	0.00000	-0.29983	-0.19577

*. The mean difference is significant at the 0.05 level.

Significantly Better | Better | Inferior | Significantly Inferior

	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00464	-0.17632	-0.08652	-0.00505	0.00235	-0.04600	-0.00472	-0.24545
Buckley		-0.17168	-0.08188	-0.00041	0.00699	-0.04137	-0.00008	-0.24081
Chang			0.08980	0.17127	0.17867	0.13032	0.17160	-0.06913
FAM				0.08147	0.08887	0.04052	0.08180	-0.15893
FGM					0.00740	-0.04096	0.00033	-0.24040
FICSM						-0.04835	-0.00707	-0.24780
FRSM							0.04128	-0.19945
Laarhoven								-0.24073

Fig. 4. Heat map – Mean CIV differences between nine FAHP methods. * *Sample read from the heat map: mean CIV of Boender is lower by* 0.00464 *as compared to Buckley and this difference is not significant mean CIV of Boender is lower by* 0.17632 *as compared to Chang and this difference is significant*.

Table 7
Welch ANOVA analysis between different matrix sizes.

Test	Statistic[a]	df1	df2	Sig.
Welch	27.866	3	3053.866	.000

[a] *Asymptotically F distributed.*

Since the homogeneity of variance assumption is violated, Games-Howell post hoc test is conducted instead of commonly used LSD or Tuckey test. Results of Games-Howell post hoc test corresponding to FICSM vs other FAHP methods are tabulated in Table 6. Similar analysis is conducted for each one of the algorithms and results of these analyses are graphically presented in Fig. 4 as heat map.

Fig. 4 shows that the mean CIV for FICSM algorithm is lower when compared with other algorithms and these differences are significant for FAM, FRSM, Chang and Wang.

We employed a similar methodology to determine if the above conclusions are also valid under various experimental conditions i.e., size of the matrix (n), fuzzification parameter (α) and inconsistency levels $(C.R)$ which are discussed below.

4.2. Matrix size

Welch one way ANOVA test (Table 7) shows that mean CIV is statistically different for selected four different matrix sizes, $F(3, 3053.866) = 27.866$, $p < .005$.

Games Howell post hoc test is conducted for each FAHP algorithm separately in order to investigate these differences more effectively. Mean CIV differences are illustrated as heat map in Fig. 5.

Where as I(n) and J(n) refer to mean CIV with different matrix sizes. Result shows that in general, mean CIV is lowest for lower matrix sizes and it increases as the matrix size is increased with the only exception for FEA (Chang

Significantly Better | Better | Inferior | Significantly Inferior

(I) n	(J) n	Boender	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
3	7	**-0.03499**	**-0.04622**	0.03251	**-0.09474**	**-0.04631**	**-0.03335**	**-0.06340**	**-0.03869**	**0.20329**
3	11	**-0.04375**	**-0.05755**	-0.10877	**-0.16414**	**-0.05571**	**-0.04857**	**-0.10690**	**-0.04420**	0.08642
3	15	**-0.04733**	**-0.06128**	-0.05394	**-0.21140**	**-0.05946**	**-0.05199**	**-0.14427**	**-0.04727**	**0.17501**
7	11	-0.00876	-0.01133	**-0.14128**	**-0.06940**	-0.00939	**-0.01522**	**-0.04351**	-0.00551	**-0.11687**
7	15	-0.01234	**-0.01506**	-0.08646	**-0.11666**	-0.01314	**-0.01864**	**-0.08087**	-0.00859	-0.02829
11	15	-0.00358	-0.00373	0.05482	**-0.04727**	-0.00375	-0.00342	**-0.03736**	-0.00307	0.08858

Fig. 5. Post hoc test – Mean CIV differences (I–J) at different matrix sizes for nine FAHP methods.

