Please use the following reference regarding this paper:

Wang, X., and E. Triantaphyllou, **"Ranking Irregularities When Evaluating Alternatives by Using Some ELECTRE Methods,"** Omega, Vol. 36, No. 1, pp. 45-63, February 2006.

Ranking Irregularities When Evaluating Alternatives by Using Some ELECTRE Methods

Xiaoting Wang
Department of Industrial Engineering
3128 CEBA Building, Louisiana State University
Baton Rouge, LA 70803, U.S.A.
Email: xwang8@lsu.edu
Webpage: http://bit.csc.lsu.edu/~xiaoting

Evangelos Triantaphyllou
Department of Computer Science
298 Coates Hall, Louisiana State University
Baton Rouge, LA 70803, U.S.A.
Email: trianta@lsu.edu
Webpage: http://www.csc.lsu.edu/trianta

Abstract:
The ELECTRE II and III methods enjoy a wide acceptance in solving multi-criteria decision-making (MCDM) problems. Research results in this paper reveal that there are some compelling reasons to doubt the correctness of the proposed rankings when the ELECTRE II and III methods are used. In a typical test we first used these methods to determine the best alternative for a given MCDM problem. Next we randomly replaced a non-optimal alternative by a worse one and repeated the calculations without changing any of the other data. Our computational tests revealed that sometimes the ELECTRE II and III methods might change the indication of the best alternative. We treat such phenomena as rank reversals. Although such ranking irregularities are well known for the additive variants of the AHP method, it is the very first time that they are reported to occur when the ELECTRE methods are used. These two methods are also evaluated in terms of two other ranking tests and they failed them as well. Two real-life cases are described to demonstrate the occurrence of rank reversals with the ELECTRE II and III methods. Based on the three test criteria presented in this paper, some computational experiments on randomly generated decision problems were executed to test the performance of the ELECTRE II and III methods and an examination of some real-life case studies are also discussed. The results of these examinations show that the rates of the three types of ranking irregularities were rather significant in both the simulated decision problems and the real-life cases studied in this paper.

Keywords: Multi-criteria decision-making, ranking irregularities, ELECTRE methods, the Analytic Hierarchy Process (AHP), multiplicative AHP.

1. Introduction

Multi-criteria decision-making (MCDM) is one of the most widely used decision methodologies in the sciences, business, government and engineering worlds. MCDM methods can help to improve the quality of decisions by making the decision-making process more explicit, rational, and efficient. It is not a coincidence that a simple search (for instance, by using google.com) on the web under the key words "multi criteria decision making” returns more than one million hits. Some applications of MCDM in engineering include the use on flexible manufacturing systems [Wabalickis, 1988], layout design [Cambron and Evans, 1991], integrated manufacturing systems [Putrus, 1990], and the evaluation of technology investment decisions [Boucher and Mcstravic, 1991].

The typical MCDM problem is concerned with the task of ranking a finite number of decision alternatives, each of which is explicitly described in terms of different characteristics (also often called attributes, decision criteria, or objectives) which have to be taken into account simultaneously. Usually, the performance values a_{ij} and the criteria weights w_j are viewed as the entries of a *decision matrix* defined as in Figure 1. The a_{ij} element of the decision matrix represents the performance value of the *i-th* alternative in terms of the *j-th* criterion. The w_j represents the weight of the *j-th* criterion. Data for MCDM problems can be determined by direct observation (if they are easily quantifiable) or by indirect means if they are qualitative [Triantaphyllou, *et al.*, 1994].

		Criteria		
	C_1	C_2	...	C_n
	$(w_1$	w_2	...	$w_n)$
Alternatives				
A_1	a_{11}	a_{12}	...	a_{1n}
A_2	a_{21}	a_{22}	...	a_{2n}
	.	.	.	
	.	.		.
	.	.		.
A_m	a_{m1}	a_{m2}	...	a_{mn}

Figure 1. Structure of a Typical Decision Matrix.

Another term that is also used frequently to mean the same type of decision models is multi-criteria decision analysis (MCDA). There is a subtle difference between these two terms. The term MCDM is often used to mean finding the best alternative in a continuous environment. However, in the setting of MCDA, the alternatives are not known a priori but they can be determined by calculating the values of a number of discrete and/or continuous variables. Usually, an MCDA method aims at one of the following four goals, or “problematics” [Roy, 1985], [Jacquet-Lagreze and Siskos, 2001]:

Problematic 1: Find the best alternative.
Problematic 2: Group the alternatives into well-defined classes.
Problematic 3: Rank the alternatives in order of total preference.
Problematic 4: Describe how well each alternative meets all the criteria simultaneously.

Many interesting aspects of MCDA theory and practice are discussed in [Hobbs, 2000; and 1986], [Hobs, *et al.*, 1992], [Stewart, 1992], [Triantaphyllou, 2000], [Zanakis, *et al.*, 1995], and [Zanakis, *et al.*, 1998]. However, the terms MCDM and MCDA may also be used to denote the same class of models.

A prominent role in MCDM methods is played by the Analytic Hierarchy Process (AHP) method which is based on pairwise comparisons as it was proposed by Saaty [1980 and 1994]. According to that method the decision maker compares two decision entities (pair of alternatives considered in terms of a single criterion or a pair of criteria) at a time and elicits his/her judgment with the help of a scale. Such a scale assigns numerical values to linguistic expressions and later these numerical values are analyzed mathematically and the a_{ij} values are determined.

Many methods have been proposed to analyze the data of a decision matrix and rank the alternatives. Often times different MCDM methods may yield different answers to exactly the same problem [Triantaphyllou, 2000]! Some of the methods use additive formulas to compute the final priorities of the alternatives. Representatives of such methods are the weighted sum model (WSM) [Fishburn, 1967], and the Analytic Hierarchy Process (AHP) [Saaty, 1980 and 1994] and its variants (such as the Revised or Ideal Mode AHP [Belton and Gear, 1983]). Some multiplicative versions of these methods have also been developed. Examples are the weighted product model (WPM) [Bridgman, 1922; Miller and Starr, 1969] and a later version of it; the multiplicative AHP [Barzilai and Lootsma, 1994; Lootsma, 1999]. In some earlier research it was found that the previous additive models might often exhibit cases of irregular ranking reversals under certain tests [Triantaphyllou, 2000 and 2001; Triantaphyllou and Mann, 1989; Belton and Gear, 1983]. However, the previous multiplicative models are immune to most of these ranking irregularities.

Another family of MCDM models uses what is known as "outranking relations" to rank a set of alternatives. A prominent role in this group is played by the ELECTRE method and its derivatives. The ELECTRE approach was first introduced in [Benayoun, *et al.*, 1966]. It is a comprehensive evaluation approach in that it also tries to rank a number of alternatives each one of which is described in terms of a number of criteria. The main idea is the proper utilization of what is called "outranking relations". Soon after the introduction of the first version known as ELECTRE I [Roy, 1968], this approach has evolved into a number of other variants. Today the most widely used versions are known as ELECTRE II [Roy and Bertier, 1971, 1973] and ELECTRE III [Roy, 1978]. Another variant of the ELECTRE approach is the TOPSIS method [Hwang and Yoon, 1981]. The TOPSIS method was also found to suffer of the ranking irregularities related to the AHP and its additive variants (according to some unpublished results by the authors).

In contrast to the above approaches, there is a different type of analysis based on value functions. These methods use a number of trade-off determinations which form what are known as value functions [Kirkwood, 1997]. A value function attempts to map changes of values of performance of the alternatives in terms of a given criterion into a dimensionless value. Some key assumptions are made in the process for transferring changes in values into these dimensionless quantities [Kirkwood, 1997]. The roots to this type of analysis can be found in [Edwards, 1977], [Edwards and Barron, 1994], [Edwards and Newman, 1986], and [Dyer and Sarin, 1979].

Of the above MCDM methods a few stand out as being the most widely used, for example, the AHP and the ELECTRE approaches. Cases of ranking irregularities when the AHP is used have been reported by many researchers for a number of years [Belton and Gear, 1983; Dyer and Wendell, 1985; Triantaphyllou, 2000 and 2001]. The ELECTRE method is a well known method, especially in Europe. It has been widely used in civil and environmental engineering [Hobbs and Meier, 2000]. Applications include the assessment of complex civil engineering projects, selection of highway designs, site selection for the disposal of nuclear waste, water resources planning [Raj, 1995] and waste water [Roger, *et al.*, 1999] or solid waste management [Hokkanen and Salminen, 1997a] etc. However, to the best of our knowledge, this is the first time that ranking irregularities are reported for the ELECTRE approach.

This paper is organized as follows. The next section discusses the three test criteria that have been used in this study to test the performance of the ELECTRE II and III methods. The third section describes two examples that are based on two real-life decision problems for which rank reversals occurred under test criterion #1 by using the ELECTRE II and III methods. The fourth section presents some empirical results on randomly generated decision problems according to the three test criteria described in section 2. The fifth section discusses the results based on the test criteria of some real-life case studies. Some concluding comments are presented in the last section.

2. Some Test Criteria

An intriguing problem with decision-making methods which rank a set of alternatives in terms of a number of competing criteria is that oftentimes different methods may yield different answers (rankings) when they are fed with exactly the same numerical data. Thus, the issue of evaluating the relative performance of such methods is naturally raised. This, in turn, raises the question how can one evaluate the performance of such methods? Since it is practically impossible to know which one is the best alternative for a given decision problem, some kind of testing procedures need to be determined. The above subject, along with some rank irregularity issues, has been discussed in detail in [Triantaphyllou, 2000 and 2001]. In those studies, three test criteria were established to test the relative performance of various MCDM methods. These test criteria are as follows:

Test Criterion #1:

An effective MCDM method should not change the indication of the best alternative when a non-optimal alternative is replaced by another worse alternative (given that the relative importance of each decision criterion remains unchanged).

