Estimating Errors in Concept Selection

Kevin N. Otto, Assistant Professor
Engineering Design Research Laboratory
Massachusetts Institute of Technology
77 Massachusetts Avenue, Room 3-449
Cambridge, Massachusetts 02139

Kristin L. Wood, Associate Professor
Department of Mechanical Engineering
The University of Texas
ETC 5.160
Austin, TX 78712-1063

January 31, 1995

Abstract

Numerical concept selection methods are used throughout industry to determine which among several design alternatives should be further developed. The results, however, are rarely believed at face value. Uncertainties (or errors) in subjective choices, modeling assumptions, and measurement errors are fundamental causes of this disbelief. This paper describes a methodology developed to predict overall final error ranges and to estimate a confidence measure in the numerical evaluation results, based upon assigning error ranges to the individual ratings and weightings. Each numerical assignment is given an associated error range, and then treated as a probability error to create a simple means to propagate the errors into an overall error range on the final rating value. Further, a degree of confidence is derived, similar to a statistical t-test, to indicate an induced confidence level in the final decision. Two preliminary concept selections are shown, to illustrate the methodology. Results from these concept selections indicate that (1) uncertainties can be suitably captured and quantified; (2) critical design questions are addressed during the process of numerical concept selection with error propagation; and (3) designers can make more informed and confident decisions through error estimation.

1. Introduction

A designer has many different methods available to help select a product concept. Concept selection charts are used to help to choose among designs. Perhaps the most common and simple selection chart is a decisions matrix, where a number of concepts are rated against a finite set of criteria or objectives, then each criterion is subjectively weighted, and the results are summed to provide a "score" for each concept (Pahl and Beitz, 1984). More recent techniques, and perhaps more advanced, include Pugh's method, with extensions (Pugh, 1990). Pugh's method assigns a datum product or concept, and then all others are rated as better or worse on each product objective. Numerical scoring procedures extend this to assign degrees of better or worse with numbers that are weighted and summed (Otto, 1995). More complex methods such as the *analytic hierarchy process* (Saaty, 1980) use hierarchical groupings of criteria.

These methods are designed to help make a determination, yet experience shows that many designers, students, and instructors do not believe in the results of the methods. When using a concept selection chart, the tendency is to feel uncomfortable with adopting the highest scoring result. Central to this feeling of discomfort is the question: What do the differences in numbers mean? The methods, however, are usually effective; they can quickly form group consensus on major issues, and they provide a systematic approach for identifying and answering important design questions, such as design feasibility, expected quality, and relative performance. They especially help in overcoming social/teaming issues in the design process. Nonetheless most

Corresponding Author: Kevin Otto, knotto@mit.edu, (617)□253-8199.

persons involved do not put faith in the numerical result, and have no quantitative means to estimate errors.

In this paper, we propose and develop a simple means to determine an error range for the final score in a numerical concept evaluation. We propose a method whereby a designer assigns error ranges to the individual ratings for each concept on each objective. Also, an error range is assigned to the importance rating on each objective. Given these data, we propose the application of first order probability mathematics to combine the error ranges into an error range on the final overall rating. This range can be compared with the differences in overall rating among the concepts, and thereby a confidence level in the result is determined.

Related Work

Many researchers and practitioners have developed methods for selecting concepts in design. (Otto, 1995) presents a detailed analysis of these methods using measurement theory. Four classes of selection methods are distinguished: ordinal, interval, ratio, and extensively measurable scales. Pro/Con assignments and Pugh's method (Pugh, 1990) use ordinal measurement for a preliminary comparison of design concepts. Ordinal measurements in this context implies that design concepts are rank ordered with respect to each other, but with no quantification of the differences in value of the comparison. Concepts are only ordered. Numerical tolerances, or error ranges, therefore have little real meaning in ordinal measurement.

Quantitative numerical assignments in a selection chart additionally represent information on how much better different concepts satisfy the designer and/or defined customer requirements, perhaps mapped from a quality function deployment chart (Hauser and Clausing, 1988; Breyfogle, 1992; Ullman, 1992). Such numerical values form a quantitative scale across the options for each performance objective, thereby defining an interval scale. Interval scales are the structurally simplest decision methods on which error ranges can be applied. Use of quantitative concept selection scales is also sometimes called concept scoring (Ulrich and Eppinger, 1995). Two basic methods, the *better/worse method* and the *lottery method* (Otto, 1995), can be used to scale quantitative ratings between two reference options. These methods are based on the ideas that a design concept is negatively or positively ranked relative to an assigned base-point design (better/worse method), or is ranked relative to the worst design as compared to the difference between the best and worst designs (lottery method). Utility theory (Raiffa and Keeney, 1976, Thurston, 1991) and the method of imprecision (Otto and Antonsson, 1991; Wood *et al.*, 1992; Otto and Antonsson, 1994) are based upon the lottery method. In this paper, we apply an error analysis to the better/worse method, but the analogy to the lottery method is clear. Thurston (1991) also reviews decision matrices similar to that proposed here, and notes their difficulty when using unreferenced linear value functions, and the difficulty in formulating weighting coefficients.

Siddall (1983, Chapter 10) describes a similar but more simplified approach by adding value functions defined over performance metric values. The value functions are unweighted additive value functions, which presents difficulty in formulation, see (Krantz et al, 1971). A unit of value on one metric must equal one unit of value on every other metric. Formulating the value functions while satisfying this constraint can prove difficult. Rather, we choose to separate the process into two steps, formulating a "raw value" and a weighting. Following the standard utility theory approach, Siddall also describes evaluating decision uncertainty due to uncertain delivered performance metric values, as we also do. However, he applies the uncertainties differently than the work described in this paper, to define the best decision as the one with maximum expected overall additive value across the performance metric noise, entirely analogous to utility theory. We rather accept the crisp decision, and use the uncertainty to quantify the confidence in this crisp result.

We have found no work quantifying the subjective uncertainty in formal value assignments, only work using value as a quantification of subjective uncertainty. In part, this is due to many utility theory advocates rejecting the frequency principle, confidence levels and statistical confidence tests as a means to manipulate and understand uncertainty (French, 1988, Savage, 1954).

More structured numerical concept selection methods are also developed in the literature. A ratio scale exists when the performance objectives used have well understood zero values. Such is the case with noise and vibration (no noise, no vibration), or importance (no importance), for example. Methods which apply ratio scales include the analytic hierarchy process, (Saaty, 1980; Marsch *et al.*, 1994; Putrus, 1989) and goal programming (Bascaran *et al.*, 1989). Methods more structured than ratio scales have extensively measurable units, where there exists a well understood zero value and completely calculable or derivable values for all concepts in the evaluation scale. These methods fall into the regime of standard optimization, where error propagation methods have been in existence for some time (Sveshnikov, 1968).

While it is clear that extensive efforts have been devoted to developing concept selection methods, very little work exists in estimating selection chart uncertainties and errors. Without such estimation, the designer is faced with the dilemma iterated above: what belief can be derived from the numerical results, if any? In the remainder of this paper, we, optimistically, seek to resolve this dilemma. The next section reviews an explicit numerical concept selection method, following precise and systematic steps. This is necessary so that each step can be assigned an error range. A simple material selection example from a real industrial case study is demonstrated. Section 3 then develops the error analysis method, and demonstrates it with the material example. Section 4 subsequently presents two preliminary design concept evaluations as examples. Finally, Section 5 draws a number of conclusions concerning error estimation during concept selection.

2. Numerical Concept Evaluation

Figure 1 schematically illustrates an overview of the concept evaluation methodology used in this research. As shown in the figure, the concept selection methodology includes five composite steps: (1) determining the concepts' performance levels, based on unbiased selection criteria assigned early in the design process; (2) selecting an appropriate measurement scale; (3) performing a relative comparison of the concepts; (4) weighting the selection criteria based on preassigned weights; and (5) determining a cumulative rating of all concepts. Each of these composite steps include a number of tasks or substeps. These are shown in Figure 1 according to the action required by the designer. Actions include assignment (or subjective choice), derivation, and application of required information. By necessity, iteration is shown between assignment, choice, and application tasks to highlight their subjective nature.

In the remainder of this section, the concept selection method of Figure 1 is detailed, including a motivating material selection example. Before constructing the detailed concept selection method, however, we must revisit our ultimate goal and consider the implicit requirements of estimating error during decision making. One such requirement concerns the need to propagate error sources. To develop a means to propagate individual errors into an overall error analysis, the mapping among the variables must be made explicit. This presents a problem, in that there are multiple methods to assign values within a concept selection chart. We therefore adhere to a systematic approach that is consistent with mathematical measurement theory, the field of mathematics devoted to the study of measuring any quantity, including non-physical entities like design concepts. The righthand segment of Figure 1 summarizes the mappings to be used in our approach.

