

An MCDA Approach for Establishing GPRA Metrics: Oil Spill Response

I LINKOV, A TKACHUK, A LEVCHENKO

Intertox, Inc.

83 Winchester Street, Brookline, MA 02446, USA

TP SEAGER, SP TULER, R KAY.

Social and Environmental Research Institute

278 Main Street, Room 404, Greenfield, MA 01301, USA

JR FIGUEIRA, T TERVONEN

CEG-IST, Center for Management Studies, Instituto Superior Técnico

Technical University of Lisbon

TagusPark, Av. Cavaco Silva

2790-990 Porto Salvo, Portugal

Abstract

The Government Performance and Results Act (GPRA) of 1993 requires Federal agencies to establish objective, quantifiable goals describing the outcomes of agency programs. The purposes of the Act are fairly straightforward: to improve confidence, performance, accountability, feedback, decision-making and management in Federal agencies. However, implementation of the Act in different agencies has been complicated by several factors including a lack of expertise in strategic planning, difficulty reconciling multiple missions or divergent perspectives, a lack of structured approaches to engaging relevant stakeholder groups and the presence of multiple confounding factors that obscure the relationship between agency efforts and outcomes. Currently, there is no reliable, structured approach to establishing GPRA metrics that can be transferred from one agency to the next. Successful case studies have largely resulted from *ad hoc* efforts within agencies with different degrees of expertise related to performance measurement, public or stakeholder participation and decision support. Consequently, the requirements of the Act remain a source of concern to many federal agency managers. This paper hypothesizes that many of the challenges posed by the Act are characteristic of the types of problems that multi-criteria decision analysis (MCDA) is appropriate for and proposes an MCDA framework for establishing a suite of GPRA metrics for gauging the effectiveness of oil spill response efforts. The key advantage of this framework is the visualization and quantification of the uncertainties involved in planning or decision-making processes and the potential to model multiple stakeholder responses. Moreover, we expect an MCDA approach to be generalizable to multiple agencies engaged in environmental protection missions and/or crisis response.

1. Introduction

Government agencies can rarely be characterized by a single, focused mission. Most government agencies work towards multiple objectives and serve many ‘customers’ with divergent goals. Alternatively, a single purpose (e.g., protecting human health) may be served by multiple agencies. Consequently, agencies are often required to partner with each other and non-governmental organizations to accomplish goals that seem to always be increasing in complexity. No single outcomes metric can reliably capture the efficacy of any single agency (Kravchuk and Schack 1996). Nonetheless, the Government Performance and Results Act (GPRA) of 1993 requires Federal agencies to set objective, measurable performance goals for the purpose of assessing the effectiveness of individual agency programs. Ostensibly, program budgets are revised in accordance with their perceived effectiveness in relation to the Act. A more efficient allocation of government resources is expected to result. However, in reality implementing both the spirit and letter of the Act has been complicated by a surfeit of challenges related to strategic planning, performance measurement, stakeholder engagement and disentangling the relationship between individual agency effectiveness and measurable program outcomes.

In retrospect, passage of the Act may seem like the inevitable intersection of two important socio-political trends of the early 90’s: the increasing adoption of measurable performance management systems in industry and the Clinton-Gore administration’s “reinventing government” reform initiative (e.g., Gore 1993). In industry during the early 1990’s, there was an increasing movement towards adoption of standardized systems for managing quantitative information (as exemplified by International Standards Organization specifications for quality and environmental management) and application of quantitative management systems to foster continuous quality improvements (such as the Six Sigma approach -- Harry 1994). There is no doubt that the framers of the Act envisioned making government more business-like by increasing responsiveness to ‘customers’, improving efficiency, reducing bureaucratic procedures (i.e., ‘red tape’) and fostering a culture of accountability. However, unlike business units that receive continuous feedback and capital from financial markets, government agencies operate in an environment comparatively isolated from market forces relative to political. While connecting budgets and performance assessment may seem consistent with the goal of building a business-like culture in government, in reality government agencies operate in a different milieu. Consequently, federal agencies must develop unique approaches to both information management systems and performance assessments.