Suppose that an MCDM method has ranked a set of alternatives in some way. Next, suppose that a non-optimal alternative, say A_k, is replaced by another alternative, say $A_k^{/}$, which is less desirable than A_k. Then, according to the test criterion #1 the indication of the best alternative should not change when the alternatives are ranked again by the same method. The same should also be true for the relative rankings of the rest of the unchanged alternatives.

Test Criterion #2:

The rankings of alternatives by an effective MCDM method should follow the transitivity property.

Suppose that an MCDM method has ranked a set of alternatives of a decision problem in some way. Next, suppose that this problem is decomposed into a set of smaller problems, each defined on two alternatives at a time and the same number of criteria as in the original problem. Then, according to this test criterion all the rankings which are derived from the smaller problems should satisfy the transitivity property. That is, if alternative A_1 is better than alternative A_2, and alternative A_2 is better than alternative A_3, then one should also expect that alternative A_1 is better than alternative A_3.

The third test criterion is similar to the previous one but now one tests for the agreement between the smaller problems and the original un-decomposed problem.

Test Criterion #3:

For the same decision problem and when using the same MCDM method, after combining the rankings of the smaller problems that an MCDM problem is decomposed into, the new overall ranking of the alternatives should be identical to the original overall ranking of the un-decomposed problem.

As before, suppose that an MCDM problem is decomposed into a set of smaller problems, each defined on two alternatives and the original decision criteria. Next suppose that the rankings of the smaller problems follow the transitivity property. Then, according to this test criterion when the rankings of the smaller problems are all combined together, the new overall ranking of the alternatives should be identical to the original overall ranking before the problem decomposition.

We used these three test criteria to evaluate the performance of the ELECTRE II and the ELECTRE III methods. Both of them failed in terms of each one of these three test criteria. Next we demonstrate two rank reversal examples which occurred with the ELECTRE II and III methods under the first test criterion. The other two test criteria were also applied as they had been stated above.

3. Illustration of Rank Reversals with ELECTRE II and III

For most ELECTRE methods, there are two main stages. These are the construction of the outranking relations and the exploitation of these relations to get the final ranking of the alternatives. Different ELECTRE methods may be different in how they define the outranking relations between alternatives and how they apply these relations to get the final ranking of the alternatives. This is true with the ELECTRE II and III methods. However, the essential difference between these two methods is that they use different types of criteria. ELECTRE II uses the true criteria where no thresholds exist and the differences between criteria scores are used to determine which alternative is preferred. In this preference structure, the indifference relation is transitive [Rogers, *et al.*, 1999]. The criteria used by ELECTRE III are pseudo-criteria which involve the use of two-tiered thresholds. One is the indifference threshold q, beneath which the decision maker shows clear indifference, and the other one is the preference threshold p, above which the decision maker is certain of strict preference [Rogers, *et al.*, 1999]. The situation between the above two is regarded as weak preference for alternative *a* over alternative *b* which indicates the decision maker's hesitation between indifference and strict preference [Rogers, *et al.*, 1999]. The following two rank reversal examples demonstrate how both of the two methods work and how the rank reversals may happen when using them to rank a set of decision alternatives.

3.1 An Example of Rank Reversal with the ELECTRE II Method

This example is based on a real-life case study where the ELECTRE II method was used to help find the best location for a wastewater treatment plant in Ireland [Rogers, *et al.*, 1999]. The decision problem is defined on 5 alternatives and 7 criteria. Note that here all the criteria are *benefit* criteria, that is, the higher the score the better the performance is. The decision matrix, that is, the performances of the alternatives A_i in terms of the criteria C_j, is as follows:

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
A_1	1	2	1	5	2	2	4
A_2	3	5	3	5	3	3	3
A_3	3	5	3	5	3	2	2
A_4	1	2	2	5	1	1	1
A_5	1	1	3	5	4	1	5

The weights of the criteria are:

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
Weight	0.0780	0.1180	0.1570	0.3140	0.2350	0.0390	0.0590

The ELECTRE methods are based on the evaluation of two indices, the *concordance* index and the *discordance* index, defined for each pair of alternatives. The concordance index for a pair of alternatives *a* and *b* measures the strength of the hypothesis that alternative *a* is at least as good as alternative *b*. The discordance index measures the strength of evidence against this hypothesis [Belton and Stewart, 2001]. There are no unique measures of concordance and discordance indices. In ELECTRE II, the concordance index *C(a, b)* for each pair of alternatives *(a, b)* is defined as follows:

$$C(a,b) = \frac{\sum_{i \in Q(a,b)} w_i}{\sum_{i=1}^{m} w_i},$$

where *Q (a, b)* is the set of criteria for which *a* is equal or preferred to (i.e., at least as good as) *b*, and w_i is the weight of the *i-th* criterion. For instance, the concordance indices for this example are as follows:

	A_1	A_2	A_3	A_4	A_5
A_1	1.0000	0.3730	0.4120	0.8430	0.5490
A_2	0.9410	1.0000	1.0000	1.0000	0.7060
A_3	0.9410	0.9020	1.0000	1.0000	0.7060
A_4	0.6670	0.3140	0.3140	1.0000	0.5490
A_5	0.8430	0.7650	0.7650	0.8820	1.0000

The discordance index *D (a, b)* for each pair of alternatives *(a, b)* is defined as follows:

$$D(a,b) = \frac{\max_i [g_i(b) - g_i(a)]}{\delta},$$

Where $g_i(a)$ represents the performance of alternative *a* in terms of criterion C_i, $g_i(b)$ represents the performance of alternative *b* in terms of criterion C_i, and $\delta = \max_i |g_i(b) - g_i(a)|$ (i.e., the maximum difference on any criterion). This formula can only be used when the scores for different criteria are comparable. When the above formula is used, it turns out that the discordance indices for this example are as follows:

	A_1	A_2	A_3	A_4	A_5
A_1	0.0000	0.7500	0.7500	0.2500	0.5000
A_2	0.2500	0.0000	0.0000	0.0000	0.5000
A_3	0.5000	0.2500	0.0000	0.0000	0.7500
A_4	0.7500	0.7500	0.7500	0.0000	1.0000
A_5	0.2500	1.0000	1.0000	0.2500	0.0000

After computing the concordance and discordance indices for each pair of alternatives, two types of outranking relations are built by comparing these indices with two pairs of threshold values: (C^*, D^*) and (C^{-}, D^{-}). The pair (C^*, D^*) is defined as the concordance and discordance thresholds for the *strong* outranking relation and the pair (C^{-}, D^{-}) is defined as the thresholds for the *weak* outranking relation where $C^* > C^{-}$ and $D^* < D^{-}$. Next the outranking relations are built according to the following two rules:

(1) If $C(a, b) \geq C^*$, $D(a, b) \leq D^*$ and $C(a, b) \geq C(b, a)$, then alternative *a* is regarded as strongly outranking alternative *b*.

(2) If $C(a, b) \geq C^{-}$, $D(a, b) \leq D^{-}$ and $C(a, b) \geq C(b, a)$, then alternative *a* is regarded as weakly outranking alternative *b*.

The values of (C^*, D^*) and (C^{-}, D^{-}) are decided by the decision makers for a particular outranking relation. They may be varied to yield more or less severe outranking relations. The higher the value of C^* and the lower the value of D^*, the more severe the outranking relation becomes. That is, the more difficult it is for one alternative to outrank another [Belton and Stewart, 2001].

For this example, two pairs of thresholds for the strong outranking relation and one pair of thresholds for the weak outranking relation were chosen to be as follows: $C_1^*=0.85$, $D_1^*=0.50$; $C_2^*=0.75$, $D_2^*=0.25$; and $C^{-}=0.65$, $D^{-}=0.25$. According to the above rules and the three pairs of thresholds, the outranking relations for this example were derived to be as follows:

	A_1	A_2	A_3	A_4	A_5
A_1	—			S^F	
A_2	S^F	—	S^F	S^F	
A_3	S^F		—	S^F	
A_4				—	
A_5	S^F			S^F	—

In the above notation S^F stands for the strong outranking relation. For example, $A_1\ S^F\ A_4$ means that alternative A_1 strongly outranks alternative A_4. We use S^f (i.e., the superscript now is low case "f"; not present on the above table) to stand for the weak outranking relation. The weak outranking relation would happen later in this example.

On the basis of the outranking relations, next the descending and ascending distillation processes are applied to obtain two complete pre-orders of the alternatives. The details of the distillation processes can be found in [Belton and Stewart, 2001] and [Rogers, *et al.*, 1999]. The descending pre-order is built up by starting with the set of "best" alternatives (those which outrank other alternatives) and going downward to the worse one. On the contrary, the ascending pre-order is built up by starting with the set of "worst" alternatives (those which are outranked by other alternatives) and going upward to the best one. The distillation results for this example are as follows: the pre-order from the descending distillation is $\boldsymbol{A_2 = A_5 > A_3 > A_1 > A_4}$; the pre-order from the ascending distillation is $\boldsymbol{A_2 > A_5 = A_3 > A_1 > A_4}$.

The last step is to combine the two complete pre-orders to get either a partial or a complete final pre-order. Whether the final product is a partial pre-order (not containing a relative ranking of all of the alternatives) rather than a complete pre-order depends on the level of consistency between the rankings from the two distillation procedures [Rogers, *et al.*, 1999]. The partial pre-order allows two alternatives to remain incomparable without affecting the validity of the overall ranking, which differentiates from the complete pre-order. A commonly used method for determining the final pre-order is to take the intersection of the descending and ascending pre-orders. The intersection of the two pre-orders is defined such that alternative *a* outranks alternative *b* if and only if *a* outranks or is in the same class as *b* according to the two pre-orders. If alternative *a* is preferred to alternative *b* in one pre-order but *b* is preferred to *a* in the other one, then the two alternatives are incomparable in the final pre-order [Rogers, *et al.*, 1999]. By following the above rules, the intersection of the two pre-orders for this example resulted in the following complete pre-order of the alternatives: $A_2 > A_5 > A_3 > A_1 > A_4$ and obviously A_2 is the optimal alternative at this point.

