ASSESSING OIL SPILL RISK IN PORT TANKER OPERATIONS USING A MULTIATTRIBUTE UTILITY APPROACH TO RANKING AND SELECTION

John C. Butler

Jason R. W. Merrick

Center for Energy Management and Innovation
McCombs School of Business
The University of Texas at Austin
Austin, TX 78712-1175, USA

Statistical Sciences and Operations Research
Virginia Commonwealth University
Richmond, VA 23284, USA

Douglas J. Morrice

Department of Information, Risk and Operations Management
The University of Texas at Austin
Austin, Texas 78712-0212, USA

ABSTRACT

In this paper we apply multiattribute ranking and selection to the management of a port facility in the environmentally sensitive Prince William Sound area of Alaska. The approach allows tradeoffs between the disparate performance measures associated with the operation of the port and non-linear scoring of the attributes, including the notion of satisfying a target level of performance. The example considered is based on real data from a risk analysis of the port that has been simplified for ease of exposition, but the methods employed generalize to situations with real data and larger number of performance measures.

1 INTRODUCTION

Like many industrial projects, the management of a large commercial port involves the use of multiple, conflicting objectives, e.g. minimize cost vs. maximize safety. In addition there is uncertainty regarding the performance of these attributes and often the different system configurations have to be simulated so that they can be compared. In this paper we use multiattribute utility theory (MAU) to convert multiple performance measures into a single scalar performance measure and Bayesian ranking and selection (R&S) procedures to identify the best of the simulated system configurations. Our example analysis also features the intuitive concept of a target level of performance on one or more of the performance attributes. The methodology we develop is quite general and not restricted to port management. In fact, it can be used in any logistics application where the objective is to select the best system configuration from a finite number of alternatives based on multiple performance measures.

In Section 2, we introduce the problem setting for a decision maker attempting to select an optimal port configuration. In Section 3 we briefly summarize multiattribute utility theory, and describe how performance on multiple criteria can be captured with a single measure and used in an R&S procedure to address some of issues associated with the analysis in Section 2. Section 4 describes an analysis of the simplified port design problem using a multiattribute utility function and a Bayesian ranking and selection procedure. Finally, we conclude and discuss future research in Section 5.

2 OPTIMIZING THE OPERATION OF A PORT

The Prince William Sound (PWS) Risk Assessment was performed in 1996 to assess the level of oil spill risk in this environmentally important and sensitive area (Merrick et al. 2000; Merrick et al. 2002). The

Exxon Valdez had run aground seven years earlier spilling 11 million gallons of crude oil which impacted 1,500 miles of shoreline. The spill has had long-term effects on fish and wildlife, as well as economic effects on industry and local communities. All told, the spill cost Exxon 13.7 billion, not including loss of business and brand image effects (Harrald et al. 1990). A simulation model to test the effect of proposed policy changes on the level of risk in the sound was developed by Merrick et al. (2000) and Merrick et al. (2002).

Risk was measured in terms of the expected number of accidents involving oil tankers per year and the throughput of the system was measured by the number of tanker "calls" (opportunities to load and carry crude oil). Six types of accidents were modeled:

- Collisions: An underway tanker colliding with or striking another underway vessel as a result of human error or mechanical failure and lack of vigilance (inter-vessel collision) or striking a floating object, for example ice.
- Drift Groundings: A drifting tanker out of control because of a propulsion or steering failure making contact with the shore or bottom.
- Powered Groundings: An underway tanker under power making contact with the shore or bottom because of navigational error or steering failure and lack of vigilance.
- Foundering: A tanker sinking because of water ingress or loss of stability.
- Fire/Explosion: A fire occurring in the machinery, hotel, navigational, or cargo space of a tanker or an explosion occurring in the machinery or cargo spaces.
- Structural Failure: The hull or frame cracking or eroding seriously enough to affect the structural integrity of the tanker.

The simulation model (shown in Figure 1) represents oil tanker arrivals, other traffic arrivals, wind speed, visibility, and the presence of icebergs with probability distributions. On the left is a display of the dynamic behavior of the Prince William Sound marine transportation system including traffic patterns and environmental conditions, such as wind speed and direction. On the right, the analysis shown is broken into seven locations, with estimates of the probability of an opportunity for an incident, the probability of an accident given such an opportunity and finally the dynamic variation in the expected frequency of accidents for the whole region. Both the number of accidents and the number of tanker calls vary for each iteration of a simulated year of operation.

The simulation model counted situations where an accident could occur and then the probability that a specific situation would lead to an accident was estimated using accident data and expert judgment (Merrick et al. 2000, 2002). Clearly, an oil tanker passing a fleet of fishing boats has a higher potential for an accident if there is low visibility or high wind. Data is collected on vessel arrivals and environmental conditions, but the combinations of these events are not. Traffic rules, such as a one-way zone, mean that the movements of vessels are dependent, while weather-based closure restrictions cause dependence between vessel movements and environmental conditions. A simulation of the system captures the complex dynamic nature of the system and accurately models the interactions between the vessels and their environment.

The simulation model was created using data on the movement of various types of vessels obtained from the Coast Guard's Vessel Traffic Service and data on the movements and specifications of oil tankers from the Ship Escort/Response Vessel System. Environmental data is publicly available from the National Oceanographic and Atmospheric Administration weather buoys. It is more difficult to obtain data on small vessels that are not required to log in with the Vessel Traffic Service, so community surveys were used to model the movements of such vessels. Based on the data, traffic arrival models and weather models were developed. In addition, as deep-draft vessels participate in the Vessel Traffic Service and they must follow defined traffic rules, such as weather-based closure restrictions, one-way zones, the tug escort scheme, and docking procedures. So these rules were programmed into the simulation.

