

This is a PDF file of an unedited manuscript that has been accepted for publication in *Applied Energy*. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## **On multi-criteria sustainability assessment: spider-gram surface and dependence biases**

Luis C. Dias<sup>ab\*</sup>, Ana Rita Domingues<sup>a</sup>

<sup>a</sup>INESC Coimbra, Rua Antero de Quental, 3000-033 Coimbra, Portugal

<sup>b</sup>University of Coimbra, FEUC, Av Dias da Silva 165, 3004-512 Coimbra, Portugal

\*Corresponding author. Tel.: +351 239 790500; fax: +351 239 403511

E-mail address: [lmcdias@fe.uc.pt](mailto:lmcdias@fe.uc.pt) (L. C. Dias)

### **Abstract.**

Multi-criteria evaluation methods are often used for sustainability assessment. One such method is the Multi-criteria Spider-gram Cumulative Surface Area (MCSA score) recently used for this purpose (Nzila *et al.*, Multi-criteria sustainability assessment of biogas production in Kenya. *Applied Energy* 93:496–506, 2012). This paper presents results illustrating that the MCSA method results and rankings might be biased by the arbitrary order of the criteria. This paper also addresses the way the comparison of two alternatives can be biased by the presence of a third (possibly irrelevant) alternative. Such dependence bias is due to the use of an internal normalization operation, a problem shared by some multi-criteria analysis methods. The paper concludes with a few suggestions to avoid such biases.

**Keywords:** Multi-criteria, spider-gram cumulative surface area, sustainability assessment, life-cycle approaches

### **Highlights:**

- The criteria order can bias the Spider-gram Cumulative Surface Area results.
- Changing one alternative can influence the relative position of the remaining alternatives.
- This latter bias is due to the use of an internal normalization operation.

## 1 Introduction

Nzila *et al.* ([1]) compared three different biogas digester plants, analysing thoroughly the environmental, technical and economic aspects of this highly relevant problem. As the authors note, biogas production is particularly relevant in developing countries because it reduces logging for charcoal or firewood, which are commonly used fuels, avoiding losses of carbon sequestered in forests. Due to the decreased use of charcoal and firewood it promotes air quality and consequently diminishes human health problems. In addition, biogas production enables reducing agriculture costs due to the production of a biofertiliser during the process and it reduces losses in the supply chain, since it is a decentralized power production.

In their work, Nzila *et al.* ([1]) followed a Life Cycle Sustainability Assessment approach and evaluated the three alternative designs (floating-drum, fixed dome, and tubular type) on nine criteria: three environmental criteria (Resource depletion, Global warming reduction, Energy demand), three technical criteria (Energy breeding ratio, Energy payback, Reliability), and three economic criteria (Total investment, Energy autonomy, Labour cost).

To select the best design the authors used the Multi-criteria spider-gram Cumulative Surface Area (MCSA score) as a multi-criteria aggregation method that considers all criteria have the same weight. This allowed the authors to conclude the best option would be the tubular digester, followed by the fixed dome digester, and finally the floating-drum digester. However the method the authors used might bias the results based on the relative order of the criteria. Although this bias does not affect the main conclusion of Nzilla *et al.*'s work concerning which would be the top alternative, with different data the conclusions might be arbitrary.

In this paper we show the MCSA should not be used as a multi-criteria aggregation method, since it produces a ranking of the alternatives that may be contingent on the order the criteria are considered. This is exemplified on a pedagogical example and also on variants of the same case addressed by [1].

This paper also addresses the way the comparison of two alternatives can be biased by the presence of a third (possibly irrelevant) alternative. This dependence bias means that changes in one alternative may influence the relative position of the remaining alternatives. This is a problem not only with the MCSA, but also with other much more popular multi-criteria aggregation methods such as the weighted sum (the interested reader will find in [2] other examples of the difficulties involved in the use of evaluation models).

## 2 Methods

The field of Multi-Criteria Decision Analysis (MCDA) [3] [4] (also known as MCDM, where the letter M stands for Making) [5] has developed rapidly in the last 40 years and offers a large

array of methods to aggregate evaluations made on multiple criteria. Such methods have been used in many application areas, including energy (e.g., see [6] [7]).

