Simulation techniques for the sensitivity analysis of multi-criteria decision models

John Butler, Jianmin Jia*, James Dyer

MSIS, CBA 5.202
The Graduate School of Business
University of Texas at Austin
Austin, Texas 78712

*Department of Marketing
The Chinese University of Hong Kong
Shatin, N.T. , Hong Kong

April, 1996

revised: September, 1996

Accepted for Publication in *European Journal of Operational Research*

Simulation techniques for the sensitivity analysis of multi-criteria decision models

Abstract

This paper presents a simulation approach for high dimensional sensitivity analysis of the weights of multi-criteria decision models. This approach allows simultaneous changes of the weights and generates results that can easily be analyzed statistically to provide insights into multi-criteria model recommendations. In this study we consider three cases: no information, order information, and partial information regarding the weights. Our approach also allows investigation of sensitivity to the form of multi-criteria decision models. The simulation procedures we propose can also be used to aide in the actual decision process, particularly when the task is to select a subset of superior alternatives.

Keywords: Multi-criteria analysis; Utility theory; Weights in decision models

1. Introduction

In this paper we will present a new method for testing the sensitivity to the weights assigned to the measures of multi-criteria decision models. Our methodology utilizes Monte Carlo simulation and provides the flexibility to vary all of the weights of a multi-criteria model simultaneously and, in addition, investigate the impact of varying the functional form of the multiattribute aggregation. Three general classes of simulation models will be presented: random weights, rank order weights and response distribution weights. In some instances these simulation approaches can render a formal weight assessment unnecessary, particularly when attempting to select a subset of superior alternatives. The insights that can be obtained from these approaches will be demonstrated using an example problem from the literature.

2. Background

Multi-criteria models have been used in a variety of settings to solve real problems, such as siting an electricity generation facility (Keeney, 1980), choosing among vendors for the commercial generation of electricity by nuclear fusion (Dyer and Lorber, 1982), and selecting a nuclear waste clean up strategy (Keeney and von Winterfeldt, 1994). These and other applications have utilized multiattribute utility (MAU) theory because it provides a logical and tractable means to make trade-offs among conflicting objectives (Keeney and Raiffa, 1976).

One of the primary tasks in the application of MAU is the assignment of weights to the objectives or measures so that component scores can be aggregated. The weights of a multi-criteria decision model are measures of the importance of the *increase* from the worst to the best level of performance on one objective compared to the *increase* from the worst to the best level of performance on another objective. Therefore, weights must be assessed carefully to ensure that the results of the evaluation are consistent with the preferences of the decision maker or decision makers.

After the weights have been assessed, the component scores of the multi-criteria model can be aggregated. The two predominant aggregation methods for multiattribute utility models are the additive and multiplicative forms. The general form of the additive model is:

$$u(X) = \sum_{i=1}^{n} k_i u_i(x_i) \tag{1}$$

where x_i is the performance on attribute i; $u_i(\cdot)$ is a single attribute von-Neuman Morgenstern utility function over attribute i; and $0 \leq k_i \leq 1$ are the scaling constants (weights) for the n attributes such that $\sum_{i=1}^{n} k_i = 1$.

The additive model is appropriate only when the decision maker's preferences satisfy additive independence (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986). If additive independence is not satisfied, a multiplicative form can be used for aggregation if utility independence, a weaker preference condition, is satisfied. The multiplicative model can be represented as:

$$1+ku(X)=\prod_{i=1}^{n}\left[1+kk_i u_i\left(x_i\right)\right] \tag{2}$$

where x_i and $u_i(\cdot)$ are defined as in the additive case; $0\leq k_i \leq 1$ are the scaling constants (weights) for the *n* attributes where $\sum_{i=1}^{n} k_i \neq 1$; and *k*, hereafter referred to as the "common *k*", is an additional scaling constant defined such that $1+k=\prod_{i=1}^{n}\left[1+kk_i\right]$ (see Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986). The additive model is a special case of the multiplicative form when $\sum_{i=1}^{n} k_i = 1$.

A great deal of behavioral research has focused on the correct procedure to assess these weights (for a review see Weber and Borcherding, 1993). Unfortunately, experimental studies have revealed numerous sources of inconsistencies rather than a single, superior assessment technique (see for example Schoemaker and Waid, 1982; Wainer, 1976; Borcherding, Eppel, and von Winterfeldt, 1991). Normally the form of the utility function is determined as part of the weight assessment procedure. For example, an additional trade-off question can identify the presence and strength of an interaction among the attributes, and determine the value of the common *k* for the multiplicative model (2). Thus, the assessed form of the utility function may be affected by errors in the determination of the weights.

Once a set of weights has been assessed, conducting a sensitivity analysis of the weights is often insightful, but current techniques typically vary a single weight and observe the effect on the results of the model. A method of simultaneously varying all, or at least a large subset, of the weights would be useful. This paper will explore three simulation techniques that allow this type of high dimensional sensitivity analysis. The first of the techniques, random weighting, requires no weight assessments and yet may aide the decision maker both before and after assessing the weights. The second, random rank-order weights, requires an importance ranking which may be easier to elicit from a decision maker than numerical weights[1]. The third approach requires weight assessments, but recognizes that these assessed weights may be subject to response error.

These three techniques offer several advantages over conventional sensitivity analysis. Each may be useful when there is a single decision maker or a committee of stakeholders making the necessary judgments. While we develop the simulation approaches in a multiattribute utility setting, they are applicable to any multi-criteria method that employ weights in an aggregation scheme; e.g. the Analytic Hierarchy Process (Saaty, 1980) or with value functions in choice under certainty (Dyer and Sarin, 1979). In addition, they can be used to conduct a sensitivity analysis of the functional form of the aggregation.

The first two simulation approaches can be applied before a numerical assessment of the weights has been completed. In some cases, the use of these two approaches may result in the identification of a single, most preferred alternative and make further weight assessment unnecessary. These situations are rare, but the two techniques may help to reduce the number of alternatives in the choice set and thereby simplify the task of assessing weights, and prove particularly useful when attempting to select a subset of superior alternatives.

In the next section of this paper, we will present a simple three attribute example to demonstrate the power of simultaneously varying the weights of a multi-criteria model. Then our simulation approaches are outlined and applied to a larger six attribute example problem. The final section offers a summary.

---

[1] For example, determining the rank order of the criteria in terms of importance is the first step in the swing weight procedure. It is important to emphasize that by "importance" we are referring to the relative importance of an increase from the worst to best level of performance for a criteria.

3. Coal power plant site selection -- the three attribute case

Site selection is a natural application of multiattribute utility theory (Keeney, 1980; Merkhofer and Keeney, 1987). These siting decisions are typically based on cost, environmental concerns and other technology specific features. In choosing the location of a coal power plant, these notions can be captured by the measures cost, air quality and site biology (Sarin, 1979).

For this simple example, established decision analysis techniques can be applied to derive the utility functions and scaling constants, or "weights", associated with each measure. Initially, we will assume that the three attributes combine in an additive fashion. Table 1 provides the utility values of the choice set for this hypothetical siting problem. Based on Table 1 alone, it appears that Site 1 is the best alternative.

Attribute	*Cost*	*Air Quality*	*Site Biology*	
Weight	0.60	0.30	0.10	Score
Site 1	0.8000	0.3000	0.4000	**0.6100**
Site 2	0.7000	0.2000	0.7000	**0.5500**
Site 3	0.5000	0.3000	0.6500	**0.4550**
Site 4	0.3000	0.7000	0.2000	**0.4100**
Site 5	0.3500	0.5000	0.6000	**0.4200**

Table 1 **--** Choice Set for Coal Power Plant Sites

The next step in most applications of multiattribute utility theory would focus on a series of sensitivity analyses to establish a sense of the robustness of the recommendations generated by the multi-criteria model. Often, a "one-dimensional" sensitivity to the weights is performed similar to those presented in Figure 1. When performing a one-dimensional analysis on a given weight, the ratios among the other weights are held constant. This single attribute approach can be misleading as it ignores

