

## RETHINKING WEIGHT ELICITATION FOR ANALYTIC HIERARCHIES: PRACTICAL IMPLICATIONS OF THE MODIFICATIONS

**Kiyotada Hayashi**

The Ministry of Agriculture, Forestry and Fisheries of Japan  
Tohoku Natl. Agr. Expt. Stn., 4 Akahira, Shimo-Kuriyagawa, Morioka, Iwate 020-0198, Japan  
hayashi@affrc.go.jp

**Abstract:** This paper reviews the weighting procedures proposed as revisions of the analytic hierarchy process from a practical standpoint. First, the modifications of the 1-to-9 scale are discussed as a way to cope with the criticisms of the AHP. The balanced scale and the 8-based multiplicative scale are treated and the measures to express the inconsistencies of judgments are examined. Second, since the modifications of the scale are not necessarily resolve the problems, a hierarchical weighting procedure is explained as a method to explicitly incorporate the ranges of attribute levels into the procedure. The inconsistencies are expressed as the intervals of values in this approach.

### Introduction

The AHP (Analytic Hierarchy Process) is a well-known multicriteria method and widely used in practice for selecting discrete alternatives. The reasons for the acceptance of practitioners are considered to be simplicity of pairwise comparisons in eliciting preferences, the ability to check the consistency of judgments, and the usefulness of hierarchical problem structuring. The popular software, Expert Choice, makes the AHP a user-friendly tool.

There are, however, debates on the appropriateness of the AHP (Dyer, 1990; Saaty, 1990; Salo and Hamalainen, 1997a). In addition, textbooks on decision analysis explain the criticisms of the AHP (French, 1986; Watson and Buede, 1987; Marshall and Oliver, 1995; Mollaghasemi and Pet-Edwards, 1997; Goodwin and Wright, 1998). Since the criticisms will be serious in practicing the methodology, it is important to summarize the criticisms in the first place; they are grouped into the following six categories:

- (i) The definition of the numerical scale is vague and giving any substantive operational meaning to the verbal expressions is difficult (French, 1986). The verbal statements constitute a nominal (Likert) scale whereas numerical weights are a ratio scale (Stewart, 1992).
- (ii) The 1-to-9 scale imposes the inconsistencies of responses. One reason is that the scale is not closed under multiplication (French, 1986). The other is the uneven dispersion of local weights (Poyhonen, Hamalainen and Salo, 1997).
- (iii) The notion of weights is ambiguous; that is, the questions based on the relative importance of attributes without reference to attribute levels (the scales on which attributes are measured) are less well defined and they may, therefore, be interpreted differently by decision makers (Belton, 1986).
- (iv) Although the AHP can check the consistency of judgments by overspecification of the model, it requires a large number of comparisons from the decision maker.
- (v) The phenomenon called rank reversal may occur. This means the AHP does not satisfy the principle of 'independence of the irrelevant alternatives'. Recently, the phenomenon is claimed to be an advantage of the AHP. Nevertheless, there is no evidence as to the correspondence between rank reversals in the

behavior observed descriptively and the decision making supported by the AHP (Hamalainen and Salo, 1997).

(vi) The axioms of the AHP are not founded on testable descriptions of rational behavior, though the axioms provide the basis for rational decision making (Dyer, 1990).

The purpose of this paper is to review the methods to cope with these criticisms from a practical viewpoint. In Section 2 modifications of the 1-to-9 scale, which are proposed as the techniques to overcome the criticisms, are discussed. In Section 3 a hierarchical weighting procedure is illustrated as a method to cope with the range effect as well as the inconsistencies of judgments. Finally, some practical issues will be discussed in Section 4.

### Modifications of the 1-to-9 Scale

Modifying the 1-to-9 scale is proposed as a way to resolve the problems discussed earlier; especially, the criticisms (i) and (ii). The fifth criticism (rank reversal) is also related to the modifications. Although several scales have been proposed (Ma and Zheng, 1991; Lootsma, 1993; Dodd, Donegan and McMaxter, 1995; Poyhonen, Hamalainen and Salo, 1997), we concentrate on the following two representative scales (see Table 1).