Next, alternative A_3 was randomly selected to be replaced by a worse one, say A_3', in order to test the stability of the alternatives' ranking under the first test criterion. The new decision matrix now is as follows:

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
A_1	1	2	1	5	2	2	4
A_2	3	5	3	5	3	3	3
A_3'	**2**	**4**	**2**	**4**	**2**	**1**	**1**
A_4	1	2	2	5	1	1	1
A_5	1	1	3	5	4	1	5

Please note that alternative A_3 is replaced by a less desirable one denoted as A_3' which is determined by subtracting the value 1 from the performance values of the original alternative A_3 (the subtracted value was selected randomly by a computer program to make sure it will cause the chosen alternative to become worse than the one it replaces).

The rest of the data are kept the same as before. The intermediate results during the ranking process are as follows:

The concordance indices are:

	A_1	A_2	A_3'	A_4	A_5
A_1	—	0.3730	0.6470	0.8430	0.5490
A_2	0.9410	—	1.0000	1.0000	0.7060
A_3'	0.5880	0	—	0.6860	0.2350
A_4	0.6670	0.3140	0.5690	—	0.5490
A_5	0.8430	0.7650	0.8040	0.8820	—

The discordance indices are:

	A_1	A_2	A_3'	A_4	A_5
A_1	—	0.7500	0.5000	0.2500	0.5000
A_2	0.2500	—	0	0	0.5000
A_3'	0.7500	0.5000	—	0.2500	1.0000
A_4	0.7500	0.7500	0.5000	—	1.0000
A_5	0.2500	1.0000	0.7500	0.2500	—

The outranking relations are:

	A_1	A_2	A_3'	A_4	A_5
A_1	—			S^F	
A_2	S^F	—	S^F	S^F	
A_3'			—	S^f	
A_4				—	
A_5	S^F			S^F	—

where S^f (in location (3, 4)) stands for the weak outranking relation.

When the descending and ascending distillation processes are applied again, the descending pre-order now is $A_2 = A_5 > A_3 = A_1 > A_4$ while the ascending pre-order is $A_2 = A_5 > A_3 = A_1 > A_4$. After combining the two pre-orders together, a new complete pre-order is got as follows: $A_2 = A_5 > A_3 = A_1 > A_4$. Now the best ranked alternatives are A_2 and A_5 together, a contradiction from the previous result that had A_2 as the only optimal alternative.

3.2 An Example of Rank Reversal with the ELECTRE III Method

This illustrative example is also based on a real-life decision problem which was defined on 11 alternatives and 11 decision criteria. The goal of this case is to choose the best waste incineration strategy for the eastern Switzerland region [Rogers, *et al.*, 1999]. In this example, the first test criterion reveals a case of rank reversal when the ELECTRE III method is used. The main data are as follows:

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}
A_1	125	866	9.81	218	1.41	542	483	23	1.5	1	1
A_2	11,980	900	11.45	189	1.45	452	303	12	1.5	6	6
A_3	31,054	883	9.86	172	1.82	341	311	0	0	3	3
A_4	28,219	840	10.38	171	1.95	339	318	0	0	3	3
A_5	31,579	903	10.74	165	1.7	312	281	0	0	5	5
A_6	39,364	922	13.87	167	1.65	287	269	0	0	8	7
A_7	125	769	9.33	182	1.64	458	180	0	1.5	1	1
A_8	8,075	896	9.82	172	1.7	408	121	0	1.5	6	6
A_9	3,089	770	9.39	177	1.9	430	228	0	1	2	2
A_{10}	6,449	766	7.22	172	1.65	401	157	0	1	4	4
A_{11}	12,074	897	10.61	169	1.65	378	162	0	1	7	6

Please note that in this example, criteria C_2, C_6 and C_7 are *benefit* criteria, which means the higher the score of a given criterion is, the more preferable it is. The other criteria are *cost* criteria, which means the lower the score of a given criterion is, the more preferable it is.

The weights W, the indifference thresholds Q and the preference thresholds P of the criteria are as follows:

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}
W	0.16	0.033	0.033	0.097	0.097	0.16	0.097	0.16	0.033	0.033	0.097
Q	$\pm$1,000	10%	10%	$\pm$5	10%	10%	10%	$\pm$2	0.2	$\pm$0	$\pm$0
P	$\pm$2,000	20%	20%	$\pm$10	20%	20%	20%	$\pm$4	0.4	$\pm$1	$\pm$1

*Please note that in this example, no veto threshold is specified as it is the case in the referenced paper.

Next the concordance index $C_i\ (a, b)$ calculated for each pair of alternatives (a, b) in terms of each one of the decision criteria according to the following formula:

$$C_i(a,b)=\begin{cases}1, & if\ \ z_i(a)+q_i(z_i(a))\geq z_i(b)\\ 0, & if\ \ z_i(a)+p_i(z_i(a))\leq z_i(b)\end{cases} \tag{3-1}$$

or by linear interpolation between 0 and 1 when $z_i\,(a)+q_i\,(z_i\,(a))<z_i\,(b)<z_i(a)+p_i\,(z_i\,(a))$, where $q_i\,(.)$ and $p_i\,(.)$ are the indifference and preference threshold values for criterion C_i [Belton and Stewart, 2001]. For instance, the concordance indices in terms of the first decision criterion, which is $C_1\ (a, b)$, are as follows:

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
A_1	—	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
A_2	0.00	—	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00
A_3	0.00	0.00	—	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00
A_4	0.00	0.00	1.00	—	1.00	1.00	0.00	0.00	0.00	0.00	0.00
A_5	0.00	0.00	1.00	0.00	—	1.00	0.00	0.00	0.00	0.00	0.00
A_6	0.00	0.00	0.00	0.00	0.00	—	0.00	0.00	0.00	0.00	0.00
A_7	1.00	1.00	1.00	1.00	1.00	1.00	—	1.00	1.00	1.00	1.00
A_8	0.00	1.00	1.00	1.00	1.00	1.00	0.00	—	0.00	0.37	1.00
A_9	0.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	—	1.00	1.00
A_{10}	0.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	0.00	—	1.00
A_{11}	0.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	—

The next step is to calculate the discordance index $D_i\ (a, b)$ for all the alternatives in terms of each one of the decision criteria according to the following formula:

$$D_i(a,b)=\begin{cases}0, & if \;\; z_i(b)\le z_i(a)+p_i(z_i(a)) \\ 1, & if \;\; z_i(b)\ge z_i(a)+t_i(z_i(a))\end{cases} \tag{3-2}$$

or by linear interpolation between 0 and 1 when $z_i(a)+p_i(z_i(a))<z_i(b)<z_i(a)+t_i(z_i(a))$, where $t_i(.)$ is the veto threshold for criterion C_i [Belton and Stewart, 2001]. If no veto threshold is specified, then $D_i(a, b)=0$ for all pairs of alternatives. For instance, in this example, since no veto thresholds are specified, the discordance indices in terms of each decision criterion are all equal to zero.

The next step is to calculate the overall concordance index $C(a, b)$ of all the alternatives by applying the following formula:

$$C(a,b)=\frac{\sum_{i=1}^{m} w_i C_i(a,b)}{\sum_{i=1}^{m} w_i}. \tag{3-3}$$

Finally, the credibility matrix $S(a, b)$ of all the alternatives is calculated by applying the following formula:

$$S(a,b)=\begin{cases}C(a,b), & if \;\; D_i(a,b)\le C(a,b)\forall i \\ C(a,b)\prod_{i\in J(a,b)}\frac{(1-D_i(a,b))}{(1-C_i(a,b))}, & otherwise\end{cases} \quad , \tag{3-4}$$

where $J(a, b)$ is the set of criteria for which $D_i(a, b)>C(a, b)$. The credibility matrix is a measure of the strength of the claim that "alternative *a* is at least as good as alternative *b*". For this case, the credibility matrix is equal to the concordance matrix since no veto thresholds are assigned so the discordance matrices are all zero matrices, which results to $S(a, b) = C(a, b)$ and both are as follows:

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
A_1	0.00	0.74	0.71	0.71	0.71	0.71	0.74	0.74	0.71	0.68	0.71
A_2	0.44	0.00	0.57	0.58	0.58	0.71	0.48	0.57	0.39	0.39	0.71
A_3	0.36	0.58	0.00	0.84	0.96	1.00	0.55	0.69	0.55	0.69	0.83
A_4	0.36	0.58	1.00	0.00	0.95	0.95	0.49	0.65	0.55	0.62	0.76
A_5	0.38	0.63	0.86	0.68	0.00	1.00	0.54	0.68	0.54	0.52	0.68
A_6	0.38	0.48	0.48	0.47	0.68	0.00	0.52	0.52	0.52	0.52	0.52
A_7	0.72	0.86	0.76	0.77	0.75	0.74	0.00	0.88	0.87	0.84	0.85
A_8	0.38	0.84	0.74	0.74	0.70	0.87	0.58	0.00	0.58	0.61	0.87
A_9	0.35	0.78	0.85	0.85	0.74	0.73	0.67	0.97	0.00	0.94	0.89
A_{10}	0.40	0.81	0.72	0.74	0.81	0.84	0.60	0.98	0.61	0.00	0.98
A_{11}	0.41	0.70	0.74	0.74	0.74	0.87	0.53	0.81	0.55	0.68	0.00

Next the descending and ascending distillations [Belton and Stewart, 2001; Rogers, *et al.*, 1999] based on the credibility matrix are applied to construct two pre-orders for the alternatives. The pre-order obtained from the descending distillation is as follows: $A_9 > A_4 > A_7 > A_{10} > A_3 = A_5 = A_8 = A_{11} > A_1 > A_2 > A_6$. The pre-order obtained from the ascending distillation is as follows: $A_1 = A_7 > A_9 > A_4 > A_{10} > A_2 = A_5 > A_3 = A_{11} > A_8 > A_6$. Then the two pre-orders are combined to get the final overall ranking of the alternatives as shown in Figure 2.