Many of these methods require data to be normalized in order to obtain comparable scales. Examples of such methods used in practice are the MCSA [1], the weighted sum [8][9][10], and TOPSIS [11][12] (these citations refer to examples of their use in life-cycle assessment). A common method is the ratio normalization that attributes value 1 to the best performance on a criterion and a proportional value to the other performances. If  $x_{ij}$  denotes the performance (on the original scale) of an alternative indexed by  $i$  on a criterion indexed by  $j$ , then the normalized performance value is:

$$y_{ij} = \begin{cases} x_{ij}/\max_i x_{ij}, & \text{if criterion } j \text{ is to maximize} \\ \min_i x_{ij}/x_{ij}, & \text{if criterion } j \text{ is to minimize} \end{cases} \quad (1)$$

Thus, for a maximization (benefit) criterion a normalized performance of 0.5 means the alternative yields half of what the best alternative according to the same criterion yields; for a minimization (cost) criterion a normalized performance of 0.5 means the alternative incurs the double of what the best alternative according to the same criterion does.

The MCSA method addressed in this paper aggregates the performance values of multiple criteria as follows (let  $n$  denote the number of criteria). For an alternative indexed by  $i$  this method computes the area of the spider-gram formed by the values  $y_{i1}, \dots, y_{in}$ , which corresponds to summing the area of  $n$  triangles. The MCSA score is:

$$SA_i = \sum_{j=2}^n \frac{1}{2} y_{ij} y_{i(j-1)} \sin \frac{\pi}{n} + \frac{1}{2} y_{i1} y_{i(n-1)} \sin \frac{\pi}{n} \quad (2)$$

According to eq. (1), the normalized values depend on the set of alternatives being evaluated. If an alternative's performance data is changed, or an alternative is removed from the set, or an alternative is added to the set, then this may change the normalized values of all other alternatives. Thus, the MCSA method is potentially subject to a dependence bias, leading to potential rank reversals. Let us note that the optional normalization step in Life-Cycle Assessment does not introduce this bias since it is based on external references and not on the contingent alternatives, although it is subject to other types of problems [13]. The MCSA method, however, is subject to an even more debatable type of bias: the score can change depending on the way the criteria are ordered, meaning that relabeling the criteria may lead to rank reversals. These biases are illustrated in the following section.

### 3 Results and discussion

#### 3.1 Criteria order bias in the spider-gram cumulative surface area

The spider-gram cumulative surface area can be biased by the criteria order: the position of the criteria influences the area. Table 1 presents a pedagogical example to exemplify this bias. If criteria are relabelled (Table 1, on the right) then the area of alternative A1 more than doubles

while the area of A2 is drastically reduced, which leads to reverse the ranking of these alternatives. Fig. 1 allows visualizing the reason for this bias: if by coincidence good performances alternate with poor performances when going from  $y_1$  to  $y_n$ , then this penalizes the alternative significantly as the MCSA will be low.

Let us consider the study of Nzila *et al.* [1], who compared three alternatives on nine criteria ranked as in Table 2. The order of the criteria is arbitrary, except for their classification in three groups: Environmental, Technical, and Economic criteria. Indeed, there are  $3!=6$  ways of positioning the three groups and  $3!=6$  ways of positioning the three criteria within each group, which means there are  $(3!)^4 = 1296$  different ways of assigning the nine criteria to labels 1,2,...,9 while respecting the 3-group classification. Otherwise there are  $9!=362880$  different ways of assigning the nine criteria to labels 1,2,...,9 freely. Some of these ways of defining a criteria order lead to the same area because a rotation of the criteria labels (e.g., 234567891 instead of 123456789) does not change the polygon's area. The same holds in case of exchanging a clockwise for a counter clockwise direction (e.g. 198765432 instead of 123456789). However there are still many combinations potentially yielding different results. A notable exception is the case with only three criteria, since the permutations 231 and 312 are rotations of 123, whereas 213, 321, and 132 correspond to reversing the direction of 231, 312, and 123, respectively.