In this paper, we will consider a hypothetical but realistic decision. At the time of the study, oil tankers used four docks in the Port of Valdez to unload ballast water and load crude oil for transportation. We consider a "Faster" alternative that adds a dock and a "Slower" alternative that removes a dock.

Renewal processes modeled the tanker arrivals with inter-arrival time distributions fitted to historical data. The Slower alternative also increases the average inter-arrival times for tankers by 1% while the Faster alternative decreases them by 1% relative to the Status Quo. Table 1 shows the three alternatives considered.
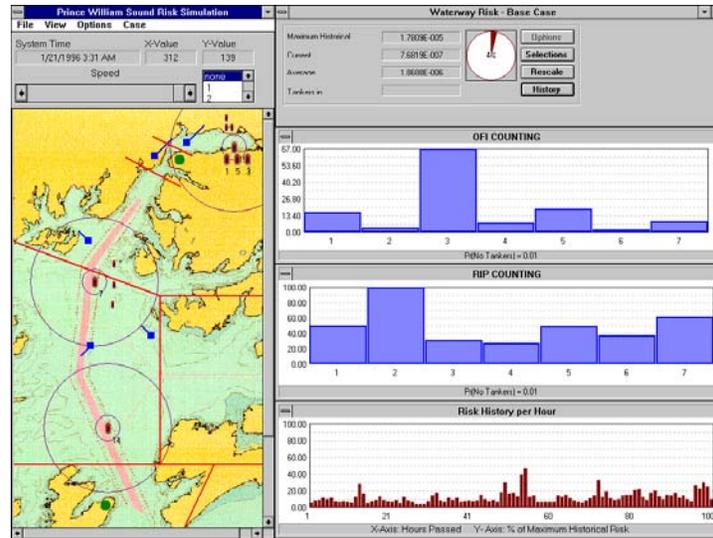

Figure 1: The Prince William Sound Risk Assessment Simulation Model

Table 1: The three systems considered in the example

System	Nos. Docks	Inter-arrival Times
Faster	5 (+1)	−1%
Status Quo	4	no change
Slower	3 (−1)	+1%

For the purposes of this paper we simulated a hypothetical year of data 1000 times so that we would have a sense of the true distribution of performance for each alternative on each performance measure and we will use the data from this simulation analysis to determine the true best alternative when working with the R&S approaches. As shown in Figure 2, the average number of accidents for 1000 simulated years was 0.0828 and the average number of tanker calls per simulated year was 641.89 under the Status Quo. Figure 2 also shows the average performance of each system on each performance measure, as well as boxplots of the performance for each system on each performance measure. The boxplot diagrams are constructed so that the preference for each attribute is increasing as one moves up the y-axis. The average of each measure under the base case (Status Quo) is represented with the dashed line in each panel of Figure 2. The "box" of the boxplots represents the 25th and 75th percentiles while the whiskers represent the 5th and 95th percentiles. The mean of each distribution is represented with a diamond in the box. As evident in Figure 2 this situation requires a tradeoff: Faster processing leads to more Calls but increases the Accident Risk while the opposite is true for the Slower configuration.
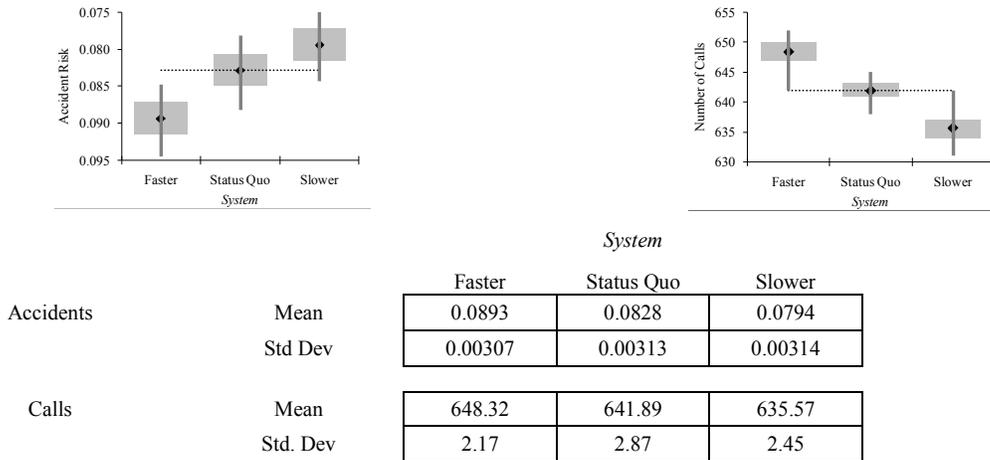
In this example, there are other attributes that might impact the evaluation of the system configurations, particularly the costs associated with constructing and operating the utilized docks. Utility theory can easily accommodate more than two attributes; e.g., Butler et al. (2001) used six attributes to capture system performance. Even when more than two attributes are used to capture performance it is usually possible to group the performance measures into two "super-attributes": benefits and costs.