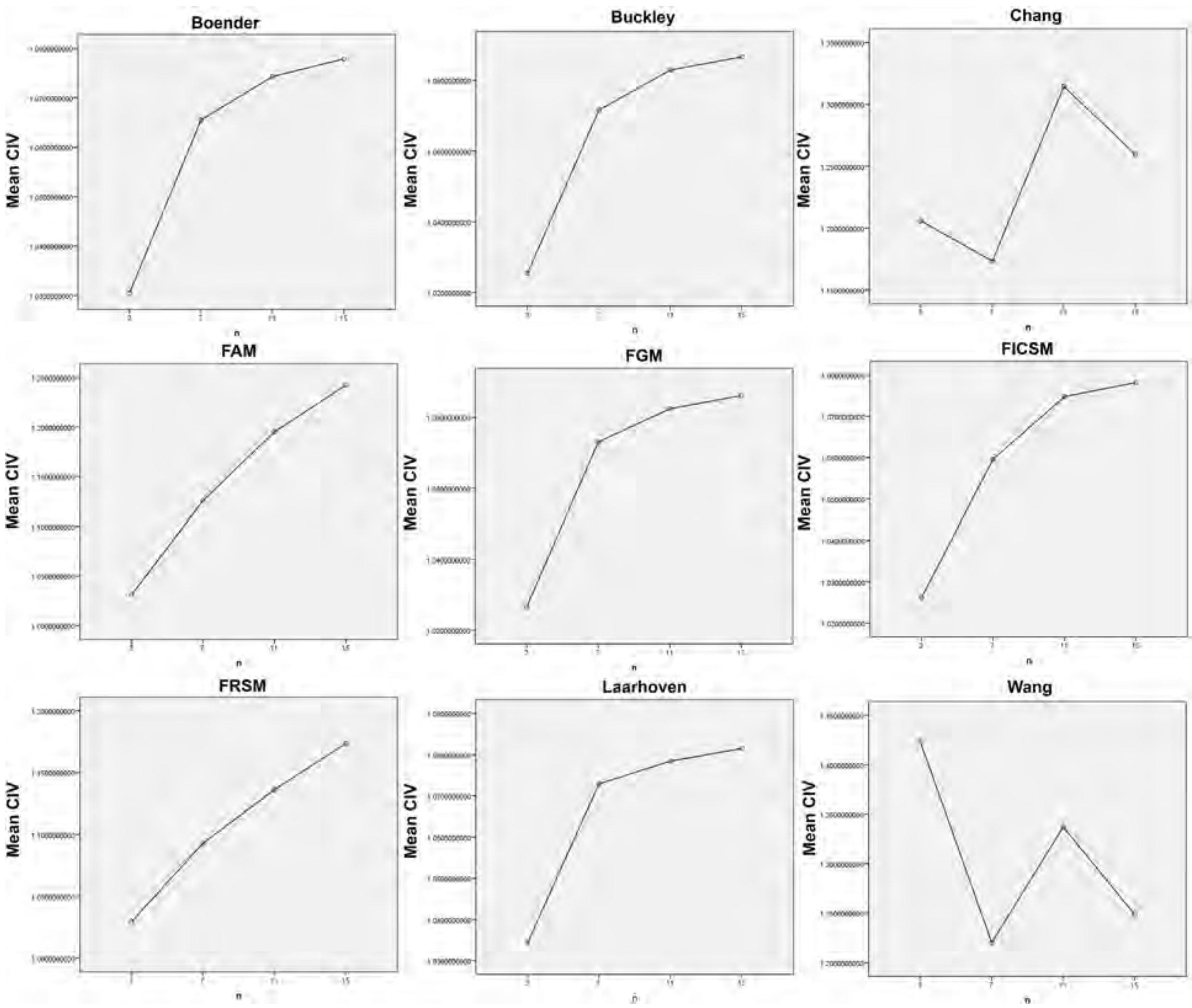

Fig. 6. Estimated marginal means of CIV.

and Wang). Otherwise, results are consistent for all other FAHP algorithms as illustrated in Fig. 6. That is to say, for all FAHP algorithms (except Chang and Wang), mean CIV increases as matrix size is increased.

Next we compared performance of selected nine FAHP methods under different matrix sizes. Results are graphically illustrated as heat map in Fig. 7 which shows that at $n = 3$ the best performing algorithm is Buckley, while at $n = 7$ FICSM is the best performing algorithm. Boender is the best performing algorithm at higher matrix sizes i.e., $n = 11$, $n = 15$. This result will serve as a useful guideline for practitioners in their decision making problems.

Please cite this article in press as: F. Ahmed, K. Kilic, Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.08.009

Significantly Better | Better | Inferior | Significantly Inferior

(a)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	0.00511	**-0.17529**	-0.00047	0.00381	0.00431	0.00112	-0.00370	**-0.39315**
Buckley		**-0.18040**	-0.00558	-0.00130	-0.00080	-0.00399	-0.00881	**-0.39826**
Chang			**0.17482**	**0.17909**	**0.17960**	**0.17641**	**0.17159**	**-0.21786**
FAM				0.00428	0.00478	0.00159	-0.00323	**-0.39268**
FGM					0.00050	-0.00269	-0.00750	**-0.39695**
FICSM						-0.00319	-0.00800	**-0.39746**
FRSM							-0.00482	**-0.39427**
Laarhoven								**-0.38945**

(b)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00612	**-0.10778**	**-0.06022**	-0.00752	0.00595	**-0.02729**	-0.00739	**-0.15486**
Buckley		**-0.10166**	**-0.05409**	-0.00140	0.01207	**-0.02116**	-0.00127	**-0.14874**
Chang			0.04757	**0.10027**	**0.11374**	**0.08050**	**0.10039**	-0.04708
FAM				**0.05270**	**0.06617**	**0.03293**	**0.05282**	**-0.09465**
FGM					0.01347	**-0.01977**	0.00012	**-0.14735**
FICSM						**-0.03324**	-0.01335	**-0.16081**
FRSM							**0.01989**	**-0.12758**
Laarhoven								**-0.14747**

(c)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00869	**-0.24030**	**-0.12085**	-0.00815	-0.00051	**-0.06203**	-0.00415	**-0.26297**
Buckley		**-0.23161**	**-0.11216**	0.00054	0.00818	**-0.05334**	0.00455	**-0.25428**
Chang			0.11945	**0.23215**	**0.23979**	**0.17827**	**0.23616**	-0.02267
FAM				**0.11270**	**0.12035**	**0.05882**	**0.11671**	**-0.14212**
FGM					0.00764	-0.05388	0.00400	**-0.25482**
FICSM						-0.06152	-0.00364	**-0.26246**
FRSM							**0.05788**	**-0.20094**
Laarhoven								**-0.25883**