Table 3 presents the range of MCSA results for each alternative for the different ways of positioning the 9 criteria while respecting the 3-group classification. For these data there is no rank reversal. The tubular digester alternative, which is the best alternative on all criteria except two, is always the one with highest area. However, with different data the bias might strongly influence the results. The last row of Table 3 presents the average results considering all the possible criteria orders. This might be a defensible result in face of the arbitrariness of choosing the “right” order for the criteria. However, it may be cumbersome to compute these averages (especially when  $n$  is large) and it might be difficult to convey a meaning for it when dialoguing with decision-makers or communicating with the general public.

Let us suppose that the tubular digester alternative did not exist (or it was excluded due to its reliability score) and a new technology was available, with the data presented on Table 4 (the reader will note that due to the normalization equation (1) the normalized values for the other two alternatives also change). The MCSA scores corresponding to this data are provided on the “Before change” row of Table 5. The fixed dome digester has the best area, whereas the floating-drum digester has the worst area.

Now, suppose that the order of the criteria was not the one presented on Table 4, but a small variant of it respecting the three groups: Global warming reduction, Resource depletion, Energy demand, Energy payback, Energy breeding ratio, Reliability, Total investment, Energy autonomy, and Labour costs (Order 2). In other words, Resource depletion and Global warming reduction exchanged their positions, and Energy breeding ratio exchanged position with Energy payback (all other positions unchanged). Then, the results would be those presented on the “After change” row of Table 5. It is noteworthy that the best alternative becomes the worst one and vice-versa. A comparison of the two situations is provided in Fig. 2.

### 3.2 The bias of dependence of other alternatives

Let us now examine the possibility that the relative ranking of two alternatives (to conclude which one of the two is better) may be biased by the presence of other alternatives. Table 6 presents a situation that considers the same alternatives as of Nzila *et al.* [1] plus the new technology already present in Table 4. The order of the criteria corresponds to the “After change” situation in Table 5.

Table 7 depicts the MCSA results for the situation with 4 alternatives (corresponding to Table 6), for a situation with 3 alternatives (corresponding to Table 4 but keeping the criteria order 2 presented in Table 6), and for a situation with only two alternatives. With four alternatives the floating-drum digester would be considered the worst choice. However, if the tubular digester was not considered, then it would become the best choice (instead of the new technology, as one might expect if the bias did not occur).

In a similar way, if we started with just two technologies (floating-drum digester and fixed dome digester as in the last row of Table 7) then the fixed dome digester would have the largest MCSA score. Adding a third alternative (new technology) would logically lead to either keeping the same winner or declaring the new technology is even better. However, in the situation with three alternatives the previously worst one (floating-drum digester) would become the new winner.

The bias of dependence of other alternatives can appear whenever the other alternatives cause a change in the anchors (maximum or minimum performances) used for the normalization according to eq. (1). This means that the bias can occur not only when an alternative is inserted (or removed), but also when the data for a third alternative changes in a way that changes the normalization anchors.

In parenthesis, let us note that the bias of dependence of other alternatives can appear in many methods that perform normalizations anchored on the performances of the actual set of alternatives being evaluated. The well-known weighted sum method is one such case. Consider for instance equal weights  $w_1, \dots, w_n = 1/9$ , i.e., the weighted sum is equal to  $(y_{i1} + \dots + y_{i9})/9$ . Table 8 depicts weighted sum results for the situation with four alternatives (Table 6) and for the situation with three alternatives (Table 4). With four alternatives the tubular digester would be considered the best choice, followed by the new technology. However, if the tubular digester was not considered, then the best choice would become the fixed dome digester (instead of the new technology, as one might expect if the bias did not occur).

## 4 Conclusions and perspectives

Sustainability assessment calls for evaluation of multiple criteria addressing environmental, economic, technical and social issues. The MCDA field offers many methods to aggregate multiple criteria, most of which assign a global value to each alternative and allow ranking the alternatives. This paper presents results illustrating that the MCSA method results and rankings might be biased by the arbitrary order of the criteria, even when the criteria are all deemed

equally important. In addition, this paper presents results that illustrate the MCSA method is subject to a bias of dependence of other alternatives. Therefore, this method should be used with a lot of caution, if used at all.