Consider the decision frame of a federal decision maker who is bound by National Environmental Protection Act (NEPA) to allow continued operation of the business, but cannot allow changes that would increase risk to the environment. In other words, the Status Quo provides a target level of performance for the selected criteria. It might be tempting to formulate this decision problem as the following optimization: Choose system m such that

$$\text{Argmax}_{m=1...9} \, \text{E}[X_{mC}]$$
$$\text{s.t.} \quad \text{E}[X_{mA}] \le 0.0828 \tag{1}$$

where X_{mA} and X_{mC} denote the number of accidents and number of tanker calls for a simulated year, respectively, when system m is implemented. This constrained approach is consistent with the two-staged formulation of Andradóttir and Kim (2010) and is discussed extensively in Merrick, Butler, and Morrice (2011).

Applying their approach to the averages presented in Figure 2, the Faster configuration would be removed in phase I due to a mean accident level above 0.0828. In phase II, the Status Quo is selected as the top performer as it has the highest number of average calls of the systems that meet the accident constraint.

		Faster	Status Quo	Slower
Accidents	Mean	0.0893	0.0828	0.0794
	Std Dev	0.00307	0.00313	0.00314
Calls	Mean	648.32	641.89	635.57
	Std. Dev	2.17	2.87	2.45

System Performance : Accidents and Calls

Figure 2: Boxplots of the Number of Accidents and Number of Tanker Calls Across 1000 Simulated Years of Each Alternative

While it is true that the average number of calls for the Slower configuration is about 6.31 less than the Status Quo, the Slower configuration has 0.0034 less expected accidents per year which might offset the decrease in average number of calls (see Figure 2). In other words, there is no tradeoff between Accident risk and tanker Calls for the systems that satisfy the desired level of performance.

Rather than impose constraints, we prefer to identify target levels of performance and include that information in the scoring of the attributes. Further, the intent of the constraints is to prevent the decision maker from realizing poor performing alternatives and this can achieved by allowing tradeoffs between the attributes in the objective function

$$\text{Argmax}_{m=1...M} \, \text{E}\left[u\left(X_{m1}, X_{m2}\right)\right] \tag{2}$$

where $u()$ is a multiattribute utility function.

As discussed in Merrick, Butler, and Morrice (2011), our view is that the tradeoff implied in the following argument is likely to capture the views of many decision makers. Given the size of current

tankers, each carries crude oil valued in the tens of millions of dollars. However, the Exxon Valdez disaster in 1990 has been estimated to directly cost Exxon $13.7 billion, excluding lost of business, good will, and brand image (Harrald and Wallace 1990). A tanker with a dead weight of 125,000 tons can carry 929,000 barrels of crude oil. At the time of writing, a barrel of crude oil is valued at approximately $80. Thus, our hypothetical tanker carries approximately $74 million of crude oil and, based on a 5% profit margin, the company can expect $3.7 million in profits from each tanker. The Exxon-Valdez accident cost (at least) $13.7 billion. Using that as a proxy for the cost of an accident there is an implied trade-off of 13,700 / 3.7 ≈ 3,703 tanker calls to one accident, or one tanker call is equivalent to 0.00027 accidents. This "costing-out" procedure can be generalized to accommodate the decision maker's preferences via the weights of a multiattribute utility theory as outlined in Section 3.

3   MULTIATTRIBUTE DECISION ANALYSIS AND RANKING AND SELECTION

3.1   Multiattribute Utility Theory

In what follows we assume that there are $M > 1$ system configurations, each having two uncertain attributes $X_m = (X_{m1}, X_{m2})$. A utility function $u(x_1, x_2)$ assigns a number capturing preference to each vector of consequences $x = (x_1, x_2)$ where x_i is a level of attribute X_i. Each attribute is defined on a range from its least preferred level, x_i°, to its most preferred level, x_i^{*}. Further, the attribute utility functions $u_i(x_i)$ are scaled so that $u_i(x_i^{\circ}) = 0$ and $u_i(x_i^{*}) = 1$. We also assume that the worst possible alternative receives a score of 0 and the best possible alternative receives a score of 1; i.e. $u(x_1^{\circ}, x_2^{\circ}) = 0$ and $u(x_1^{*}, x_2^{*}) = 1$.

Attribute X_1 is utility independent of X_2 if preference for lotteries (i.e. outcomes of random variables) of X_1 given $X_2 = x_2$ does not depend on the level of x_2 (Keeney and Raiffa 1976). If a more restrictive preference condition called additive independence is also satisfied then it is the marginal, *not the joint*, distribution of each performance measure that determines preference. In this case we can represent the decision maker's preferences with an additive MAU model:

$$u(x) = w_1 u_1(x_1) + w_2 u_2(x_2) \tag{3}$$

where $0 \leq w_i \leq 1$ and $w_1 + w_2 = 1$.

While we focus on two attribute additive models in this paper the reader is encouraged to consult the extensive literature on multiattribute utility theory (e.g. Keeney and Raiffa 1976) and multiattribute ranking and selection (Butler et al. 2001; Morrice and Butler 2006) for examples with more performance attributes and more general forms of multiattribute aggregation functions.

3.2   Ranking and Selection Procedures

Let us denote the ordering of the means of the systems' true performance, μ_i, from lowest to highest by

$$\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[M]} \tag{4}$$

where $[M] = \text{Argmax}_{m=1,\ldots M} \ \mu_1 , \mu_2, \ldots \mu_m$. If the R&S procedure accurately identifies the true best system configuration, we will say that a "correct selection" (CS) is made. In our setting the mean level of performance is the average expected utility and the most preferred configuration has the highest value $\mu_{[M]} = \text{E}[u(X_{[M],1}, X_{[M]2})]$.