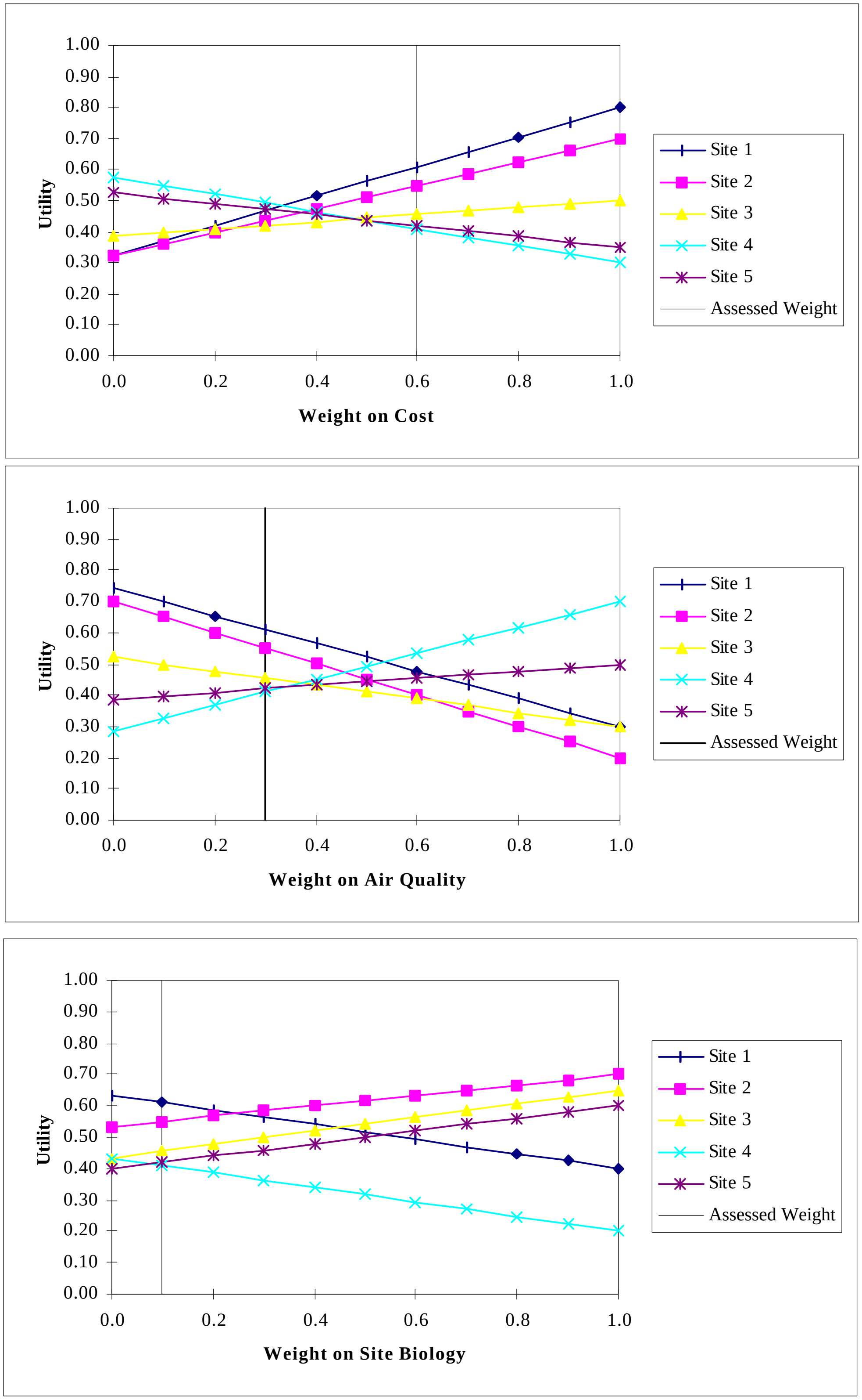


Figure 1 - Example of one way sensitivity analysis of the weights

the potential interaction that can result from simultaneous manipulations of multiple weights.

This concern can be alleviated by evaluating all possible combinations of weights via simulation or algebraic manipulation. The results of such an analysis are fairly easy to interpret in a graphic format such as Figure 2 when there are only two or three attributes. In Figure 2, the x and y-axes represent the weights for the Cost and Air Quality attributes respectively; the implied weight on the attribute Site Biology for any (x,y) pair is simply 1-x-y. At each (x,y) pair, the alternative with the highest score is plotted. For example, using the assumed weights from Table 1 of 0.6, 0.3, and 0.1 for Cost, Air Quality, and Site Biology respectively, Site 1 would be the best choice as a location for the power plant.

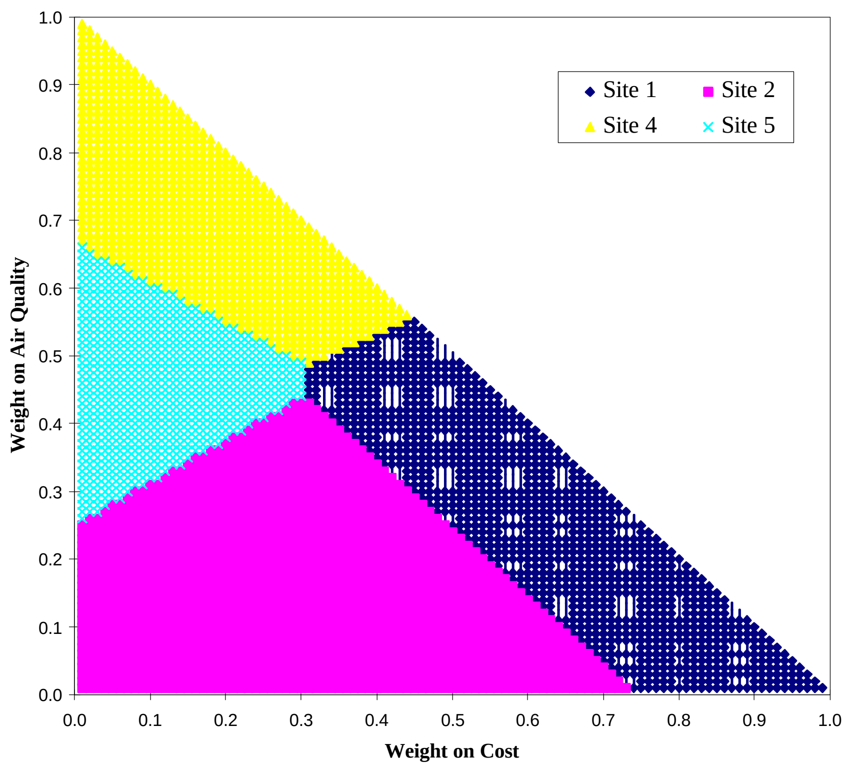


Figure 2 - Most preferred alternative at all possible weight combinations

One of the most important insights that can be obtained from Figure 2 is that Site 3 is "structurally dominated"[2]. Inspection of Table 1 reveals that no alternative dominates any other; however, Site 3 is never the most preferred alternative for any weight combination. This is particularly interesting because Site 3 is the third most preferred site based on the assumed weights (see Table 1).

Another interesting observation is that while Site 5 is ranked fourth (Table 1) and is never the preferred alternative based on the one dimensional sensitivity analysis (Figure 1), Figure 2 indicates that Site 5 can be the superior choice if the weights on Cost and Site Biology are relatively low. This demonstrates the insights gained from manipulation of multiple weights without constraining the ratios between weights. It is important to recognize that these observations about Sites 3 and 5 could have been made without any assessment of the weights of the attributes.

While Figure 2 displays all possible attribute weight combinations, Figure 3 is based on the constraint that the three attributes have a rank order of importance consistent with the weight assumptions presented in Table 1. In other words, the weight on Cost is larger than the weight on Air Quality, and the weight on Air Quality is larger than the weight on Site Biology. Given this importance ordering, then the choice set is reduced to Site 1 and Site 2. Site 1, the original recommendation, is the best choice in the majority of weight combinations. Further, using Figure 3, it may be possible to determine that the decision maker's trade-off among the alternatives excludes Site 2 from being the superior alternative; i.e. the decision maker indicates that the weights will not fall in the gray region of Figure 3.

[2] An alternative is structurally dominated if it is not deterministically dominated by any alternative and yet there is no combination of weights that leads to the alternative being the most preferred.