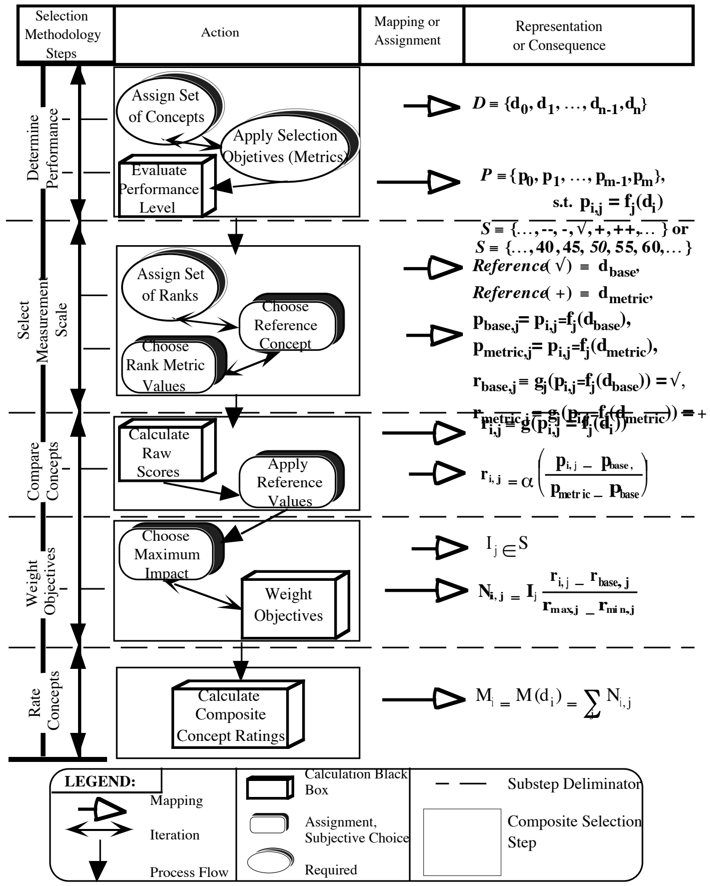

Figure 1. Overview of Concept Selection Method.

For our concept selection approach, alternative concepts are listed along the columns of a matrix. The objectives (or criteria or performance metrics) that form the basis of the selection are listed along the rows. Entries in the matrix are the ratings of each alternative concept on each objective, and are given ranks from the set

$$S = (--, -, \surd, +, ++)$$

where the √ denotes a base point or “zero” ranking, (–) denotes one unit of negative rank compared with √, and (+) denotes one unit of positive rank. The set S may be assigned with additional or less multiple (+)'s and (–)’s, depending on the level of discrimination in rank that is desired. An alternative set of symbols or numbers may be used for S, depending on the desires of the designer. In the evaluation process, the ranks are normalized and summed across the different objectives for each design concept, and the design with the highest sum is nominally determined to be the most promising candidate.

When analyzing purely the mechanics of the method, we can clarify the need for five separate assignment, application, and calculation tasks (Figure□1) when constructing the ranks in the chart (Otto, 1995):

1. Identify a set of concepts and apply objectives,
2. Select a base point and metric configuration or a scale for each objective,
3. Construct a ranking across the concepts for each of the objectives,
4. Normalize the different objectives’ measuring scales, and
5. Sum the normalized ratings for an overall evaluation of each design concept.

Table 1: Performance Values for the Material Selection Example.

Objective	(units)	Steel 1020	Stainless 304	Aluminum 5052	Copper	Bronze
Thickness	inches	0.107	0.107	0.407	0.205	0.205
Conductivity	Btu ft/hr °F ft^2	27	9.4	80	200	109
Thermal mass	Btu/°F	2.93E-04	3.43E-04	7.33E-04	5.45E-04	5.26E-04
Diffusivity	ft^2/hr	909	270	3749	6751	3809
Hardness	Brinnell	111	240	47	2	1
Yield	psi	30000	110000	13000	40000	37000
Machinability	[0-100]	65	90	30	20	20

Each of these separate activities can give rise to errors in the evaluation. A simple, yet demonstrative design example is used below to clarify these activities and associated error sources.

Example: Material Selection

Consider the selection of a material for a surface which must transfer heat quickly and uniformly, and must also support a load. This material selection example has benefit of measurable values for the performance objectives reflecting each of these needs, making it a useful case study. In particular, the required thickness for proper heat transfer, the conductivity, thermal mass, diffusivity, hardness, yield strength, and machinability are all metrics with tabulated numerical values for different available materials. Suppose the decision has been reduced to considering 1020 steel, 304 stainless steel, 5052 aluminum, copper and bronze. The values of the performance metrics for each of these materials are shown in Table□1.

This provides the first mapping where errors can enter. The level of performance provided for each objective at each configuration must be determined, and we will denote it as given by

$$p = f(d) \tag{1}$$

where p is the level of provided performance, d is a design concept (a material in this example), and f is the mapping which provides the level of performance, given the configuration d (a lookup in a table for this example).

We can apply the mechanics of numerical concept selection method to this material selection problem. Consider the first task. When constructing the chart, best design practice calls for a designer to first select a reference *base point* design concept (Pugh, 1990; Otto, 1995). In re-design applications, this reference case is usually the existing product or perhaps a competitor's product (Hauser and Clausing, 1988). This case is given √ ranks in every objective. In this example, the base point can be chosen as the "1020 steel" option. It gets a √ for every objective. Next, a *metric* option must be chosen that will be given a (+) if it is better than the base point, or a (–) if it is worse than the base point. Then every other option is ranked in separation from the base point as compared to the separation of the metric relative to the base point by asking

How many time the difference of metric option from the base point do you think the option is better/worse than the base point? (Q1a)

This approach is the *better/worse method* of constructing an interval scale (Otto, 1995), and is applied in the first concept selection example in Section□4.

Rather than creating an interval scale, if we have available supplementary information such as performance levels, we can apply this to create a ratio scale. For example, as opposed to using a

Table□2. Interval Measurements for the Material Selection Example.

Objective	Steel 1020	Stainless 304	Aluminum 5052	Copper	Bronze
thickness	√	√	– – –	–	–
Conductivity	√	-0.5	+	3.5+	1.5+
Thermal mass	√	√	– –	–	–
Diffusivity	√	√	+	++	+
Hardness	√	++	–	-1.5	-1.5
Yield	√	11.5+	-2.5	1.5+	+
Machinability	√	–	1.5+	++	++

metric option, we might choose a metric value for machinability as 10 units, where we know the manufacturing cost or cycle time associated with 10 units of machinability. This would apply a ratio scale (using a metric value of 10 units), rather than an interval scale with no units (Otto, 1995). If performance metric values are known, the question can change slightly to

How much value do you assign to p_i, given p_{base} is a reference of zero value and p_{metric} is a reference of value one? (Q1b)

When the difference between the metric and base point is the smallest discernible difference, this corresponds to choosing a *resolution* in performance levels. For the material selection example, we can choose the performance delivered by 1020 steel as the base point, and 5052 aluminum for a metric material, and bronze or stainless steel when aluminum is at an extremum of the range (so we do not have to deal with fractional units). This approach is adopted in the remainder of the materials selection example. Note that the results (proportional differences) do not change with the selection of the base and metric, since the raw scores can always be proportionally scaled.

Whether using an interval scale (Q1a) or a ratio scale (Q1b), a transformation to raw scoring values is the next mapping where errors can enter:

$$r = g(p) \tag{2}$$

where r is the raw value score, p is a performance level from the last mapping, and g is a mapping reflecting the designer's judgment of the relative value of p, given the base point and metric option values. A linear relationship is given by

$$r = \alpha\left(\frac{p - p_{\text{base point}}}{p_{\text{metric}} - p_{\text{base point}}}\right) \qquad \text{where} \quad \alpha = \begin{cases} +1 & \text{metric better than basepoint} \\ -1 & \text{metric worse than basepoint} \end{cases} \tag{3}$$

The results of the raw score mapping are shown in Table□2. This mapping is not always necessarily a linear function. For example, stress levels below the ultimate tensile stress (UTS) may provide a monotonic value relationship with stress, but beyond UTS, the value is zero. Nonetheless, for the linearized error analysis, we will assume this relationship is valid, as it usually proves a sufficiently accurate representation.

We must now combine these raw value measurements across the different objectives into a final rating for each option. Although the rated values for any objective are unitless and proportional, we cannot simply sum across the different objectives since each objective has different unitless ranges. Thus, we must "normalize" or "weight" the objectives to reduce/expand their ranges until each can be added with the others. An approach we have often used effectively in industrial practice is to limit each objective to a maximum impact in the evaluation, which we will refer to as the *maximum impact normalization* (Otto, 1995). Maximum impact limits for the

Table 3. Maximum Allowed Effect for the Material Selection Example.

Objective	Maximum Effect
thickness	+
Conductivity	++
Thermal mass	++
Diffusivity	+
Hardness	+
Yield	0.5+
Machinability	0.5+

Table 4. Normalized ratings for the material selection.

Objective	Steel 1020	Stainless 304	Aluminum 5052	Copper	Bronze
Thickness	√	√	–	-0.5	-0.5
Conductivity	√	-0.5	0.5+	++	+
Thermal mass	√	√	– –	–	–
Diffusivity	√	√	0.5+	+	0.5+
Hardness	√	0.5+	-0.5	-0.5	-0.5
Yield	√	0.5+	√	√	√
Machinability	√	√	0.5+	0.5+	0.5+

material example are shown in Table 3. These limits range from (+1/2) to (++), implying that the objectives are considered to be one to four times as important relative to each other. Ideally, only integral values (+) and (++) would be used. However, sometimes this provides an excessively course resolution, and so half points are required. In theory, any real valued number could be used, determined with a marginal rate of substitution question (Otto, 1995). That is, choose one objective ϕ_1 and define its impact as (+). Then for each other objective ϕ ask

By how much should ϕ be increased to compensate for a loss of one unit in ϕ_1? (Q2)

This produces real numbers for the impact on each objective. Past practical experience, however, suggests starting by rounding to either (+) and (++), and iteratively halving the resolution until satisfactory weightings arise.

