

An MCDM for a Large Set of Criteria

Dr. Hilja Lisa Huru and M.Sc. Erlend Hoff

Norwegian Defence Research Establishment

P.O.Box 25, NO-2027 Kjeller

Norway

hilja-lisa.huru@ffi.no / erlend.hoff@ffi.no

ABSTRACT TITLE

We present a method for multi-criteria decision analysis (MCDA) capable of dealing with a large number of criteria. The issue with the most common methods for MCDA is that the number of pairwise comparisons grows quickly with the number of criteria. We have developed a method which reduces the number of pairwise comparisons to a small fixed number. This produces an incomplete judgement matrix from which we obtain a ranking and a weighting of the criteria. The way of doing this is similar to methods based on the geometric or arithmetic mean. A common problem with MCDA is inconsistency, and with a large number of criteria is inconsistency even more abundant. This method is designed to overcome the inconsistency which is bound to occur and extract the decision makers' preferences. The result is a method which is time-saving and which minimizes the workload while sufficient level of accuracy and quality is secured.

Furthermore, an interesting result of applying the method is that it acts as a structuring tool. In our applications the formulation of the criteria was improved, new criteria were added and superfluous ones were removed. We present the method, its mathematical foundation and demonstrate a simple tool developed for executing an analysis using the method

1.0 INTRODUCTION

Decision making in large investment projects is undoubtedly a challenge. Having to consider both how each alternative covers the specified needs and making sure that it performs at a sufficient level to a reasonable cost, demands a coherent and transparent method. In investments of large complexity where numerous aspects have to be taken into account, which leads to a large number of criteria applying to the alternatives, it is challenging even to establish which criteria should matter the most. Presented here is a method for exactly this, constructed specially for a large number of criteria, where traditional methods for Multi-Criteria Decision Analysis (MCDA) come short.

Over the last decades, it has been developed several methods whose aim is to extract the decision makers (DM) preferences and to give an outcome by evaluating how much impact each criteria should have on the choice. One example is the Analytical Hierarchy Process (AHP), see [14], which is one of the most extensively used methods, but also one of the most extensively debated and criticized methods, see e.g. [16] (for a summary) and [3, 4, 5, 6, 9, 11].

In working with investment projects of large complexity we saw the need for an improvement of the existing methods that would better suit our needs. In earlier projects it has been a problem with inconsistent weighting. One goal, or requirement, in MCDA is consistency, i.e. if attribute i get the score a_{ij} in comparison with attribute j which then again get the score a_{jk} when compared with attribute k , then i should get the score $a_{ik} = a_{ij} + a_{jk}$ compared with k . In large and complex processes it is always a problem for the DM to keep track of all the scores and to give steady, consistent weights. Inconsistency *will* happen and the challenge is how this should be treated. The method presented here is constructed to deal with inconsistencies.

Report Documentation Page

*Form Approved
OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE OCT 2009	2. REPORT TYPE N/A	3. DATES COVERED -			
4. TITLE AND SUBTITLE An MCDM for a Large Set of Criteria		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S)		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Norwegian Defence Research Establishment P.O.Box 25, NO-2027 Kjeller Norway		8. PERFORMING ORGANIZATION REPORT NUMBER			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES See also ADA562565. RTO-MP-SAS-080 Decision Support Methodologies for Acquisition of Military Equipment (Methodologies d'aide a la decision pour l'acquisition d'equipements militaires). Meeting Proceedings of Systems Analysis and Studies Panel (SAS) Specialists Meeting held in Brussels, Belgium on 22 and 23 October 2009., The original document contains color images.					
14. ABSTRACT We present a method for multi-criteria decision analysis (MCDA) capable of dealing with a large number of criteria. The issue with the most common methods for MCDA is that the number of pairwise comparisons grows quickly with the number of criteria. We have developed a method which reduces the number of pairwise comparisons to a small fixed number. This produces an incomplete judgement matrix from which we obtain a ranking and a weighting of the criteria. The way of doing this is similar to methods based on the geometric or arithmetic mean. A common problem with MCDA is inconsistency, and with a large number of criteria is inconsistency even more abundant. This method is designed to overcome the inconsistency which is bound to occur and extract the decision makers preferences. The result is a method which is time-saving and which minimizes the workload while sufficient level of accuracy and quality is secured. Furthermore, an interesting result of applying the method is that it acts as a structuring tool. In our applications the formulation of the criteria was improved, new criteria were added and superfluous ones were removed. We present the method, its mathematical foundation and demonstrate a simple tool developed for executing an analysis using the method					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 6	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