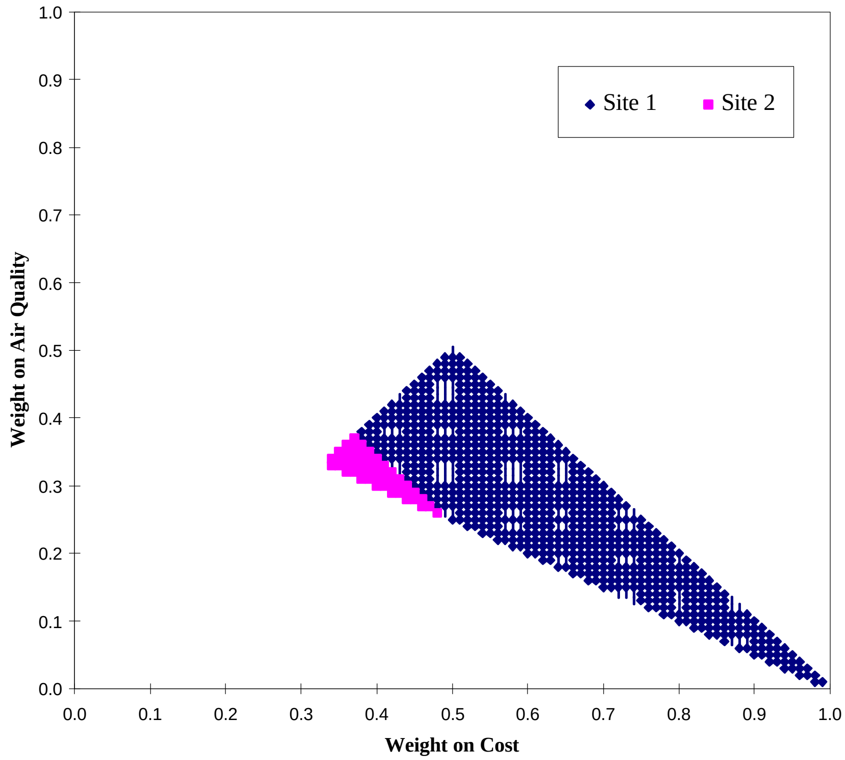


Figure 3 - Most preferred alternative if rank order of criteria weights is enforced

Based on these simple analyses and the assumption of an additive multiattribute utility function, it is possible to narrow the field of competing alternatives to two. In fact, it may be possible to aide the decision maker in choosing a site without conducting a formal weight assessment.

Often a multi-criteria analyses is performed to select a subset of the available alternatives rather than a single, "best" alternative. Figure 2 illustrates that Site 3 is never the most preferred alternative. However, it is possible the Site 3 performs relatively well relative to the other alternatives at other weight combinations. As an extreme example, Site 3 could be the second most preferred alternative at every possible weight combination. Therefore, eliminating Site 3 from further consideration may be premature.

Figure 4 provides some guidance with respect to this issue. Figure 4 presents the ranking of Site 3 at every weight combination. For example, if the weights on Cost, Air Quality and Site Biology are 0.50, 0.10, 0.40 respectively, then Site 3 is the third most preferred alternative. Based on Figure 4, it may be possible to make a judgment about eliminating Site 3. Of course, it is also possible to duplicate Figure 4 enforcing any assessed rank orderings of the weights.

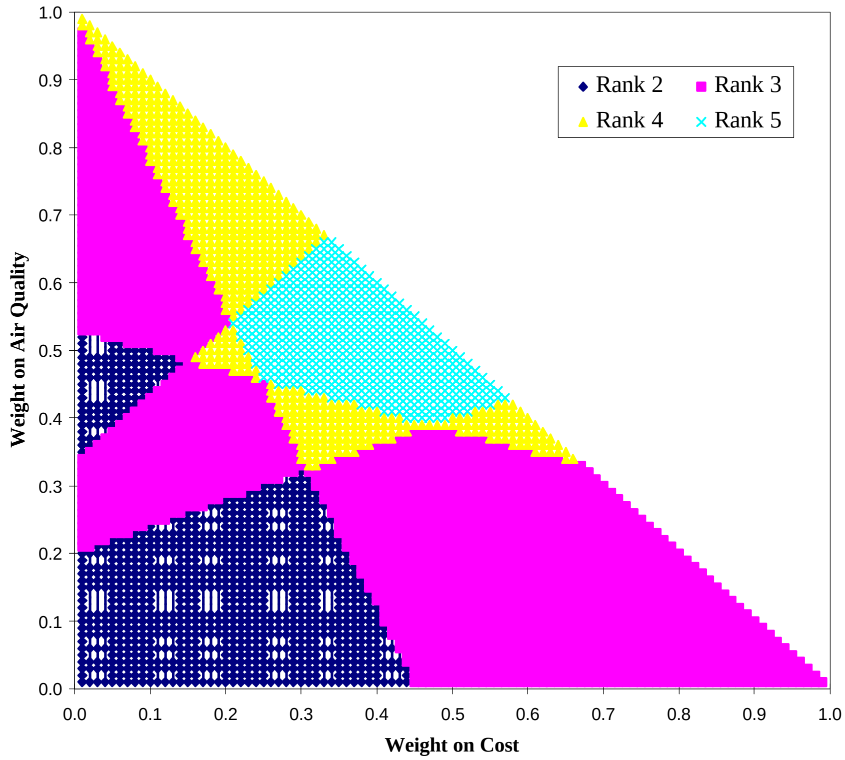


Figure 4 - Site 3's potential rankings

The preceding techniques work relatively well in this simplified setting, but may not be adequate in many "real" decision problems. Obviously, if there are more than three attributes in the multi-criteria decision model, the analysis that generated Figures 2, 3, and 4 becomes tedious and difficult, if not impossible, to explain to the decision maker. Most realistic multi-criteria problems utilize many more than three attributes, so it is desirable to accommodate high dimensional sensitivity analysis. Further, Figures 2, 3 and 4, are based on the additive form (1). Testing the degree to which a multiplicative form (2) would change the model recommendations with the previous methods reduces to applying these techniques at various fixed levels of the common k in equation (2).

4. High dimensional sensitivity analysis via simulation

As discussed in the previous sections, it should be useful to explore the results of changing all of the weights of a multiattribute utility function, perhaps simultaneously, in order to explore in more detail the robustness of the rankings of the alternatives in multiattribute models. However, it would be extremely tedious to try to explore all reasonable combinations of values for the weights one at a time.

Several other approaches could be used to perform a high dimensional sensitivity analysis. For example, Olson (1996) uses a combination of the Centroid (Solymosi and Dombi, 1986) and the SMARTER (Edwards and Barron, 1994) methods of weight assessment to perform such an analyses. The centroid approach was originally developed as a means to assign the weights to an n dimensional multi-criteria model. Olson (1996) determines the grid of possible criteria weights and then uses the centroid weights for all possible grid weight combinations to determine the set of possible rank orders.

Kirkwood and Sarin (1983) provide specific conditions under which partial information about the scaling constants or weights can lead to a rank ordering. For example, the decision maker may be unable to respond to a series of trade-off questions, but he or she may be able to rank order the weights in terms of their magnitude. However, their approach often leads to ranked *subsets* of alternatives rather than a complete rank-ordering.

We propose selecting the weights at random using a computer simulation program so that the results of many combinations of weights, including a complete ranking, can be explored in an efficient manner. In addition, this simulation technique provides a convenient means of testing the robustness of the form of a decision model. We will now discuss three classes of simulation procedures based on randomly generated weights, random weights preserving a rank-order of importance, and random weights from a hypothetical response distribution.

4.1 Random weights

As an extreme case, weights for the measures can be generated completely at random. This approach implies no knowledge whatsoever of the relative importance of the measures. In many multi-criteria settings, the scores of the alternatives significantly limit the subset of potential rankings. It is easy to find dominating or dominated alternatives by inspection; however, subtle relationships that determine "structural dominance" are often difficult to see. By using randomly generated weights it may be possible to uncover these ranking relationships between the alternatives.

To generate the additive weights for *n*-attribute case, we first select *n* - 1 random numbers from a uniform distribution on (0, 1) independently, then rank these numbers. Suppose the ranked numbers are: $1 > r_{(n-1)} \geq \ldots \geq r_{(2)} \geq r_{(1)} > 0$. The first differences of these ranked numbers (including the bounds 0 and 1) can be obtained as $k_n = 1 - r_{(n-1)}$, $k_{n-1} = r_{(n-1)} - r_{(n-2)}$, ... , $k_1 = r_{(1)} - 0$. Then, the set of numbers (k_1, k_2, ..., k_n) will sum to one and be uniformly distributed on the possible domain of weights for an additive model.

For a multiplicative model, the possible range of each attribute weight is from 0 to 1. Thus, weights could be generated at random using a uniform distribution on (0, 1). Once the set of weights is generated, the common *k* can be calculated for each set of weights. However, this approach will not thoroughly explore the possible range of the common *k* in equation (2). In other words, the majority of the common *k*'s will be less than 0 because $\mathrm{E}\left[\sum_{i=1}^{n} k_i\right] = n/2$, which is greater than 1 for *n* > 2.