These maximum impact values from S, as shown in Table 3, are used to modulate the relative values determined from the better/worse construction. The better/worse method constructs a relative scale among the concept options. To combine the ranks into an overall figure-of-merit, the scale ranges must be modulated so they are "appropriate," i.e., weighted according to the desired impact by the designer or customers. To do this when using the maximum impact normalization, we multiply each value by the maximum allowed effect (impact), and divide by the difference between the highest and lowest score. This introduces another mapping

$$N = h(r, I) \tag{4}$$

where r is the raw score, and I is the maximum allowed impact. A typical linear transformation is

$$h(r, I) = I \frac{r - r_{base}}{r_{max} - r_{min}} \tag{5}$$

where I is the impact, $r_{base} = 0$, r_{min} is the minimum raw score, and r_{max} is the maximum raw score value. The results of this mapping are shown in Table 4. Note that the linearized value approximation (Equation (3)) can lead to some results that might not be intuitive (Table 4). For example, the yield strength of stainless steel is much higher than all other concepts, and so

Table 5. Overall ratings for the material selection.

Objective	Steel 1020	Stainless 304	Aluminum 5052	Copper	Bronze
Thickness	√	√	–	-0.5	-0.5
Conductivity	√	-0.5	0.5+	++	+
Thermal mass	√	√	– –	–	–
Diffusivity	√	√	0.5+	+	0.5+
Hardness	√	0.5+	-0.5	-0.5	-0.5
Yield	√	0.5+	√	√	√
Machinability	√	√	0.5+	0.5+	0.5+
Sum	0.0	0.5	-2.0	1.5	0.0

obscures the differences in all other yield-strength normalized ratings in Table□4. If this relationship is not acceptable, a different value transformation (Equation□(2)) can be applied.

Summing the results in Table□4 produces the overall numerical rating, M_i, for the evaluation:

$$M_i = M(d_i) = \sum_j N_{ij} \tag{6}$$

where i is an index for the number of design concepts. The overall ratings from this calculation are shown in Table□5. The analysis would apparently indicate a selection of copper. But the designer is left wondering whether the uncertainty in each assignment causes sufficient errors to invalidate this result. The next section develops an error analysis to address this issue.

3. Error Analysis

Having clearly represented each step required to complete a numerical concept selection as a mapping, we can assign and propagate "tolerance" (or error) ranges of uncertainty. The final overall ranking for each concept can be represented as the composition of each mapping, representing each operation the designer must complete:

$$\begin{aligned} M_i = M(d_i) &= \sum_j h_{ij}(I_j) \circ g_{ij} \circ f_{ij}(d_i) \\ &= \sum_j h_{ij}(I_j, g_{ij}(f_{ij}(d_i))) \end{aligned} \tag{7}$$

where $\circ$ denotes a composition operator. Observing this relationship, one can incorporate uncertainty information within each mapping, and propagate the uncertainties into the final rating.

Uncertainty Specification

To model uncertainty for each mapping, we propose to use tolerances about the nominal selection values. For each value, these tolerances are represented as intervals. There are three such tolerance specifications required for each of the three mappings in the numerical evaluation. These tolerances are:

- *Delivered Performance Uncertainty* $p \pm \delta p$. This is the uncertainty in the level of performance that the design will deliver, irrespective of how much value the designer interprets in the performance metric value or the importance placed on the performance metric.
- *Value Ranking Uncertainty* $r \pm \delta r$. This is the uncertainty in the magnitude of value the designer assigns to the level of each performance metric as compared to the other levels of the same performance metric. It corresponds to the error associated with the rank of a

Table 6: Performance Metric Value Uncertainty for the Material Selection Example.

Objective	(units)	Steel 1020	Stainless 304	Aluminum 5052	Copper	Bronze
Thickness	inches	±0.01	±0.01	±0.04	±0.02	±0.02
Conductivity	Btu ft/hr °F ft^2	±3	±1	±8	±20	±10
Thermal mass	Btu/°F	±3E-05	±3E-05	±7E-05	±5E-05	±5E-05
Diffusivity	ft^2/hr	±90	±30	±400	±700	±400
Hardness	Brinnell	±10	±10	±5	±0.5	±0.5
Yield	psi	±5000	±5000	±5000	±5000	±5000
Machinability	[0-100]	±5	±5	±5	±5	±5

concept as assigned by a designer, and is independent with respect to other performance metrics.

- *Impact Scaling (Weighting) Uncertainty* $I \pm \delta I$. This is the uncertainty in the level of importance (weight) that the designer prescribes to the performance metric as compared to the other performance metrics.

Consider the first uncertainty in delivered performance. Such uncertainties arise from modeling and measurement errors in the calculations or experiments completed to attain the tabulated values. Table 6 lists representative performance uncertainties for the material selection example. As an example, consider the yield strength data in the table. A tolerance of ±12,000 psi (σ=±4,000 psi) for steel may be indicative of the measurement error of a tensile test or the variance associated with a replicated design of experiments.

The second uncertainty in value is a subjective designer judgment. It reflects how certain a designer is in the value of the performance as compared to the other values of the performance. Thus, these error values must be compiled by examining each performance metric one at a time. Note that we define the relative value of the base point to be zero, the comparison point, and the relative value of the metric point to be one. There is no uncertainty in these two assignments. The uncertainty is in comparing all other levels to these two. Table 7 contains the uncertainties in value for each metric by each material. Notice that for each performance metric, the base point and metric materials have no uncertainty in their relative value.

To estimate the value ranking uncertainties, a number of approaches might be used. A practical approach for specifying this uncertainty is to ask each member of a design team or to ask a focus group of customers (of appropriate sample size) the better/worse question (Q1), and use the ensemble statistics to define the uncertainty. Alternatively, the following question might be tried:

For each interval measurement, what is the uncertainty in the ranking as compared to the difference of the metric from the base point? (Q3)

The results of such questioning can be tabulated, and the value ranking uncertainty can be estimated based on the average. It should be directly apparent that any element in Table 2 that is a base point or a metric value will have zero uncertainty with respect to the interview question (Q1). For example, the question: "what is the uncertainty of the difference of steel from steel relative to the difference of bronze from steel?" will obviously result in the trivial result of a zero tolerance. Table 7 reflects this approach for the material selection example, where each of the base and metric points have a zero tolerance value. The results tabulate the metric points from each member of a design team or from customers through a semantic inquiry how much difference in an objective is significant. In Table 7, for each objective the error values are the same across all the concepts. This need not be the case.

Table 7. Interval Measurement Tolerances for the Material Selection Example.

Objective	Steel 1020	Stainless 304	Aluminum 5052	Copper	Bronze
thickness	±0	±0.5	±0.5	±0.5	±0
Conductivity	±0	±0.5	±0.5	±0	±0.5
Thermal mass	±0	±0.5	±0.5	±0.5	±0
Diffusivity	±0	±0.5	±0.5	±0.5	±0
Hardness	±0	±0.5	±0.5	±0	±0.5
Yield	±0	±0.5	±0.5	±0.5	±0
Machinability	±0	±0	±0.5	±0.5	±0.5

The third uncertainty, impact scaling or weighting, is also a subjective designer judgment. It reflects how certain a designer is in the importance (weighting) of each performance metric as compared to the others. One approach for specifying impact scaling uncertainties is to choose one of the maximum allowed effects of Table 3 and compare all other effects to it (Q2). Importance needs only one of the values defined since there is a well defined zero of "no importance". For the materials selection example, we choose the thickness objective, arbitrarily, as a basis for comparing the importance (weighting) of all other objectives. All other metric maximum effects are defined relative to the standard impact of thickness. Table 8a tabulates the uncertainty in impact for each metric, where thickness is shown to have a zero tolerance since it is used for comparison.

The designer may use an alternative approach to specify the impact scaling uncertainty, in a similar manner to the value ranking uncertainty. For example, each member of a design team, or each member of a group of customers, can be asked to independently list the maximum allowed effects (weights) for a concept selection problem. The average of the tabulated results of this inquiry corresponds to the default maximum allowed effect, in the form of Table 3, and the standard deviation corresponds to a third of the tolerance for the impact scaling uncertainty. As an example, consider the materials concept selection example. Assume that the materials selection problem is part of a broader product design, where the design team is composed of six members. Each team member assigns the maximum allowed effects (weights) of each objective, where there is an agreed upon upper bound. Table 8b shows the results of this process, with the corresponding average effects and impact scaling uncertainty.

Given these specifications of uncertainties, each must be combined into a final overall error using an appropriate algebra to produce an overall uncertainty in the final scores. The next section develops this overall error estimation.

Linearized Uncertainty Characterization

Given that tolerances are specified for each judgment that composes the numerical ratings, the next task is to combine these into an overall error. There are many different and possible schemes to perform this combination. To determine an appropriate mathematics to apply, an interpretation must be assigned to the tolerances previously elicited. What does "$\pm\delta X$" mean? If it implies an uncertainty in what choice a designer should prefer, fuzzy set mathematics can be argued as appropriate. If it implies a random uncertainty, probability mathematics can be argued as appropriate.

We choose to apply a probabilistic interpretation to the uncertainty ranges. This probabilistic approach aligns with historical subjective probability theory (Savage, 1954). Other applicable methods might include fuzzy set mathematics (Zimmerman, 1985; Wood *et al.*, 1992; Wood and Antonsson, 1989) or interval mathematics (Ward and Seering, 1991). Though these other methods might be argued as more appropriate, we choose not to use these approaches due to their current lack of derivable confidence measures in results.

Table 8a. Tolerances in Maximum Allowed Effect for the Material Selection Example.