The criteria position bias in the MCSA method has no simple solution. It is not realistic to expect that there is only one possible order for the criteria, according to some undisputed argument. One possibility is to compute the average scores for all possible orders, but these may be cumbersome to compute when the number of criteria is high, and it might be difficult to convey a meaning for such average scores when communicating results.

The bias of dependence of other alternatives occurs for the MCSA method as well as other methods that use an internal normalization anchored on the performance values of the alternatives being evaluated. Using normalization anchors based on an external reference system (as is often the case in Life-Cycle Assessment [13]) or based on external reference performances might be a way to avoid this bias. If weights are used, then the weighting coefficients should be discussed *after* the normalization transformation, in order to be aware of the range of the resulting scales and the implied trade-offs.

There are many possibilities to avoid the bias of dependence of other alternatives. A possibility close in spirit to the Weighted Sum method is to use an additive aggregation of value functions, as proposed by, e.g., Seppälä *et al.* [14]. This method requires to convert the criteria original units into value units (or utility units) taking into account the judgement of a decision maker, which can be done without being anchored on the maximum and minimum performance values. If the decision makers wish to partition the set of alternatives into classes (e.g., “very good”, “good”, etc.) then a classification method such as ELECTRE TRI (see [3], Chapter 7.5, for a general description, and [15] for an application to the area of energy using an approach that does not require setting precise weights [16]) can be used. This method can use references to define the boundaries of the classes that are external to the set of alternatives being evaluated, hence avoiding the dependence bias.

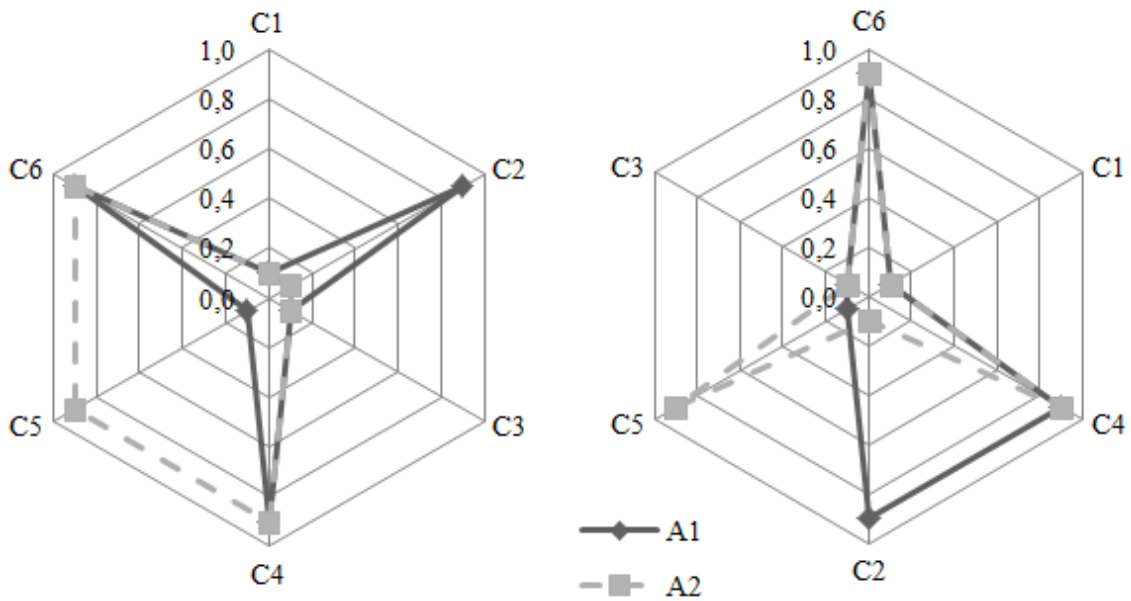
## Acknowledgements

The authors gratefully acknowledge support by the FCT/FEDER Programs (COMPETE MIT/MCA/0066/2009, MIT/SET/0014/2009, PEst-C/EEI/UI0308/2011, CENTRO-07-0224-FEDER-002004).