One must further make sure that the workload does not become insurmountable. When a large amount of alternatives or criteria is considered, it is extremely time-consuming if a pairwise comparison is required for every possible pair of attributes. Our method is designed to be time-saving and to decrease the amount of work required; it limits the amount of pairwise comparisons as each criterion is matched with a fixed (small) number of criteria with whom pairwise comparisons are done. Gathered in a matrix, the base for further analysis is then an *incomplete* judgment matrix from which the weight vector is calculated.

To sum up, the main advantages of the method are as follows

- the method does not require consistent weighting, but is able to overcome inconsistency while extracting the DM's wishes and preferences
- the method minimizes the workload by limiting the number of pairwise comparisons while accuracy and quality is secured

Our method is connected to those based on the geometric or arithmetic mean, see [2, 4]. This is described in section 1.1. The rest of this paper is organized as follows; in section 2 the method itself is presented. The method is described additively, but can easily be transferred to be multiplicative. In section 3 we describe how the method was applied with examples and valuable experiences from the process.

1.1 Pairwise comparisons and arithmetic mean

We base the method on pairwise comparisons which are frequently used in MCDA, see e.g. [14], which is a user-friendly technique the decision maker is comfortable with.

A complete judgment matrix involves $n(n - 1)/2$ pairwise comparisons and for decision of some complexity this will soon get to thousands. For example, if 300 objects are involved the DM will have to do 44850 pairwise comparisons. Such an enormous amount makes this extremely time-consuming and impossible for the DM to keep up the concentration and to be consistent.

It is a fact that only $(n - 1)$ comparisons in one connected chain are necessary to rank n objects, see [1, 9, 15], but every little inaccuracy will be crucial for the outcome. An increase in the number of comparisons can be used to improve the accuracy and overall consistency. One way to do this is to collect the objects in more or less homogeneous clusters with pivots connecting them, see [9]. Another method is Minimal Pairwise Comparison (MIPAC) in [15]. There also exists a method based on geometric least square presented in [11] which demands neither complete nor consistent judgment matrices, and there is a method for incomplete matrices based on the eigenvalue method and AHP, described in [8].

In simulations, see for example [7], the most common methods for determining weight vectors have been tested, but none of the methods performed significantly better than the others. However, the geometric or arithmetic mean methods have the advantage that they can be used on incomplete judgment matrices, [16]. The method presented in this paper differs from the methods mentioned above by using the (arithmetic) mean for finding the weight vector. If there exists a complete set of pairwise comparisons the arithmetic mean method, as described in [2, 4], produces a weight vector as follows: Assume there are n objects to be ranked and weighted. The judgment matrix $A = \{a_{ij}\}$ is a complete $n \times n$ matrix where the entry a_{ij} represents how attribute i is judged by DM when compared with j . A is obviously skew-symmetric, that is $A = -A^T$, $a_{ij} = -a_{ji}$. The weight vector is $W = (w(1), \dots, w(n))^T$ where $w(i) = (\sum_{j=1}^n a_{ij})/n$ is the weight obtained for attribute i .

2.0 THE METHOD

Let's describe our method. As already mentioned, to reduce the workload there is for every attribute (criteria) done a relatively small number of pairwise comparisons. All attributes are still connected, either directly or indirectly. A weight vector is produced taking the arithmetic mean both of the direct and indirect assessments recursively, where the weight vector is adjusted by the score for the connecting attributes.

Assume that s comparisons are done for each attribute, where $s < n$ (preferably $s \ll n$). These s attributes are drawn randomly and evenly distributed such that each attribute is compared pairwise with some other attribute $2s$ times. The result of DM's assessments is then an $n \times n$ matrix A where $2sn$ of the entries are filled in. In other words, assume the attribute i is directly compared with the attributes j_1, \dots, j_s (where the attributes j_1, \dots, j_s are assigned to i and j_{s+1}, \dots, j_{2s} are the ones to which i is assigned). The entry a_{ji} in A is the score i achieved when compared with j and $a_{ij} = -a_{ji}$. The entries in row i in A then are all a_{ij} for $j = j_1, \dots, j_{2s}$, denoted by $a_{i(j_1)}, \dots, a_{i(j_{2s})}$. The remaining entries are not filled in, except $a_{ii} = 0$.