Objective	Maximum Effect
thickness	±0
Conductivity	±0.5
Thermal mass	±0.5
Diffusivity	±0.5
Hardness	±0.5
Yield	±0.5
Machinability	±0.5

Table 8b. Uncertainty in Maximum Allowed Effect — Design Team Averaging.

Objective	Designer #1	Designer #2	Designer #3	Designer #4	Designer #5	Designer #6	Average	s
thickness	1.00	1.00	1.00	1.00	1.00	1.00	1.0	0.0
Conductivity	1.50	2.00	1.75	2.50	1.75	2.00	1.9	0.3
Thermal mass	1.75	1.75	2.00	2.00	1.50	2.50	1.9	0.3
Diffusivity	1.00	1.50	0.75	1.00	0.75	1.00	1.0	0.3
Hardness	0.50	1.00	0.50	0.50	1.00	1.00	0.8	0.3
Yield	0.25	0.50	0.75	0.50	0.25	0.50	0.5	0.2
Machinability	0.50	0.50	0.75	0.50	0.50	0.25	0.5	0.2

With probability mathematics, a "$\pm\delta X$" implies that the true value of an independent variable X behaves as a normal distribution, centered at X with a standard deviation of σ. The standard deviation relates to the tolerance δX by the following relationship:

$$\delta X = n\sigma, \quad n = 3,\ldots,6 \tag{8}$$

By convention, n is chosen such that $\delta X = 3\sigma$, meaning that a tolerance at any step of the concept selection method is 3σ from the nominal value specified (Bjorke, 1989).

With the subjective probability assumption, the resulting uncertainty range on the final evaluation can be readily calculated through application of probability mathematics. Given an invertible mapping $y = f(x)$ and a random variable X, it can be shown (Sveshnikov, 1968) that the induced distribution for Y is given as

$$pdf_y(y) = \left|\frac{\partial f^{-1}}{\partial y}\right| pdf_x(f^{-1}(y)) \tag{9}$$

and for problems of more than one variable,

$$pdf_y(y) = \frac{d}{dy}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} pdf_x(x_1,\ldots,f_{x_i}^{-1}(y),\ldots,x_n)\,dx_1\ldots,dx_{i-1},dx_{i+1},\ldots,dx_n \tag{10}$$

where $f_{x_i}^{-1}(y)$ returns a value of x_i given y, holding all other values x_j fixed. This forms the basis for propagating the errors into an overall uncertainty range for the final rating.

If we make a first order approximation, it can be shown (Sveshnikov, 1968) that the variances are related by:

$$\sigma^2(y) = \left(\frac{\partial f}{\partial x}\right)^2 \sigma^2(x) \tag{11}$$

Applying this approximation to Equation (7), we have our result:

Table 9a. Uncertainty in Overall Ratings for the Material Selection Example.

Objective	Steel 1020	Stainless 304	Aluminum 5052	Copper	Bronze
Thickness	±0.0	±0.1	±0.1	±0.1	±0.0
Conductivity	±0.0	±0.1	±0.1	±0.2	±0.1
Thermal mass	±0.1	±0.2	±0.4	±0.3	±0.2
Diffusivity	±0.0	±0.1	±0.1	±0.2	±0.1
Hardness	±0.0	±0.1	±0.1	±0.1	±0.1
Yield	±0.0	±0.2	±0.1	±0.0	±0.0
Machinability	±0.0	±0.1	±0.1	±0.1	±0.1
RMS Error (σ)	±0.1	±0.3	±0.5	±0.4	±0.3
Tolerance (3σ)	±0.3	±1.0	±1.4	±1.3	±0.9

$$\sigma_M^2(d) = \sum_j \sigma_{N_j}^2 \qquad (12)$$

where

$$\sigma_r^2 = \left(\frac{\partial r}{\partial p}\right)^2 \left(\frac{\delta p}{3}\right)^2 + \left(\frac{\partial r}{\partial p_{\text{base}}}\right)^2 \left(\frac{\delta p_{\text{base}}}{3}\right)^2 + \left(\frac{\partial r}{\partial p_{\text{metric}}}\right)^2 \left(\frac{\delta p_{\text{metric}}}{3}\right)^2 + \left(\frac{\delta r}{3}\right)^2$$
$$\sigma_N^2 = \left(\frac{\partial N}{\partial I}\right)^2 \left(\frac{\delta I}{3}\right)^2 + \left(\frac{\partial N}{\partial r}\right)^2 \sigma_r^2 + \left(\frac{\partial N}{\partial r_{\text{base}}}\right)^2 (\sigma_{r_{\text{base}}})^2 + \left(\frac{\partial N}{\partial r_{\max}}\right)^2 (\sigma_{r_{\max}})^2 + \left(\frac{\partial N}{\partial r_{\min}}\right)^2 (\sigma_{r_{\min}})^2 \qquad (13)$$

where δx is the tolerance the designer has in prescribing a value x.

Equations (12) and (13) form the basis for a first order error analysis of a numerical concept selection. If we use the mappings (3) and (5), we can apply Equations (12) and (13) to derive an overall error formula. Equation (9) for the linearized mappings (3) and (5) become

$$\sigma_r^2 = \left(\frac{1}{p_{\text{metric}} - p_{\text{base}}}\right)^2 \left(\frac{\delta p}{3}\right)^2 + \left(\frac{p - p_{\text{metric}}}{(p_{\text{metric}} - p_{\text{base}})^2}\right)^2 \left(\frac{\delta p_{\text{base}}}{3}\right)^2 + \left(\frac{p_{\text{base}} - p}{(p_{\text{metric}} - p_{\text{base}})^2}\right)^2 \left(\frac{\delta p_{\text{metric}}}{3}\right)^2 + \left(\frac{\delta r}{3}\right)^2 \qquad (14)$$

for each raw score r_j, and

$$\sigma_N^2 = \left(\frac{r - r_{\text{base}}}{r_{\max} - r_{\min}}\right)^2 \left(\frac{\delta I}{3}\right)^2 + \left(\frac{I}{r_{\max} - r_{\min}}\right)^2 \sigma_r^2 + \left(\frac{-I}{r_{\max} - r_{\min}}\right)^2 (\sigma_{r_{\text{base}}})^2 + \left(\frac{-I(r - r_{\text{base}})}{r_{\max} - r_{\min}}\right)^2 (\sigma_{r_{\max}})^2 + \left(\frac{I(r - r_{\text{base}})}{r_{\max} - r_{\min}}\right)^2 (\sigma_{r_{\min}})^2 \qquad (15)$$

for each normalized score N_j. This approach is applied to the material selection problem, and the results are shown in Table 9a. Note that the approach remains valid even without the linearity assumption; however, the calculations become more difficult. Validity of the uncertainty characterization is elaborated in the higher-order uncertainty characterization section below.

Examining Table 9a, it becomes somewhat less clear which concept should be chosen. The copper rating is 1.5 ± 1.3, and the stainless steel rating is 0.5 ± 1.0. To within 3 standard deviations, we cannot be assured of the copper recommendation.

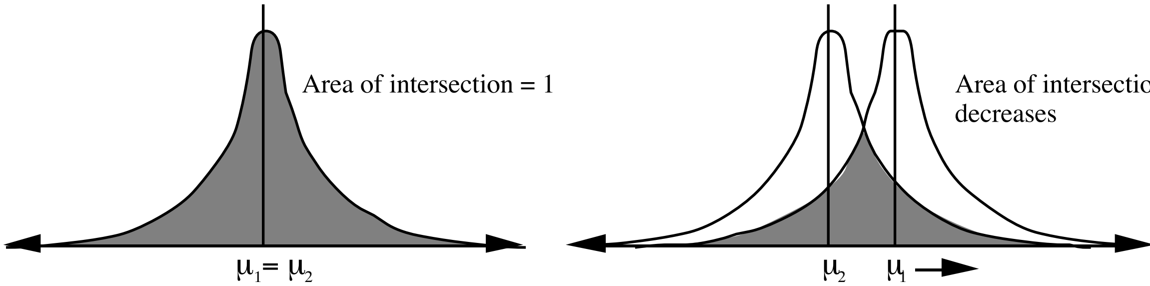

Figure 2. The confidence against is equal to the intersection of the density functions.

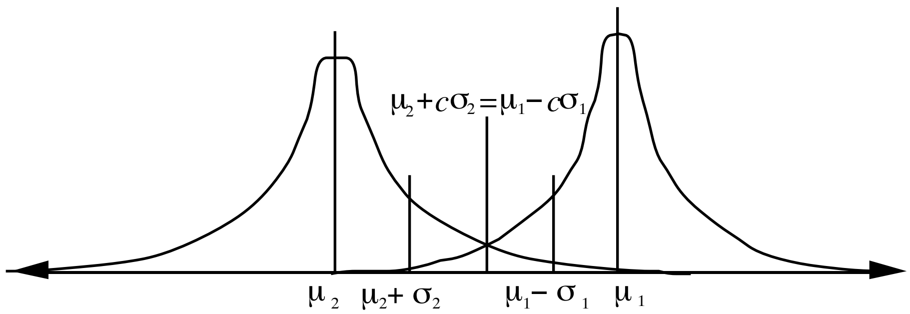

Figure 3. Definition of c.

Confidence Levels

The previous decision which seemed clear perhaps now is not. A quantification of the decision certainty would be helpful. What must be ascertained is if the maximum sum is indeed larger then the other concepts, and the next highest in particular.

In the proposed error analysis methodology, the nominal concept selection ratings (e.g., Table 5) represent the average of a normal probability distribution. Denote the maximum rating μ_1 and the next highest rating μ_2. These are the means of two normal distributions X_1 and X_2. The posed question is then whether $\mu_1 > \mu_2$, a well formed statistical hypothesis.