(d)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00884	**-0.18190**	**-0.16454**	-0.00832	-0.00035	**-0.09582**	-0.00364	**-0.17081**
Buckley		**-0.17306**	**-0.15570**	0.00052	0.00850	**-0.08697**	0.00520	**-0.16197**
Chang			0.01736	**0.17358**	**0.18155**	0.08609	**0.17826**	0.01109
FAM				**0.15622**	**0.16419**	**0.06872**	**0.16090**	-0.00627
FGM					0.00797	**-0.08750**	0.00468	**-0.16249**
FICSM						**-0.09547**	-0.00329	**-0.17046**
FRSM							**0.09217**	**-0.07499**
Laarhoven								**-0.16717**

Fig. 7. Heat map – Mean CIV differences of nine FAHP methods at different matrix sizes (a) $n = 3$, (b) $n = 7$, (c) $n = 11$, (d) $n = 15$.

Table 8
Welch ANOVA analysis between different levels of fuzzification.

Test	Statistic[a]	df1	df2	Sig.
Welch	8.689	3	2936.801	0.000

[a] *Asymptotically F distributed.*

4.3. Fuzzification parameter

One-way Welch ANOVA test for different levels of fuzzification is tabulated in Table 8 which shows that mean CIV is statistically different for four different levels of fuzzification, Welch $F(3, 2936.801) = 8.689$, $p < .005$.

Games Howell post hoc test at different fuzzification levels for nine FAHP models is presented in the form of heat map which is illustrated in Fig. 8.

Fig. 8 shows that performance of all FAHP algorithms except Chang and Wang decreases as the fuzzification level are increased. This trend is more clearly illustrated in Fig. 9 which shows that, the performance of FEA (Chang and Wang) increases at higher levels (0.75, 1.00) of fuzzification, whereas, for all other FAHP methods, performance decreases as fuzzification levels are increased. Review of literature on FAHP shows that FEA (Chang and Wang) has been extensively applied in various decision making environments. Therefore, FEA is the preferred choice of practitioners, then this study recommends to use higher levels of fuzzification in triangular fuzzy numbers.

Please cite this article in press as: F. Ahmed, K. Kilic, Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.08.009

		Significantly Better		Better		Inferior		Significantly Inferior		
(I) alpha	**(J) alpha**	**Boender**	**Buckley**	**Chang**	**FAM**	**FGM**	**FICSM**	**FRSM**	**Laarhoven**	**Wang**
0.25	0.50	-0.00756	-0.01097	0.12466	**-0.05050**	-0.01115	-0.00372	-0.01876	-0.01041	**0.16503**
0.25	0.75	-0.00778	-0.01008	**0.20704**	**-0.06202**	-0.01042	-0.00405	-0.02443	-0.01073	**0.27610**
0.25	1.00	**-0.02024**	**-0.02841**	**0.23764**	**-0.12255**	**-0.02894**	-0.00883	**-0.06140**	**-0.02914**	**0.35003**
0.50	0.75	-0.00022	0.00089	0.08238	-0.01152	0.00073	-0.00033	-0.00568	-0.00032	0.11107
0.50	1.00	-0.01267	**-0.01744**	0.11299	**-0.07205**	**-0.01779**	-0.00511	**-0.04265**	**-0.01873**	**0.18500**
0.75	1.00	-0.01245	**-0.01833**	0.03061	**-0.06054**	**-0.01852**	-0.00478	**-0.03697**	**-0.01841**	**0.07393**

Fig. 8. Post hoc test – Mean CIV differences at different levels of α for nine FAHP methods.