As an alternative, we offer another approach for generating the random weights associated with the multiplicative models. This approach randomly generates the sum of the scaling parameters, and then randomly divides this interval into the individual scaling constants. This shift in methodology is necessary to ensure that $0 < \mathrm{k_i} < 1, \ \forall \mathrm{i},$ and to fully explore all possible values of the common *k*. Our proposed method can be summarized in four steps:

Step 1: generate *n* random weights as in the additive case;

i.e., $0 < k_i < 1, \ \forall i$ and $\sum_{i=1}^{n} k_i = 1$

Step 2: generate a random number *c* on the interval (0,1)
generate a random number s_1 on the interval (0,1)
generate a random number s_2 on the interval $(1,n)$

Step 3: if c <0.5, $k_i^* = s_1 k_i$

else $k_i^* = k_i + \left[(s_2 - 1)(1 - k_i)\right] \Big/ \sum_{j=1}^{n} \left(1 - k_j\right)$

where k_i^* is the multiplicative scaling constant associated with attribute *i*.

Step 4: calculate[3] the common *k* such that $1 + k = \prod_{i=1}^{n} \left(1 + k k_i^*\right)$.

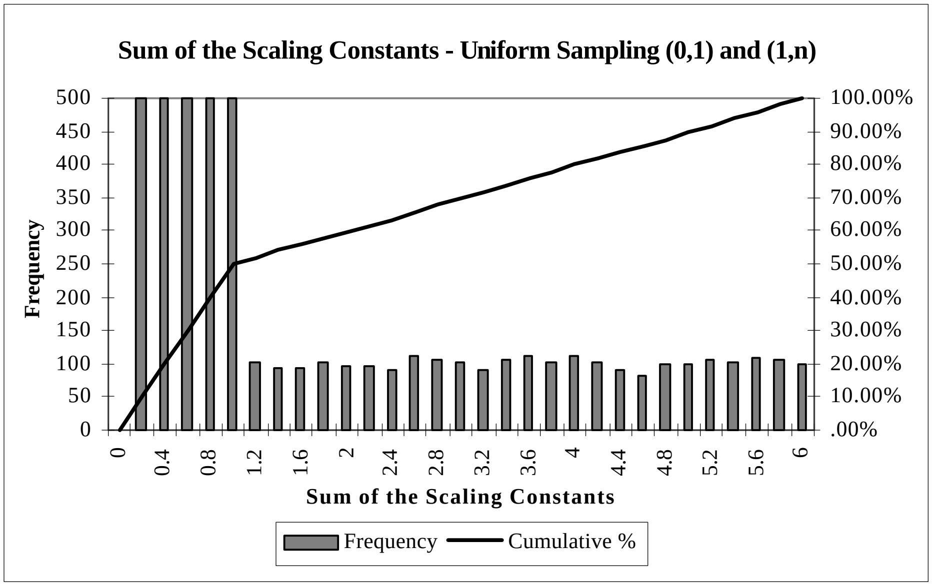
Step 3 causes half of the common *k*'s to be determined when $\sum_{i=1}^{n} k_i^* < 1$ and the other half when $\sum_{i=1}^{n} k_i^* > 1$. This results in a non-symmetric sampling distribution that is "uniform" in the range (0,1) and uniform in the range $(1,n)$. Several factors may cause this constraint to be questioned, but the general technique is still valid.

[3] The iterative approach in Keeney and Raiffa (1976), Section 6.5, was implemented using a simple bisection routine to solve for the common *k*'s.

For example, if n is large in magnitude it is possible that the range $(1,n)$ will be under sampled. In other words, in some contexts it may be desirable to assume a uniform distribution from $(0,n)$. This can be achieved by substituting 0.5 with the quantity $1/n$ in Step 3. Figure 5 presents a graphic comparison of $\sum_{i=1}^{n} k_i^*$ based on 5000 iterations using both of the methods previously discussed for a six attribute multi-criteria problem.

In some situations, it may be reasonable to assume that the multiplicative form implies "substitutability" $\left(\sum_{i=1}^{n} k_i^* > 1\right)$ or "complementarity" $\left(\sum_{i=1}^{n} k_i^* < 1\right)$ among the criteria. In this case, one simply incorporates the relevant condition in Steps 2 and 3.

Using randomly generated weights is analogous to using Figure 2 to explore the entire domain of possible weight combinations. This type of analysis may be particularly useful when attempting to identify a subset of alternatives for a more detailed analysis. Contexts with multiple decision makers (each with a unique set of weights) may also benefit from this type of analysis. It may be possible to eliminate some alternatives from further consideration without conducting an assessment of the weights.



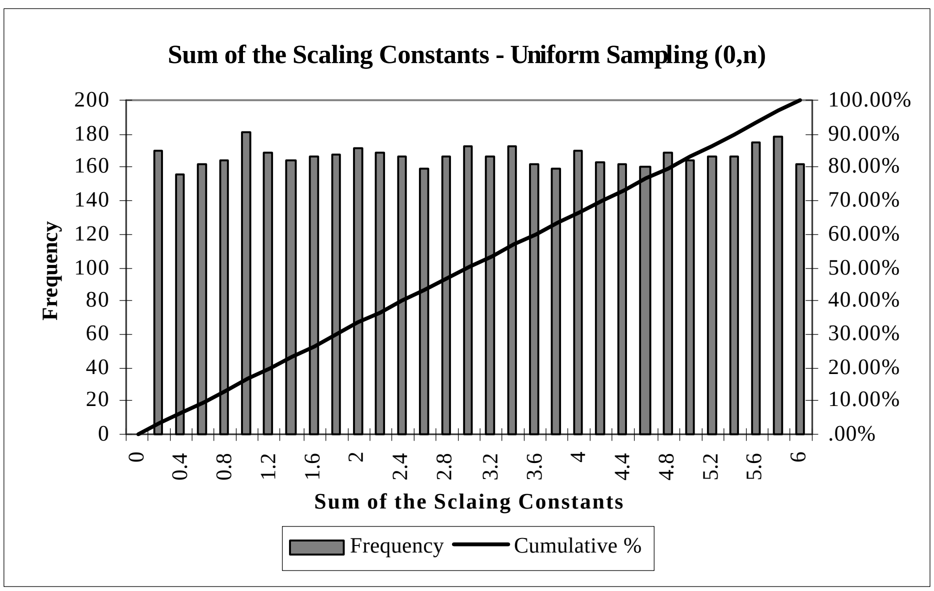


Figure 5 - Comparison of sampling schemes for multiplicative weights

4.2 Rank order weights

The idea of random weights ignores all judgments regarding the tradeoffs among the measures, which may be considered unreasonable if some objectives are “more important” than others. On the other hand, the exact weights from an assessment procedure may be legitimately questioned. The numbers attached to these weights may imply an accuracy that is not appropriate for the analysis.

While the exact magnitude of the weights may be called into question, the relative importance ranking of the attributes may be less controversial. By importance ranking, we are referring to a rank ordering where the highest ranked objective is the one the decision maker would most prefer to increase from the worst to the best level of performance. For a single decision maker, obtaining this rank order information is often easier and subject to less error than assessing numerical weights. In the case of multiple decision makers, it may prove easier to gain agreement on a rank-ordering than obtain consensus on a series of weight assessments (i.e., a series of trade-offs).

It is straightforward to randomly generate the weights based on the importance rank ordering of the criteria. The random weights (of either the additive model or multiplicative model) generated above can be ranked and applied to the corresponding measures. In this analysis, the rank order weights on the measures is maintained, but the weights are otherwise generated at random. Using this simulation model is analogous to confining the domain of possible weights. The impact of this restriction can be readily seen in the case of three criteria by comparing Figures 2 and 3.

4.3 Response distribution weights

The third type of sensitivity analysis using simulation recognizes that the weight assessment procedure is subject to variation. For a single decision maker, this variation may be in the form of response error associated with the weight assessment. In a group setting, different decision makers may have different opinions and preferences. In our approach, the idea is to consider the assessed weights as responses obtained from a distribution of possible responses.

In order to use simulation to explore the implications of response variation, the assessed weights are treated as means of probability distributions of responses, and weights are then generated from these distributions. Although this approach will result in choices of weights that are relatively close (depending on the assumption regarding the underlying distribution) to the actual weights, it allows for the selection of weights that violate the rank order of the assessed weights.