The way to combine the two pre-orders is the same as that of ELECTRE II, which has been described in the first example. The arrow line in Figure 2 means 'outrank'. For example, A_9 outranks A_4. Two alternatives are incomparable if there is no direct or indirect arrow line to link them together. For example, A_7 and A_9 are incomparable. Alternatives are indifferent if they are at the same level. For example, A_3 and A_{11} are indifferent with each other. We can see now that A_7 and A_9 are both located at the top level and they are incomparable with each other. Incomparability may be caused by the lack of the criterion information of the alternatives. This means that there is no clear evidence in favor of either A_7 or A_9. In real-life applications of ELECTRE II and III methods, the decision analysts will need to find more information about such alternatives and do a further study to decide which one is the best one. However, for the simplicity of the current test, A_7 and A_9 are both regarded as the best-ranked alternatives because both of them are ranked first in the final partial pre-order. As a result, the rest of the alternatives were regarded as the non-optimal ones.

Next, according to the first test criterion, we randomly selected one of the non-optimum alternatives; say alternative A_1 to be replaced by a worse one $A_1^{/}$ to test the reliability of the alternatives' ranking. Since the performance value of A_1 in terms of each criterion was [125 866 9.81 218 1.41 542 483 23 1.5 1 1], we subtracted [-2,000 0 0 0 0 90 0 0 0 0 0] (the subtracted value was selected randomly by a computer program to make sure it will make the chosen alternative to become worse than before) from A_1 to get $A_1^{/}$ which is less desirable than A_1. Please recall that the first criterion is a *cost* criterion which means the bigger a score the less desirable it is. So the performance values of $A_1^{/}$ are [2,125 866 9.81 218 1.41 452 483 23 1.5 1 1]. The new decision matrix with the a_{ij} values of the alternatives (after alternative A_1 is replaced by $A_1^{/}$) is as follows:

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}
$A_1^{/}$	**2,125**	**866**	**9.81**	**218**	**1.41**	**452**	**483**	**23**	**1.5**	**1**	**1**
A_2	11,980	900	11.45	189	1.45	452	303	12	1.5	6	6
A_3	31,054	883	9.86	172	1.82	341	311	0	0	3	3
A_4	28,219	840	10.38	171	1.95	339	318	0	0	3	3
A_5	31,579	903	10.74	165	1.7	312	281	0	0	5	5
A_6	39,364	922	13.87	167	1.65	287	269	0	0	8	7
A_7	125	769	9.33	182	1.64	458	180	0	1.5	1	1
A_8	8,075	896	9.82	172	1.7	408	121	0	1.5	6	6
A_9	3,089	770	9.39	177	1.9	430	228	0	1	2	2
A_{10}	6,449	766	7.22	172	1.65	401	157	0	1	4	4
A_{11}	12,074	897	10.61	169	1.65	378	162	0	1	7	6

The rest of the data are kept the same. When the previous steps are applied on the modified problem again, we get that the descending pre-order is $A_7 > A_9 > A_4 > A_{10} > A_3 = A_5 = A_8 = A_{11} > A_1 > A_2 > A_6$ and the ascending pre-order is $A_7 > A_1 = A_9 > A_4 > A_{10} > A_2 = A_5 > A_3 = A_{11} > A_8 > A_6$. The overall ranking of the alternatives is as shown in Figure 3. This time it turns out that the best-ranked alternative now is only A_7 which is different from the original conclusion which had A_7 and A_9 as the best-ranked alternatives.

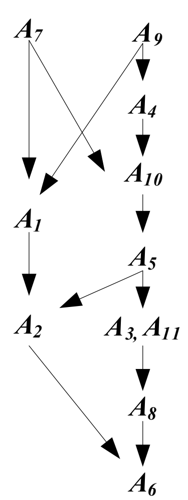


Figure 2. Ranking for the original example.

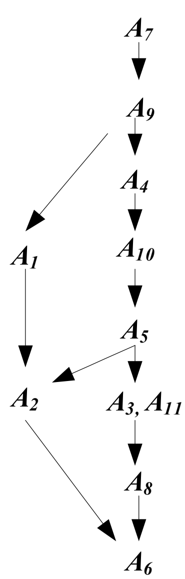


Figure 3. Ranking for the changed example.

Why did the above contradictions occur? When A_1 was replaced by a worse one, it is reasonable to assume that some alternatives which originally are ranked lower than A_1 may become more preferable than it. However, there is no legitimate reason why the optimal alternative should also be changed and why the original incomparable relation between two equally ranked alternatives should also be changed.

4. An Analysis of the Causes of the Rank Reversals under the ELECTRE II and III Methods

After analyzing the ranking processes of the ELECTRE II and III methods and some rank reversal cases which occurred when these methods were used, it was found that the main reason for the above rank reversals lies in the exploitation of the pairwise outranking relations. That is, the upward and downward distillation processes of ELECTRE II and ELECTRE III. The basic idea behind the distillation processes is to decide the rank of each alternative by the degree of how this alternative outranks all the other alternatives. When a non-optimal alternative in an alternative set is replaced by a worse one, the pairwise outranking relations related to it may be changed accordingly and the overall ranking of the whole alternative set, which depends on those pairwise outranking relations, may also be changed. The first change is reasonable when considering the fact that a non-optimal alternative has been replaced by a worse one. However, the second change is unreasonable and may cause undesirable rank reversals as in the examples presented in section 3.

As it is demonstrated in the next section when one decomposes a decision problem into smaller problems and analyzes them by using the ELECTRE II or III method, the rankings of the smaller problems may not follow the transitivity property. This fact along with the above rank reversal examples reveals that there is not an priori ranking of the alternatives when they are ranked by the ELECTRE II or III methods because the ranking of an individual alternative derived by these methods depends on the performance of all the other alternatives currently under consideration. This causes the ranking of the alternatives to depend on each other. Thus, it is likely that the optimal alternative may be different and the ranking of the alternatives may be distorted to some extent if one of the non-optimal alternatives in the alternative set is replaced by a worse one.

This can be further explained by means of a simple example. Given three alternatives: A_1, A_2, and A_3, suppose that originally A_1 strongly outranks A_3, A_2 weakly outranks A_3 and A_1 and A_2 are indifferent with each other. The ranking of these three alternatives will be $A_1 > A_2 > A_3$ when using the ELECTRE II method. Next, if the non-optimal alternative A_3 is replaced by a worse one, then A_2 may strongly outrank A_3 while A_1 is still strongly outranking A_3 and A_1 is still indifferent with A_2. Nothing is wrong so far. But now the ranking of the three alternatives will be $A_1 = A_2 > A_3$ by using the same method since both A_1 and A_2 now strongly outrank A_3 and they are indifferent with each other. It can be seen that A_1 and A_2 are ranked equally now because A_3 becomes less desirable. This is exactly what happened in the first example: A_2 and A_5 are ranked equally after A_3 has been replaced by a less desirable alternative. This kind of irregular situation is undesirable for a practical decision-making problem though it is reasonable in terms of the logic of the ELECTRE II method. It could leave the ranking of a set of alternatives to be manipulated to some extent.

The ranking irregularity in the above example is very likely to occur when using the ELECTRE II or III method to rank a set of alternatives. If the number of alternatives of a decision problem is more than 3, there will be more than C_3^2 (=6) pairwise outranking relations between them. Then the situation may become worse by totally changing the indication of the best ranked alternative. It was once pointed out in [Belton and Stewart, 2001] that the results of the distillations are dependent on the whole alternative set, so that the addition or removal of an alternative can alter some of the preferences between the remaining alternatives. A similar situation occurs with the PROMETHEE method which is another variant of the outranking method. In [Keyser and Peeters, 1996], it was pointed out that the complete pre-orders from the PROMETHEE method are based on an all-to-all comparison between the alternatives; adding or deleting an alternative can put the previous pre-orders upside down. From the study reported in this paper, now it can be seen that a similar situation also occurs with the ELECTRE II and III methods. That is, even without addition or removal of alternatives, the best ranked alternative might be altered and the previous pre-order between the remaining alternatives might be changed to some degree by just replacing a non-optimal alternative by a worse one.

It must be pointed out here that there is another factor that may contribute to rank reversals. During the construction of the pairwise outranking relations, both ELECTRE II and III need to use a value or a threshold which is also dependent on the performance values of all the currently considered alternatives. For ELECTRE II, it is the parameter δ (i.e., the maximum difference of any criterion) in the discordance index formula. For ELECTRE III, it is the parameter λ used to decide the $\lambda - preference$ relations between the alternatives during the distillations. These δ and λ values may be altered when a non-optimal alternative is replaced by a worse one. Then the previous outranking relations between the other unchanged alternatives may be distorted to some degree, which finally may alter the indication of the best ranked alternative or the overall ranking of the alternatives. According to some experimental analysis, the above two factors may function together or separately to cause rank reversals.

From the above analysis, it can be seen that the ranking processes of the ELECTRE II and III are not reliable and robust enough to offer a firm answer to a decision problem. Usually, decision makers undertake some kind of sensitivity analysis to appreciate the sensitivity of the final rankings and the robustness of the ranking procedures to changes in the criteria weights and thresholds when they use ELECTRE methods to solve decision problems. However, the above ranking irregularities can warn decision analysts that they should be cautious in accepting the ranking recommendations of the ELECTRE methods even after a careful sensitivity analysis is undertaken.

5. An Empirical Study

This section describes an empirical study that focused on how often these ranking irregularities may happen under the ELECTRE II and III methods. Some computer programs were written in MATLAB in order to generate simulated decision problems and test the performance of ELECTRE II and III under the three test criteria described in section 2. In these test problems, the number of the alternatives was equal to the following ten different values: 3, 5, 7, 9, 11, 13, 15, 17, 19, and 21. However, there is not a common range of the criteria for all the tests. Compared with the tests of ELECTRE II, a wider range of criteria for the tests of ELECTRE III was needed in these experiments in order to clearly show how the ranking irregularity rates under ELECTRE III will fluctuate with the increase on the number of the criteria. For the three tests of ELECTRE II, the number of criteria was equal to 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31. Thus, a total of 150 different cases were examined with 10,000 randomly generated decision problems (in order to derive statistically significant results) per case. For the three tests of ELECTRE III, the number of criteria was equal to the odd numbers between 3 and 61. Thus, a total of 300 different cases were examined with 10,000 random decision problems per case. Each random decision problem was analyzed first by using the ELECTRE II or III method and then was analyzed again by using the same method after one of the non-optimal alternatives was replaced by a worse one or the whole decision problem was decomposed into smaller problems as described in the last two test criteria. Any occurred ranking irregularity was recorded. Figures 4 to 9 summarize these test results. In these figures, different curves correspond to cases with different numbers of alternatives; the X axis stands for the number of criteria and the Y axis is the rate of ranking irregularities that occurred in the 10,000 simulated decision problems.