In this paper, we argue that many of the challenges posed by the GPRA are amenable to multi-criteria decision analytic approaches such as those that have been adopted in some instances for social problems such as environmental decision making (Table 1). Multi-criteria decision analysis refers to a group of methods used to impart structure to the decision-making process. Generally, these decision analysis methods consist of four steps: (1) creating a set of criteria relevant to the decision at hand, for use in evaluating the decision alternatives, (2) weighting the relative importance of the criteria, (3) scoring how well each alternative performs on each criteria, and (4) combining scores across criteria to produce an aggregate score for each alternative or pair-wise comparisons of

alternatives. The goal of the process is often to select a single best alternative, but a ranking of alternatives is very useful. Most MCDA methodologies share similar steps 1 and 3, but diverge on their processes for steps 2 and 4 (Yoe, 2002). A detailed analysis of the theoretical foundations of different MCDA methods and their comparative strengths and weaknesses is presented in Belton and Stewart (2002).

Table 1: Comparison of GPRA Challenges and MCDA Capabilities

GPRA	MCDA
Requires expertise in strategic planning	Facilitates structured examination of objectives, alternatives
Must balance multiple missions and perspectives	Criteria result from multiple objectives and may be weighed differently by different groups
Must engage stakeholders and be responsive to feedback	Structures and facilitates stakeholder value elicitation and non-expert knowledge
Requires expertise in performance measurement	Accommodates multiple sources and types of information including qualitative and semi-quantitative

2. Performance Metrics for Oil Spill response

Measuring the effectiveness of an oil spill response is extremely challenging. Until very recently there has been little research on what constitutes “good” environmental performance metrics within the context of environmental crisis, what different types of metrics there are, how these could apply to oil spill response, and how they might relate to stakeholder or public attitudes, expectations or objectives. Incident-specific strategies must be identified early and on a case-by-case basis that usually must be implemented by an *ad hoc* association of many organizations that may be unfamiliar with working with one another (Grabowski et al. 1997). To further complicate matters, little guidance is available to responders on how to incorporate stakeholder or public views into the initial assessment of priorities. Consequently, methods for setting objectives, tracking progress, and communicating or determining success depend largely on the experience of the on-scene coordinator and the level of interaction with state, local, or other nonfederal government groups outside the command structure, including the media. Even in the case that the response is closely coordinated among agencies and planning documents are scrupulously adhered to, public perceptions can be that the response has failed—partly because it is not apparent what normative standards of success should be applied or how the measures of success employed by decision makers will be interpreted by the public or intermediaries (such as journalists or nongovernment organizations; see Harrauld 1994; Chess et al. 2005). In short, there currently is no comprehensive consensus approach to assessing the success of response efforts (Seager et al. 2007).

2.1 DIFFICULTIES IN PERFORMANCE MEASURES

Oil spill response, like many emergency or crisis response situations, displays many characteristics of a government function that is extremely difficult to assess with respect to the GPRA:

- Every oil spill is unique. Although planning efforts are typically intense with regard to pre-positioning of equipment, establishing communication and leadership protocols and personnel training, the response *tactics* are virtually impossible to plan ahead of time. Optimal alternatives are highly dependent upon the circumstances of individual spills. Therefore, compliance with the GPRA – in which strategic planning is an essential aspect – is complicated by the inability to establish a universal vision of how to achieve a successful response.
- Even with generalizable goals for oil spill response in place, performance measurement requires a benchmark to determine the sensitivity of outcomes to agency efforts. In the case of oil spills (and other crisis response situations), it may be impossible to say what the level of measurable outcomes *would* be under alternative response scenarios. The unique nature of the crisis precludes comparative assessments to control or ‘normal’ circumstances. Moreover, the time-sensitive nature of the crisis typically prohibits dynamic modeling of alternative scenarios in real time. Consequently, decision outcomes must be judged without the benefit of context.
- Oil spills inevitably engage multiple stakeholder groups including industry (e.g., fishing, tourism), non-governmental organizations, local government agencies, journalists and the greater public. In many cases, the goals of these groups may be at odds, complicating the process of incorporating key stakeholder concerns into GPRA measures.
- Effective oil spill response requires close coordination between multiple public and private agencies (such as the responsible party and response contractors). The perceived success or failure of a response is a function of the collective efforts of the wide-ranging *ad hoc* administrative structure. Disentangling the contribution of each of these agents to the eventual outcome may be impossible, if not explicitly counter-productive to the extent that organizations may fail to cooperate effectively if they become preoccupied with their own performance assessment at the expense of overall system effectiveness.