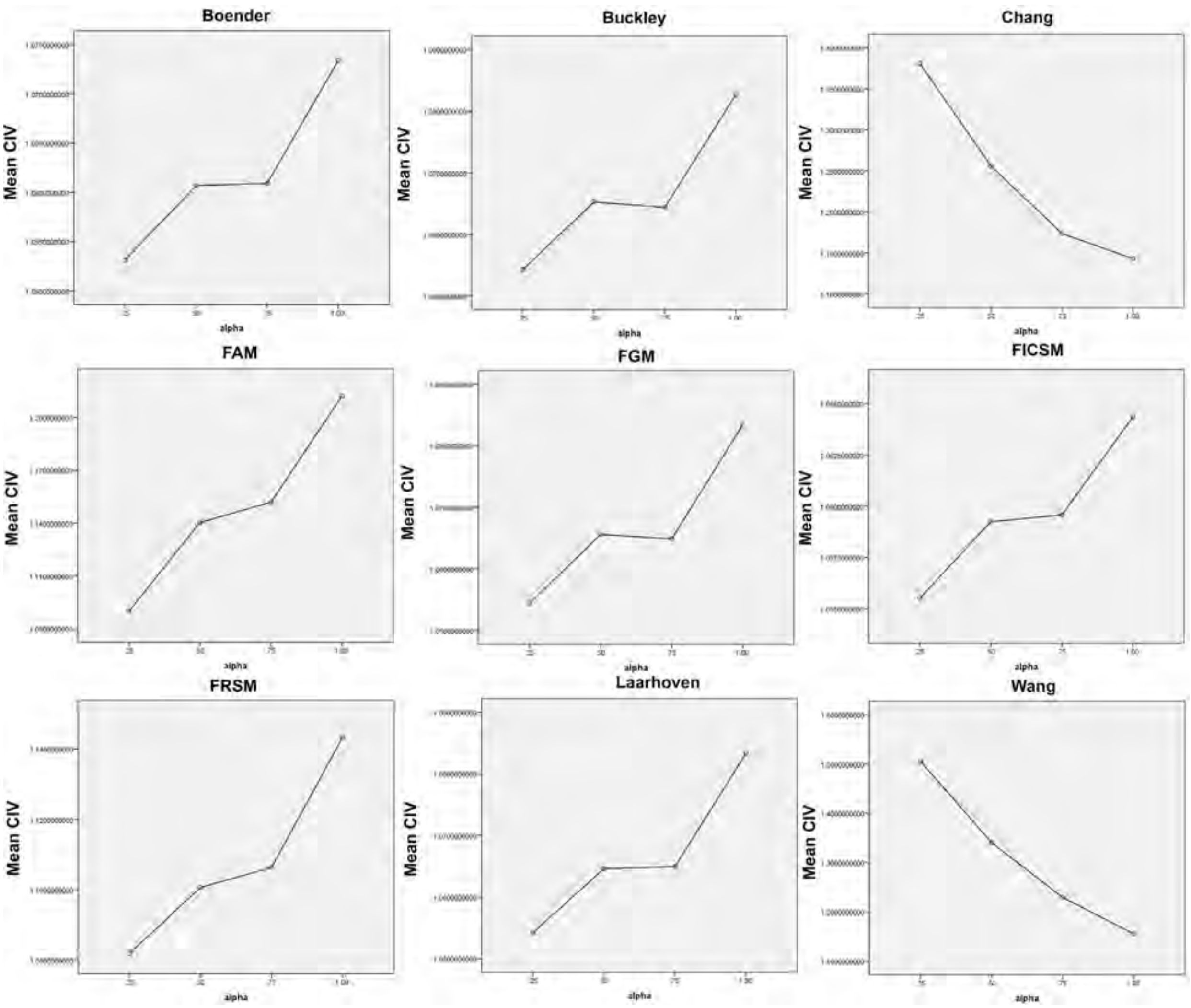

Fig. 9. Estimated marginal means of CIV.

We also conducted comparison of nine FAHP methods under different levels of fuzzification and results are graphically illustrated as heat map in Fig. 10 which shows that at lower levels of fuzzification ($\alpha = 0.25$), Boender outperforms all other algorithms. As the fuzzification levels are increased, FICSM is the best performing algorithm. As stated before, performance of FEA (Chang and Wang) seems to improve at higher levels of fuzzification, however, this improved performance is not significant and its performances remains inferior when compared with other FAHP algorithms.

Please cite this article in press as: F. Ahmed, K. Kilic, Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.08.009

Significantly Better | Better | Inferior | Significantly Inferior

(a)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00117	-0.32755	-0.03665	-0.00131	-0.00240	-0.02875	-0.00105	-0.45213
Buckley		-0.32638	-0.03548	-0.00014	-0.00123	-0.02758	0.00012	-0.45097
Chang			0.29090	0.32624	0.32515	0.29880	0.32650	-0.12459
FAM				0.03533	0.03425	0.00790	0.03560	-0.41549
FGM					-0.00108	-0.02744	0.00027	-0.45082
FICSM						-0.02635	0.00135	-0.44974
FRSM							0.02770	-0.42338
Laarhoven								-0.45109

(b)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00457	-0.19533	-0.07959	-0.00490	0.00145	-0.03994	-0.00389	-0.27954
Buckley		-0.19076	-0.07501	-0.00033	0.00602	-0.03537	0.00068	-0.27497
Chang			0.11574	0.19043	0.19678	0.15539	0.19144	-0.08421
FAM				0.07468	0.08104	0.03964	0.07569	-0.19996
FGM					0.00635	-0.03504	0.00101	-0.27464
FICSM						-0.04139	-0.00534	-0.28099
FRSM							0.03605	-0.23960
Laarhoven								-0.27565

(c)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00346	-0.11273	-0.09088	-0.00395	0.00134	-0.04540	-0.00399	-0.16825
Buckley		-0.10926	-0.08742	-0.00049	0.00480	-0.04194	-0.00053	-0.16479
Chang			0.02185	0.10878	0.11407	0.06733	0.10874	-0.05552
FAM				0.08693	0.09222	0.04548	0.08689	-0.07737
FGM					0.00529	-0.04145	-0.00004	-0.16430
FICSM						-0.04674	-0.00533	-0.16959
FRSM							0.04141	-0.12285
Laarhoven								-0.16426

(d)	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00934	-0.06967	-0.13896	-0.01002	0.00901	-0.06992	-0.00995	-0.08187
Buckley		-0.06033	-0.12962	-0.00068	0.01835	-0.06058	-0.00061	-0.07253
Chang			-0.06930	0.05965	0.07868	-0.00025	0.05972	-0.01220
FAM				0.12895	0.14798	0.06905	0.12902	0.05710
FGM					0.01903	-0.05990	0.00007	-0.07185
FICSM						-0.07893	-0.01896	-0.09088
FRSM							0.05997	-0.01195
Laarhoven								-0.07192

Fig. 10. Heat map – Mean CIV differences of nine FAHP methods at different fuzzification levels (a) $\alpha = 0.25$, (b) $\alpha = 0.50$, (c) $\alpha = 0.75$, (d) $\alpha = 1.00$.