We denote the ordering of the *observed* sample means of performance by $\hat{\mu}_{(1)} \leq \hat{\mu}_{(2)} \leq \ldots \leq \hat{\mu}_{(M)}$; thus (m) denotes the m-th largest observed sample mean and $[m]$ denotes the m-th largest true mean. At the end of any R&S algorithm, we have correctly selected the best simulated system if the system selected based on sample means, D, has the true highest mean: $D = (M) = [M]$.

The decision maker is asked to specify some value δ^* such that $E[u(X_{[M],1}, X_{[M]2})] - E[u(X_{[M-1],1}, X_{[M-1]2})] \geq \delta^*$ is practically significant; i.e., we are willing to consider $E[u(x_{[M-1],1}, x_{[M-1]2})]$ indistinguishable from $E[u(X_{[M],1}, X_{[M]2, ..., X_{[M]n})]} - \delta^*$.

Chick and Inoue (2001) introduce the concept of Bayesian Expected Opportunity Cost (EOC_{Bayes}) based on the linear loss function (or opportunity cost), $L_{OC}(D,\mu) = \mu_{[M]} - \mu_{(M)}$. In our case, this is the amount of utility that is given up by making a choice based on the observed sample means; when a correct selection is made $D = (M) = [M]$, $\mu_{[M]} = \mu_{(M)}$, and $L_{OC}(D,\mu) = 0$. Let $C = \{ u(x_{m1k}, x_{m2k}, ... x_{mnk}), i = 1$ to M, $k = 1$ to $k_m\}$ be the set of all simulation replicates performed thus far. We can estimate EOC_{Bayes} using

$$
\begin{aligned}
EOC_{Bayes} \quad &= E[L_{OC}(D,\mu) \mid C] \\
&= E\left[\max_{i=1,...m}(\hat{\mu}_i - \hat{\mu}_{(M)} \mid C) \right] \\
&\leq \sum_{i \neq s} \lambda_{is}^{-1} \Psi_{v_{(i)(s)}}\left[d_{is}^* \right],
\end{aligned}
$$

where $\lambda_{is}^{-1} = \sigma_{(i)}^2 /k_i + \sigma_{(s)}^2 /k_s$, d_{is}^* is the standardized difference and $\Psi_v[a] = (v+a^2)/(v-1) \times \phi_v(a) - a\Phi_v(-a)$ and $\phi_v()$ and $\Phi_v()$ are the probability density function and cumulative distribution respectively function of the standard t-distribution with v degrees of freedom.

4   MAKING DECISIONS ABOUT TANKER PRODUCTIVITY AND SAFETY

4.1   Optimal Port Configurations

It is important to emphasize that utility functions should be assessed for each unique decision context and decision maker. Here we illustrate the flexibility of multiattribute R&S by selecting a few representative utility function combinations. Our hypothetical analysis of the Prince William Sound simulation focuses on the use of two additive utility models displayed in Figure 3.

The utility function in the left panel of Figure 3 is predisposed to alternatives with good scores on Accident Risk as it not only allocates most of the weight to this attribute, it is also risk seeking (convex) in Accident Risk and risk averse (concave) with regards to Calls. Based on the analysis in the left panel of Figure 3, and confirmed with EOC_{Bayes}, the Slower was the identified alternative with an initial sample (n_0) of 50 and $\delta^* = 0.001$ without any additional required observations.

We invoke a binary target-like single attribute utility functions for the Accident Risk (see Figure 4 for an example) in the right panel of Figure 3. The quantity b is a binary target. When Accident Risk is less than or equal to b, then the utility equals 1, otherwise it equals zero. In Figure 3, b equals 0.0828. Note that the weight on Accident Risk is lower (0.4 vs 0.6) but the use of the binary utility function helps to avoid alternatives that perform poorly. As shown in right panel of Figure 3, and confirmed with EOC Bayes, $n_0 = 50$, $\delta^* = 0.001$, the Slower configuration is still preferred. The weight on Accident risk has to decrease to about 0.23 before the Status Quo configuration becomes the most preferred with this set of attribute utility functions; if the weight on Accident Risk is below 0.05 then the Faster configuration is preferred.

The boxplots in the right panel of Figure 3 also highlights one important consequence of using hard, binary constraints in the evaluations. First the distribution of the Status Quo performance has a "wide" boxplot because there is about an equal likelihood that the constraint on Accidents is achieved or not which leads to high variance. Note that Faster and Slower configurations utility distributions are highly skewed because the Faster (Slower) alternative typically does not (does) satisfy the constraint and the mean utility is low (high) with a long right (left) tail pulled towards the few times the constraint is met (not met) and the configuration utility is high (low).

Additive Utility Model, Risk Seeking for Accidents
$w_A = 0.6, w_C = 0.4$
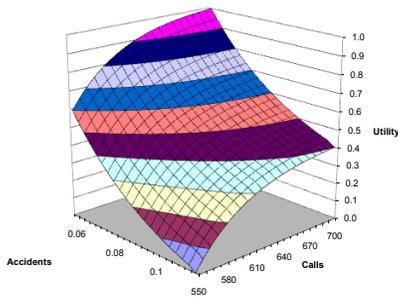
Additive Model :Binary for Accidents
$w_A = 0.4, w_C = 0.6$
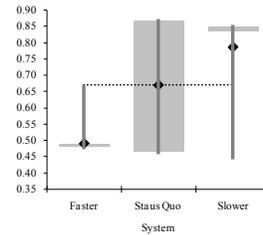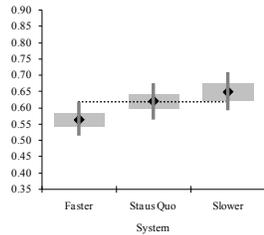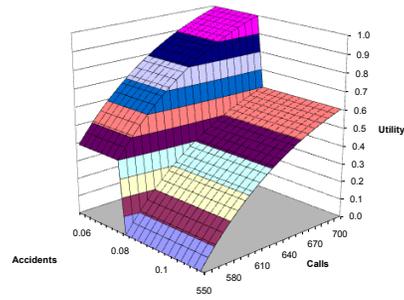
The Distribution of Utility for the Alternatives