Table 1: Definition of scales

| Verbal statement | Scale  |             |              |                   |                 |
|------------------|--------|-------------|--------------|-------------------|-----------------|
|                  | 1-to-9 | Balanced    |              | MAHP <sup>a</sup> |                 |
|                  | $t$    | $\delta(t)$ | $\varphi(t)$ | $\phi(t)$         | $\exp(\phi(t))$ |
| Equal            | 1      | 0.00        | 1.00         | 0.00              | 1.00            |
| —                | 2      | 0.20        | 1.22         | 0.12              | 1.13            |
| Moderate         | 3      | 0.40        | 1.50         | 0.25              | 1.29            |
| —                | 4      | 0.60        | 1.86         | 0.39              | 1.48            |
| Strong           | 5      | 0.80        | 2.33         | 0.54              | 1.72            |
| —                | 6      | 1.00        | 3.00         | 0.72              | 2.06            |
| Very strong      | 7      | 1.20        | 4.00         | 0.96              | 2.60            |
| —                | 8      | 1.40        | 5.67         | 1.32              | 3.73            |
| Extreme          | 9      | 1.60        | 9.00         | 2.63              | 13.93           |

<sup>a</sup> 8-based multiplicative scale.

The balanced scale (Poyhonen, Hamalainen and Salo, 1997) or the 10/10-18/2 scale (Ma and Zheng, 1991) is proposed as follows to equalize the distribution of weights:

$$\varphi(t) = \begin{cases} \frac{t+9}{-t+11}, & \text{if } t \geq 1, \\ \frac{11t-1}{9t+1}, & \text{if } t < 1. \end{cases} \quad (1)$$

In this case, the difference of importance is:  $\delta(t) = 2(w_i - w_j) / (w_i + w_j) = 2(\varphi(t) - 1) / (\varphi(t) + 1) = (t - 1) / 5$ , where  $t \geq 1$ , that is  $w_i \geq w_j$ .

The 8-based multiplicative scale for the MAHP (the Modified AHP) (Dodd, Donegan and McMaxter, 1995) is defined as  $\exp(\phi(t))$ , where  $\phi(t)$  means the difference of importance as follows:

$$\phi(t) = \begin{cases} \tanh^{-1} \frac{t-1}{H-1}, & \text{if } t \geq 1, \\ -\tanh^{-1} \frac{t^{-1}-1}{H-1}, & \text{if } t < 1. \end{cases} \quad (2)$$

The numerical value assigned to  $H$  is considered to be dependent on the decision maker; in this scale the value of  $H$  is determined so that  $8 * 8 = 9$ , where  $x * y \equiv \phi^{-1}[\phi(x) + \phi(y)]$ . This scale is almost closed under multiplication (except for  $9 * 9 = 9.082$ ). In the 7-based multiplicative scale, the value of  $H$  is determined so that  $7 * 7 = 9$ .

In order to apply these scales in practice, however, it is necessary to redefine the measure of consistency because the consistency ratio (C.R.) is dependent on the specific scale. One way is to calculate the random index (R.I.), the mean of the consistency index (C.I.) of the randomly generated comparison matrix. C.I. is defined as  $(\lambda_{\max} - n) / (n - 1)$ , where  $\lambda_{\max}$  is the maximum eigenvalue and  $n$  is the size of the matrix. So far the indexes are not shown explicitly despite their importance. Another way is to use the consistency measure (C.M.), which is proposed as a scale-invariant measure (Salo and Hamalainen, 1997a). This is defined as follows:

$$\text{C.M.} = \frac{2}{n(n-1)} \sum_{i < j} \frac{u_{ij} - l_{ij}}{(1 + u_{ij})(1 + l_{ij})}, \quad (3)$$

where  $u_{ij} = \max_k a_{ik} a_{kj}$  (Salo, 1993), and  $l_{ij}$  is the inverse of  $u_{ji}$ .

Table 2 shows the means (R.I.) and the standard deviations of C.I. and C.M. C.M. has the preferable nature that it is more stable than C.I.; the standard deviations of C.M. are smaller than those of C.I. However, as the size of comparison matrices increases, the mean of the C.M. becomes larger as well as C.I. Consequently the finding that the inconsistency (C.M.) between statements increases as the number of attributes becomes higher (Poyhonen and Hamalainen, 1997) has to be modified; that is, C.M. divided by the mean of randomly generated matrices does not necessarily becomes larger as the size of comparison matrices increases. The inconsistency measured by C.M. comes from the nature of the comparison matrix as well as the judgments of the decision maker.