2.2 REVIEW OF PERFORMANCE METRICS USED FOR OIL SPILLS

To examine performance metrics in oil spill response, we interviewed key stakeholders to study performance metrics used to assess two recent oil spill responses: the *Bouchard-120* spill response that began on April 27th, 2003, as the tugboat *Evening Tide* ran aground and released No. 6 home heating fuel oil just at the entrance to Buzzards Bay, Massachusetts, and the Chalk Point spill response that began on April 7, 2000 when an intrastate pipeline that transports oil to the Potomac Electric Power Company’s (PEPCO) Chalk Point electrical generating facility in Prince George’s County released No. 2 and No. 6 home heating fuel oil into Swanson Creek and subsequently into the Patuxent

River. Those interviewed represented a diverse group involved in the spill response, including federal, state, and local government officials, NGO representatives, and local community members. The Federal on-scene coordinator and State on-scene coordinators from each response were included in the interview samples.

We found that the stakeholders were collectively able to generate a large number of performance metrics. A sampling is listed in Table 3. Each metric can be characterized by where in the organization the metrics applies, how (and how well) the information is expressed, and what type of information is collected (Table 4) according to the typology of environmental performance metrics described in previous work (Seager et al 2007).

Table 3: Example Oil Spill Response Metrics by Type of Information Measured

<u>Economic</u>	<u>Thermodynamic</u>	<u>Environmental</u>	<u>Ecological</u>	<u>Human Health</u>	<u>Socio-Political</u>
Clean up costs. Property & ecosystem damage. Ecosystem damages or lost services. Lost marginal profits. Volunteer opportunity costs.	Volume of oil spilled, recovered, destroyed, or contained. Slick area and thickness. Mass of clean up wastes generated. Volume cleaning agent deployed.	Chemical concentration & toxicity. Habitat suitability, e.g., acres shellfish bed. Length of oiled shoreline. Degradation rates. Residual Risk	Wildlife deaths or populations, fecundity and recovery rates. Biodiversity. Catch sizes. Plantings, seedings. Habitat Suitability	Threatened population Quality-adjusted-life-years (QALYS) Disability-adjusted-life-years (DALYS) Life-expectancy Injuries	Newspaper column inches, minutes TV coverage, web hits. Volunteerism. Public meeting attendance. Critical sites protected Historic sites protected

Table 4: Characterization of Performance Metrics

HOW		WHERE			WHAT
MATHEMATICAL	QUALITY	DECISION LEVEL	CAUSAL CHAIN		TYPE
quantitative	verifiability	strategic	inputs	planning	economic
semi-quantitative	cost	tactical	processes	response	environmental
qualitative	communicability	operational	outputs	recovery	ecological
	sensitivity		outcomes	restoration	socio-political
	credibility				human health
	scalability				thermodynamic
	relevancy				

The *how* column in Table 4 characterizes the mathematical form and quality of the metric. Mathematically, all metrics could be classified as either quantitative (e.g., cardinal measurement), semi-quantitative, or qualitative. In the GPRA, a clear preference is expressed for quantitative measures -- although they may not always be available. Regarding quality, different metrics may have multiple attributes to different degrees. Therefore, a judgment regarding the quality of any metric may depend upon the relative importance of each attribute such as cost or verifiability to the decision maker.

The *where* column describes the relationship of a metric to both organizational structure and sequence events in crisis response. A different suite of metrics may be applicable at the strategic level of thinking than at the tactical or operational. Similarly, different

metrics are applicable for an accounting of resources, processes or outcomes depending upon whether the concern is spill response, systems recovery or restoration. In some cases, outcome metrics for response (e.g., rescued birds) may be viewed as input metrics for recovery (e.g., viable breeding population) or restoration. For example, an overarching strategy may be to reduce the severity of oil spills (as measured by wildlife deaths, for example). One tactical approach may be to contain and remove slicks. In pre-spill planning, a resource metric might be to measure capital equipment expenditures for purchase and pre-positioning of additional equipment. In the event of an oil spill, the on-scene coordinator might track tactical measures such as the time required to deploy the prepositioned equipment during the response. At the operational level, the effectiveness of the deployed equipment must be tracked to ensure success of the overall strategy. The importance of tracking resource and process (or output) measures must be emphasized in cases where outcomes are significantly delayed or disconnected in time from the actual decisions. In many instances, mid-course corrections must be made before final outcomes can be tallied.