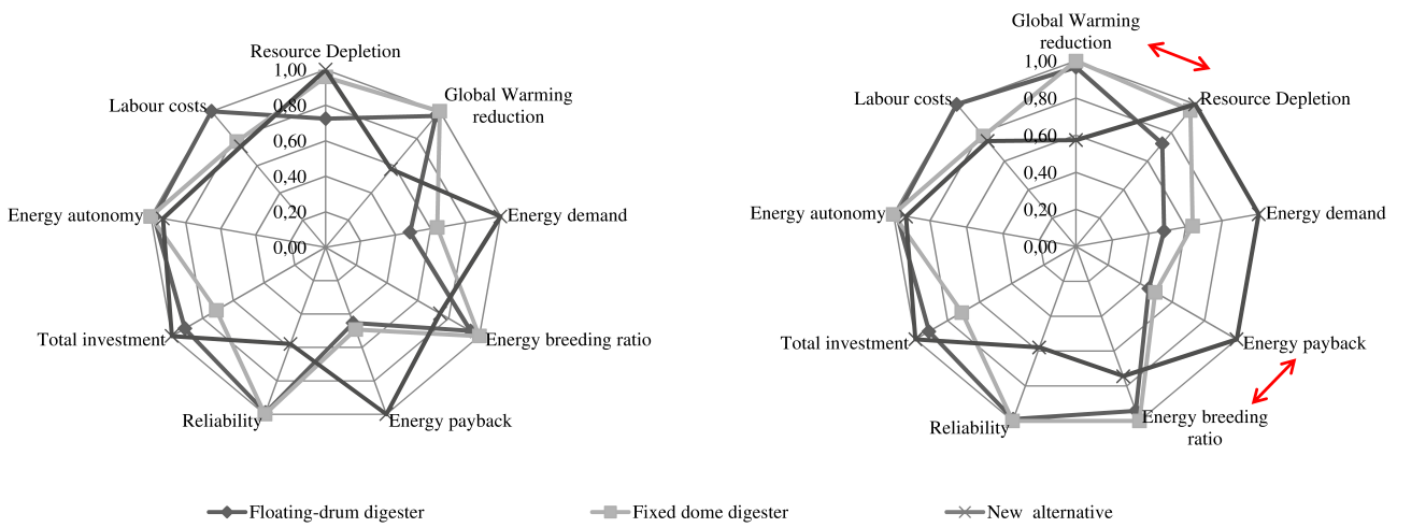
## References

- [1] Nzila C, Dewulf J, Spanjers H, Tuigong D, Kiriamiti H, Van Langenhove H. Multi criteria sustainability assessment of biogas production in Kenya. *Applied Energy* 2012;93:496–506.
- [2] Bouyssou D, Marchant T, Pirlot M, Perny P, Tsoukiàs A. *Evaluation and Decision models: A critical perspective*. Kluwer; 2000.
- [3] Ishizaka A, Nemery P. *Multi-criteria Decision Analysis: Methods and Software*. Chichester: Wiley; 2013.