	Faster	Status Quo	Slower		Faster		
Mean	0.5616	0.6193	0.6491				
Std Dev	0.0285	0.0342	0.0369			Status Quo	Slower
				Mean	0.4908	0.6695	0.7860
				Std Dev	0.0537	0.1996	0.1405

Statistics for Utility of Alternatives

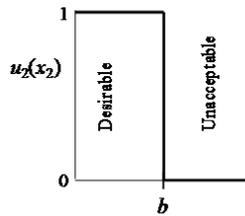Figure 3: u(X_{mA}, X_{mC}) with Target Attribute Utility Functions

Figure 4: Single Attribute Utility Functions for Constrained Attribute (x_2)

4.2 The Use of a Target and Its Impact on Variance

The recent popularity of Constraints and Targets in both decision analysis (e.g. Bordley and Kirkwood 2004) and ranking and selection (Andradóttir and Kim 2010) suggest that the variance in performance exhibited in the right panel of Figure 3 is likely to become more common in the future. To gain some insight into the impact of the use of binary target on the outcome variance we perform some additional analysis in our two attribute setting assuming that the X_i are independent and identically distributed N(μ_i, σ_i) random variables. In practice, the X_i are often statistically dependent. The latter is the subject of future research.

In our setting, we have $u(X_1,X_2) = w_1 u_1(X_1) + w_2 u_2(X_2)$. For simplicity we will assume that the $u(X_1) = X_1$ and $u(X_2) = l(X_2 \leq b)$, where the indicator function l(·) takes the value 1 when its argument is true and 0 when it is false. Under these circumstances,

$$E[u(X_1,X_2)] = E[w_1 u_1(X_1) + w_2 u_2(X_2)] = w_1 E[u_1(X_1)] + w_2 E[u_2(X_2)] = w_1 \mu_1 + w_2 E[l(X_2 \leq b)]$$

and

$$\text{Var}[u(X_1,X_2)] = \text{Var}[w_1u_1(X_1) + w_2u_2(X_2)] = w_1^2 \text{Var}[u_1(X_1)] + w_2^2\text{Var}[u_2(X_2)]$$
$$= w_1^2\sigma_1^2 + w_2^2\text{Var}[l(X_2 \leq b)]$$

The random variable $l(X_2 \leq b)$ takes the value 1 with probability $\Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)$ and 0 with probability

$\left(1 - \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)\right)$, where $\Phi(\cdot)$ is the cdf of the standard normal distribution.

Using these probabilities, $\text{E}[l(X_2 \leq b)] = 1 \times \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right) + 0 \times \left(1 - \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)\right) = \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)$

and , $\text{E}[l(X_2 \leq b)^2] = 1^2 \times \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right) + 0^2 \times \left(1 - \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)\right) = \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right).$

Thus, $\text{Var}[l(X_2 \leq b)] = \text{E}[l(X_2 \leq b)^2] - \text{E}[l(X_2 \leq b)]^2 = \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right) - \left(\Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)\right)^2$

$$= \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)\left(1 - \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)\right).$$

Using these quantities we can specify $\text{E}[u(X_1,X_2)] = w_1\mu_1 + w_2\Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)$, and

$$\text{Var}[u(X_1,X_2)] = w_1^2\sigma_1^2 + w_2^2\ \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)\left(1 - \Phi\left(\dfrac{b - \mu_2}{\sigma_2}\right)\right). \tag{5}$$

The $\text{Var}[u(X_1,X_2)]$ in (5) is graphed in Figure 5 assuming that $X_1 \sim N(1, 1)$, $u(x_1)$ is linear and the weight on X_1 (w_1) is fixed at 0.4. Further, we assume $X_2 \sim N(1, \sigma_2)$ with σ_2 varying from 0.25 to 1.75, and the target level on X_2 varying from zero to two. As expected, $\text{Var}[u(X_1,X_2)]$ is highest when the standard deviation of the variable with a target is high and the target level is close to the mean level of performance. Under these conditions the decision makers' preferences are the most sensitive to the target and, therefore, provide the most interesting case studies. For example, if the decision maker desires to achieve a level of less than 4 for $X_2 \sim N(1, 1)$, the target will not be satisfied in about 0.135% of the simulated cases and will not have a major impact in determining the top performer. Another way to frame the problem is that a decision maker is likely to include a target on situations where that target is less likely to be satisfied. This high variance could compromise the performance of any R&S procedure and we test the performance of our example utility functions in the next section.
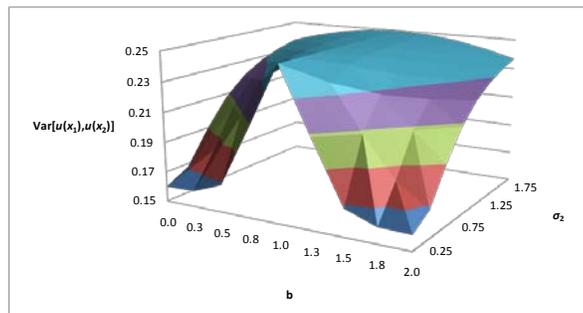