To formulate the solution to this hypothesis, consider the intersection of the two distributions X_1 and X_2. When X_1 and X_2 are identical distributions, our hypothesis fails, $\mu_1 = \mu_2$. We have no confidence that $\mu_1 > \mu_2$. The probability that our hypothesis is false is one, as happens to be the intersection of the two probability distributions X_1 and X_2. Now as μ_1 moves away from μ_2, our confidence in the statement that $\mu_1 > \mu_2$ increases, and the area of intersection of X_1 and X_2 decreases. Thus, the area of intersection of the distributions is a measure of confidence against our hypothesis. This is shown in Figure 2.

To derive the equation for the confidence level, consider a variable c which uniformly scales both σ_1 and σ_2 until $\mu_1 - c\sigma_1 = \mu_2 + c\sigma_2$. The equation for c is thus

$$c = \frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2} \tag{16}$$

This value of c is the value required to stretch the two normal distributions until their areas of intersection are one standard deviation. This is shown in Figure 3. Note that c is always positive.

Now when $c = 1$, the distributions are exactly one standard deviation apart each. Then the intersection of the density functions is twice of one minus the cumulative density function. That is,

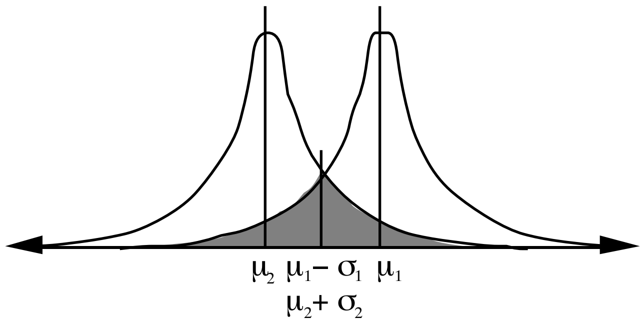

Figure 4. The confidence against is equal to the intersection of the density functions.

Table 9b. Confidence Against Each Material for the Material Selection Example.

Objective	Steel 1020	Stainless 304	Aluminum 5052	Copper	Bronze
c value $(c\sigma)$	2.80	1.27	3.88	N/A	2.03
Confidence	0.99	0.80	1.00	N/A	0.96

$$\text{Area of intersection} = 2(1 - \text{Cu}(c)) \tag{17}$$

where

$$\text{Cu}(c) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{c} e^{\frac{(c-\mu)^2}{2\sigma^2}} dc$$

as shown in Figure 4.

Equation (17) represents the solution to calculating a degree of belief in the concept selection chart solution. We want the probability that the means are not the same, or one minus the above probability. Therefore,

$$\Pr(\text{decision}) = 2\text{Cu}(c) - 1 \tag{18}$$

This equation represents a confidence level in the decision, given the specified error estimates. Note that this derivation is analogous to a statistical t-test which gives the confidence that the means of two sample distributions are the same (Montgomery, 1991). If the input uncertainties were characterized sampled data and this data were propagated through Equation (7), the output sample distributions would be compared with a t-test, not with Equations (17-18), which are derived with normal distributions. The t-test results approach Equations (17-18) as the sample sizes become large and normal. The c value, therefore, represents the number of standard deviations of error and will be called the *confidence factor*.

For the material selection example, we apply Equation (16) to the copper and stainless steel evaluation values and errors. This results in $c = 1.27$, or 1.27σ error. Then applying Equation (18) results in $\Pr = 0.80$. That is, we are about 80% confident that the choice for copper over stainless steel is correct, based upon our known errors and a first order model. Table 9b lists the confidence factor c values and the confidence level against each of the non-selected materials, as compared to the recommended Copper material.

Higher-Order Uncertainty Characterization

Before presenting more detailed design examples of concept selection with error analysis, it is important to validate the linearized approach presented above. A higher order, yet more computationally expensive, error propagation approach is developed in this section to perform this validation. This approach does not assume a first order approximation to the probability calculations.

As stated above, three types of uncertainty are considered for the concept selection method. These uncertainties manifest in the mappings given by Equations□(1-7). To estimate the propagation of errors, each of the mappings may be represented as a composition of binary operations of two probabilistic variables, x_1 and x_2:

$$y = \not{f}(x_1, x_2) \tag{19}$$

where $\not{f}()$ is a binary operation, e.g., addition, subtraction, multiplication, and division. The function $\not{f}()$ combines the probabilistic variables, where y is the probabilistic result of the binary operation. As an example, Equation (5) includes four binary operations: a subtraction between r and r_{base}; a subtraction between $r_{\max}$ and $r_{\min}$; a division between the results of the subtractions; and a multiplication by I. Based on axiomatic probability theory (Wood, Antonsson, and Beck, 1990), it is possible to develop propagation relations for the probability density functions of binary operations. These relations are summarized below:

$$\begin{aligned}
pdf_{\text{add}}(y) &= \int_{-\infty}^{\infty} pdf_{x_1}(y - x_2) \cdot pdf_{x_2}(x_2)\,dx_2 \\
pdf_{\text{sub}}(y) &= \int_{-\infty}^{\infty} pdf_{x_1}(y + x_2) \cdot pdf_{x_2}(x_2)\,dx_2 \\
pdf_{\text{mul}}(y) &= \int_{-\infty}^{\infty} \frac{1}{x_2}\, pdf_{x_1}\left(\frac{y}{x_2}\right) \cdot pdf_{x_2}(x_2)\,dx_2 \\
pdf_{\text{div}}(y) &= \int_{-\infty}^{\infty} pdf_{x_1} x_2 (y \cdot x_2) \cdot pdf_{x_2}(x_2)\,dx_2
\end{aligned} \tag{20}$$

The relations given by Equation□(20) may now be applied to each operation in the composition mapping, Equation□(7). The probability density functions for the uncertainty variables are obtained from the tolerances δp, δr, and δI. Applying Equation□(20) to Equation□(3) gives

$$\not{p}\!f_r = \int_{p_{m,b}} \int_{p_b} \not{p}\!f_{m,b} \cdot \not{p}\!f_{p_\alpha}\left(r \cdot p_{m,b} + p_{b,\alpha}\right) \cdot \not{p}\!f_{p_{b,\alpha}}\left(p_{b,\alpha}\right) \cdot \not{p}\!f_{p_{m,b}}\left(p_{m,b}\right)\; dp_{b,\alpha}\; dp_{m,b} \tag{21}$$

where

$$\not{p}\!f \equiv pdf \qquad p_m \equiv p_{\text{metric}}$$

$$p_b \equiv p_{\text{base}} \qquad p_{b,\alpha} = \alpha \cdot p_b = \alpha \cdot p_{\text{base}}$$

$$p_\alpha = \alpha \cdot p \qquad p_{m,b} = p_m - p_b$$

Similarly, applying Equation□(20) to Equations□(4-5), we have

$$\begin{aligned}
\not{p}\!f_N = \int_I \int_{x_2} \int_{\underline{r}} \int_{r_b} \frac{x_2}{I} \cdot \not{p}\!f_R\left(\frac{N \cdot x_2}{I} + r_b\right) \cdot \not{p}\!f_{r_b}\left(r_b\right) \cdot \not{p}\!f_{\underline{r}}\left(x_2 + \underline{r}\right) \cdot \\
\not{p}\!f_{\underline{r}}\left(\underline{r}\right) \cdot \not{p}\!f_I\left(I\right)\; dr_b\; d\underline{r}\; dx_2\; dI
\end{aligned} \tag{22}$$

where

Table 10a. Higher-Order Uncertainty Analysis: Overall Ratings for the Material Selection Example.

Objective	Steel 1020	Stainless 304	Aluminum 5052	Copper	Bronze
Thickness	±0.023	±0.060	±0.150	±0.077	±0.052
Conductivity	±0.015	±0.070	±0.084	±0.154	±0.100
Thermal mass	±0.088	±0.195	±0.388	±0.273	±0.201
Diffusivity	±0.011	±0.089	±0.134	±0.227	±0.105
Hardness	±0.026	±0.103	±0.066	±0.079	±0.090
Yield	±0.021	±0.269	±0.061	±0.040	±0.032
Machinability	±0.024	±0.071	±0.087	±0.114	±0.114
RMS Error (σ)	±0.100	±0.383	±0.462	±0.420	±0.293
Tolerance (3σ)	±0.304	±1.14	±1.39	±1.26	±0.880

Table 10b. Higher-Order Uncertainty Analysis: Confidence Levels for the Material Selection Example.