It has often been written that *what* gets measured will be managed. Therefore, it is essential to understand the relationship of any metric to the multiple objectives of the spill response. Table 4 proposes a fairly comprehensive – albeit far from exhaustive – list of different dimensions by which oil spill response success might be gauged. However, where two different dimensions are incompatible with one another (e.g., additional expense is required to remove additional oil from the environment, although the ecological benefits may be dubious) disagreement may exist about which measures are most important.

- The **ecological** dimension refers to biological aspects of living organisms such as populations, health or incidence of injury or disease. Presumably, human populations could be lumped together with wildlife in this category. However, the preeminence of **human health** protection in oil spill response likely justifies a separate category specifically tracking human health measures, such as worker injuries or exposures.
- The **environmental** dimension refers to the chemical or physical characteristics of the impacted area, such as oil or toxic chemical concentrations. These are distinguished from ecological measures because the relationship between environmental conditions and biological endpoints may be complex, indirect or difficult to reveal. Moreover, environmental endpoints (such as presence of oil in shoreline soils) that may have limited ecological relevance may nonetheless be held in high regard by some stakeholder groups.
- The **economic** dimension may itself have several important aspects, such as the direct costs to the responsible party and the indirect costs to volunteers, impacted industries, or even potential beneficiaries of the spill expenses such as contractors hired to participate in response.
- The **thermodynamic** dimension characterizes all the mass and/or energy measures that may be relevant process outputs (or resource inputs). For example, the mass of oiled litter recovered from beaches, the volume of oil recovered from shipboard, or the total gallons (barrels) of oil spilled into the environment.

- The **socio-political** dimension describes the quality of human relationships and how they are impacted by the spill. For example, the incidence of litigation may be one symptom of a deterioration of trust. Criminal proceedings may be an indication of an ethically bankrupt culture. However, high incidence of volunteerism may be indicative of the quality or effectiveness of community outreach programs.

Ideally, an agency will populate the entire *where* and *what* dimensions of the typology with high quality metrics that are applicable to every level of decision-making within the organization. For example, one important strategic goal listed in *A Guide For Spill Response Planning In Marine Environments* is to "... minimize adverse environmental and socioeconomic effects by coordinating all containment and removal activities to carry out a timely, effective response." (NOAA 2001). This single statement cuts across several dimensions of the environmental performance metrics typology. First, the notion of environmental "effects" is a qualitative, outcomes-based, environmental and ecological metric that should inform strategic decision-making -- while the notion of socioeconomic effects is also qualitative, outcomes-based and strategic. It seemingly suggests aggregating the economic and socio-political information that are treated separately in the typology of Table 4. Second, specific suggestions are provided to tactical decision-makers thinking about achieving the strategic goal, namely containment and removal. In this regard, the emphasis is on tactical processes, rather than strategic outcomes and the type of information to be gathered would presumably be thermodynamic (i.e., volume of oil contained or removed). Lastly, criteria for judging the spill response are suggested: specifically, the response should be "timely" and "effective." However, it should also be apparent that this statement alone is insufficiently detailed to allow responders to gauge the effectiveness of their efforts. For example, how should containment operational effectiveness be measured? How would timeliness be measured quantitatively? A more comprehensive elicitation process such as interviews, literature review and consultation with experts is called for. This exercise will allow identification of gaps that may concern managers or stakeholder groups. For example, they may discover a paucity of strategic or quantitative metrics, which may cause them to return to the step of metrics elicitation or invent new metrics to fill the gaps they consider important.

Following elicitation, the agency must reorganize the information in multiple directions to discover whether a suitable cross-section of different types of feasible, high-quality metrics are available for each organizational objective (strategic, tactical, operational) or at each step in the agency's chain of influence (inputs, process outputs, and outcomes) or for each type (economic, environmental, ecological, socio-political, human health and thermodynamic). Once again, significant deficiencies may be corrected by further elicitation or invention.