Figures 4 and 5 describe how often rank reversal happened to ELECTRE II and III methods under test criterion #1 in this empirical test. That is, how often the indication of the best alternative is changed when a non-optimal alternative is replaced by another worse alternative (given that the relative importance of each decision criterion remains unchanged). The basis of any ELECTRE method is to decide the pairwise outranking relations between alternatives. Given *n* alternatives, when a non-optimal alternative was replaced by a worse one, the number of pairwise outranking relations that might be changed is at most *(n-1)*. This indicates that the higher the number of the alternatives is the more possible becomes that the variation of a single alternative may have a noticeable influence on the outranking relations. It is the change of the outranking relations that results in the rank reversals. This is why the rank reversal rates usually increase with the increase on the number of alternatives.

It should be clarified here that even if a case passed test criterion #1, this does not mean that this case is immune to the rank reversal situation described in test criterion #1. When applying test criterion #1, one non-optimal alternative needs to be pick up and replaced by a worse one. Which non-optimal alternative will be selected and how worse it could be to trigger the rank reversal to happen were all randomly chosen by the program. When replacing a non-optimal alternative by a worse one, the program only makes the selected non-optimal alternative to be worse than before to a certain degree to test if it is enough to trigger the rank reversal to occur. If no rank reversal happens, the case will be released and marked as having passed test criterion #1. It is not possible to test all the possibilities in terms of given single case. Therefore, even if a case passed test criterion #1 in a single experiment, that does not mean it is immune to the type one rank reversal.

Figures 6 and 7 depict how the ranking irregularity rates of ELECTRE II and III varied with the increase on the number of alternatives and the number of criteria in terms of test criterion #2. One can see from these figures that the rates generally increase with the increase on the number of alternatives. This happens because the higher the number of alternatives is, the higher is the number of smaller problems that a decision problem was decomposed into, and then the more likely it is for a contradiction between the smaller problems to happen.

In order to apply test criterion #3, 10,000 random decision problems whose rankings follow the transitivity property by using the ELECTRE II or III method must be generated and then be examined under test criterion #3. However, as one can see from Figures 6 and 7, when the number of alternative is up to 7 or 9, the rankings of the random decision problems almost never follow the transitivity property when the number of criteria is in some range. It is difficult to find 10,000 random decision problems per case that can be used in the third test. Thus, only the cases where the number of alternatives was equal to 3, 5 or 7 were tested. Figures 8 and 9 show how often the ranking irregularity will happen to these cases under test criterion #3. The reason why this rate increases with the number of alternatives is the same as that of the experiments under test criterion #2.

What is the relationship between these ranking irregularity rates with the number of the decision criteria? From these figures one can see that, in general, the ranking irregularity rates will first increase with the increase on the number of the criteria but then decrease when the number of criteria increased beyond a certain value for each case. Please recall that the pairwise outranking relations between each pair of alternatives are decided by the concordance and discordance indices which are computed by their performance under each criterion. For a fixed number of alternatives, with the increase on the number of criteria, if the number of criteria is beyond some value, the pairwise outranking relations and the subsequent ranking of the alternatives is more likely to become more stable than before. Just like it will be more believable if one decides that Car A is better than Car B in terms of 7 decision criteria than in terms of 3 decision criteria.

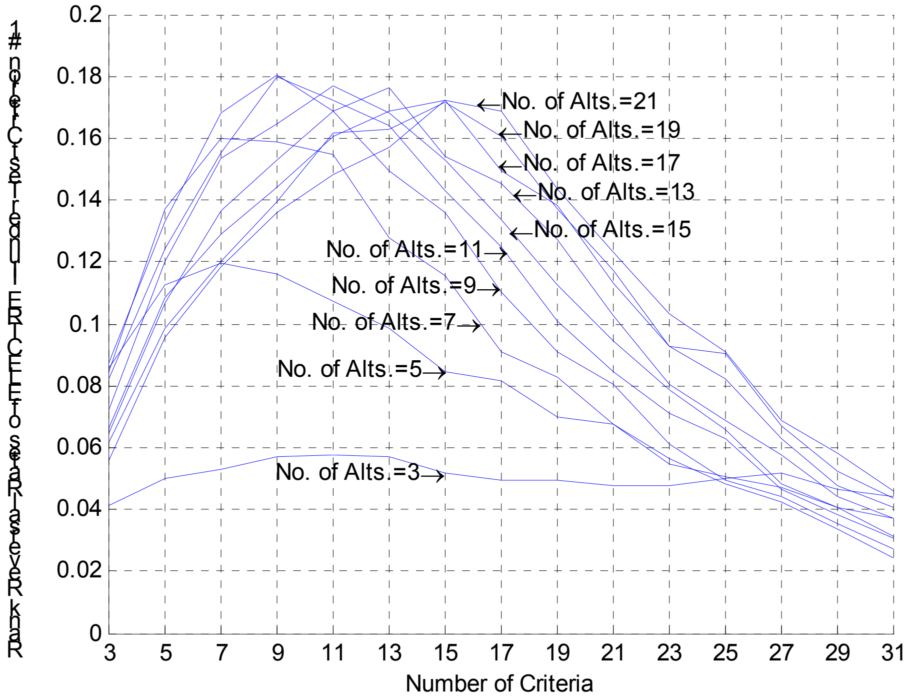


Figure 4. Rank Reversal Rates of ELECTRE II under Test Criterion #1.

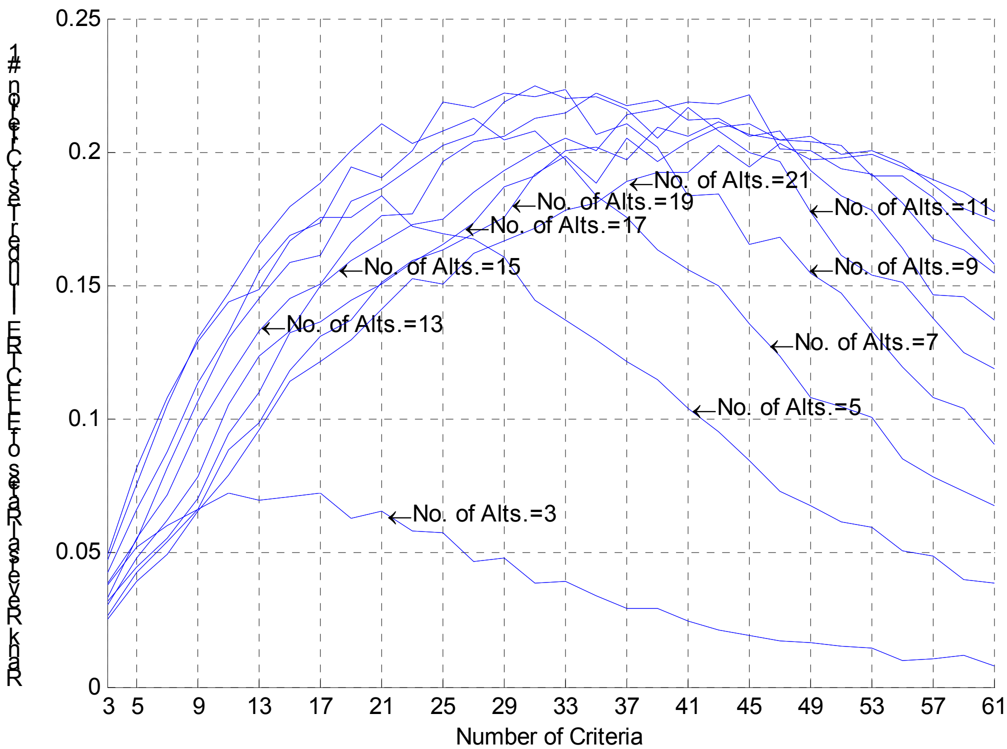


Figure 5. Rank Reversal Rates of ELECTRE III under Test Criterion #1.

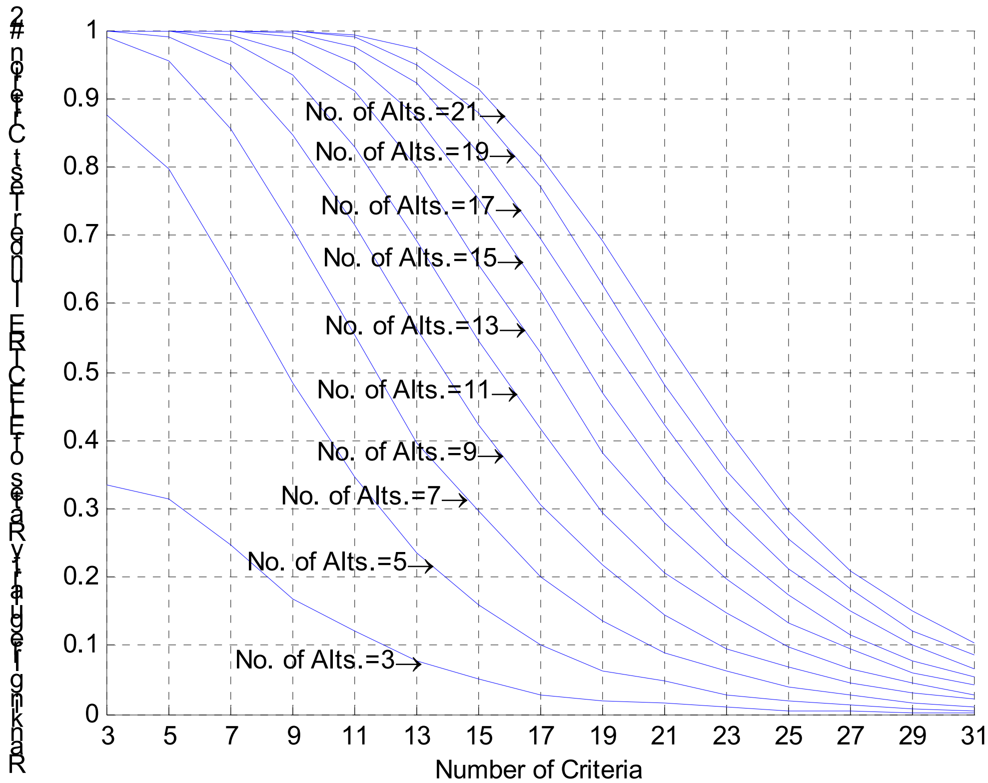


Figure 6. Ranking Irregularity Rates of ELECTRE II under Test Criterion #2.