Objective	Steel 1020	Stainless 304	Aluminum 5052	Copper	Bronze
Confidence	0.998	0.801	1.00	N/A	0.967

$$r_b \equiv r_{\text{base}} \qquad \bar{r} \equiv r_{\max}$$

$$\underline{r} \equiv r_{\min} \qquad x_2 = \bar{r} - \underline{r}$$

$$R = r + \delta r$$

and

$$\wp_R = \wp_{r+\delta r}$$

Using the results of Equations (21-22) and applying Equation (20) to Equation (6), the overall uncertainty characterization is given by

$$\wp_{M_i} = \int_{N_{i,L}} \cdots \int_{N_{i,2}} \wp_{N_{i,1}}\left(M_i - N_{i,2} - \cdots - N_{i,L}\right) \cdot \wp_{N_{i,2}}\left(N_{i,2}\right) \cdots \wp_{N_{i,L}}\left(N_{i,L}\right)\ dN_{i,2} \cdots dN_{i,L} \tag{23}$$

where

$$j = 1, \ldots, L \quad \text{in (6).}$$

A numerical scheme (Wood, Antonsson, and Beck, 1990) may be used to solve Equations (21-23), based on the tolerances given in Tables 6-8a. Tables 10a and 10b show the results of the higher-order characterization. A number of observations are apparent from this analysis. For example, the non-linear operations, i.e., multiplication and divisions in Equations (3) and (5), result in skewed distributions and deviations in the means when propagated with Equations (21-22). While the distributions are skewed, the standard deviations of each objective for each concept (Table 10a) are surprisingly close to the linear approximations in Table 9a. The important question, then, is how the skewed distributions and mean deviations will propagate to the overall concept ratings, based on Equation (6). The bottom of Table 10a clearly shows that the overall uncertainties are not affected significantly by the higher-order terms. In fact, the confidence limits calculated from the higher-order distributions (Table 10b) deviate by less than 1% compared to the linear approximations. The reason for these small deviations may be explained from the Central Limit Theorem. Because Equation (6) includes only addition operations, the resulting probability density functions will quickly approach normal distributions, minimizing the skewing and other non-linear effects of the error propagations. We conclude from

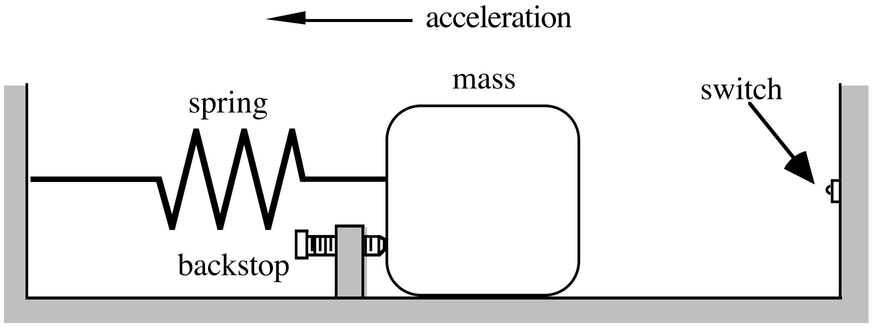

Figure 5: Mass-spring accelerometer switch.

this study that the linear-uncertainty characterization is suitable for propagating and investigating errors during concept selection.

4. Concept Selection Examples

In this section, our concept selection method, including error estimation, is applied to two conceptual design cases. The first case focuses on the design of an accelerometer switch, where the overall objective is to choose a switch configuration that satisfies actuation time and manufacturing requirements. This case also focuses on design decision making when no extensive measurements exist for the design objectives. In the second case study, the purpose is to design a bilge pump consumer product. Four concepts are provided, including engineering modeling results for the design objectives. This second case extends the concept selection method into a form similar to that of classical decision matrices (Pahl and Beitz, 1984), with the error estimation of Section 3 included.

Case I: Design of an Accelerometer Switch

As a preliminary concept selection example, consider the design of an accelerometer switch. The switch uses a mass that pulls against a mechanical spring. With sufficient acceleration, the spring stretches and allows the mass to actuate a switch, as illustrated in Figure 5. The performance objective of the switch designers is to ensure a precise actuation time under a specified acceleration.

This design has a manufacturing problem. It proves difficult to ensure a precise actuation time due to assembly errors. In particular, it proves difficult to make the spring constant precise, due to the positioning and pre-loading operations. Suppose the designer has determined possible actions or operations to remedy the situation on the factory floor. These possible operations may include:

- Improve the assembly operation so that a precise spring constant is achieved on every accelerometer,
- Adjust the mass of each accelerometer so that the spring errors are accounted for,
- Adjust the backstop position of each accelerometer so that the spring errors are accounted for,
- Do nothing and inspect out the bad parts.

These actions can be condensed into a set of labels representing each of the concepts as envisioned by the designer:

$$D = \{\text{Nothing, Mass, Spring, Length}\}$$

The objectives that the designer could use to make an evaluation over these options may be:

Table 11. Interval Measurement Chart of Product Options.

Objective	Do Nothing	Adjust Length	Improve Spring	Adjust Mass
Timing	√	+	++	+
Quality	√	+	+	+
Production Cost	√	–	––––	––
Complexity	√	–	–––	–
Cleanliness	√	–	–	––

Table 12. Objective Maximum Allowed Effect.

Objective	Maximum Effect
Timing	++
Quality	++
Production Cost	++
Complexity	+
Cleanliness	+

Table 13. Scaled Evaluation of Product Options.

Objective	Do Nothing	Adjust Length	Improve Spring	Adjust Mass
Timing	√	+	++	+
Quality	√	++	++	++
Production Cost	√	-0.5	––	–
Complexity	√	-0.5	–	-0.5
Cleanliness	√	-0.5	-0.5	–
Sum (+)	0	3	4	3
Sum (–)	0	-1.5	-3.5	-2.5
Sum	0	1.5	0.5	0.5

- The ability of the operation to ensure the timing performance of the product,
- The cleanliness of the operation, as contamination can affect the product,
- The anticipated product quality,
- The projected cost of the operation,
- The simplicity of the operation.

There are therefore multiple objectives in this evaluation:

$$\Phi = \{\text{Timing}, \text{Cleanliness}, \text{Cost}, \text{Simplicity}, \text{Quality}\}$$

Suppose the designer must evaluate these alternatives and make a decision. In this concept selection, we have no real means to predict the return values of the performance metrics; it is too early in the design process. Indeed, some of the metrics have no units. In such situations, the performance metric values (Equation (1)) are not calculated or empirically derived; instead, the designer's judgment of value (Equation (2)) is used directly. These are shown in Table 11, where the current "do nothing" situation is the base point, and adjusting the length is used as a metric.

The maximum allowed effect of each performance metric must be specified. These are shown in Table 12 for the accelerometer example.

Next, the normalized ratings are calculated by applying Equations (4-6), as shown below in Table 13. The results show that the "Adjust Length" concept is apparently reasonably superior to the other concepts, and should be pursued.

Table 14. Interval Measurement Errors for Accelerometer Options.

Objective	Do Nothing	Adjust Length	Improve Spring	Adjust Mass
Timing	±0	±0	±0.5	±0.5
Quality	±0	±0	±0.5	±0.5
Production Cost	±0	±0	±0.5	±0.5
Complexity	±0	±0	±0.5	±0.5
Cleanliness	±0	±0	±0.5	±0.5

Table 15. Objective Maximum Allowed Effect.

Objective	Max Effect Error
Timing	±0
Quality	±1.0
Production Cost	±1.0
Complexity	±1.0
Cleanliness	±1.0

Table 16a. Errors for Accelerometer Options.

Objective	Do Nothing	Adjust Length	Improve Spring	Adjust Mass
Sum	0.0	1.5	0.5	0.5
RMS Error (σ)	±0.0	±0.4	±0.7	±0.7
Tolerance (3σ)	±0.0	±1.2	±2.2	±2.0

Table 16b. Confidence against choosing each concept for the accelerometer.

Objective	Do Nothing	Adjust Length	Improve Spring	Adjust Mass
c value ($c\sigma$)	3.66	N/A	0.87	0.94
Confidence	1.00	N/A	0.61	0.65

Having this concept selection exercise completed, we can apply the same error analysis methodology to this preliminary evaluation. First, we must specify tolerances in the value assignments, denoted as value ranking uncertainties. These tolerances are shown in Table 14. Notice in the table that the “Do Nothing” and “Adjust Length” concepts have no uncertainty since all other rankings for the “Improve Spring” and “Adjust Mass” concepts are compared to the base and metric points.

Next, we must specify tolerances in the maximum allowed effects (weights). These tolerances are shown in Table 15, where the “timing” objective is chosen arbitrarily as a basis for comparing the relative effects or weights (Table 12) for quality, production cost, complexity, and cleanliness.

Having these tolerances specified, Equation (8) can be applied to derive the errors in each final evaluation result. These are shown below in Table 16a.

From this analysis, it becomes less clear if we would make the right decision to pursue the length adjustment concept. Quantifying this, Equation (16) results in a value of 0.87, or 0.87σ error. This corresponds through Equation (18) to a confidence level of 61%. This is better information for the designer to consider, rather than using the differences in the sums as an informal measure of confidence. A 60% confidence level may be sufficient for the designer(s) to proceed with the single “Adjust Length” concept. This confidence level, on the other hand, may be too low for the designer(s), leading them to invest more resources to develop performance data for

Equation (1), perhaps by pursuing several of the options. Overall, the concept selection method has systematically guided the designers to ask critical performance and manufacturing questions about the design concepts, and quantify both judgments and uncertainties relative to these questions.

Case II: Design of Bilge Pump Consumer Product

As a second concept selection example, consider the design of a bilge pump consumer product, as motivated by (Hubka, 1988). The engineering problem here is to design a device to remove water (due to leakage, rain, or waves) from the bilges of small pleasure craft. These pleasure craft, or boats, are usually docked in marinas for lengthy periods of time, without humans present. An important customer demand for this design problem is that only natural energy sources be used to power the bilge-water removal device.

Based on problem clarification, quality function deployment, functional decomposition, and solution principle generation (Pahl, 1984, Ulrich and Eppinger, 1994, Hubka, 1982), four concepts are under consideration for the design problem (Wood, 1994). These concepts are illustrated in Figure 5. For representation purposes, the following set of labels is used to describe the design concepts, in the order shown in Figure 5:

$$D = \{\text{Mooring Line Pump}, \text{Wind Pump}, \text{Pendulum Pump}, \text{Float Pump}\}$$

Two levels of customer objectives are used by the design team to evaluate the concepts:

(A) Binary feasibility with respect to the design objectives (demands).