Figure 5: $\text{Var}[u(X_1,X_2)]$ as a Function of the Level of b and σ_2, $X_1 \sim N(1, 1)$, $X_2 \sim N(1, \sigma_2)$, and $w_1 = 0.4$

4.3   Performance of the Multiattribute Ranking and Selection Algorithm

Merrick, Butler, and Morrice (2011) demonstrate the impact of the use of targets on R&S procedures ability to make a correct selection. Here we change the focus to look at the effect of the target level and the variance of the attribute on which we place a target, using the PWS simulation and the three systems discussed previously. We performed 1000 runs of the simulation under each system and fitted a bivariate log-normal distribution to the two outputs of the simulation, namely the number of calls and the number of accidents. In the following testing, we sampled from the fitted distributions, not the original simulation as this was not computationally feasible for repeated iterations of the R&S algorithm.

We use the target utility function shown in the right pane of Figure 3. The target is specified for the number of accidents. The target level is 0.0828 as shown in equation (1). We modify this target by using the following multiples of 0.0828: 0.8, 0.9, 1.0, 1.1, and 1.2. We also modified the variance of the distribution of the number of accidents. The variances for each of three systems were also modified by 0.8, 0.9, 1.0, 1.1, and 1.2.

Table 2 and Figure 6 are based on 1000 samples from the distribution of the utility of each of the three systems under each of the twenty five combinations of target multiplier and variance multiplier. In Figure 6, the best system under each combination is circled. One can observe in Figure 6 that the "best" system changes as we vary the target level and the variance of the accident distribution. Table 2 illustrates the insight gleaned from Section 4.2 that when the target b is near the mean level of performance, the variance is the greatest (see columns corresponding to $b = 0.07452$ and $b = 0.0828$).

To help understand the results in Figure 6, consider any row of graphs (corresponding to a particular level of variance). When the level of b is small, all the configurations score poorly on Accident Risk because all the observations do not satisfy the target level. Hence, the number of tanker Calls determines the configuration with the best utility and Faster wins out. As the level of b increases, certain configurations (Slower and Status Quo) start to satisfy the target level of Accident Risk more than Faster. Hence, Slower and Status Quo take top spots because they do better on Accident Risk than Faster. Finally, when b gets large enough for all configurations satisfy the target level of Accident Risk and the determining factor is once again Calls which favors Faster.

Table 2: Mean and Standard Deviation of the Utility Distributions for Each System as the Target Accident Level and Variance of the Accident Distribution are Modified

|			Level of b									
			0.06624		0.07452		0.08280		0.09108		0.09936	
			Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
$Var(X_2)=q^2\sigma^2$: Level of q	0.8	Faster	0.4939	0.0121	0.5021	0.0573	0.8097	0.1554	0.8857	0.0533	0.8937	0.0122
		Status Quo	0.4908	0.0083	0.5000	0.0604	0.8429	0.1272	0.8906	0.0091	0.8905	0.0084
		Slower	0.4891	0.0067	0.4994	0.0636	0.8576	0.1069	0.8890	0.0067	0.8889	0.0067
	0.9	Faster	0.4937	0.0122	0.5093	0.0768	0.8016	0.1615	0.8839	0.0587	0.8938	0.0121
		Status Quo	0.4905	0.0084	0.5076	0.0802	0.8336	0.1379	0.8906	0.0084	0.8906	0.0084
		Slower	0.4889	0.0067	0.5079	0.0847	0.8453	0.1242	0.8889	0.0068	0.8890	0.0067
	1.0	Faster	0.4937	0.0122	0.5163	0.0911	0.7960	0.1653	0.8826	0.0622	0.8937	0.0122
		Status Quo	0.4905	0.0084	0.5153	0.0955	0.8279	0.1435	0.8903	0.0125	0.8905	0.0084
		Slower	0.4889	0.0067	0.5164	0.1007	0.8384	0.1322	0.8889	0.0086	0.8889	0.0067
	1.1	Faster	0.4937	0.0123	0.5209	0.0995	0.7862	0.1710	0.8819	0.0642	0.8936	0.0133
		Status Quo	0.4905	0.0084	0.5209	0.1051	0.8158	0.1539	0.8902	0.0151	0.8904	0.0083
		Slower	0.4889	0.0067	0.5219	0.1094	0.8268	0.1440	0.8890	0.0087	0.8888	0.0067
	1.2	Faster	0.4937	0.0122	0.5263	0.1077	0.7797	0.1745	0.8810	0.0666	0.8935	0.0164
		Status Quo	0.4905	0.0084	0.5269	0.1137	0.8078	0.1601	0.8895	0.0224	0.8908	0.0084
		Slower	0.4888	0.0067	0.5287	0.1188	0.8184	0.1515	0.8882	0.0187	0.8891	0.0067
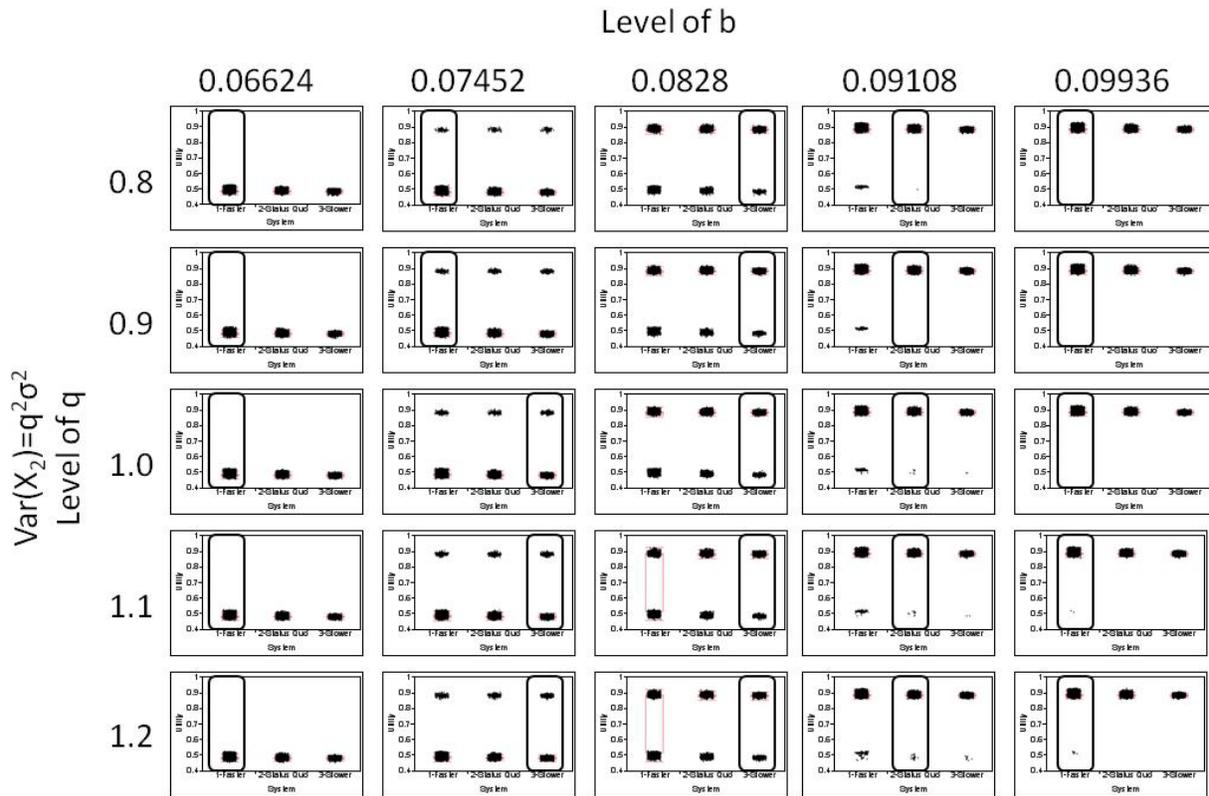
Figure 6: Boxplots of the Utility Distributions for Each System as the Target Accident Level and Variance of the Accident Distribution are Modified