**Table 2: Means of C.I. and C.M.**

| Size <sup>a</sup> | C.I.            |                 |                   | C.M.            |                 |                   |
|-------------------|-----------------|-----------------|-------------------|-----------------|-----------------|-------------------|
|                   | 1 to 9          | Balanced        | MAHP <sup>b</sup> | 1 to 9          | Balanced        | MAHP <sup>b</sup> |
| 3                 | 0.5245 (0.6959) | 0.2687 (0.3851) | 0.2345 (0.4032)   | 0.3801 (0.2524) | 0.3142 (0.2119) | 0.2802 (0.1922)   |
| 4                 | 0.8863 (0.6280) | 0.4402 (0.3653) | 0.3863 (0.4189)   | 0.5744 (0.1812) | 0.4846 (0.1601) | 0.4361 (0.1599)   |
| 5                 | 1.1065 (0.5097) | 0.5496 (0.3217) | 0.4920 (0.3982)   | 0.6852 (0.1276) | 0.5876 (0.1197) | 0.5359 (0.1310)   |
| 6                 | 1.2475 (0.4075) | 0.6242 (0.2799) | 0.5706 (0.3730)   | 0.7567 (0.0915) | 0.6566 (0.0917) | 0.6057 (0.1090)   |
| 7                 | 1.3418 (0.3310) | 0.6770 (0.2427) | 0.6281 (0.3396)   | 0.8059 (0.0670) | 0.7059 (0.0715) | 0.6580 (0.0908)   |
| 8                 | 1.4039 (0.2780) | 0.7142 (0.2144) | 0.6725 (0.3119)   | 0.8402 (0.0510) | 0.7414 (0.0581) | 0.6971 (0.0776)   |
| 9                 | 1.4507 (0.2375) | 0.7438 (0.1908) | 0.7097 (0.2865)   | 0.8656 (0.0391) | 0.7691 (0.0478) | 0.7289 (0.0668)   |
| 10                | 1.4851 (0.2085) | 0.7648 (0.1716) | 0.7352 (0.2627)   | 0.8845 (0.0309) | 0.7906 (0.0403) | 0.7538 (0.0582)   |
| 11                | 1.5141 (0.1855) | 0.7836 (0.1563) | 0.7597 (0.2437)   | 0.8991 (0.0246) | 0.8083 (0.0345) | 0.7751 (0.0512)   |
| 12                | 1.5357 (0.1664) | 0.7976 (0.1425) | 0.7783 (0.2247)   | 0.9104 (0.0198) | 0.8227 (0.0298) | 0.7926 (0.0452)   |
| 13                | 1.5555 (0.1513) | 0.8101 (0.1312) | 0.7942 (0.2087)   | 0.9194 (0.0161) | 0.8349 (0.0261) | 0.8077 (0.0401)   |
| 14                | 1.5707 (0.1383) | 0.8203 (0.1214) | 0.8073 (0.1949)   | 0.9265 (0.0132) | 0.8451 (0.0230) | 0.8204 (0.0358)   |
| 15                | 1.5839 (0.1273) | 0.8284 (0.1129) | 0.8177 (0.1821)   | 0.9322 (0.0109) | 0.8539 (0.0205) | 0.8315 (0.0322)   |

Note: The numbers in the parentheses are standard deviations.

<sup>a</sup> Random samples of 100,000 matrices were produced (by M-sequence random numbers on Sun SparcStation 5) for  $4 \times 4$  to  $15 \times 15$  matrices.

<sup>b</sup> The 8-based multiplicative scale.

<sup>c</sup> All possible comparison matrix ( $17 \times 17 \times 17 = 4,913$ ) are used for the calculation instead of generating the random samples.