- (i) Minimum flow rate of 8 l/hr.
- (ii) Sufficient energy to drive the pump, from the given natural energy choice.
- (iii) Sufficient force to drive the pump variants (as applicable).
- (iv) Assure high reliability (quantified with number of components).
- (v) Minimize deformation of boat (including permanent deformation and aesthetics).

(B) Quantitative concept evaluation of feasible concepts with respect to the customer demands and wishes.

- (i) Ergonomics and Manufacturability — mass (kg), storage volume/greatest linear dimension (m), and assembly (steps/time)
- (ii) Cost — materials ($) and manufacturing ($).
- (iii) Available energy, implying improved flow rate beyond minimum.

For context and as a precursor to concept selection, we first concentrate on the feasibility of the concepts. Two primary technical issues are critical in this study, i.e., the “energy” needed to overcome the bilge pressure head, and the satisfaction of the flow rate specification. Based on the five objectives in category (A), engineering models of the four concepts show that the wind pump (concept 2) is not feasible. For nominal wind conditions, the wind cups are too large and the rotary pump too small for practical consideration. The other three concepts are feasible, and should be carried forward to concept selection. It should be noted that the pendulum pump (concept 3) is carried forward on a marginal basis despite the high number of components and required boat deformation due to attachments to the bilge.

The category (B) objectives may now be considered. Three global objectives exist, with a total of six metrics as subcategories:

$$\Phi = \{\text{Mass}, \text{Volume}, \text{Assembly}, \text{Materials Cost}, \text{Mfg. Cost}, \text{Energy}\}$$

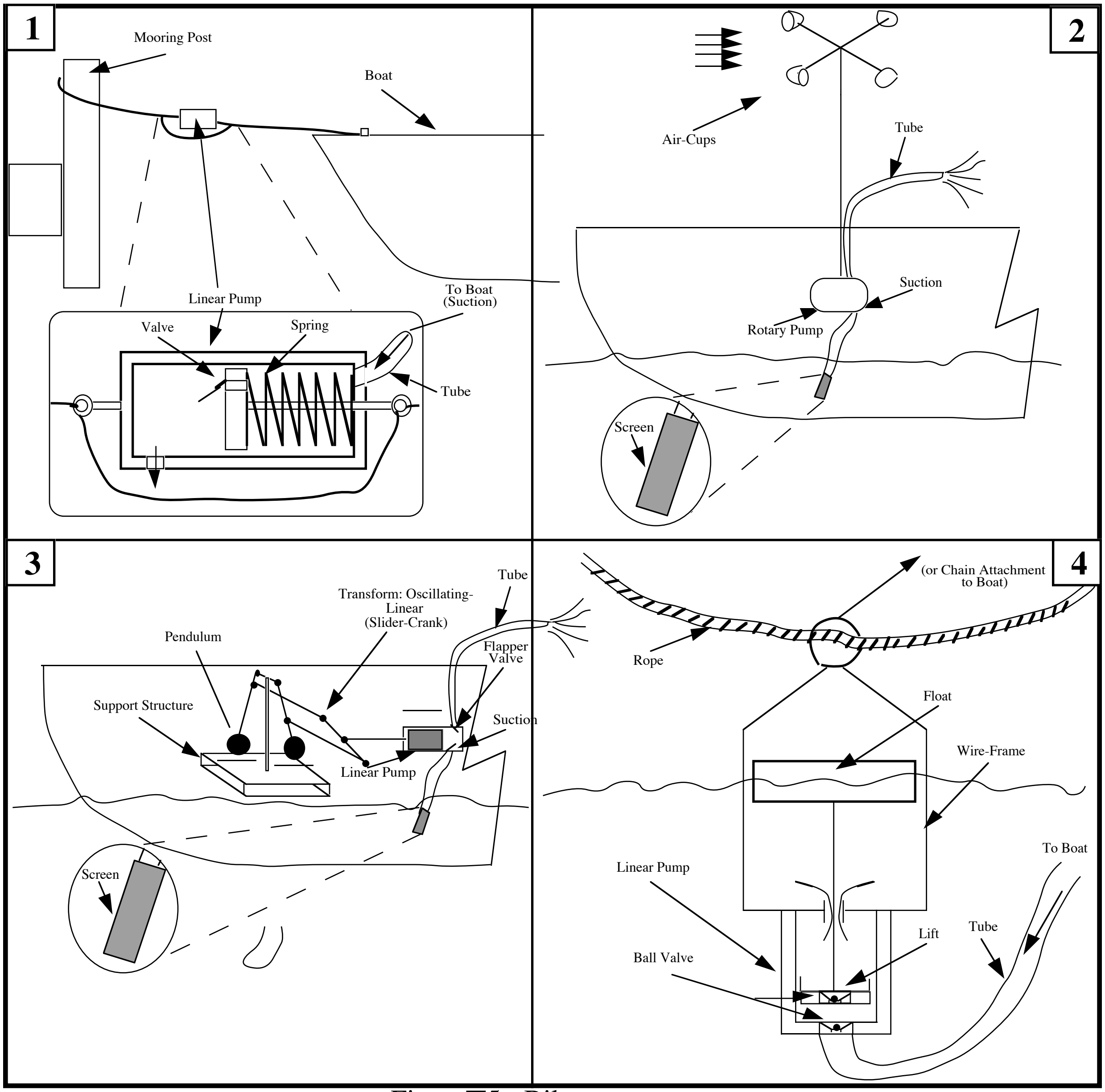

Figure 5. Bilge pump concepts.

The design team must evaluate the concepts with respective to these objectives. In this concept selection, we have engineering models to estimate the performance metrics. Table 17 lists the performance values for the metrics. Volume is measured by the greatest linear dimension, and energy is measured by the volumetric flow rate.

Given the design concepts and metric values of Table 17, we now define a metric scale for rating each concept with respect to each objective. Instead of using an interval scale as in previous examples, a ratio scale is defined for the bilge pump selection problem. By defining a ratio scale, we may transform our concept selection method into a form similar to, but more complete than, a common decision matrix approach (Pahl, 1984). The ratio scale to be used here is given by

$$S = (0, 5, \ldots, 95, 100)$$

Table 17. Metric Values for the Bilge Pump Concepts.

Objective	(units)	Mooring-Line Pump	Pendulum Pump	Float Pump
Mass	kg	3.5	15.5	5.0
Volume	m	0.25	0.50	0.60
Assembly	steps/time	5	13	4
Mfg. Cost	$	18	36	15
Material Cost	$	17	41	20
Energy	l/hr	80	50	45

Table 18. Base and Metric Performance Values for Bilge Pump Objectives

Objective	(units)	Base Value	Metric Value
Mass	kg	2.5	5
Volume	m	0.15	0.25
Assembly	steps/time	0	1
Mfg. Cost	$	15	20
Material Cost	$	15	20
Energy	l/hr	80	72

Table 19. Measurement Chart of Bilge Pump Concepts.

Objective	Mooring-Line Pump	Pendulum Pump	Float Pump
Mass	98	74	95
Volume	95	82.5	77.5
Assembly	75	35	80
Mfg. Cost	97	79	100
Material Cost	98	74	95
Energy	100	81	78

where "$S_{\text{base}} = 100$" is the reference, or base value, corresponding to a design concept of "complete customer satisfaction", and "$S_{\text{metric}} = 5$" is a metric change in the scale. Using this scale (set of ranks), we may define performance base and metric values. These values are known *a priori* from customer and design team interviews, as opposed to an interval scale where one of the design concepts is chosen as a reference design. Table 18 lists the base and metric values for each objective. A metric change from the base value indicates a negative value ($\alpha = -1$), since a ratio scale is used with a base value rank of 100.

Using the performance values, base values, and metric values, the relative measurements of the bilge pump concepts, per design objective, may be determined from (24) below, the ratio scale analogy to Equation (3). Table 19 shows the results of this mapping. Notice in (24) that the base and metric values of the measurement scale have been included since they do not equal zero and one, respectively, as defined in (3).

$$r = S_{\text{metric}} \cdot \alpha \left(\frac{p - p_{\text{base point}}}{p_{\text{metric}} - p_{\text{base point}}} \right) + S_{\text{base}} \qquad (24)$$

The maximum allowed effect, or weighting, of each performance metric must be specified. These are shown in Table 20 for the bilge pump example, where the sum of the weights must equal 100 (analogous to a decision matrix approach).

Table 20. Objective Weights.

Objective	Weight
Mass	15
Volume	10
Assembly	15
Mfg. Cost	30
Material Cost	20
Energy	10

Table 21. Weighted Evaluation of Bilge Pump Concepts.

Objective	Mooring-Line Pump	Pendulum Pump	Float Pump
Mass	14.7	11.1	14.3
Volume	9.5	8.3	7.8
Assembly	11.3	5.3	12.0
Mfg. Cost	29.1	23.7	30
Materials Cost	19.6	14.8	19.0
Energy	10.0	8.1	7.8
Sum	94.2	71.3	90.9

Table 22. Delivered Performance Tolerances for the Bilge Pump Concepts.