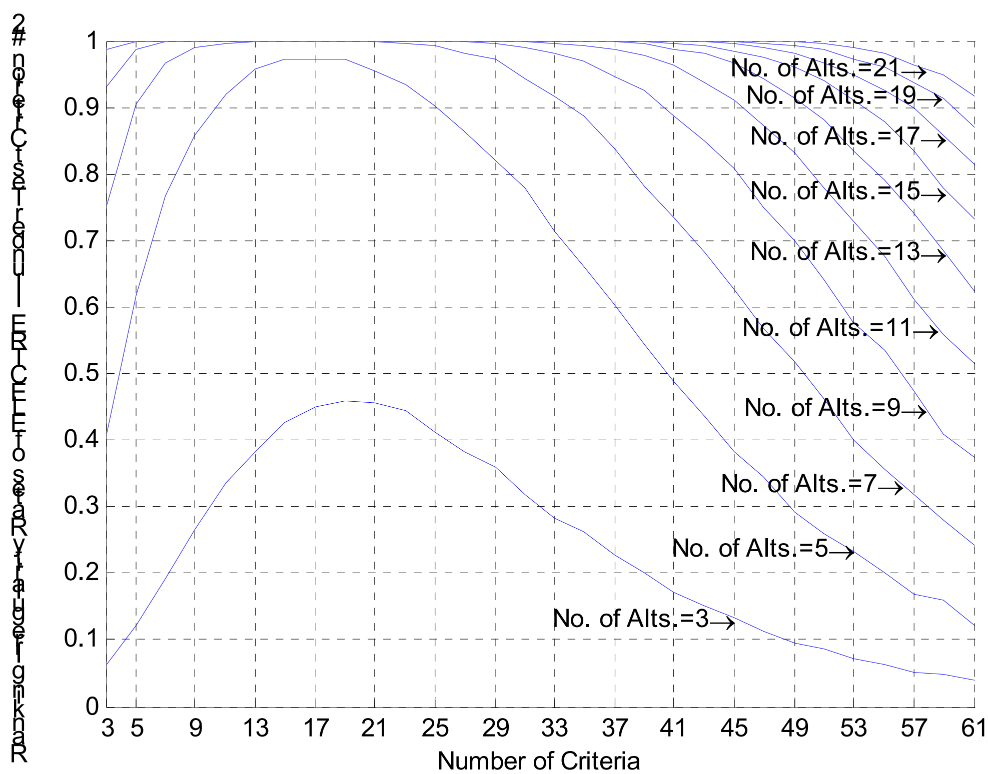


Figure 7. Ranking Irregularity Rates of ELECTRE III under Test Criterion #2.

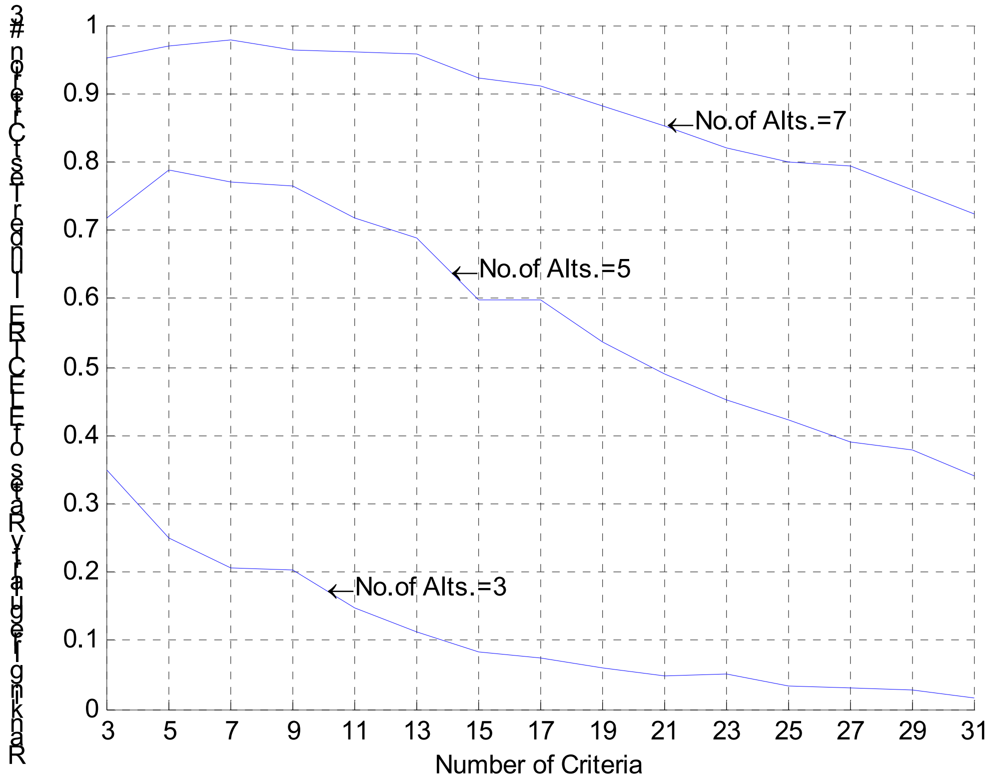


Figure 8. Ranking Irregularity Rates of ELECTRE II under Test Criterion #3.

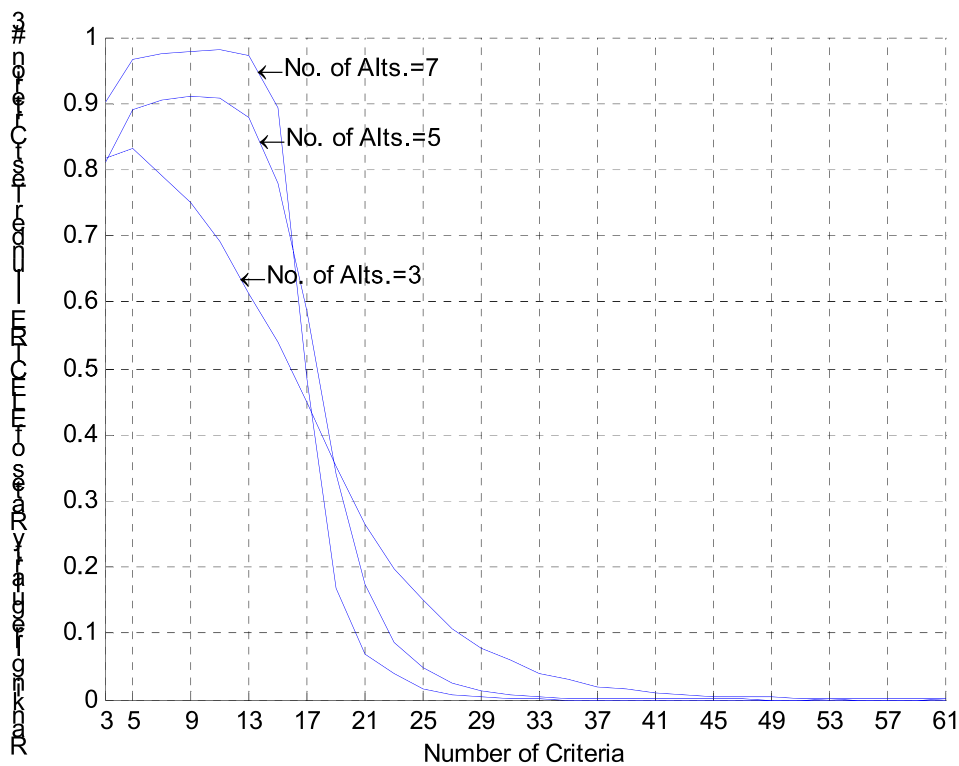


Figure 9. Ranking Irregularity Rate of ELECTRE III under Test Criterion #3.

A different type of experiments was run as well. The goal now was to examine if there are any explicit connections between the results under the three test criteria, especially between test criterion #1 and test criterion #2. The experimental tests were executed as follows. First, a large number (i.e., 10,000) of randomly generated decision problems were examined by using the ELECTRE II or III method in terms of test criterion #1 and the random test problems were divided into two groups. One group had the problems that passed test criterion #1 and the other group had those that did not pass test criterion #1. Next, the problems within each of these two groups were examined in terms of the test criterion #2 and the rates of how often they passed or failed to pass this test criterion were recorded and plotted for each one of the two groups.

Next, a test process similar to the above one was performed. Again, a large number of randomly generated decision problems were examined by using the ELECTRE II or III method but now the process started by first testing for behavior under the test criterion #2. The problems were divided into two groups indicating passing or not passing this test criterion. Then the problems within each one of these two groups were examined in terms of the test criterion #1 and the rates of how often they will pass or fail to pass this test criterion are recorded and plotted as before for each one of these two groups.

Similar tests as the above ones were also performed between test criterion #1 and test criterion #3. From the above experimental test results, no clear tendency was found to indicate that failure in one test criterion would have a tendency to lead to failure in terms of another test criterion. That is, not any explicit connection between the results under the three different test criteria was found from these types of experiments.