Now we do some consecutive repeated calculations to find a final weight vector. The "first" weight vector is given by the arithmetic mean of these scores. That is, the attribute i gets the following weight;

$$w_1(i) = w(i) = (a_{i(j_1)} + \dots + a_{i(j_{2s})})/2s$$

for $i = 1, \dots, n$, and the weight vector is $W_1 = (w_1(1) \dots w_1(n))^T$.

This is then repeated with the weights of the directly connected attributes taken into account, that is, $w_l(j_l)$, $l = 1, \dots, 2s$ is added to the value $a_{i(j_l)}$. The score attained for each of the directly connected attributes will enhance the relative score and the "second" weight for i is

$$w_2(i) = (a_{i(j_1)} + w_1(j_1) + \dots + a_{i(j_{2s})} + w_1(j_{2s}))/2s = w_1(i) + (w_2(j_1) + \dots + w_2(j_{2s}))/2s.$$

This is done repeatedly until the weight vector stabilizes, that is, W_r does not differ significantly from W_{r+1} and the ranking does not change. Let us assume this happens after r iterations. Then the attribute i has the score/weight

$$w_r(i) = w_1(i) + (w_{r-1}(j_1) + \dots + w_{r-1}(j_{2s}))/2s,$$

and the weight vector is $W_r = (w_r(1), \dots, w_r(n))^T$. In our tests the weight vector did stabilize after a relatively small number of iterations (about 20).

An attribute will only relate to its closest neighbours, but will be pushed up or down on the ranking according to the score attained by the direct neighbours. For example can one attribute which is expected to score average compared with the others, do really well in the first round and end up towards the top of the ranking if its direct neighbours typically all obtain low ratings. The score for this attribute is then adjusted in the subsequent rounds according to how its neighbours score. Hence, these adjustments will spread out and draw scores from all the attributes and by repeating a satisfactory number of times give the desired result, which in this example would be to pull the score for this particular attribute down.

3.0 EXECUTION OF THE METHOD

The method was developed to defeat some of the most common causes for divergences observed in decision making and to take into consideration some well known aspects causing "errors" in such processes. We use expert groups in the role of DM. To be able to control and identify disalignments and inconsistencies, to avoid that strong members of the expert group dominates and to ensure that the opinion of the experts as a whole comes through, the collection of experts are divided into groups that do the same pairwise comparisons. The mean of these groups' weightings makes the base for the calculations.

3.1 Scale and GUI

With a sufficient number of connections within the set of attributes we aim on being able to deal with the more or less “natural” inconsistency. Furthermore, each attribute is simultaneously compared with a number of other attributes, in our case 3, as seen in figure 1 showing the GUI. The experts are able to see the attribute in a wider setting and hence do a better evaluation. It also turned out that it is more effective to do the evaluation in this way. The experts evaluate the attributes following the scale;

- critically more important
- much more important
- more important
- slightly more important
- equal importance
- slightly less important
- less important
- much less important
- critically less important