To specify the additive model, *n* weights are generated and normalized via:

$$\frac{X_i}{X_1+\ldots+X_n} \quad \text{where } X_i \sim \text{Gamma}(\alpha_i, \beta) \qquad (3).$$

(3) implies that $\boldsymbol{X}$ ~ Dirichlet with $\alpha = (\alpha_1 \ldots \alpha_m)$ [4]. As the marginal distribution of each k_i is a Beta distribution, the weights are well behaved (i.e., all weights range from 0 to 1) (DeGroot, 1970). The Gamma distributions are parameterized so that the mean of each distribution is the assessed weight. The amount of uncertainty (or response variation) in weights can be controlled by adjusting the sum of Dirichlet parameters.

The multiplicative model uses Beta distributions to simulate the weights. The parameters of the marginal Beta distributions can be determined as $\text{Beta}\{Sa_i, S(1-a_i)\}$ where a_i is the assessed weight for attribute *i* and *S* is a scaling parameter determining the variance of the simulated weights. As the scaling constants will not always sum to 1, the joint distribution of the weights is no longer Dirichlet.

[4] For a review of the Dirichlet distribution see Degroot, 1970, and Wilks, 1962.

5. Selecting a site for a coal power plant

Now, we will apply our simulation techniques to a more realistic example problem. Sarin (1979) and Sarin, Dyer, Nair and Keshavan (1978) describe the use of multiattribute utility theory to select one of thirteen sites as the location for a coal power plant. This study was selected for analysis because it provided a complete set of weights that could be used to test our methods of sensitivity analysis. Table 2 provides the complete alternatives by objectives matrix.

5.1 Random weight simulation results

We will begin by applying a completely random weighting scheme to the scores in Table 2. This approach will allow us to test for interrelationships between potential sites that partially determine their rank-orderings irrespective of preference judgments reflected in the weights. This type of analysis is particularly useful before any assessments are conducted. The results of both the additive and multiplicative random simulations are provided in Table 3. It should be noted that the statistics in Table 3 are based on the ranks resulting from simulated scores in 5000 independent random trials.

Several observations are apparent in Table 3. It seems clear that Site 11 is inferior relative to the others. In both the additive and multiplicative rankings, Site 11's best performance was sixth out of thirteen. Modal performance was eleventh in the additive case and twelfth in the multiplicative case. If we were conducting a preliminary analysis rather than a historical evaluation, there is evidence to suggest removing Site 11 from further consideration. Similar logic indicates that sites 2, 7, 10 and 12 may also be candidates for removal.

	Cost	Air Quality	Site Biology	Socioeconomic	Impact on Fish	Line Biology	
	Annualized differential capital and yearly operating and maintenance costs measured in 1978 dollars	Calculated concentrations as a percentage of allowable standard	Subjective scale capturing the biological impact at the site	Subjective scale measuring the potential short and long term impacts caused by construction and operation of the plant	Loss of salmonids and other fish	Environmental impact of transmission interconnection, water line and railroad	
Weight[5]	0.52	0.19	0.17	0.07	0.03	0.02	**Score**
Site 1	1.0000	0.7731	0.7400	0.8234	0.7211	0.4375	0.8807
Site 2	0.9167	0.4088	0.7600	0.7831	0.7548	0.5750	0.7725
Site 3	0.9333	0.5333	0.9650	0.7380	0.7188	1.0000	0.8439
Site 4	0.8500	0.9539	0.9300	0.9295	0.7188	1.0000	0.8880
Site 5	0.9833	0.9211	0.9300	0.7569	0.7188	0.1413	0.9218
Site 6	0.8333	0.9737	0.9300	0.8748	0.8577	1.0000	0.8834
Site 7	0.9333	0.0000	0.9300	0.9250	0.5969	0.8500	0.7431
Site 8	0.9333	0.6833	0.9650	0.9160	0.6328	1.0000	0.8823
Site 9	0.9000	0.0000	1.0000	0.5360	0.5558	1.0000	0.7122
Site 10	0.5333	0.8092	0.0000	0.9385	0.7188	0.1957	0.5222
Site 11	0.4000	0.2700	0.9300	0.6588	0.7188	0.8500	0.5021
Site 12	0.2833	0.6667	0.9300	0.1450	0.7188	1.0000	0.4838
Site 13	0.4667	0.8882	0.9000	0.9340	0.9000	0.6750	0.6703

Table 2 **--** Alternatives by objectives matrix, siting a coal power plant (Sarin, 1979)

[5] The original study featured many sets of weights. This particular set was chosen because it was representative of all sets.

Additive Random Weights								
	Mode	Minimum	25th Percentile	50th Percentile	75th Percentile	Maximum	Mean	Standard Deviation
Site 1	6	1	6	7	8	12	7.14	2.22
Site 2	9	3	8	9	10	12	8.91	1.30
Site 3	4	1	4	5	6	10	4.97	1.55
Site 4	2	1	2	2	2	8	2.13	0.89
Site 5	12	1	5	7	10	13	7.22	3.15
Site 6	1	1	1	1	2	9	1.56	1.01
Site 7	6	2	6	8	10	13	8.28	2.50
Site 8	3	1	3	3	4	11	3.42	1.39
Site 9	12	1	7	10	12	13	9.58	2.61
Site 10	13	4	11	13	13	13	11.61	1.94
Site 11	11	6	9	11	11	13	10.35	1.55
Site 12	13	3	8	11	12	13	10.24	2.57
Site 13	3	1	3	5	7	12	5.59	2.50

Multiplicative Random Weights								
	Mode	Minimum	25th Percentile	50th Percentile	75th Percentile	Maximum	Mean	Standard Deviation
Site 1	7	1	6	7	8	12	7.19	1.64
Site 2	10	3	8	9	10	12	9.24	1.23
Site 3	4	1	4	4	5	10	4.60	1.10
Site 4	2	1	2	2	2	8	1.98	0.50
Site 5	6	1	6	6	8	13	6.95	2.43
Site 6	1	1	1	1	1	8	1.21	0.57
Site 7	8	2	7	8	9	13	8.13	1.95
Site 8	3	1	3	3	3	8	3.20	0.82
Site 9	9	1	8	10	11	13	9.40	2.23
Site 10	13	4	12	13	13	13	12.26	1.47
Site 11	12	6	10	11	12	13	10.90	1.33
Site 12	11	4	10	11	12	13	10.61	1.96
Site 13	5	1	4	5	6	12	5.33	1.79

Table 3 **--** Ranking results from the random weights models

Looking towards potential winners, Site 6 performs well in both simulations. Site 6 is the top ranked site in fifty percent of the additive and seventy five percent of the multiplicative simulations. Sites 4, 8 and 13 also appear to have promise. These observations can be summarized in a graphic format as shown in Figure 6.

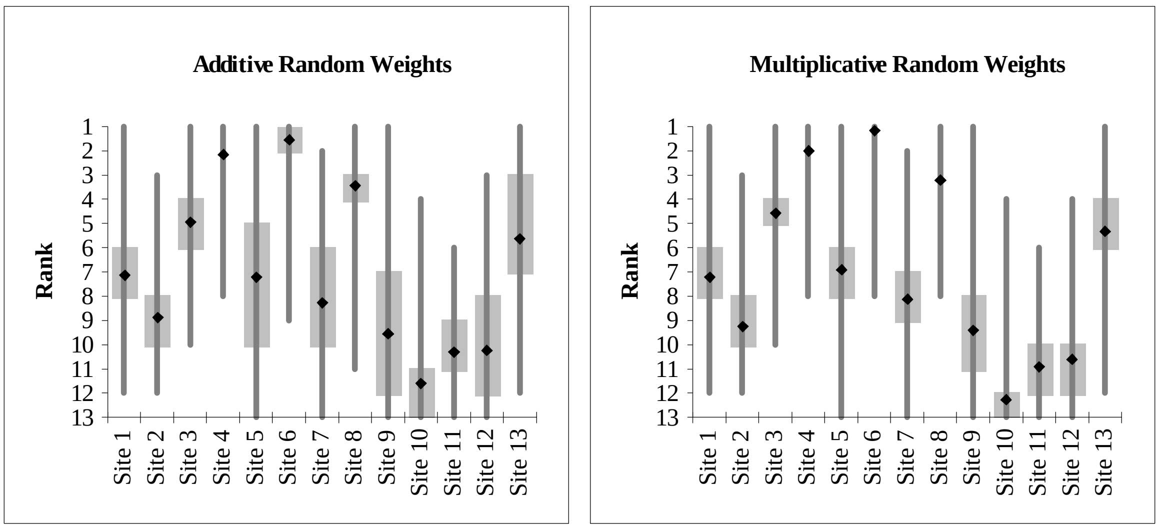


Figure 6 - Ranking comparison for random weights

KEY: The black diamonds correspond to the mean ranking, the gray box encloses the middle 50% of the ranking distribution (if there is no gray box, the first and third quartiles are identical), and the minimum and maximum ranks are the endpoints of the gray lines..