- [4] Belton V, Stewart T. *Multiple Criteria Decision Analysis. An Integrated Approach*. Boston: Kluwer; 2002.
- [5] Yoon K, Hwang C-L. *Multiple attribute decision making: an introduction*. Sage Publications; 1995.
- [6] Pohekar SD, Ramachandran M. Application of multi-criteria decision making to sustainable energy planning—A review. *Renewable and Sustainable Energy Reviews* 2004;8:365–81.
- [7] Zhou P, Ang BW, Poh KL. Decision analysis in energy and environmental modeling: An update. *Energy* 2006;31:2604–22.
- [8] Bouwman ME, Moll HC. Environmental analyses of land transportation systems in The Netherlands. *Transportation Research Part D: Transport and Environment* 2002;7:331–45.
- [9] Hirschberg S, Dones R, Heck T, Burgherr P, Schenler W, Bauer C. *Sustainability of Electricity Supply Technologies under German Conditions: A Comparative Evaluation*. 2004.
- [10] Zhou Z, Joang H, Qin L. Life cycle sustainability assessment of fuels. *Fuel* 2007;86:256–63.
- [11] Kuo T, Chang S, Huang SH. Environmentally conscious design by using fuzzy multi-attribute decision-making. *International Journal of Advanced Manufacturing Technology* 2006:419–25.
- [12] Huang H, Zhang L, Liu Z, Sutherland JW. Multi-criteria decision making and uncertainty analysis for materials selection in environmentally conscious design. *The International Journal of Advanced Manufacturing Technology* 2010;52:421–32.
- [13] Heijungs R, Guinée J, Kleijn R, Rovers V. Bias in normalization: Causes, consequences, detection and remedies. *The International Journal of Life Cycle Assessment* 2007;12:211–6.
- [14] Seppälä J, Basson L, Norris GA. Decision Analysis frameworks for Life-Cycle Impact Assessment. *Journal of Industrial Ecology* 2002;5:45–68.
- [15] Neves LP, Martins AG, Antunes CH, Dias LC. A multi-criteria decision approach to sorting actions for promoting energy efficiency. *Energy Policy* 2008;36:2351–63.
- [16] Dias L, Mousseau V, Figueira J, Clímaco J. An aggregation/disaggregation approach to obtain robust conclusions with ELECTRE TRI. *European Journal of Operational Research* 2002;138:332–48.



**Fig. 1.** Pedagogical example with six criteria: the MCSA is low when good performances alternate with poor performances. Changing the order of the criteria can change the MCSA significantly.



**Fig. 2.** A variation of the problem addressed in [1]. Changing the order of the criteria can change the MCSA significantly.



**Table 1.**

Pedagogical example with six criteria: changing the order of the criteria leads to different spider-gram cumulative surface areas and even to a preference reversal.

Criteria	A1	A2	Criteria	A1	A2
C1	0.1	0.1	C1'=C6	0.9	0.9
C2	0.9	0.1	C2'=C1	0.1	0.1
C3	0.1	0.1	C3'=C4	0.9	0.9
C4	0.9	0.9	C4'=C2	0.9	0.1
C5	0.1	0.9	C5'=C5	0.1	0.9
C6	0.9	0.9	C6'=C3	0.1	0.1
Area ( $SA_i$ )	0.23	1.10	Area ( $SA_i$ )	0.51	0.20
Ranking	2 <sup>nd</sup>	1 <sup>st</sup>	Ranking	1 <sup>st</sup>	2 <sup>nd</sup>

**Table 2.**

Data from the study of [1] keeping the order of appearance of the criteria. Values marked with an asterisk (\*) are approximations derived from the charts in [1]. Normalized values are shown in parentheses.

	Criteria	Floating - drum digester	Fixed dome digester	Tubular digester
Environmental	Resource Depletion	1.92 (0.76)	1.45* (1.00)	1.88* (0.77)
	Global Warming reduction	0.96 (0.88)	0.99* (0.91)	1.09 (1.00)
	Energy demand	1.89 (0.61)	1.43* (0.81)	1.16* (1.00)
Technical	Energy breeding ratio	9.50 (0.66)	10.08* (0.70)	14.40 (1.00)
	Energy payback	25.40 (0.65)	23.38* (0.71)	16.60 (1.00)
	Reliability	0.94* (0.99)	0.95 (1.00)	0.40 (0.42)
Economic	Total investment	2.95* (0.61)	3.80 (0.47)	1.80 (1.00)
	Energy autonomy	42.18 (0.94)	43.00* (0.96)	44.79 (1.00)
	Labour costs	7.27 (0.78)	9.35 (0.61)	5.70 (1.00)

**Table 3.**

Range of MCSA results for each alternative for the different ways of positioning the 9 criteria while respecting the 3-group classification considering the data from [1].

MCSA ( $SA_i$ )	Floating - drum digester	Fixed dome digester	Tubular digester
Min	1.65	1.77	2.37
Max	1.71	1.88	2.42
Average	1.68	1.83	2.38

**Table 4.**

A variation of the problem addressed in [1]. Normalized values are shown in parentheses.