One other interesting feature of Figure 6 is the impact of variance in some situations. For $q < 1$, the top performing alternative changes from Faster to Slower at b = 0.08280. In contrast, for higher variance levels ($q \geq 1$) the transition from Faster to Slower occurs at a lower target level 0.07452. This happens because the larger variance allows more of the Slower observations to satisfy the target level for lower targets, e.g. p($X2<b \mid q = 1$) > p($X2<b \mid q = 0.9$) for a fixed level of b. In other words an alternative's score on Accident Risk can improve in one of two ways: (i) the target level of performance can be increased so that it is easier to satisfy the target or (ii) the variance of performance when the mean performance is far from the target can increase so that the probability of satisfying the target in some cases increases.

Figure 7 shows the effect on the performance of the R&S algorithm. The left panel of Figure 7 shows the empirical probability that the algorithm correctly selected the best system for each combination of target multiplier and variance multiplier. To obtain the empirical PCS, we ran the R&S algorithm 1000 times for each combination and counted the number of times it selected the best system indicated by the analysis in Figure 6. The right pane of Figure 7 shows the average number of iterations in the R&S algorithm necessary to reach the specified stopping rule.

Figure 7 results are quite consistent with Figure 6. As b increases and the determining factor for the best configuration starts to switch away from Calls toward Accident Risks, it becomes hard to detect the best configuration, especially for higher levels of variation (e.g., q = 1, 1.1, and 1.2 at target level b=0.07452). Another factor contributing to the poor correct selection performance, especially for the case q=1 and b = 0.07452, is the closeness of the utility values for the three configurations (note that mean values for two of the configurations differ only at the fourth decimal place). Not surprisingly, the average number of runs increases for these "hard to detect" cases. For low levels of variation and higher values of

b, the R&S procedure does reasonably well making the correct selection. It is important to remember that in addition to b and the level of variation, the weights and the shapes of the single attribute utility functions also play a role in determining the location of a region (or regions) in which the procedure lacks robustness.

The Chick and Inoue (2001) R&S procedure we use is based on certain assumptions that do not hold for the data in Figure 7 (e.g., normality of the performance measures – in many cases the results are obviously bimodal). We suspect that this could be a contributing factor to the lack of robustness in some of the results. Consequently, our point is not to critique the performance of the R&S procedure but to illustrate some of the potential problems associated with the use of binary targets in practice.
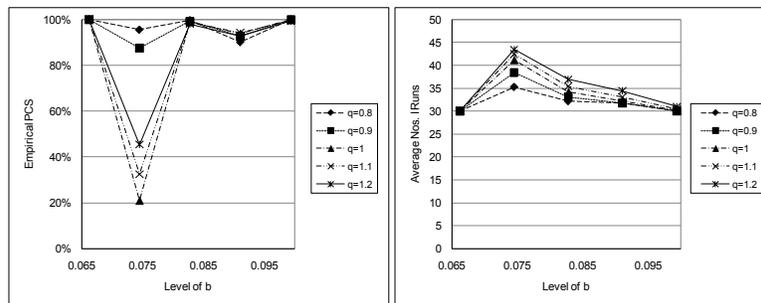

Figure 7: The Empirical Probability of Correct Selection and the Average Number of Runs as a Function of the Target Accident Level for Varying Multiples of the Variance of the Accident Distribution

5   CONCLUSIONS

Multiattribute utility theory explicitly includes tradeoffs in the objective function and so it avoids the shortcomings associated with the other approaches: it will always find the best alternative and it avoids a negative value of sample information. Further, target-oriented utility functions can be used as an alternative way to assess or interpret multiattribute utility functions which may be particularly useful when there is some desired level of performance. The ability of utility theory to appropriately capture a wide variety of preferences leads to a general approach with appealing implications.