Table 9
Welch ANOVA analysis between different levels of inconsistency.

Test	Statistic[a]	df1	df2	Sig.
Welch	92.232	2	3663.228	0.000

[a] *Asymptotically F distributed.*

4.4. Inconsistency levels

One-way Welch ANOVA test is conducted for various levels of consistency and results are tabulated in Table 9 which shows that mean CIV is statistically different for three levels of inconsistency level, Welch $F(2, 3663.228) = 92.232$, $p < .005$.

Results from Games Howell post hoc test for different levels of inconsistency for all FAHP methods is graphically illustrated as heat map in Fig. 11 which shows that performance of all algorithms is higher at lower inconsistency levels as expected. Same is illustrated in Fig. 12.

Furthermore, performance of all FAHP methods is analyzed at different levels of inconsistency and results are illustrated as heat map in Fig. 13.

Please cite this article in press as: F. Ahmed, K. Kilic, Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.08.009

Significantly Better	Better	Inferior	Significantly Inferior

(I) CR	(J) CR	Boender	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Low	Medium	-0.03664	-0.03511	-0.01550	-0.05118	-0.03518	-0.03979	-0.04684	-0.03532	-0.03991
Low	High	-0.08276	-0.08076	-0.14290	-0.11336	-0.08038	-0.09095	-0.10008	-0.08120	-0.12882
Medium	High	-0.04611	-0.04565	-0.12740	-0.06217	-0.04520	-0.05115	-0.05325	-0.04587	-0.08890

Fig. 11. Post hoc test – Mean CIV differences at different levels of inconsistency (C.R) for nine FAHP methods.

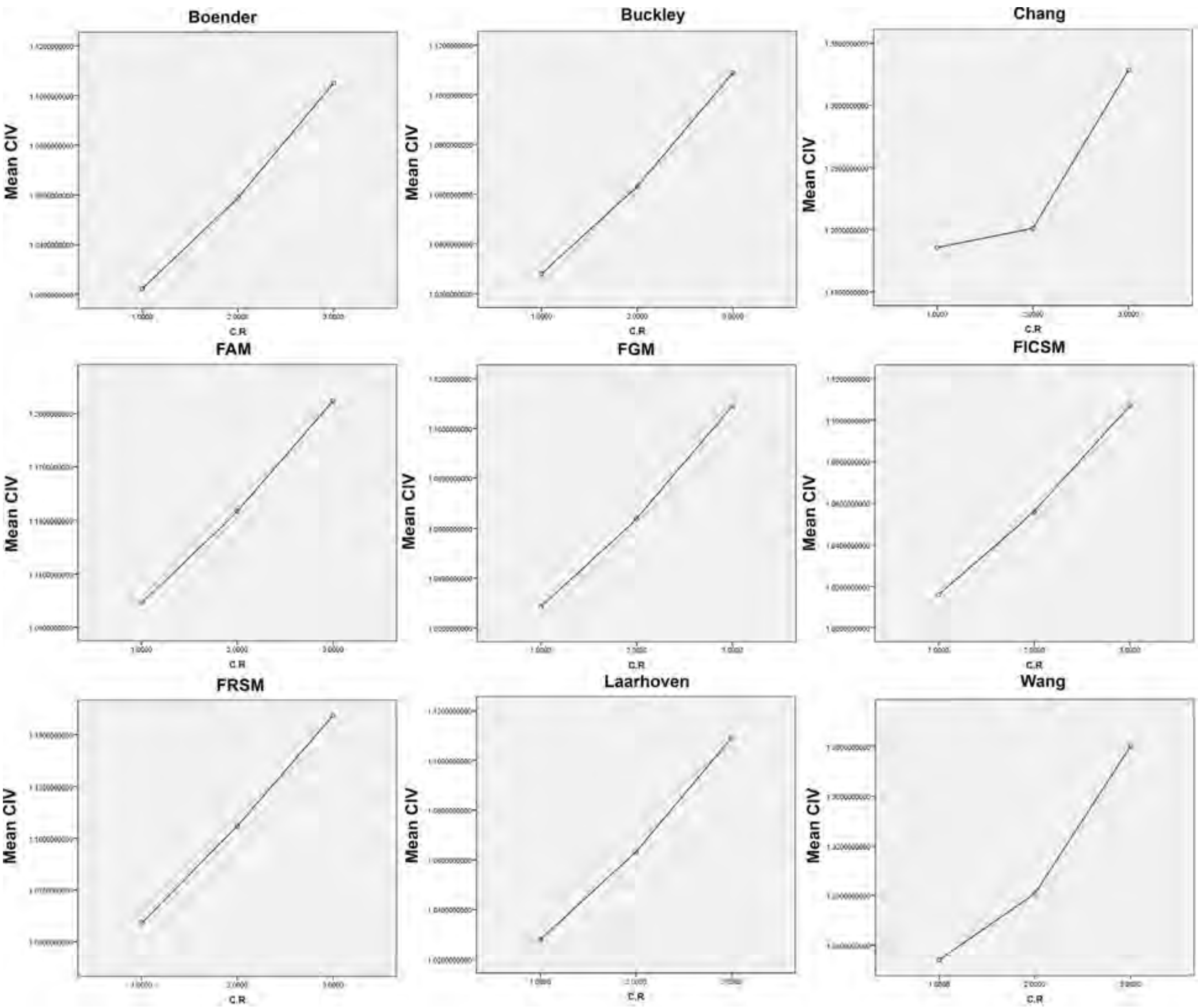

Fig. 12. Estimated marginal means of CIV.