6. Case Studies

The previous computational results revealed that the ranking irregularities studied in this paper may occur frequently in simulated decision problems. This raised the question whether the same could be true with real-life decision problems. In order to enhance the understanding of this situation, ten real-life cases were studied. These cases were selected randomly from the published literature. That is, no special screening was performed. The only requirement was to be able to extract the numerical data needed to form a decision matrix and the weights of the criteria. It is better if threshold values could be given in the published case to avoid the inconvenience with the using of the newly defined thresholds. In these experiments, the required thresholds for case 1 to case 8 have been specified in the referenced publications. For the last two cases, the thresholds were specified appropriately according to the score range of each criterion. After getting the data, every case was tested by using the ELECTRE II or III method as in the referenced publication. Then the three types of ranking irregularities were recorded whenever they occurred. Please refer to Table I for the summary of the experimental results. Actually, the two examples presented in section 3 are among these 10 tested cases.

Under test criterion #1, that is, when replacing one of the non-optimal alternatives by a worse one, there are mainly two types of rank reversal situations:

1. The optimal alternatives of the changed decision problem are partially different from that of the original problem. The number of the optimal alternatives of the changed problem is more or less than that of the original problem. For example, in terms of case 7, originally the optimal alternative is $\boldsymbol{A_8}$. Next the optimal alternatives may become $\boldsymbol{A_7}$ and $\boldsymbol{A_8}$ under test criterion #1; for case 2, originally the optimal alternatives are $\boldsymbol{A_6}$ and $\boldsymbol{A_3}$, and then it may be just $\boldsymbol{A_6}$ after one of the non-optimal alternatives was replaced by a worse one.

2. The optimal alternative of the new problem is totally different from that of the original problem. For example, for case 8, originally the optimal alternative is $\boldsymbol{A_9}$, and then it becomes $\boldsymbol{A_7}$ when one of the non-optimal alternatives was replaced by a worse one; for case 6, originally the optimal alternative is $\boldsymbol{A_4}$, it may become $\boldsymbol{A_{10}}$ and $\boldsymbol{A_{18}}$ under the first test criterion #1.

In terms of the same case, the above two situations might both happen or just one of them happened in the tests. The emphasis is that the indication of the best alternative had been changed for those cases if any of the two situations occurred to them. Then one can conclude that rank reversals occurred to those cases and they failed to pass test criterion #1. From Table I, it can be seen that 6 out 10 cases failed to pass test criterion #1. Also, 9 out 10 cases failed to pass test criterion #2. For the only case which could be tested under test criterion #3, it failed to pass it too.

7. Conclusions

Although MCDM plays a critical role in many real-life problems, it is hard to accept an MCDM method as being accurate all the time. The present research results complement previous ones and reveal that even more MCDM methods suffer of ranking irregularities. The ELECTRE methods are widely used today in practice. However, the ranking irregularities should function as a warning in accepting ELECTRE's recommendations without questioning their validity. Previous and current research indicates that such ranking irregularities tend to occur when the alternatives appear to be very close to each other. If, on the other hand, the alternatives are very distinctive from each other, then it is less likely that these ranking irregularities will take place. However, one needs more powerful MCDM method when alternatives are closely related with each other.

In previous studies [Triantaphyllou, 2000 and 2001], it was found that the Multiplicative AHP does not suffer of the previous three types of ranking irregularities. The way in which alternatives are ranked by the Multiplicative AHP utilizes less information than the ELECTRE methods. In Table II, we compared the alternatives' rankings of some of the previous ten real-life cases by using the Multiplicative AHP and the ELECTRE II or III method, respectively. From Table II, one can see that the majority of the situations will result in the same optimal alternative but the rankings of the non-optimal alternatives under both methods have a significant difference. Since the ranking irregularities under the three test criteria will not happen under the Multiplicative AHP, it is tempting to try to see if a new method can be designed, which combines qualities from both the Multiplicative AHP and the ELECTRE methods or some other MCDM methods, and still not to suffer of these ranking irregularities.

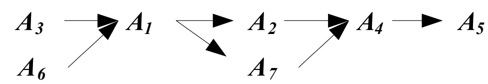
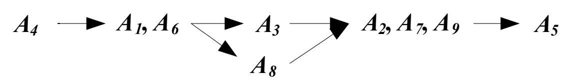
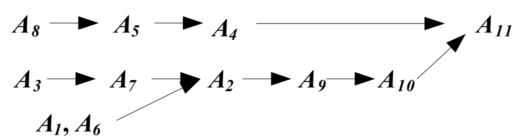
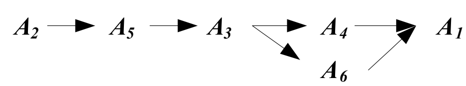
Another direction for future research is to define more test criteria against which existing and future MCDM methods can be evaluated. Some interesting work in this area has been conducted. For example, in [Kujawski, 2005], the author proposed three properties for a desirable MCDM approach. They are about independence of dominated alternatives, no imposed rank reversal and negative side effects associated with inferior substitutions. Clearly, this is a fascinating area of research and it is of paramount significance to both researchers and practitioners in the multi-criteria decision-making field.

Table I. Summary of Case Studies.

Case number	Reference	Domain of application and method used	Size of decision problem		Did it fail T. C. #1?	Did it fail T. C. #2?	Did it fail T. C. #3?
			No. of alternatives	No. of criteria			
1	Hokkanen, J., and P. Salminen, [1997a]	Choosing a solid waste management system (ELECTRE III)	22	8	Yes	Yes	¯
2	Belton, V., and T.J. Stewart, [2001]	Business location problem (ELECTRE III)	7	6	No	Yes	¯
3	Rogers, M., and M. Bruen, [1996]	Environmental appraisal (ELECTRE II)	9	9	No	No	Yes
4	Rogers, M.G., M. Bruen, and L.-Y. Maystre, [1999]	Site selection for a wastewater treatment plant (ELECTRE II)	5	7	Yes	Yes	¯
5	Raj, P.A., [1995]	Water resources planning (ELECTRE II)	27	6	Yes	Yes	¯
6	Buchanan, J., P. Sheppard, and D.V. Lamsade, [1999]	Project ranking (ELECTRE III)	5	5	No	Yes	¯
7	Hokkanen, J., and P. Salminen, [1997b]	Choosing a solid waste management system (ELECTRE III)	11	8	Yes	Yes	¯
8	Rogers, M.G., M. Bruen, and L.-Y. Maystre, [1999]	Choosing a waste incineration strategy (ELECTRE III)	11	11	Yes	Yes	¯
9	Poh, K.L., and B.W. Ang, [1999]	Choosing an alternative fuel system for land transportation (ELECTRE II)	4	6	No	Yes	¯
10	Leyva-López, J.C., and E. Fernández-González, [2003]	Selection of an alternative electricity power plant (ELECTRE III)	6	6	Yes	Yes	¯

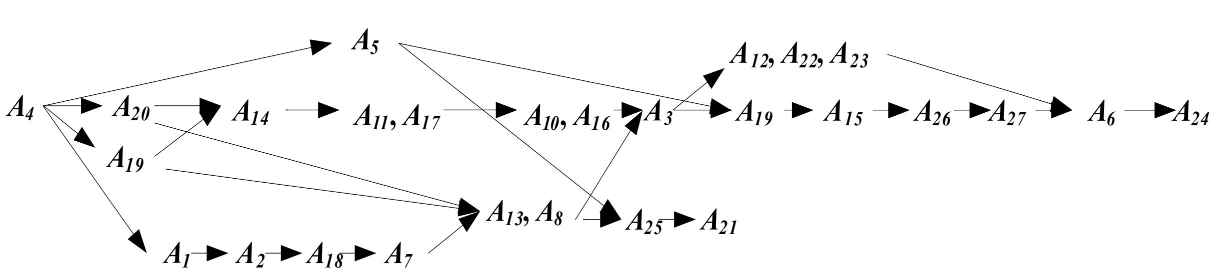
* "T. C." stands for "Test Criterion".
* If a case failed under test criterion #2, it will not be able to get an overall ranking of the alternatives from the smaller problems. That means it will not be able to apply test criterion #3 to that case. The symbol "¯" was used in the corresponding cell to stand for this situation.

Table II. Ranking Comparison

No. Of Cases	Ranking from the Multiplicative AHP method	Ranking from the ELECTRE method
Case 1	$A_9 > A_{12} > A_6 > A_{21} > A_{15} > A_{18} > A_{22} > A_{11} > A_8 > A_5 > A_{14} > A_{17} > A_{20} > A_3 > A_7 > A_{10} > A_4 > A_{13} > A_{16} > A_2 > A_{19} > A_1$	$A_6 = A_9 = A_{12} = A_{15} = A_{21} > A_5 = A_8 = A_{11} = A_{14} = A_{18} = A_{20} > A_4 = A_7 = A_{10} = A_{13} = A_{16} = A_{17} = A_{19} > A_{22} > A_3 > A_1 = A_2$ (ELECTRE III)
Case 2	$A_7 > A_1 > A_3 > A_6 > A_4 > A_2 > A_5$	 (ELECTRE III)
Case 3	$A_4 > A_1 > A_6 > A_3 > A_2 > A_5 > A_7 > A_8 > A_9$	 (ELECTRE II)
Case 4	$A_2 > A_3 > A_5 > A_1 > A_4$	$A_2 > A_5 > A_3 > A_1 > A_4$ (ELECTRE II)
Case 5	$A_2 > A_4 > A_1 > A_{18} > A_{10} > A_7 > A_3 > A_{11} > A_{16} > A_{17} > A_{15} > A_{12} = A_{22} > A_5 > A_{14} > A_{26} > A_{27} > A_8 = A_{13} > A_9 = A_{20} > A_6 > A_{21} > A_{19} > A_{25} > A_{24}$	Becaue of the space constraint, the ranking of case 5 is shown as a footnote to this table. (ELECTRE II)
Case 7	$A_8 > A_5 > A_4 > A_6 > A_{11} > A_1 > A_7 > A_3 > A_2 > A_{10} > A_9$	 (ELECTRE III)
Case 9	$A_1 > A_3 > A_2 > A_4$	$A_1 > A_2 > A_3 = A_4$ (ELECTRE II)
Case 10	$A_2 > A_3 > A_5 > A_6 > A_4 > A_1$	 (ELECTRE III)

Note: Because some of the performance values in the decision matrices of case 6 and case 8 are zeros and the Multiplicative AHP method forbids the use of zero performance values during its ranking process, we did not apply the Multiplicative AHP methods to these two cases.