Table 3 and Figure 6 may also prove useful in discussions with decision makers. It may be possible to ask decision makers questions like: “Site 11’s best performance is sixth; in fact, in 75 % of the simulations it was ranked 9th or lower. Would you like to remove it from consideration?” It may also be possible to define more general rules about when to remove alternatives based on decision maker preferences.

Of course, removing any alternative from consideration that had been ranked first in at least one simulation could be dangerous as there exists a combination of weights and functional form that led to it being the most preferred. However, this combination may be unrealistic in terms of the decision maker’s preferences. While it is possible to examine the simulation data and determine exactly what combination of weights caused a particular ranking to occur, it may be more expedient to assess a rank ordering of the criteria and apply the rank order weight simulation.

5.2 Rank order weight simulation results

Randomly generating the weights while preserving their criteria rank order places substantial restrictions on the domain of possible weights that are consistent with the decision maker's judgment of criteria importance. Therefore, the results from the rank order simulation may provide more meaningful results. Figures 7 and 8 present the results from the rank order simulations each featuring 5000 iterations for the coal plant siting example.

Figure 7 clearly illustrates the range over which the alternatives vary. Figure 8 can be interpreted in a manner analogous to using cumulative probability distributions to determine stochastic dominance. Rather than use a cumulative probability distribution, Figure 8 features a cumulative ranking distribution. Alternative A is said to stochastically rank dominate alternative B if alternative B is never above or to the left of alternative A in Figure 8. We argue that if an alternative is stochastic rank dominated it is inferior; however, it is also possible that an alternative that is not stochastic rank dominated is inferior[6].

As one would expect, the range of possible rankings for the alternatives is much narrower when the rank order of the criteria is imposed in the simulation. After the random simulations, it appeared the Site 13 was a good choice as a site for the power plant. From Figure 8, once rank order of the attributes is enforced in the simulation, Site 13's performance drops considerably. As indicated in Table 2, Site 13's best performance is on attributes that were ranked as less important. So, for some decision maker or stakeholder group Site 13 may be one of the better sites, but based on the assessed rank

[6] Consider a simple example: alternative 1 is ranked first in 99% of a set of simulations and ranked third in the other 1% while alternative 2 is ranked second in all the simulations. Although alternative 1 does not stochastic rank dominate alternative 2, it is realistic to consider alternative 2 as inferior.

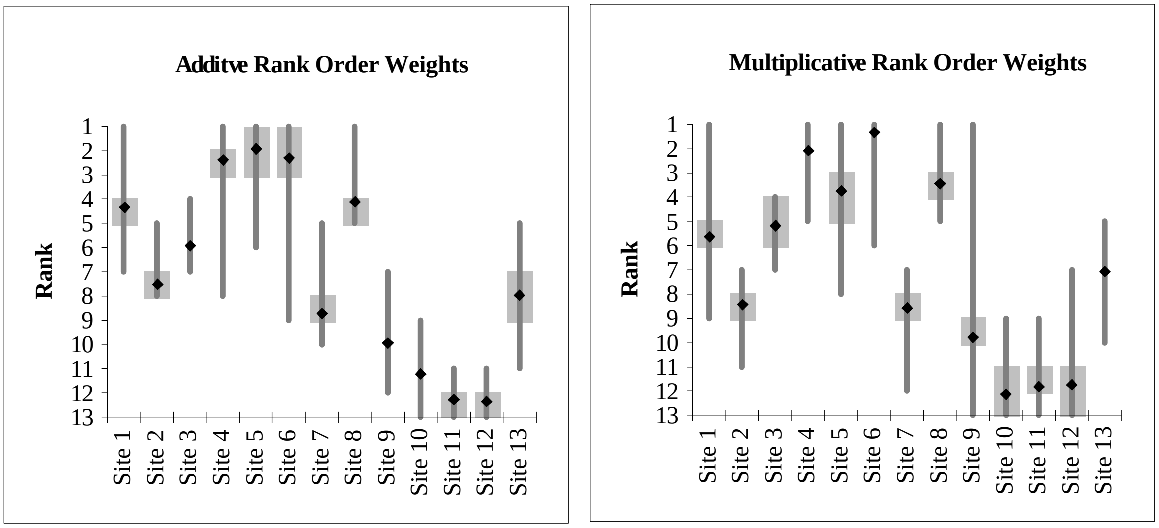
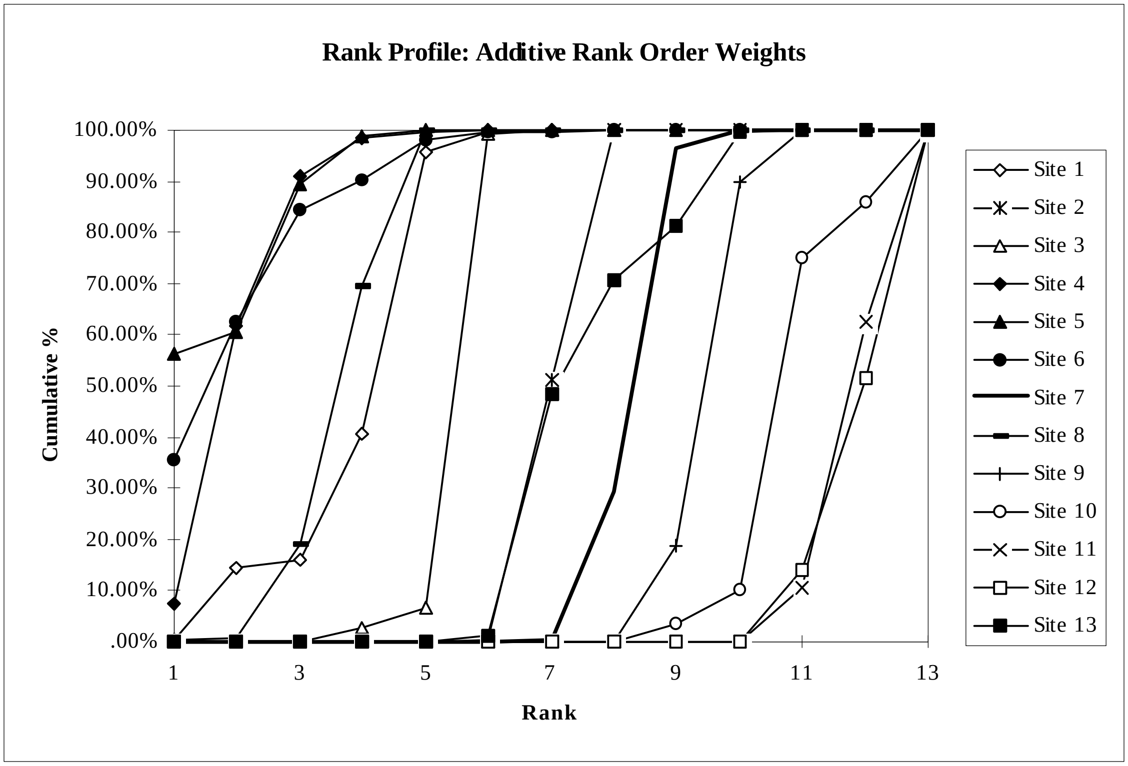


Figure 7 - Ranking comparison for rank order weights

KEY: The black diamonds correspond to the mean ranking, the gray box encloses the middle 50% of the ranking distribution (if there is no gray box, the first and third quartiles are identical), and the minimum and maximum ranks are the endpoints of the gray lines.

ordering, Site 13 could be removed from any remaining analysis for a decision maker whose weight are consistent with the rank order of weights shown in Table 2.

If the goal of this analysis were to select a subset of six alternatives, we could stop now without conducting a formal weight assessment. Using Figure 8, it is clear that regardless of the form of the utility function, Sites 4, 5 and 6 are consistent top performers. Sites 1, 3 and 8 also appear to be superior when compared to the rest of the contenders.