### Hierarchical Weighting to Cope with the Range Effect

The method discussed earlier will resolve the problems related to the scale. However, the problem that the notion of weight is ambiguous, the criticism (iii), remains unsolved. This is a crucial point in applying multicriteria analysis; because, for example, in order to evaluate the alternative farm production systems, considering the attribute ranges of profits and working hours explicitly is necessary (Hayashi, 1998; Hayashi, 1999). Moreover, it may be impossible to answer the criticism in agricultural research evaluation (assigning weights in scoring methods is a tricky procedure (Alston, Norton and Pardey, 1995)) without

relying on the theory of measurable value functions. Although Poyhonen and Hamalainen have pointed out that one cannot see the differences in weights depending on whether attribute ranges are presented or not (Poyhonen and Hamalainen, 1997), their comparisons are restricted to the AHP. Rather, the techniques which explicitly consider the attribute ranges such as swing weighting and indifference techniques should be used (von Nitzsch and Weber, 1993; Fischer, 1995).

While the methods under imprecise information about weights have been proposed (Moskowitz, Preckel and Yang, 1992; Salo and Hamalainen, 1992), which can be considered as extensions of absolute measurement of the AHP, the ratio comparisons in these approaches are not linked explicitly to the ranges of the attribute levels. Hence, a method to explicitly incorporate the attribute ranges into the weighting procedure is discussed in this section. This method is applicable unless the criteria are too subjective and for which we can not think of reasonable quantitative scales (in that case see, e.g., (Salo and Hamalainen, 1997b)).

In this method we do not have to consider the criticism (i) and (ii); moreover, since attributes and alternatives are distinguished explicitly, rank reversal, the criticism (v), is left out of consideration. However, to make the most of the advantage of the AHP, it is recommended to consider how to take into account consistency checks. At the same time, we have to pay attention to parsimony in measurement, because making all  $n(n-1)/2$  judgments would be very tedious for respondents when  $n > 4$  (the criticism (iv)); this overlaps with the study on imperfect comparison matrices.

The following procedure explains a method to cope with the range effect as well as the inconsistency of preferences. Although synthesizing either exact judgments or imprecise judgments into an overspecified system using linear constraints can be applied for estimating weight intervals (inconsistencies), the following is an example of the former judgments.

In this example decision criteria fall into tree categories; weights are denoted by  $w_1 = w_{11} + w_{12} + w_{13}$ ,  $w_2 = w_{21} + w_{22} + w_{23}$ , and  $w_3 = w_{31} + w_{32} + w_{33}$ . At the lowest level of the hierarchy, a measure (continuous or discrete scale) is identified or constructed for each criterion. Each element of the comparison matrix is determined on the basis of difference value measurement so that  $[(x_i', x_{j*}), (x_{i*}, x_{j*})] \sim [(x_{i*}, x_{j*}), (x_{i*}, x_{j*})]$ ; the relative strength of preference of the object described by  $(x_i', x_{j*})$  over the object described by  $(x_{i*}, x_{j*})$  is indifferent to the object described by  $(x_{i*}, x_{j*})$  over the object described by  $(x_{i*}, x_{j*})$ . This implies  $a_{ji} = w_j / w_i = v_i(x_i')$  (Winterfeldt and Edwards, 1986). Suppose three reciprocal matrices

$$A_1 = \begin{pmatrix} 1 & 1.25 & 2.50 \\ 0.80 & 1 & 1.67 \\ 0.40 & 0.60 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & 1.67 & 2.50 \\ 0.60 & 1 & 5.00 \\ 0.40 & 0.20 & 1 \end{pmatrix}, \text{ and } A_3 = \begin{pmatrix} 1 & 1.11 & 5.00 \\ 0.90 & 1 & 3.33 \\ 0.20 & 0.30 & 1 \end{pmatrix}$$

and the ratios  $w_{21} / w_{11} = 0.30$ ,  $w_{31} / w_{11} = 0.10$ , and  $w_{31} / w_{21} = 0.40$  are obtained.