	Criteria	Floating - drum digester	Fixed dome digester	New technology
Environmental	Resource Depletion	1.92 (0.72)	1.45 (0.96)	1.39 (1.00)
	Global Warming reduction	0.96 (0.97)	0.99 (1.00)	0.57 (0.57)
	Energy demand	1.89 (0.48)	1.43 (0.64)	0.91 (1.00)
Technical	Energy breeding ratio	9.50 (0.94)	10.08 (1.00)	7.50 (0.74)
	Energy payback	25.4 (0.45)	23.38 (0.49)	11.52 (1.00)
	Reliability	0.94 (0.99)	0.95 (1.00)	0.55 (0.58)
Economic	Total investment	2.95 (0.91)	3.80 (0.71)	2.70 (1.00)
	Energy autonomy	42.18 (0.98)	43.00 (1.00)	40.00 (0.93)
	Labour costs	7.27 (1.00)	9.35 (0.78)	9.80 (0.74)

**Table 5.**

A variation of the problem addressed in [1]. The MCSA results change contingent on the order of the criteria.

		Floating - drum digester	Fixed dome digester	New technology
Before change	Area	1.930	1.981	1.978
(Order 1)	Ranking	3 <sup>rd</sup>	1 <sup>st</sup>	2 <sup>nd</sup>
After change	Area	2.050	2.042	2.048
(Order 2)	Ranking	1 <sup>st</sup>	3 <sup>rd</sup>	2 <sup>nd</sup>

**Table 6.**

A variation of the problem addressed in [1] with a new technology and a different criteria order.

	Criteria	Floating - drum digester	Fixed dome digester	Tubular digester	New technology
Environ.	Global Warming reduction	0.96 (0.88)	0.99 (0.91)	1.09 (1.00)	0.57 (0.52)
	Resource Depletion	1.92 (0.72)	1.45 (0.96)	1.88 (0.74)	1.39 (1.00)
	Energy demand	1.89 (0.48)	1.43 (0.64)	1.16 (0.79)	0.91 (1.00)
Tech.	Energy payback	25.40 (0.45)	23.38 (0.49)	16.60 (0.69)	11.52 (1.00)
	Energy breeding ratio	9.50 (0.66)	10.08 (0.70)	14.40 (1.00)	7.50 (0.52)
	Reliability	0.94 (0.99)	0.95 (1.00)	0.40 (0.42)	0.55 (0.58)
Econ.	Total investment	2.95 (0.61)	3.80 (0.47)	1.80 (1.00)	2.70 (0.67)
	Energy autonomy	42.18 (0.94)	43.00 (0.96)	44.79 (1.00)	40.00 (0.89)
	Labour costs	7.27 (0.78)	9.35 (0.61)	5.70 (1.00)	9.80 (0.58)

**Table 7.**

A variation of the problem addressed in [1]. When one alternative is removed (or inserted) the MCSA scores and the relative ranking of the remaining alternatives are subject to change.

		Floating - drum digester	Fixed dome digester	Tubular digester	New technology
4 alternatives	Area	1.531	1.579	2.057	1.655
(Order 2)	Ranking	4 <sup>th</sup>	3 <sup>rd</sup>	1 <sup>st</sup>	2 <sup>nd</sup>
3 alternatives	Area	2.050	2.042		2.048
(Order 2)	Ranking	1 <sup>st</sup>	3 <sup>rd</sup>		2 <sup>nd</sup>
2 alternatives	Area	2.479	2.606		
(Order 2)	Ranking	2 <sup>nd</sup>	1 <sup>st</sup>		

**Table 8.**

A variation of the problem addressed in [1]. When one alternative is removed (or inserted) the Weighted Sum scores and the relative ranking of the remaining alternatives are subject to change.

		Floating - drum digester	Fixed dome digester	Tubular digester	New technology
4 alternatives	W.Sum	0.725	0.749	0.849	0.751
	Ranking	4 <sup>th</sup>	3 <sup>rd</sup>	1 <sup>st</sup>	2 <sup>nd</sup>
3 alternatives	W.Sum	0.828	0.842		0.841
	Ranking	3 <sup>rd</sup>	1 <sup>st</sup>		2 <sup>nd</sup>