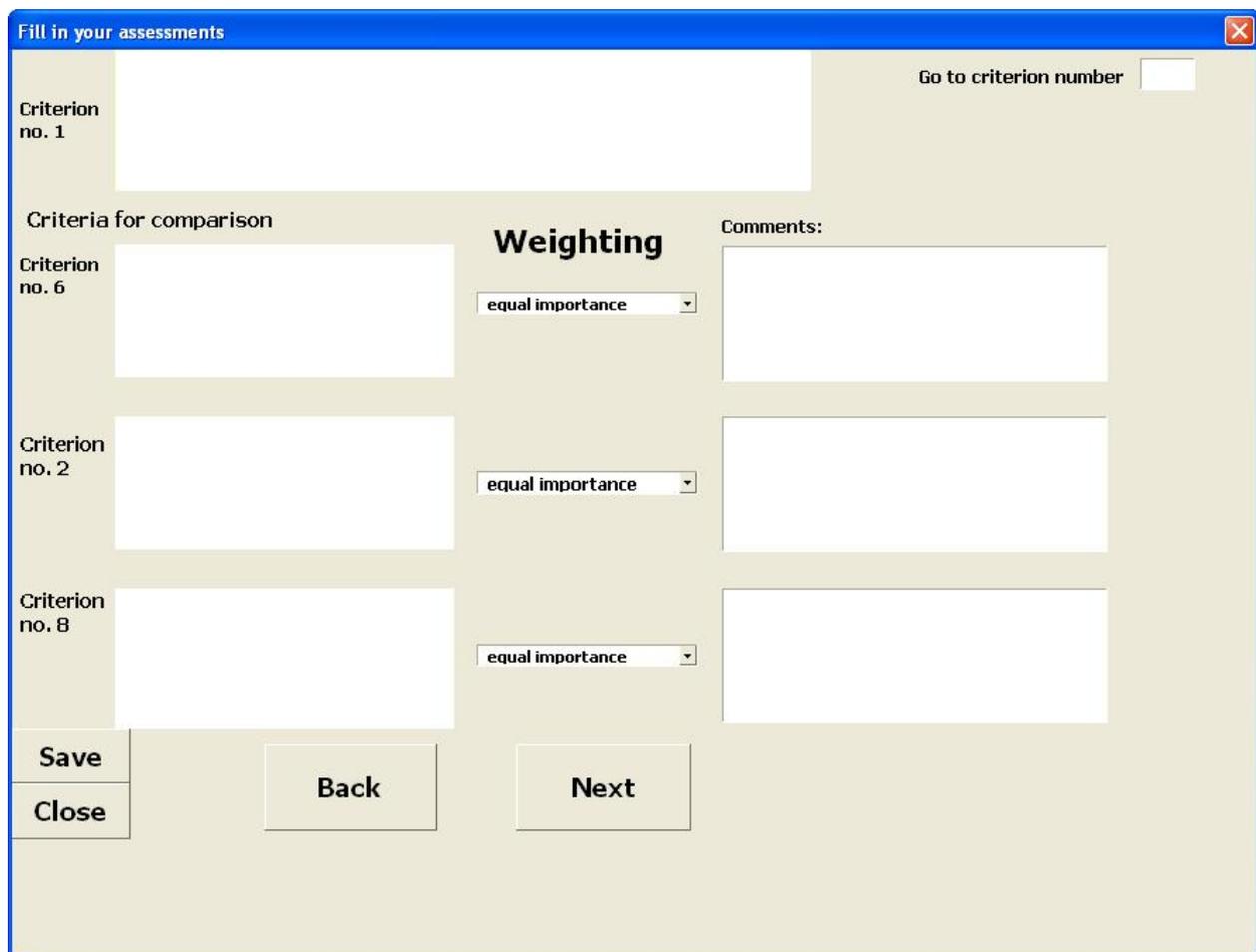


Figure 1: The GUI

In our tests the experts were instructed to picture themselves an exponential scale corresponding to the positive expressions above and similarly for the negative expressions. The choice of scale can be discussed and adjusted, but it is natural for humans to evaluate impressions exponentially. The one we used was presented for the experts and kept throughout the process.

In our execution, where we had about 300 criteria to be weighted, the experts were divided into three groups and in three days they did three pairwise comparisons for each of the 300 criteria, as described above. The number here does not have to be three, neither for the number of groups nor the number of direct comparisons. However, this is acceptable and adequate as it minimizes the workload and at the same time more or less ensures direct or indirect connections between all the attributes.

3.1 Useful experiences

As mentioned the groups of experts did the same evaluations and it was possible to pinpoint problems in the evaluation when analyzing and comparing the results. It turned out that for many of the pairwise comparisons there were large differences in the weighting between the groups. In some cases were the weightings diametrical opposites, that is, a difference of 8 steps. The largest differences were for criteria that were difficult to compare and that were connected to quite different areas of the domain of investigation. To make sure that the analysis was done on the best set of data possible, we did arrange for a reevaluation of the pairs that had a difference in the weighting of 3 steps or more. These were 33.3 % of the pairs, which confirms the discussion earlier concerning the problems that arise in such decision processes. In most of the cases the experts, now gathered in one group, did agree on the mean of the earlier evaluation. After the reevaluation 10.6 % of the new weighting of the pairs did diverge 3 steps or more from the mean of the first evaluation, 3.2 % 4 steps or more and 1.1 % diverged 5 steps or more. None of them diverged more than 5 steps. Since surprisingly few of the scores obtained when evaluating once more diverged 4 steps or more from the original mean, we can conclude that the mean is a good approximation.