In any case, it is unlikely that all stakeholders (or agency managers) agree on which metrics (and their quality) should be employed to determine the relative success of any response or the quality. Overall, a four-step process that precedes MCDA is called for. First metrics must be *elicited* from the salient stakeholder groups. Then, they must

characterized with regard to the typology in Table 4. Where important gaps are identified, a return to elicitation may be called for – or perhaps entirely new metrics must be devised. Lastly, the quality of the metrics must be *assessed*. Only then should value-based weights be assigned to any set of sub-characteristics to allow aggregation to a higher level assessment of which metrics should be adopted for any given situation. This prioritization framework could be easily coded in one of the MCDA platforms. Even then, the disaggregated information should remain available to agency managers. Finally, decision makers should *deliberate* about the results of the overall process to select a final suite of metrics considered most relevant for a particular region or problem.

2.3 USE OF MCDA TO RESOLVE DIFFICULTIES

Multi-criteria decision support processes can overcome many of the challenges of performance measurement faced by Federal agencies. The typical approach to an environmental MCDA involves identifying feasible alternatives and the criteria by which they should be judged and assessing the performance of each alternative relative to the salient criteria. While alternatives may be generated by experts, stakeholders and/or public groups, the decision *criteria* and their relative importance may justifiably be the purview of affected parties (such as stakeholder groups), who may or may not also be experts. The performance assessments are typically performed by experts. For example, in considering the problem of managing contaminated sediments, stakeholder groups may emphasize the importance of environmental quality, preservation of ecological habitat, or economic development. All of the alternatives generated by expert and stakeholder groups must then be assessed relative to measurable criteria that capture each of these performance criteria – although different groups may value each criterion differently. The results are called a *performance table* that is subsequently analyzed with different mathematical approaches including multi-attribute utility theory, outranking, pair-wise comparison or others. The advantage of MCDA is that alternatives can be ranked or prioritized for multiple decision criteria, trade-offs can be elucidated, different types of information including semi-quantitative or uncertainty information can be handled in a structured way and conflicts or opportunities for compromise can be observed that may not otherwise have been discovered (Lahdelma et al. 2000, Figueira et al. 2005). The principal disadvantage of MCDA approaches is that they can be time-intensive, especially when multiple stakeholder perspectives are important. In crisis response, there may not be time to engage in structured decision making (Tuler 1988, Klein 1999). Therefore, the typical MCDA approach to alternatives assessment is not helpful.

In the case of crisis response (such as oil spills), existing MCDA processes are at a distinct disadvantage. However, we propose a unique approach for the purposes of the GPRA. In this case, the *metrics* are the alternatives. Usually, metrics are associated with assessment criteria and relate to the underlying objectives or goals of the organization. However, the GPRA requires agencies to select a limited number of metrics from a theoretically infinite universe of potential outcomes metrics. Not all metrics would be cost-effective for the agency to track, nor can any single metric perfectly fulfill the requirements of the GPRA. Therefore, agencies need a process for generating a set of

potential metrics, vetting these with key internal and external stakeholder groups, and prioritizing among them to establish a concise suite of metrics that capture agency performance. As in a typical environmental MCDA, some agency customers may value certain metrics within the suite more highly than others. (Some may even place zero weighting on certain metrics). Consequently, despite a common set of assessment metrics it is possible that agency performance could be viewed positively by one group and negatively by others – a situation that may be all too familiar to government managers. The critical difference between the process we propose for GPRA and a typical MCDA is that the decision alternatives are *which metrics to use*. The criteria that represent the other half of the performance table are characteristics that agency and key stakeholder feel the selected metrics should have. Thus, the performance table is a guide to determining which metrics are better than others for any particular program.

3. Decision Analysis Example

As an illustrative example of MCDA application, consider the problem of prioritizing performance metrics for oil spill response according to specified management objectives. As a hypothetical example, we consider 14 alternative metrics in bold type in Table 3. These metrics represent different phases of the response (response, recovery, restoration) and the different types of metrics (economic, thermodynamic, environmental, human health, ecological, socio-political). Many of these performance metrics were mentioned by stakeholders in the case studies we examined (Tuler et al., 2007).

3.1 SMAA METHODOLOGY

One of the most time-intensive, problematic aspects of MCDA for group decision-making is the elicitation of weights that represent the values of each decision-maker. The larger the group and the greater the number of relevant criteria (or objectives), the more burdensome the problem of value elicitation becomes. Moreover, the process is complicated by the fact that decision-makers themselves may be uncertain about weights, preferences may evolve or change throughout the decision-making process, or the weights elicited may be unreliable in the sense that they depend upon the method of elicitation. Moreover, in some case the results of an MCDA may be highly sensitive to treatment of uncertainty in the weights. However, stochastic multiobjective acceptability analysis (SMAA) circumvents the difficulties of value elicitation and expedites analysis of uncertainty by using Monte Carlo simulation to explore a wide range of the weight spaces. The results can provide decision makers with a probabilistic characterization of preferred alternatives even prior to value elicitation. The results are especially helpful for understanding multiple decision maker problems by identifying when some alternatives are likely over a wide range of potential weighting schemes (Tervonen et al. 2007, Lahdelma and Salminen 2001).