Our work is consistent with recent advances in decision analysis that develop a theory showing the equivalence of utility theory and the achievement of a target (Abbas and Matheson 2005; Bordley and Kirkwood 2004; Bordley and LiCalzi 2000). The natural interpretation of attainment of some level of performance as a target will lead to the inclusion of this type of scoring rule in application, but as we have demonstrated there can be significant, potentially negative, impacts on R&S procedure performance.

In future research, we plan to gain a broader understanding of these adverse impacts by, for example, considering correlated performance measures. We will also consider ways to mitigate negative impacts by considering constraints less stringent than simple binary targets. These constraints will be modeled by more general utility functions.

REFERENCES

Abbas, A. E. and J. E. Matheson. 2005. "Normative Target-Based Decision Making." *Managerial and Decision Economics* 26:373–385.
Andradóttir, S. and S. Kim. 2010. "Fully Sequential Procedures for Comparing Constrained Systems via Simulation." To appear in *Naval Research Logistics*.
Bordley, R. F. and C. W. Kirkwood. 2004. "Multiattribute Preference Analysis with Performance Targets." *Operations Research* 52:823-0835.
Bordley, R. and M. LiCalzi. 2000. "Decision Analysis Using Targets Instead of Utility Functions." *Decisions in Economics and Finance* 53:53-74.

Butler, J. C., D. J. Morrice and P. Mullarkey. 2001. "A Multiple Attribute Utility Theory Approach to Ranking and Selection." *Management Science* 47:800-816.

Chick, S. E. and K. Inoue. 2001. "New Two-Stage and Sequential Procedures For Selecting The Best Simulated System." *Operations Research* 49(5):732-743.

Keeney, R. L., H. Raiffa. 1976. *Decisions with Multiple Objectives*. Wiley: New York.

Merrick, J. R. W., J. C. Butler and D. J. Morrice. 2011. Unconstraining Ranking and Selection: Using Utility Theory to Avoid Bad Outcomes by Focusing on the Best Alternatives. *Working Paper.*

Merrick, J. R. W., J. R. van Dorp, T. Mazzuchi, J. Harrald, J. Spahn and M. Grabowski. 2002. "The Prince William Sound Risk Assessment." *Interfaces* 32(6):25-40.

Merrick, J. R. W., J. R. van Dorp, J. Spahn, J. Harrald, T. Mazzuchi and M. Grabowski. 2000. "A Systems Approach to Managing Oil Transportation Risk in Prince William Sound." *Systems Engineering* 3(3):128-142.

Morrice, D. J. and J. C. Butler. 2006. "Ranking and Selection With Multiple "Targets"." In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 222-230. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

JOHN C. BUTLER is a Clinical Associate Professor of Finance and the Academic Director of the Energy Management and Innovation Center in the McCombs School of Business at the University of Texas at Austin. Butler received his Ph.D. in Management Science and Information Systems from the University of Texas in 1998. His research interests involve the use of decision science models to support decision making, with a particular emphasis on decision and risk analysis models with multiple performance criteria. Butler has consulted with a number of organizations regarding the application of decision analysis tools to a variety of practical problems. Most of his consulting projects involve use of Visual Basic for Applications and Excel to implement complex decision science models in a user friendly format. His email address is John.Butler2@mccombs.utexas.edu.

JASON MERRICK is a Professor of Operations Research in the Department of Statistical Sciences and Operations Research at Virginia Commonwealth University and program director for the new PhD in Systems Modeling and Analysis. He received his B.A. degree in mathematics and computation from Oxford University, England, and his D.Sc. degree in operations research from the George Washington University, Washington DC. His research is primarily in the area of decision analysis, simulation, and Bayesian statistics. He has worked on projects ranging from oil spill risk to watershed management to counter-terrorism and received grants from the National Science Foundation, the Federal Aviation Administration, the United States Coast Guard, the Department of Homeland Security, the American Bureau of Shipping, British Petroleum, and Booz Allen Hamilton, amongst others. He has also performed training for Infineon Technologies, Wyeth Pharmaceuticals, and Capital One Services. His email address is jrmerric@vcu.edu.

DOUGLAS J. MORRICE is a Professor in Operations Management at The University of Texas at Austin. He is also Chair of the Information, Risk, and Operations Management Department, and Director of the University of Texas Supply Chain Management Center of Excellence. He has an ORIE Ph.D. from Cornell University. His research interests include simulation design, modeling, and analysis, and supply chain risk management. Dr. Morrice was Co-Editor of the Proceedings of the 1996 Winter Simulation Conference, and 2003 Winter Simulation Conference Program Chair. He is currently serving as a representative for the INFORMS Simulation Society on the Winter Simulation Conference Board of Directors. His email address is morrice@mail.utexas.edu.