Objective	(units)	Mooring-Line Pump	Pendulum Pump	Float Pump
Mass	kg	±1.0	±1.5	±1.0
Volume	m	±0.025	±0.05	±0.05
Assembly	steps/time	±1	±2	±1
Mfg. Cost	$	±2.5	±2.5	±2.5
Material Cost	$	±2.5	±2.5	±2.5
Energy	l/hr	±8	±5	±5

Next, the weighted ratings are calculated using the following mapping, the ratio scale analogy to Equation (5):

$$N_{i,j} = I_j \frac{r_{i,j}}{r_{\text{base},j}} \tag{25}$$

where $r_{\text{base}} = 100$. The results of this mapping are summed according to Equation (6), as shown in Table 21. The results show that the "Mooring-Line Pump" concept is superior to the other concepts; however, the "Float Pump" is within one metric rank. Error estimation is important to determine the actual distinction between the bilge pump concepts.

The delivered performance tolerances are the first error source to consider. Table 22 lists these tolerances for the bilge pump concepts, based on predicted modeling errors for each objective.

Next, we must consider the tolerances in the performance base points, performance metric points, value rankings, and weightings. Table 23 shows the p_{base} and p_{metric} tolerances, based on customer uncertainties. The value ranking uncertainties, however, are taken as zero for this design problem. It is assumed that the design team is satisfied with the linear mapping of Equation (24),

Table 23. Base and Metric Value Performance Tolerances for the Bilge Pump Concepts.

Objective	(units)	δp_{base}	δp_{metric}
Mass	kg	±1.0	±1.0
Volume	m	±0.05	±0.05
Assembly	steps/time	±0	±1
Mfg. Cost	$	±2.5	±2.5
Material Cost	$	±2.5	±2.5
Energy	l/hr	±8	±8

Table 24. Weighting Tolerances.

Objective	Tolerance
Mass	±5
Volume	±5
Assembly	±5
Mfg. Cost	±5
Materials Cost	±5
Energy	±5

Table 25. Errors (Resolution) for Bilge Pump Concepts.

	Mooring-Line Pump	Pendulum Pump	Float Pump
Sum	94.2	71.3	90.9
RMS Error (σ)	±4.1	±4.9	±3.9
Tolerance (3σ)	±12.2	±14.7	±11.7
c value ($c\sigma$)	N/A	2.56	0.42
Confidence	N/A	99%	33%

implying no predicted error for the raw scores. Finally, the weighting tolerances are predicted for the nominal weights (Table 20) . These tolerances are listed in Table 24.

Having these tolerances specified, Equation (8) can be applied to derive the errors in each final evaluation result. These are shown below in Table 25.

Based on the concept-selection and tolerance analysis, it becomes clear that the mooring- line and float pump concepts are believed superior to the pendulum pump, within a confidence of 99%. There is only a confidence level of only 33%, though, for the mooring-line pump compared to the float pump. The designer(s) are now faced with accepting the 33% confidence, and corresponding uncertainty, and proceeding with the mooring-line pump. Alternatively, the error estimation clearly indicates that a very small distinction or no distinction exists between the mooring-line and float pump concepts, for the objectives considered. Due to this result, it seems more reasonable that the designer(s) would pursue one of three possible actions: (1) add more objectives (metrics) to the concept selection analysis based on the voice of the customer; (2) interview the customers in a focus group to compare the designs with a semantic inquiry; or (3) carry both the mooring-line and float-pump concepts forward to the first phases of embodiment design, proof-of-concept modeling, and alpha prototyping to determine more of a distinction.

5. Conclusion

A simple means to determine an error range for the final score in a numerical concept evaluation is proposed and developed. A designer assigns error ranges to the individual ratings for each concept on each metric, and these are propagated though the mathematics of subjective probability, linearized to first order, to derive an overall uncertainty range in the final results. This range can be compared with the differences in overall rating among the concepts. Given this, a confidence measure for theoretical normal distributions is derived, identical to the statistical t-test used in sampled distributions. The measure is used to quantify the confidence level in the decision results.

The uncertainties which can be modeled in the method include subjective decision uncertainty (what is preferred), model uncertainty (what will happen), and measurement errors. Remaining uncertainty not covered in the error analysis includes structural error in the evaluation mechanics. That is, linear mappings and a weighted value sum are used as the means to guide the selection. Perhaps methods other than linear mappings (e.g., nonlinear mappings such as a logarithmic one or a non-linear pointwise mapping derived with (Q1), the better/worse value construction), and combination metrics other than addition might be more appropriate for a particular decision. This issue remains an active research issue and is beyond the scope of the paper. It is also beyond the scope of this paper to determine what is the appropriate (minimum and maximum) number of objectives and concepts to be used in a concept selection matrix. The error propagation approach developed in this paper should greatly aid in answering these research questions.

A fully measurable material selection problem and two preliminary concept selection problems are presented, all with different techniques of determining raw value. In the measurable material selection case, performance metric values were known, but an interval construction was still used. In the unmeasurable accelerometer case, the performance metric map could not be constructed, and so the raw value is specified directly by the designer using an interval scale. In the last concept selection case, the performance values are again known, and so a ratio scale is implemented, thereby introducing minor changes to the concept selection method but no changes to error propagation approach. It is shown that other concept selection methods, such as decision matrices, may be easily adapted to our approach, without modifying the approach to error estimation.

Overall, the concept selection method proposed in this paper is robust and easily implemented, either by hand or in a spreadsheet. A spreadsheet template is freely available from the authors, in addition to the examples presented in this paper.

Acknowledgments

The research reported in this document was made possible in part by a Sloan initialization grant from the Department of Mechanical Engineering at the Massachusetts Institute of Technology, funded in part by the DuPont Corporation. The authors also wish to acknowledge the support of the Leaders for Manufacturing Program, a collaboration between MIT and U.S. industry, and an NSF Young Investigator Grant. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the sponsors. In addition to the sponsors, the authors wish to extend their gratitude to Prof. Ilene Busch-Vishniac at UT Austin for her probing questions and enthusiasm for this research.

References

Bascaran, E, Bannerot, R. and F. Mistree. "Hierarchical Selection Decision Support Problems in Conceptual Design," *Engineering Optimization*, 14:207-238, 1989.

Bjorke, O., *Computer-Aided Tolerancing*, ASME Press, New York, 1989.

Breyfogle, F.W., *Statistical Methods for Testing, Development, and Manufacturing*, Wiley, New York, 1992.

French, S. *Decision Theory: An Introduction to the Mathematics of Rationality*, Ellis Horwood, Chichester, 1988.

Hauser, J.R. and Clausing, D., "The House of Quality," *Harvard Business Review*, May-June, pp. 63-73, 1988.

Hubka, V., *Principles of Engineering Design*, Butterworths, London, 1982.

Hubka, V., Andreasen, M., and W. E. Ernst, *Practical Studies in Systematic Design*, Butterworths, London, 1988.

Krantz, D., Luce, R., Suppes, P. and A. Tversky, *Foundations of Measurement*, Volume I, Academic Press, New York, 1971.

Montgomery, D. *Design and Analysis of Experiments*, Wiley, 1991.

Otto, K. and E. Antonsson. "Modeling Imprecision for Product Design," in *Proceedings of the 1994 FUZZ-IEEE Conference*, Orlando FL, 1994, pages 340-351.

Otto, K. and E. Antonsson. "The Method of Imprecision Compared to Utility Theory for Design Selection Problems," in *Proceedings of the 1993 ASME Design Theory and Methodology Conference, DE-Vol. 31*, Albuquerque, NM, 1993, pages 167-173.

Otto, K. and E. Antonsson. Trade-Off Strategies in Engineering Design. *Research in Engineering Design*, 3(2):87-104, 1991.

Otto, K. Measurement Methods for Product Evaluation. *Research in Engineering Design*, 1995, to appear.

Pahl, G. and W. Beitz, *Engineering Design: A Systematic Approach*, Springer Verlag, 1984.

Pugh, S. *Total Design*. Addison-Wesley, New York, 1990.

Raiffa, H. and R. Keeney. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Wiley, New York, 1976.

Saaty, T. *The Analytic Hierarchy Process*. McGraw Hill, New York, 1980.

Savage, J. *The Foundations of Statistics*, Wiley, New York, 1954.

Siddall, J. *Probabilistic Engineering Design; Principles and Applications*. Marcel Dekker, New York, 1983.

Sveshnikov, A. *Problems in Probability Theory, Mathematical Statistics, and the Theory of Random Functions*. Dover Publications, New York, 1968.

Thurston, D. "A Formal Method for Subjective Design Evaluation with Multiple Attributes," *Research in Engineering Design*, 3(2):105-122, 1991.

Ullman, D., *The Mechanical Design Process*, McGraw-Hill, NY, 1992.

Ulrich, K. and S. Eppinger. *Product Design and Development*. McGraw-Hill, 1995.

Ward, A. and W. Seering. "Quantitative Inference in a Mechanical Design Compiler." *Artificial Intelligence in Engineering Design and Manufacturing*, 4(1):47-51, 1990.

Wood, K, and E. Antonsson, "Computations with Imprecise Parameters in Engineering Design: Background and Theory," ASME *Journal of Mechanisms, Transmissions, and Automation in Design*, 111(4):616-625, Dec. 1989.

Wood, K, Antonsson, E.K., and J. Beck, "Representing Imprecision in Engineering Design - Comparing Fuzzy and Probability Calculus." *Journal of Research in Engineering Design*, Springer-Verlag, 1(3/4):187-203, 1990.

Wood, K., Otto, K. and E. Antonsson. "Engineering Design Calculations with Fuzzy Parameters." *Fuzzy Sets and Systems*, 52(1):1-20, 1992.

Zimmerman, H. *Fuzzy Set Theory and Its Applications*, Kluwer-Nijhoff, 1985.