3.2. CRITERIA

For the purposes of illustration, 11 of the criteria from Table 4 have been chosen for an example in the use of SMAA and are depicted in the rows of Figure 1, which is an input

screen captured from a software program known as SMAA-III (Tervonen et al. 2007, Lahdelma and Salminen 2002). The criteria are: cost (e.g., of obtaining the data required to track a specific metric), ecological (Rel_Ec), environmental (Rel_Env), human health (Rel_HH), socio-political (Rel_SP), verifiability (ver), communicability (com), causality (causal), credibility (cred), scalability (scal), sensitivity (sens). The analyst must specify different aspects of these criteria, such as whether ascending or descending ordering is preferred (i.e., more or less of the characteristic), the minimum precision with which two estimates on the scale of this criteria may be distinguished (the indifference threshold) and the difference between two estimates that would result in a complete preference for one estimate over another (the preference threshold; between indifference and preference, a decision-maker may be uncertain about the extent to which one number is preferred over another). The SMAA-III software can work with different scales (including poorly defined ordinal scales), does not require precise preference information and is particularly adequate for situations where “big” uncertainties are present; these uncertainties referred to imperfect knowledge of basic data and technical and preference parameters. (In addition, SMAA-III software allows defining other model specifications that are not used in this example).

3.3 ASSESSMENT

Assessment of the performance of different metrics relative to the 11 listed criteria in general requires consultation with decision makers and technical experts. For example, the total volume of oil recovered from the environment may be very difficult to verify (because the liquids removed will be an oil-water mixture), or it may be expensive to determine. However, this measure may also be easily communicated and understood by the public. In any case, a determination must be made about the performance of a metric such as “total volume of oil recovered” relative to each criterion. It may be helpful to elicit performance estimates from a wide range of experts or stakeholders, these estimates need not agree. They may be expressed as a probabilistic distribution, or alternative analysis can be conducted for different performance assessments and the results compared. The quantities used in this example have been furnished by the authors for illustrative purposes *only*. Fourteen candidate metrics listed in bold in Table 3 are assessed: threatened human populations (TP), residual risk (RR), habitat suitability index (HSI), habitat suitability (HS), clean up costs (CLC), property damage (PD), volume of oil (VO), mass of waste (WM), length oiled shoreline (LOS), chemical concentration (CHC), injuries (IN), media attention (MA), critical sites (CUS), public meeting attendance (PUB).

Name	Type	Direction	Indifference TH	Preference TH	Veto TH	Threshold type
Cost	UNIF	DESC	2000	2500	Enable	ABS, DET
Rel_Ec	UNIF	ASC	0	0	Enable	ABS, DET
Rel_Erv	UNIF	ASC	0	0	Enable	ABS, DET
Rel_HH	UNIF	ASC	0	0	Enable	ABS, DET
Rel_SP	UNIF	ASC	0	0	Enable	ABS, DET
Ver	UNIF	ASC	0	0	Enable	ABS, DET
Com	UNIF	ASC	0	0	Enable	ABS, DET
Causal	UNIF	ASC	0	0	Enable	ABS, DET
Cred	UNIF	ASC	0	0	Enable	ABS, DET
Scal	UNIF	ASC	0	0	Enable	ABS, DET
Sens	UNIF	ASC	0	0	Enable	ABS, DET

Figure 1: Input screen from SMAA-III software describing scales in which assessment criteria operate.