Even though the criteria were thought to be final and fixed when the evaluations started, in the process the experts saw that several corrections were needed. These consisted of irrelevant criteria being removed, new ones being added and some of the criteria needed to be better formulated. However, removing attributes created the problem that not all attributes had the same number of connections, but this was more or less fixed as the added attributes took the places of the removed ones. This was an important experience as it shows that the method is useful for polishing and finalizing the criteria by putting them in a setting of evaluation. By an *early evaluation* for optimizing the criteria, the result will be closer to perfection and will prevent that more/too many changes needs to be done.

Finally, a ranked list was extracted using the method as described in section 3. The final ranking was checked and approved by the experts, still in the role of the DM, who found the result satisfactory.

4.0 CONCLUSION

This is a time-saving and accurate method which has given good results in the application described. It is well founded and deals with the problems in decision making resulting in inconsistencies. An additional gain from the method is that it improves the whole process by enforcing reflection on the importance and formulations of the criteria, on how they do in comparison with others and how they connect to the alternatives in a weighting. It is therefore useful to spend time to do a *preevaluation* for discussing the criteria in a realistic setting and should be included as a “cleaning” of the criteria for the investment object. The following main factors contributed to giving the best possible result; the mean of the weighting by several teams of experts resulted in a data set that cancels out imbalances in the opinions of the experts and gives a better input. Every attribute undergoes only a small number of pairwise comparisons and saves a lot of work for DM. The method takes into account the score of the neighbouring attributes in a sequence of rounds, such that the indirect scores spread out over the whole set of attributes and every pairwise

comparison contributes to the final score of each attribute. This helps deal with inconsistencies and extracts the DM's preferences from a minimized set of pairwise comparisons.

5.0 REFERENCES

- [1] J.-P. Briot et al., Multi-criteria decision aggregation and ranking approaches for combining multiple information strategies to decide which agents should be replicated, kap. 6.1, EuroControl Report, 2007.
- [2] J. Barzilai, Deriving weights from pairwise comparison matrices, *J. Operational Research Society*, vol. 48 nr 12, pp 1226-1232, 1997.
- [3] J. Barzilai, W. D. Cook, B. Golany, Consistent weights for judgements matrices of the relative importance of alternatives, *Operations Research Letters* Vol 6, No 3, pp 131-134, 1987.
- [4] J. Barzilai, B. Golany, Deriving weights from pairwise comparison matrices: the additive case, *Operations Research Letters* 9, pp 407-410, 1990.
- [5] V. Belton, T. Gear, On a short-coming of Saaty's method of analytical hierarchies, *Omega*, vol. 11, 1983.
- [6] V. Belton, T. J. Stewart, *Multiple criteria decision analysis: an integrated approach*, Kluwer Acad. Publ., 2002.
- [7] B. Golany, M. Kress, A multicriteria evaluation of methods for obtaining weights from ratio-scale matrices, *Eur. J. of Operational Research*, 69, 210-220, 1993.
- [8] P. T. Harker, Incomplete pairwise comparison in the analytical hierarchy process, *Mathl Modelling*, Vol. 9, No. 11, pp 837-848, 1987.
- [9] A. Ishizaka, Advantages of clusters and pivots in AHP, 15th Mini-Conference MUDSM 2004.
- [10] A. Ishizaka, M. Lusti, How to derive priorities in AHP: a comparative study, *Cent. Eur. J. of Oper. Research*, vol. 14, no4, pp387-400, Springer, 2006.
- [11] G. Islei, A. G. Lockett, Judgemental modelling based on geometric least square, *Eur. J. of Oper. Research* 36, pp 27-35, 1988.
- [12] I. Jordanger et al., Flermålsanalyser i store statlige investeringsprosjekt, Concept rapport nr 18, NTNU, 2007.
- [13] S. Malerud, T. Kråkenes, Metoder for flermålsanalyse - En oversiktsstudie fra GOAL, FFI/RAPPORT-2005/03041.
- [14] T. L. Saaty, *The analytical hierarchy process*, McGraw-Hill, 1980.
- [15] P. Sen, J.-B. Yang, Design decision making based upon multiple attribute evaluations and minimal preference Information, *Mathl. Comput. Modelling* Vol 20, No. 3 pp. 107-124, 1994.
- [16] L. Warren, Uncertainties in the analytical hierarchy process, Defence Science and Technology Organisation Information Sciences Laboratory, Department of Defence, Australian Government, 2004.