Fig. 13 shows that at high inconsistency levels, Boender outperforms all other methods while for other levels FICSM is the best performing algorithm.

4.5. Overall analysis for Boender and FICSM

From the above statistical analysis, it is observed that Boender and FICSM algorithms performed significantly better compared to the other FAHP methods under specific experimental conditions. Here, we present an overall analysis of these two best performing algorithms without considering any experimental condition and using the whole

Please cite this article in press as: F. Ahmed, K. Kilic, Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.08.009

Significantly Better | Better | Inferior | Significantly Inferior

(a)

	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00581	-0.16332	-0.07147	-0.00632	0.00613	-0.03683	-0.00568	-0.22900
Buckley		-0.15751	-0.06566	-0.00051	0.01194	-0.03102	0.00013	-0.22319
Chang			0.09185	0.15699	0.16945	0.12649	0.15764	-0.06569
FAM				0.06515	0.07760	0.03464	0.06579	-0.15753
FGM					0.01245	-0.03051	0.00064	-0.22268
FICSM						-0.04296	-0.01181	-0.23514
FRSM							0.03115	-0.19218
Laarhoven								-0.22333

(b)

	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00428	-0.14218	-0.08601	-0.00486	0.00298	-0.04702	-0.00436	-0.23227
Buckley		-0.13790	-0.08173	-0.00058	0.00726	-0.04275	-0.00008	-0.22800
Chang			0.05616	0.13731	0.14516	0.09515	0.13782	-0.09010
FAM				0.08115	0.08899	0.03899	0.08165	-0.14626
FGM					0.00784	-0.04216	0.00050	-0.22741
FICSM						-0.05001	-0.00734	-0.23526
FRSM							0.04266	-0.18525
Laarhoven								-0.22792

(c)

	Buckley	Chang	FAM	FGM	FICSM	FRSM	Laarhoven	Wang
Boender	-0.00382	-0.22346	-0.10207	-0.00395	-0.00206	-0.05416	-0.00412	-0.27507
Buckley		-0.21964	-0.09825	-0.00013	0.00176	-0.05034	-0.00030	-0.27125
Chang			0.12139	0.21951	0.22140	0.16931	0.21934	-0.05160
FAM				0.09812	0.10001	0.04792	0.09795	-0.17299
FGM					0.00189	-0.05021	-0.00017	-0.27112
FICSM						-0.05210	-0.00206	-0.27301
FRSM							0.05004	-0.22091
Laarhoven								-0.27095

Fig. 13. Heat map – Heat map – Mean CIV differences of nine FAHP methods at different inconsistency levels (a) $C.R = \text{Low}$, (b) $C.R = \text{Medium}$, (c) $C.R = \text{High}$.

Table 10
Overall analysis.

Method	Percentage of matrices in which Boender outperforms the corresponding FAHP algorithm	Percentage of matrices in which FICSM outperforms the corresponding FAHP algorithm
Boender	--------------	48%
Buckley	85%	63%
Chang	87%	91%
FAM	90%	84%
FGM	85%	63%
FICSM	52%	--------------
FRSM	87%	42%
Wang	89%	93%
Laarhoven	93%	62%

dataset of 624 matrices. Results tabulated in Table 10 represent the percentage of matrices, where Boender and FICSM algorithm outperform other FAHP methods in the experiments.

5. Utility of results and proposed framework for researchers and practitioners

This paper provides useful guidance to researchers and practitioners in their selection process of a particular FAHP technique. LLSM method proposed by Boender [19] and FICSM outperformed other selected FAHP methods under various experimental conditions and therefore should be preferred choice of FAHP method in most real life decision making environments.

For example, a company faces a multi-criteria decision problem in which it is in the process of implementing a cloud computing solution and is currently evaluating various cloud computing service providers based on number of selection criteria such as availability, security, storage capacity, acquisition and transaction cost etc. Such a decision problem will have around 7 to 9 selection criteria and thus size of comparison matrices will be 7×7 to 9×9. Fig. 7 shows that modified LLSM method proposed by Boender [19] performs significantly better under these conditions and thus should be the preferred choice of FAHP method for such an application.

Also Fig. 10 and 9 show that all FAHP algorithms excluding FEA methods [23] [21] performs significantly better at lower levels of fuzzification. Thus ideal strategy for the implementation of Fuzzy AHP in such a decision problem will be to choose modified LLSM method proposed by Boender [19] as a preferred choice of FAHP algorithm and low level of fuzzifications as a decision variable to achieve best results.