The first step is to calculate extended regions (Salo, 1993) of lower level ( $3 \times 3$  comparison) matrices. For example, the extended region of  $A_1$  is defined as  $S_1 = \{0.6 \leq w_{12} / w_{11} \leq 0.8, 0.4 \leq w_{13} / w_{11} \leq 0.48, 0.5 \leq w_{13} / w_{12} \leq 0.6, \sum_{i=1}^3 w_{1i} = 1, \text{ and } w_{1i} \geq 0\}$ ; the upper bound of  $w_{12} / w_{11}$  is  $\max \{a_{21} \times a_{11}, a_{22} \times a_{21}, a_{23} \times a_{31}\} = \max \{0.8 \times 1.0, 1.0 \times 0.8, 1.6 \times 0.4\} = 0.8$ .  $S_2$  and  $S_3$  are obtained in the same way. All the pairs are compared with each other in this example because the size of this matrix is  $3 \times 3$ . If the number of criteria is large, however, swing weighting will be a desirable technique (Hayashi, 1998). Then the procedure becomes similar to the approach taken by Keeney and Lilien (Keeney and Lilien, 1987); while in this example inconsistencies are treated explicitly by the overspecification.

The second step is the calculation of the extended region of the upper level. Instead of comparing the criteria at the upper level, the criteria at the lower level are compared with each other to consider the attribute levels explicitly. The lower and upper bounds of  $w_2 / w_1$  are, for example, obtained by minimizing or maximizing  $w_2$  subject to  $w_1 = 1$ ,  $w_1 = \sum_{i=1}^3 w_{1i}$ ,  $w_2 = \sum_{i=1}^3 w_{2i}$ ,  $w_{21} / w_{11} = 0.3$ ,  $(w_{11}, w_{12}, w_{13})$

$\in S_1 \setminus \{\sum_{i=1}^3 w_{1i} = 1\}$ , and  $(w_{21}, w_{22}, w_{23}) \in S_2 \setminus \{\sum_{i=1}^3 w_{2i} = 1\}$ . As a result, a comparison matrix with interval judgments

$$A_0 = \begin{pmatrix} 1 & [2.026, 4.419] & [9.524, 12.214] \\ [0.226, 0.494] & 1 & [1.982, 4.554] \\ [0.082, 0.105] & [0.220, 0.505] & 1 \end{pmatrix}$$

is determined. The extended region of this matrix is defined as  $S_0 = \{0.162 \leq w_2 / w_1 \leq 0.494, 0.050 \leq w_3 / w_1 \leq 0.249, 0.166 \leq w_3 / w_2 \leq 0.505, \sum_{i=1}^3 w_i = 1, \text{ and } w_i \geq 0\}$ .

The intervals of overall values are determined by maximizing or minimizing the weighted value subject to the following constraint:  $w_1 = \sum_{i=1}^3 w_{1i}$ ,  $w_2 = \sum_{i=1}^3 w_{2i}$ ,  $w_3 = \sum_{i=1}^3 w_{3i}$ ,  $(w_1, w_2, w_3) \in S_0$ ,  $(w_{11}, w_{12}, w_{13}) \in S_1 \setminus \{\sum_{i=1}^3 w_{1i} = 1\}$ ,  $(w_{21}, w_{22}, w_{23}) \in S_2 \setminus \{\sum_{i=1}^3 w_{2i} = 1\}$ ,  $(w_{31}, w_{32}, w_{33}) \in S_3 \setminus \{\sum_{i=1}^3 w_{3i} = 1\}$ . Although the number of comparisons is same as the conventional AHP, it will be reduced by combining with swing weighting at the lower level; moreover, it is possible to construct this method as interactive procedures.

### Conclusion

The discussion of this paper probably departs from the philosophy of the AHP — the ratio-judgment procedures. There is a gap between the AHP and multiattribute value theory in choosing axiom systems. This is related to the last criticism which remains unresolved. However, without including the ranges of attribute levels in weighting procedures, it may be difficult to appropriately evaluate the alternatives which have attributes expressed as numerical values (including constructed or proxy attributes) at the bottom of the hierarchy. We will encounter the same problem in environmental life cycle assessment (LCA); though the recent raising of the concern over the environment makes LCA a popular tool, there is still difficulty in weighting (the subjective part of LCA). This impracticability would be untangled only by introducing the attribute ranges into the weighting procedure.