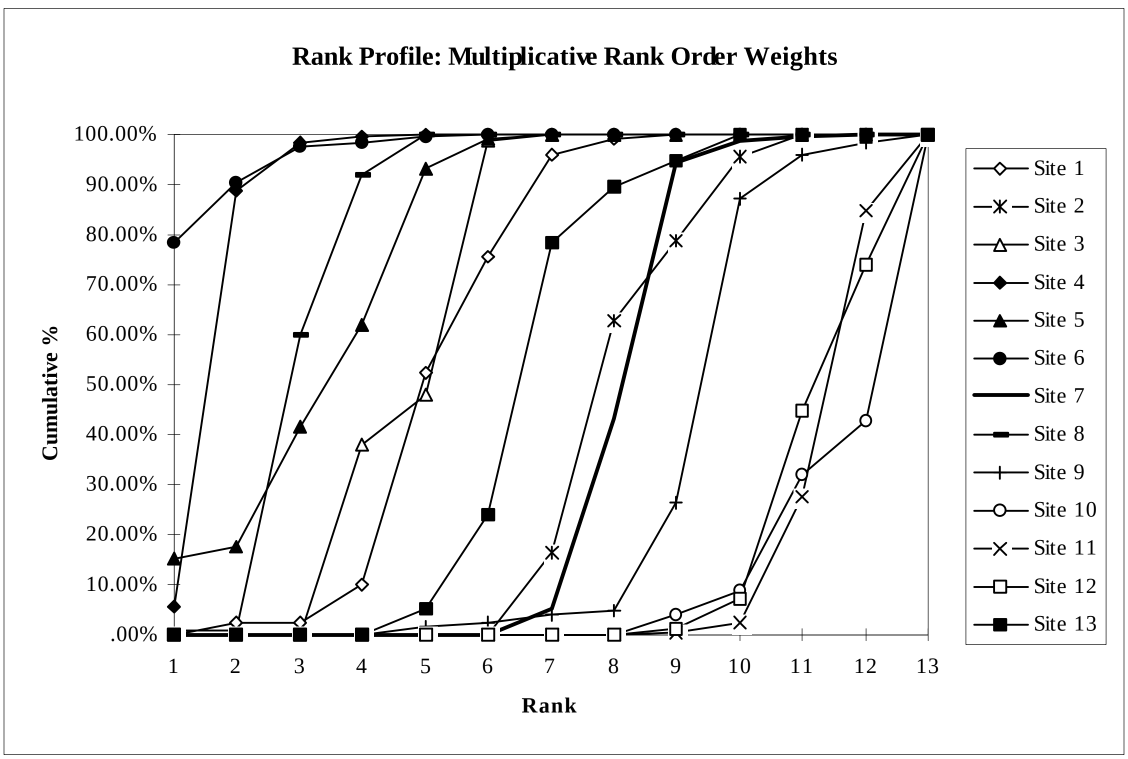


Figure 8 - Ranking profile for rank order weights

5.3 Response distribution simulation results

The results from the response distribution simulations are presented in Figures 9 and 10 and are consistent with Figures 7 and 8 in terms of the top performing alternatives. Sites 5, 4, 6, 8 and 1 are the top performers. The original additive assessment in Table 2 recommended Site 5. Site 5 stochastically rank dominates all other alternatives in the multiplicative rank profile. In the additive rank profile Site 5 is top ranked in over fifty percent of the simulations. In the additive case, Sites 4 and 6 are not rank dominated by site 5 because their range of ranks is from first to fourth compared to first to sixth for Site 5. It should be noted that the long tail on the rank distribution of Site 5 in the left panel of Figure 9 is due to the four of 5000 iterations where Site 5 was ranked worse than 6th

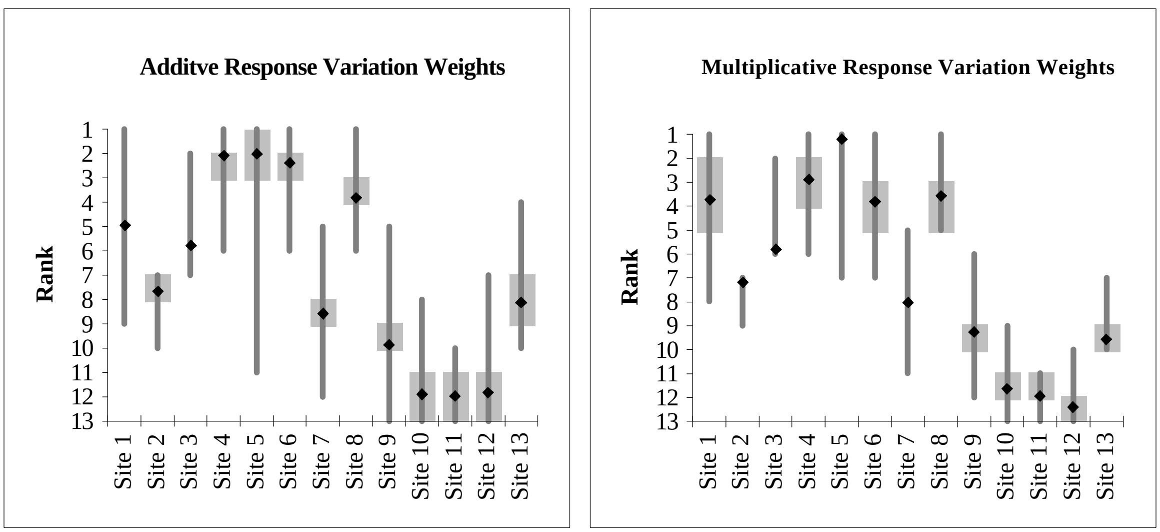
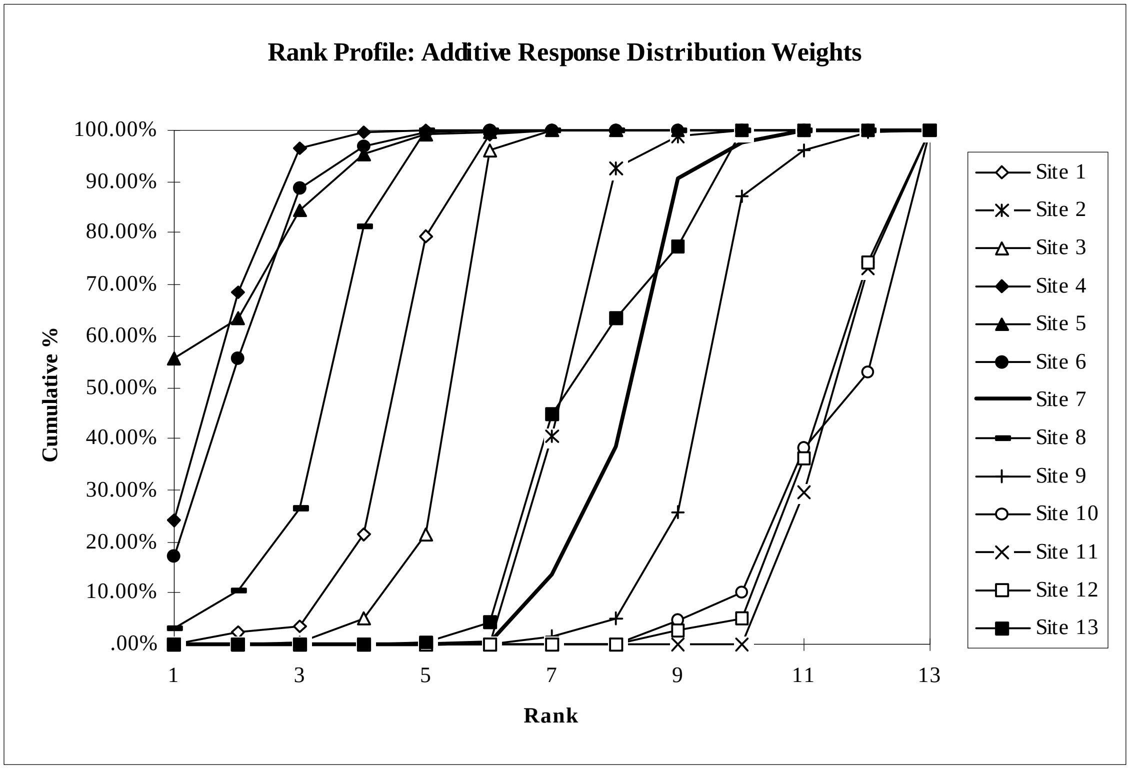


Figure 9 - Ranking comparison for response distribution weights

KEY: The black diamonds correspond to the mean ranking, the gray box encloses the middle 50% of the ranking distribution (if there is no gray box, the first and third quartiles are identical), and the minimum and maximum ranks are the endpoints of the gray lines.