Review of existing literature on FAHP shows that most practitioners utilize FEA methods in their decision problems. This comparative study shows that although FEA is not one of the best performing algorithms, but its performance slightly increases as the level of fuzzification is increased (Fig. 10 and 9). That is to say, if FEA method is the preferred choice of practitioner or researcher in their decision problem than this framework propose to use higher levels of fuzzification for better results.

6. Conclusions and future research

In this research we compared performance of nine FAHP methods among which five FAHP methods are the most popular ones in the literature. Compatibility Index Value (CIV) is used as a performance metrics to evaluate all nine FAHP methods. Three experimental conditions are considered as part of the analysis, namely, size of the matrix (n), fuzzification level (α) and inconsistency ($C.R$). For the fuzzification parameter four levels are assumed as 0.25, 0.50, 0.75 and 1.00. The fuzzification parameter is not inherent to the problem that the decision maker is facing but more a decision variable as part of the process. That is to say, the decision analysts can set the fuzzification level and conduct FAHP accordingly. On the other hand the inconsistency parameter refers to the inconsistency of the decision maker and is not a decision variable but depends on the fuzzy comparison matrices elicited from the experts. For the analysis three levels are considered for the inconsistency which is low, medium and high based on the consistency ratio (C.R) values as explained before. Finally, four different matrix sizes are considered which are 3, 7, 11 and 15. Note that one can consider 3 as the representative of small sized problems, 7 and 11 are for medium sized problems and 15 for larger cases.

As a result of this set up total of 48 ($= 4 * 3 * 4$) different experimental conditions are constructed. For each condition 13 replications are created randomly. Hence the total dataset is composed of six hundred and twenty four matrices with varying parameters for size of the matrix, fuzzification levels and inconsistency.

Main conclusions of this study are summarized as follows;

1. All three experimental parameters have significant effect on the mean CIV.
   - **Size of the Matrix:** As size of the matrix is increased, mean CIV increases and these results are consistent for all selected nine FAHP algorithms. At $n = 3$ the best performing algorithm is Buckley, while at $n = 7$ FICSM method outperforms all other methods. At higher matrix sizes i.e., $n = 11$, $n = 15$, Boender is the best performing algorithm.
   - **Fuzzification:** Performance of all FAHP algorithms except Chang and Wang decreases as the fuzzification level increases. Whereas, the performance of FEA methods (Chang and Wang) improves at higher levels of fuzzification, however this improved performance is still inferior when compared with other FAHP algorithms.
   - **Inconsistency:** Overall, as the inconsistency levels are increased, mean CIV increases and this increase is consistent over all selected nine FAHP algorithms. At low and medium level of inconsistency FICSM method outperform other methods whereas at higher inconsistency level, Boender is the best performing algorithm.
2. Among the selected nine FAHP algorithms, Boender and FICSM model performs significantly better than other models over various experimental conditions.
3. FEA methods (Chang and Wang) performed inferior compared to other methods, although this is one of the most frequently used technique in the literature.

4. If it is inevitable to use FEA method due to some reason, we propose that one must avoid low levels of fuzzification in triangular fuzzy numbers, as results shows that the performance of this methodology is significantly inferiors at lower levels of α.

Fuzzy numbers are considered as more realistic representations of the imprecise linguistic variables that are used by the decision makers during the preference elicitation stage of the analytical hierarchy process. As a result, abundant of research is being conducted that utilize FAHP and this study will help consolidate this literature. On the other hand, the value of introducing fuzziness to conventional AHP is yet to be assessed by an extensive experimental analysis similar to the one conducted in this research. We leave this as future research topic.