*The ranking derived for case 5 by using the ELECTRE III method is as follows:



REFERENCES:

Anand Raj, P., (1995), “Multi-Criteria Methods in River Basin Planning: a Case Study,” *Water Science and Technology*, Vol. 31, pp. 261-272.

Barzilai, J., and F.A. Lootsma, (1994), “Power Relations and Group Aggregation in the Multiplicative AHP and SMART,” *Proceedings of the Third International Symposium on the AHP*, George Washington University, Washington, DC, U.S.A., pp.157-168.

Belton, V., and A.E. Gear, (1983), “On a Short-Coming of Saaty’s Method of Analytic Hierarchies,” *Omega*, Vol. 11, pp. 228-230.

Belton, V., and T.J. Stewart, (2001), “Chapter 8: Outranking Methods,” *Multiple Criteria Decision Analysis: An Integrated Approach*, Kluwer Academic Publishers, Boston, MA, U.S.A.

Benayoun, R., B. Roy, and N. Sussman, (1966), “Manual de Reference du Programme Electre,” Note De Synthese et Formaton, No.25, Direction Scientifque SEMA, Paris, France.

Boucher, T.O., and E.L. Mcstravic, (1991), “Multi-attribute Evaluation within a Present Framework and its Relation to the Analytic Hierarchy Process,” *The Engineering Economist*, Vol. 37, pp. 55-71.

Bridgman, P.W., (1922), "Dimensional Analysis," Yale University Press, New Haven, CT, U.S.A.

Buchanan, J., P. Sheppard, and D.V. Lamsade, (1999), Project Ranking Using Electre III, presented at MCDM 2000, Ankara, Turkey.

Cambron, K.E., and G.W. Evans, (1991), “Layout Design using the Analytic Hierarchy Process,” *Computers and Industrial Engineering*, Vol. 20, No. 2, pp. 221-229.

Dyer, J.S., and R.K. Sarin, (1979), “Measurable Multiattribute Value Functions,” *Operations Research*, Vol. 27, pp. 810-822.

Dyer, J.S., and R.E. Wendell, (1985), “A Critique of the Analytic Hierarchy Process”, Technical Report 84/85-4-24, Department of Management, The University of Texas at Austin, Austin, TX, U.S.A.

Edwards, W., (1977), “How to Use Multiattribute Utility Measurement for Social Decision Making,” *IEEE Transactions on Man, Systems, and Cybernetics*, Vol. SMC-7, pp. 326-340.

Edwards, W., and F.H. Barron, (1994), “SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement,” *Organizational Behavior and Human Decision Processes*, Vol. 60, pp. 306-325.

Edwards, W., and J.R. Newman, (1986), “Multiattribute Evaluation,” in H.R. Arkes and K.R. Hammond (Eds.), *Judgment and Decision Making: An Interdisciplinary Reader*, Cambridge University Press, Cambridge, U.K., pp. 13-37.

Fishburn, P.C., (1967), “Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments,” *Operations Research*, 1967.

Hobbs, B.F., (1986), “What can we learn from experiments in multiobjective decision analysis,” *IEEE Trans. On Systems Management and Cybernetics*, 16, pp. 384-394.

Hobbs, B.F., V. Chankong, W. Hamadeh, and E. Stakhiv, (1992), “Does choice of multi-criteria method matter? An experiment in water resource planning,” *Water Resources Research*, Vol. 28, pp. 1767-1779.

Hobbs, B.F., and P. Meier, (2000), *Energy Decisions and the Environment: A Guide to the Use of Multicriteria Methods*, Kluwer Academic Publishers, Boston, MA, U.S.A.

Hokkanen, J., and P. Salminen, (1997a), “Choosing a solid waste management system using multicriteria decision analysis,” *European Journal of Operational Research*, Vol. 98, pp. 19-36.

Hokkanen, J., and P. Salminen, (1997b), “Electre III and IV decision aids in an environmental problem,” *Journal of Multi-criteria Decision Analysis*, Vol. 6, pp. 215-226.

Hwang, C.L., and K. Yoon, (1981), *Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, New York, NY, U.S.A.

Jacquet-Lagreze, E., and Y. Siskos, (2001), “Preference Disaggregation: 20 Years of MCDA Experience,” Invited Review, *European Journal of Operational Research*, Vol. 130, pp. 233-245.

Keyser, W. De., and P. Peeters, (1996), “A note on the use of PROMETHEE multicriteria methods,” *European Journal of Operational Research*, Vol. 89, pp. 457-461.

Kirkwood, C.W., (1997), *Strategic Decision Making*, Duxbury Press, Wadsworth Publishing Company, Belmont, Calif., U.S.A.

Kujawski, E., (2005), “A reference-dependent regret model for deterministic tradeoff studies,” *Systems Engineering*, Vol. 8, pp. 119-137.

Leyva-López, J.C., and E. Fernández-González, (2003), "A new method for group decision support based on ELECTRE III methodology," *European Journal of Operational Research*, Vol. 148, pp. 14-27.

Lootsma, F.A., (1999), *Multi-Criteria Decision Analysis via Ratio and Difference Judgment*, Kluwer Academic Publishers, Applied Optimization Series, Vol. 29, Dordrecht, The Netherlands.

Miller, D.W., and M.K. Starr, (1969), *Executive Decisions and Operations Research*, Prentice-Hall, Inc., Englewood Cliffs, NJ, U.S.A.

Poh, K.L., and B.W. Ang, (1999), Transportation fuels and policy for Singapore: an AHP planning approach, *Computers and Industrial Engineering*, Vol. 37, pp. 507-525.

Putrus, P., (1990), "Accounting for Intangibles in Integrated Manufacturing (non-financial justification based on the analytic hierarchy process)," *Information Strategy*, Vol. 6, pp. 25-30.

Rogers, M., and M. Bruen, (1996), "Using ELECTRE to rank options within a highway environmental appraisal - two case studies," *Civil Engineering Systems*, Vol. 13, pp. 203-221.

Rogers, M.G., M. Bruen, and L.-Y. Maystre, (1999), "Chapter 3: The Electre Methodology," *Electre and Decision Support*, Kluwer Academic Publishers, Boston, MA, U.S.A.

Rogers, M.G., M. Bruen, and L.-Y. Maystre, (1999), "Chapter 5: Case Study 1: Finding the Best Location for the Galway Wastewater Treatment Plant," *Electre and Decision Support*, Kluwer Academic Publishers, Boston, MA, U.S.A.

Rogers, M.G., M. Bruen, and L.-Y. Maystre, (1999), "Chapter 6: Case Study 2: Choosing the Best Waste Incineration Strategy for the Eastern Switzerland Region," *Electre and Decision Support*, Kluwer Academic Publishers, Boston, MA, U.S.A.

Roy, B., (1968), "Classement et choix en presence de points de vue multiples: La methode ELECTRE," R.I.R.O 8, pp. 57-75.

Roy, B., (1978), "ELECTRE III: Un algorithme de classements fonde sur une representation floue des preference en presence de criteres multiples," Cahiers de CERO, Vol. 20, No.1, pp. 3-24.

Roy, B., (1985), Methodologie Multicritiere d'Aide a la Decision. Econometrica, Paris, France.

Roy, B., and P. Bertier, (1971), "La methode ELECTRE II: Une methode de classement en presence de critteres multiples," SEMA (Metra International), Direction Scientifique, Note de Travail No. 142, Paris, 25p.

Roy, B., and P. Bertier, (1973), "La methode ELECTRE II: Une methode au media-planning," In Ross, M. (ed.), Operational Research 1972, North-Holland Publishing Company, pp. 291-302.

Saaty, T.L., (1994), *Fundamentals of Decision Making and Priority Theory with the AHP*, RWS Publications, Pittsburgh, PA, U.S.A.

Saaty, T.L., (1980), "The Analytic Hierarchy Process," McGraw-Hill, New York, NY, U.S.A.

Stewart, T.J., (1992), "A critical survey of the status of multiple criteria decision making theory and practice," *OMEGA*, 20, pp. 569-586.

Triantaphyllou, E., (2000), *Multi-Criteria Decision Making Methods: A Comparative Study*, Kluwer Academic Publishers, Boston, MA, U.S.A.

Triantaphyllou, E., (2001), "Two New Cases of Rank Reversals when the AHP and Some of its Additive Variants are Used that do not Occur with the Multiplicative AHP," *Multi-Criteria Decision Analysis*, (May 2001 issue) Vol. 10, pp. 11-25.

Triantaphyllou, E., and S.H. Mann, (1989), "An Examination of the Effectiveness of Multi-Dimensional Decision-Making Methods: A Decision-Making Paradox," *International Journal of Decision Support Systems*, No. 5, pp. 303-312.

Wabalickis, R.N., (1988), "Justification of FMS with the Analytic Hierarchy Process," *Journal of Manufacturing Systems*, Vol. 17, pp. 175-182.

Zanakis, S., T. Mandakovic, S.K. Gupta, S. Sahay, and S. Hong, (1995), "A Review of Program Evaluation and Fund Allocation Methods within the Service and Government Sectors," *Socio-Economic Planning Sciences*, 29 (1), pp. 59- 79.

Zanakis, S., A. Solomon, N. Wishart, and S. Dublish, (1998), "Multi-Attribute Decision Making: A Comparison of Select Methods," *European Journal of Operational Research*, Vol. 107, pp. 507-529.