### References

- Alston, J. M., Norton, G. W. and Pardey, P. G. (1995) *Science under Scarcity*, New York: Cornell University Press.
- Belton, V. (1986) "A Comparison of the Analytic Hierarchy Process and a Simple Multi-Attribute Value Function," *European Journal of Operational Research*, 26, 7-21.
- Dodd, F. J., Donegan, H. A. and McMaster, T. B. M. (1995) "Scale Horizons in Analytic Hierarchies," *Journal of Multi-Criteria Decision Analysis*, 4, 177-188.
- Dyer, J. S. (1990) "Remarks on the Analytic Hierarchy Process," *Management Science*, 36, 249-258.
- Fischer, G. W. (1995) "Range Sensitivity of Attribute Weights in Multiattribute Value Models," *Organizational Behavior and Human Decision Processes*, 62, 252-266.
- French, S. (1986) *Decision Theory*, Chichester: Ellis Horwood.
- Goodwin, P. and Wright, G. (1998) *Decision Analysis for Management Judgment* (Second ed.), Chichester: John Wiley & Sons.
- Hamalainen, R. P. and Salo A. A. (1997) "The Issue Is Understanding the Weights," *Journal of Multi-Criteria Decision Analysis*, 6, 340-343.

Hayashi, K. (1998) "Multicriteria Aid for Agricultural Decisions Using Preference Relations: Methodology and Application," *Agricultural Systems*, 58, 483-503.

Hayashi, K. (1999) "Multicriteria Analysis for Agricultural Resource Management: A Critical Survey and Future Perspectives," *European Journal of Operational Research* (to appear).

Keeney, R. L. and Lilien, G. L. (1987) "New Industrial Product Design and Evaluation Using Multiattribute Value Analysis," *Journal of Product Innovation Management*, 4, 185-198.

Lootsma, F. A. (1993) "Scale Sensitivity in the Multiplicative AHP and SMART," *Journal of Multi-Criteria Decision Analysis*, 2, 87-110.

Ma, D. and Zheng, X. (1991) "9/9-9/1 Scale Method of AHP," *Proceedings of the Second International Symposium on the Analytic Hierarchy Process*, Pittsburgh, PA, 197-202.

Marshall, K. T. and Oliver, R. M. (1995) *Decision Making and Forecasting*, New York: McGraw-Hill.

Mollaghasemi, M. and Pet-Edwards, J. (1997) *Making Multiple-Objective Decisions*, Los Alamitos, CA: IEEE Computer Society Press.

Moskowitz, H., Preckel, P. V. and Yang, A. (1992) "Multiple-Criteria Robust Interactive Decision Analysis (MCRID) for Optimizing Public Policies," *European Journal of Operational Research*, 56, 219-236.

Poyhonen, M. and Hamalainen, R. P. (1997) "On the Convergence of Multiattribute Weighting Methods," Systems Analysis Laboratory, Helsinki University of Technology.

Poyhonen, M. A., Hamalainen, R. P. and Salo, A. A. (1997) "An Experiment on the Numerical Modeling of Verbal Ratio Statements," *Journal of Multi-Criteria Decision Analysis*, 6, 1-10.

Saaty, T. L. (1990) "An Exposition of the AHP in Reply to the Paper 'Remarks on the Analytic Hierarchy Process'," *Management Science*, 36, 259-268.

Salo, A. A. (1993) "Inconsistency Analysis by Approximately Specified Priorities," *Mathematical and Computer Modelling*, 17, 123-133.

Salo, A. A. and Hamalainen, R. P. (1992) "Preference Assessment by Imprecise Ratio Statements," *Operations Research*, 40, 1053-1061.

Salo, A. A. and Hamalainen, R. P. (1997a) "On the Measurement of Preferences in the Analytic Hierarchy Process," *Journal of Multi-Criteria Decision Analysis*, 6, 309-319.

Salo, A. A. and Hamalainen, R. P. (1997b) "PRIME — Preference Ratios in Multiattribute Evaluation," Systems Analysis Laboratory, Helsinki University of Technology.

Stewart, T. J. (1992) "A Critical Survey on the Status of Multiple Criteria Decision Making: Theory and Practice," *OMEGA*, 20, 569-586.

von Nitzsch, R. and Weber, M. (1993) "The Effect of Attribute Ranges on Weights in Multiattribute Utility Measurements," *Management Science*, 39, 937-943.

von Winterfeldt, D. and Edwards, W. (1986) *Decision Analysis and Behavioral Research*, Cambridge: Cambridge University Press.

Watson, S. R. and Buede, D. M. (1987) *Decision Synthesis*, Cambridge: Cambridge University Press.