References

[1] A. Mardani, A. Jusoh, K. MD Nor, Z. Khalifah, N. Zakwan, A. Valipour, Multiple criteria decision-making techniques and their applications–a review of the literature from 2000 to 2014, Econ. Res. Ekonom. Istraž. 28 (1) (2015) 516–571.
[2] E.K. Zavadskas, Z. Turskis, S. Kildienė, State of art surveys of overviews on MCDM/MADM methods, Technol. Econ. Dev. Econ. 20 (1) (2014) 165–179.
[3] T.L. Saaty, The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, Decision Making Series, Mcgraw–Hill (Tx), 1980.
[4] O.S. Vaidya, S. Kumar, Analytic hierarchy process: an overview of applications, Eur. J. Oper. Res. 169 (1) (2006) 1–29.
[5] S.-H. Tsaur, T.-Y. Chang, C.-H. Yen, The evaluation of airline service quality by fuzzy mcdm, Tour. Manag. 23 (2) (2002) 107–115.
[6] L. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353, https://doi.org/10.1016/S0019-9958(65)90241-X, http://www.sciencedirect.com/science/article/pii/S001999586590241X.
[7] T.L. Saaty, Decision-making with the AHP: why is the principal eigenvector necessary, Eur. J. Oper. Res. 145 (1) (2003) 85–91.
[8] B. Golany, M. Kress, A multicriteria evaluation of methods for obtaining weights from ratio-scale matrices, Eur. J. Oper. Res. 69 (2) (1993) 210–220.
[9] K. Cogger, P. Yu, Eigenweight vectors and least-distance approximation for revealed preference in pairwise weight ratios, J. Optim. Theory Appl. 46 (4) (1985) 483–491.
[10] A. Ishizaka, M. Lusti, How to derive priorities in AHP: a comparative study, Cent. Eur. J. Oper. Res. 14 (4) (2006) 387–400.
[11] D.V. Budescu, R. Zwick, A. Rapoport, A comparison of the eigenvalue method and the geometric mean procedure for ratio scaling, Appl. Psychol. Meas. 10 (1) (1986) 69–78.
[12] T.L. Saaty, L.G. Vargas, Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios, Math. Model. 5 (5) (1984) 309–324.
[13] E. Takeda, K. Cogger, P. Yu, Estimating criterion weights using eigenvectors: a comparative study, Eur. J. Oper. Res. 29 (3) (1987) 360–369.
[14] F. Zahedi, A simulation study of estimation methods in the analytic hierarchy process, Socio-Econ. Plan. Sci. 20 (6) (1986) 347–354.
[15] N.A.M. Saadon, R.M. Dom, D. Mohamad, Comparative analysis of criteria weight determination in AHP models, in: Science and Social Research (CSSR), 2010 International Conference on, IEEE, 2010, pp. 965–969.
[16] G. Büyüközkan, C. Kahraman, D. Ruan, A fuzzy multi-criteria decision approach for software development strategy selection, Int. J. Gen. Syst. 33 (2–3) (2004) 259–280.
[17] R. Ureña, F. Chiclana, J.A. Morente-Molinera, E. Herrera-Viedma, Managing incomplete preference relations in decision making: a review and future trends, Inf. Sci. 302 (2015) 14–32.
[18] P. Van Laarhoven, W. Pedrycz, A fuzzy extension of Saaty's priority theory, Fuzzy Sets Syst. 11 (1) (1983) 199–227.
[19] C. Boender, J. De Graan, F. Lootsma, Multi-criteria decision analysis with fuzzy pairwise comparisons, Fuzzy Sets Syst. 29 (2) (1989) 133–143.
[20] D.-Y. Chang, Extent analysis and synthetic decision, Optim. Techn. Appl. 1 (1) (1992) 352.
[21] Y.-M. Wang, Y. Luo, Z. Hua, On the extent analysis method for fuzzy AHP and its applications, Eur. J. Oper. Res. 186 (2) (2008) 735–747.
[22] J.J. Buckley, Fuzzy hierarchical analysis, Fuzzy Sets Syst. 17 (3) (1985) 233–247.
[23] D.-Y. Chang, Applications of the extent analysis method on fuzzy AHP, Eur. J. Oper. Res. 95 (3) (1996) 649–655.
[24] T.J. Ross, Fuzzy Logic With Engineering Applications, first edition, Mcgraw–Hill College, 1995, http://amazon.com/o/ASIN/0070539170/.
[25] R. Ureña, F.J. Cabrerizo, J.A. Morente-Molinera, E. Herrera-Viedma, GDM-R: a new framework in R to support fuzzy group decision making processes, Inf. Sci. 357 (2016) 161–181.
[26] ahmedfaran/Fuzzy-AHP-Comparison: framework to conduct a comparative study of various fuzzy AHP algorithms, https://github.com/ahmedfaran/Fuzzy-AHP-Comparison.
[27] L.C. Leung, D. Cao, On consistency and ranking of alternatives in fuzzy AHP, Eur. J. Oper. Res. 124 (1) (2000) 102–113.
[28] Y.-M. Wang, J.-B. Yang, D.-L. Xu, Interval weight generation approaches based on consistency test and interval comparison matrices, Appl. Math. Comput. 167 (1) (2005) 252–273.
[29] R. Islam, M. Biswal, S. Alam, Preference programming and inconsistent interval judgments, Eur. J. Oper. Res. 97 (1) (1997) 53–62.
[30] J. Ramík, P. Korviny, Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean, Fuzzy Sets Syst. 161 (11) (2010) 1604–1613.
[31] Y. Xu, X. Liu, H. Wang, The additive consistency measure of fuzzy reciprocal preference relations, Int. J. Mach. Learn. Cybern. (2017) 1–12.

Please cite this article in press as: F. Ahmed, K. Kilic, Fuzzy Analytic Hierarchy Process: A performance analysis of various algorithms, Fuzzy Sets Syst. (2018), https://doi.org/10.1016/j.fss.2018.08.009

[32] W.-E. Yang, C.-Q. Ma, Z.-Q. Han, W.-J. Chen, Checking and adjusting order-consistency of linguistic pairwise comparison matrices for getting transitive preference relations, OR Spektrum 38 (3) (2016) 769–787.
[33] T. Saaty, A ratio scale metric and the compatibility of ratio scales: the possibility of arrow's impossibility theorem, Appl. Math. Lett. 7 (6) (1994) 51–57.
[34] G. Gamst, L.S. Meyers, A. Guarino, Analysis of Variance Designs: A Conceptual and Computational Approach with SPSS and SAS, Cambridge University Press, 2008.
[35] J.W. Osborne, A. Overbay, The power of outliers (and why researchers should always check for them), Pract. Assess. Res. Eval. 9 (6) (2004) 1–12.
[36] L.M. Lix, J.C. Keselman, H. Keselman, Consequences of assumption violations revisited: a quantitative review of alternatives to the one-way analysis of variance F test, Rev. Educ. Res. 66 (4) (1996) 579–619.