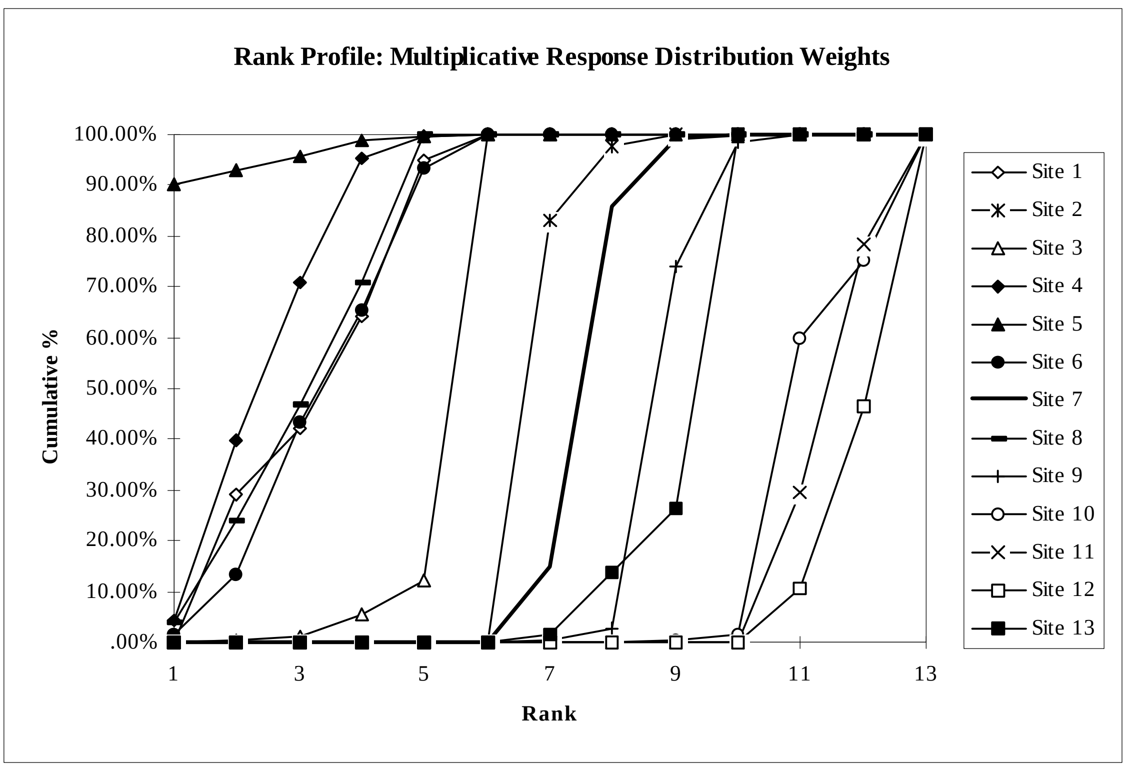


Figure 10 - Ranking profile for response distribution weights

Based on the simulation results, it appears the original recommendation of Site 5 was correct. The key factor seems to be the magnitude of the weight on cost. Sites 4 and 6 are consistent good performers across all six measures but Site 5 is the most preferred alternative because it performed extremely well on the two most "important", according to the weight assessment, measures.

6. Conclusion

Problems can be encountered when attempting weight assessments for multi-criteria decision models. This paper has presented three classes of simulation models that offer assistance in the evaluation of weights for multi-criteria models, including the possibility of a multiplicative functional form. The random weights approach requires no weight assessment; the rank order model requires a rank ordering of the weights on the criteria; and the third type of simulation allows the inclusion of response variation in a tractable fashion. All three approaches can be applied when the decision maker is a group or committee rather than an individual. Rather than require a consensus, the decision analyst can test the sensitivity of the rankings provided by the model.

The simulation techniques outlined in this paper were developed while we were involved in a project to determine the best technology to dispose of the U.S. stockpile of weapons-grade plutonium. The goal hierarchy was quite complicated and consisted of about forty measures or attributes. We developed a full multi-attribute utility model for the decision problem, including a weight assessment, and sought a means to test the robustness of rankings generated by the model (see Dyer, Edmunds, Butler and Jia, 1996).

The random weights model was useful in that it helped the decision maker(s) focus on the alternatives that were superior regardless of the relative importance attached to the attributes. It was also effective when a sub-group within the decision making body questioned the magnitude of the weights in one of the sub-goals of the MAU model. We were able to demonstrate that the rankings provided by the model were unchanged no matter what values were assigned to the weights on the criteria in question. This defused a potentially lengthy debate and returned the focus to the more consequential, discriminating attributes of the alternatives.

The group was in general agreement as to the relative importance of the remaining criteria. Using the importance rank order and response distribution models, we were able to identify a small subset of alternatives that were consistently ranked above the other options over a wide range of possible weights. The importance rank order models also proved useful when testing a partial ordering of the weights, for example $k_1 > k_2$ and $k_3 > k_4$.

In this paper we have assumed that the functional form of the model was purely multiplicative or additive. In our plutonium decision analysis application, we suspected that the form of the utility function was multiplicative for several of the sub-goals. We employed a "hybrid" aggregation using our simulation models and demonstrated that the choice of the functional form had little effect on the model.

Acknowledgments

Support for this paper was provided by the U.S. Department of Energy, Cooperative Agreement No. DE-FC04-95AL85832. However, any opinions, findings, conclusions, or recommendations expressed herein are those of the author(s) and do not necessarily reflect the views of DOE. This work was conducted through the Amarillo National Resource Center for Plutonium. We are also grateful to Thomas Edmunds of the Department of Energy for his insightful comments.

References

Borcherding, K., Eppel, T. and von Winterfeldt, D. (1991), "Comparison of weighting judgments in multiattribute utility measurement," *Management Science*, 37**,** 1603-1619.

DeGroot, M.H. (1970), *Optimal Statistical Decisions*. McGraw-Hill Book Company.

Dyer, J., Edmunds, T., Butler, J. and Jia, J. (1996), "A methodology for the analysis and selection of alternatives for the disposition of surplus plutonium," presented at American Nuclear Society 1996 Annual Meeting, Reno, NV.

Dyer, J. S. and Lorber, H. W. (1982), "The multiattribute evaluation of program-planning contractors," *OMEGA*, 6, 673-678.

Dyer, J. S. and Sarin, R. (1979), "Measurable multiattribute value functions," *Operations Research*, 27, 810-822.

Edwards, W. and Barron, F. H. (1994), "SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement," *Organizational Behavior and Human Decision Processes*, 60, 306-325.

Keeney, R. L. (1980), *Siting energy facilities,* New York: Wiley.

Keeney, R. L. and von Winterfeldt, D. (1994), "Managing nuclear waste from power plants," *Risk Analysis*, 14 , 107-130.

Kirkwood, C.W. and Sarin, R.K. (1985), "Ranking with partial information: A method and an application," *Operations Research*, 33, 38-48.

Merkhofer, M. L. and Keeney, R. L. (1987), "A multiattribute utility analysis of alternative sites for the disposal of nuclear waste," *Risk Analysis*, 7, 173-194.

Olson, D. L. (1996) "Use of the centroid approach for selection sensitivity analysis" Working paper, Texas A&M University, College Station Texas.

Sarin, R. (1979) “Ranking of multiattribute alternatives with an application to coal power plant siting,” Working Paper, Purdue University.

Sarin, R., Dyer, J. S. and Nair, Keshavan. (1978), “A comparative evaluation of three approaches for preference function assessment,” Working Paper, Purdue University.

Saaty, T. L. (1980), *The analytic hierarchy process*, McGraw-Hill, New York,.

Schoemaker, P.J., Waid, C.C. (1982) “An experimental comparison of different approaches to determining weights in additive utility models,” *Management Science*, 28, 182-196.

Solymosi, T. and Dombi, J. (1986) “A method for determining the weights of criteria: The centralized weights,” *European Journal of Operational Research*, 26, 35-41.

Wainer, H., (1976) “Estimating coefficients in linear models: It don’t make no nevermind,” *Psychological Bulletin*, 83, 213-217

Weber, M. and Borcherding, K. (1993), “Behavioral influences on weight judgments in multiattribute decision making,” *European Journal of Operational Research*, 26, 35-41

Wilks, S.S. (1962), *Mathematical statistics*. John Wiley and Sons.