Cost is measured in US dollars required to collect information for quantification. For example, we assumed a cost of \$50,000 to track changes in the threatened population. For those criteria with the “Rel” prefix (all from the “What” column of Table 3) assessments of relevance to the criteria are gauged on a scale of 1-10 (10 being highly relevant and 1 having little relevance). All the candidate measures are recognized as being relevant to more than one of the “what” criteria, which is to say that these criteria are not completely independent. Metrics performance for other criteria (credibility, communicability, etc) is assessed as merely High, Medium and Low (represented by 3, 2 and 1 in Figure 2). For example, the threatened population (TP) metric is scored high on all criteria because it is a robust measure that is easy to communicate. Even though Media Attention (MA) metric is easy to communicate and somewhat sensitive to the public perception, here it is listed as having low credibility and assessed low on other factors. The assessments are summarized in Figure 2. It bears repeating that these assessments are purely for the mathematical convenience of the authors. A more robust assessment would entail a wide elicitation of expert and stakeholder views.

Unlike some other MCDA tools, SMAA-III also allows defining the indifference and preference thresholds. Indifference thresholds could be used by decision makers to acknowledge that a small difference in assessments could be interpreted as indifference between two alternatives. In our example, indifference and preference thresholds are defined for cost criteria only. The indifference threshold of \$2000 defined in our example means that two alternatives with price difference of \$2,000 are considered as

indifferent (i.e., this amount of money is too small to worry about in spill response). However, differences of greater than \$2500 are considered to be important, which is to say that a decision maker would clearly prefer a metric that was at least \$2500 less costly to track. Between \$2000 and \$2500 savings, the decision-maker may consider the costs difference important or may not. Therefore, the preference is considered to be incomplete.

The screenshot shows the CSMAA v1.0 software interface. The window title is "CSMAA v1.0 C:\Documents and Settings\user\Desktop\DRA_USA\Software\CSMAA\Oil.csm". The interface has a menu bar with "File", "Select method", "Select #", "Alternative names", "Criteria", "Measurements", "Uncertainties", "Preferences", and "Execute". Below the menu bar is a table with 11 columns: "Cost", "Rel_Ec", "Rel_Env", "Rel_HH", "Rel_SP", "Ver", "Com", "Causal", "Cred", "Scal", and "Sens". The table contains 15 rows of data for different candidate metrics: TP, RR, HSI, HS, CLC, PD, VO, WM, LOS, CHC, INJ, MA, CUS, and PUB. Each cell in the table contains a numerical value. At the bottom of the window, there are buttons for "<- Go back" and "Continue ->".

	Cost	Rel_Ec	Rel_Env	Rel_HH	Rel_SP	Ver	Com	Causal	Cred	Scal	Sens
TP	50000	10	2	7	9	3	3	3	3	3	3
RR	100000	5	7	10	8	1	2	3	2	3	3
HSI	100000	1	10	2	3	1	2	2	2	1	2
HS	10000	5	1	1	10	3	3	1	3	1	1
CLC	30000	10	1	2	5	3	3	3	3	3	3
PD	50000	10	5	4	4	3	3	2	3	3	3
VO	20000	5	7	5	3	1	3	3	2	3	3
WM	10000	5	7	3	4	1	2	3	2	3	3
LOS	10000	7	7	3	2	3	3	3	3	3	3
CHC	50000	1	10	7	2	2	3	2	2	3	2
INJ	5000	5	1	10	7	3	3	1	2	3	2
MA	30000	7	1	2	10	1	3	1	1	1	2
CUS	30000	7	1	1	10	2	3	2	3	2	2
PUB	10000	1	1	2	10	3	3	1	2	2	1

Figure 2: Hypothetical assessment of candidate metrics relative to representative criteria.

4. Preference analysis

SMAA-III is based upon an outranking approach to preference determination rather than a utility function approach. (See Seager 2004 or Linkov et al., 2006a for additional details). There are several advantages of SMAA-III compared with other MCDA approaches:

- As depicted in Figure 1, SMAA-III works with the original and natural scales for different assessment criteria, including cardinal (e.g., cost), or semi-quantitative, or qualitative (e.g., High, Medium, Low). There is no need to translate assessments into a single-dimensional utility or value function.
- Low time requirements and cognitive efforts from decision makers.
- Uncertainty can be explicated incorporated in every aspect including assessments and weightings.

In our example, we assumed that the decision maker can not assign weights to the criteria a priori but believes that all of them are important and neither one dominates. The importance for all criteria is thus represented by uniform distribution from 5% to 49%, with 100% representing a decision criteria based upon only one criteria to the exclusion of all others and 0% representing a criterion that receives no consideration at all (Figure 3). In an actual case, decision makers may assign their own bounds specific to each criterion and different analyses conducted for individual decision maker bounds. The results may be important for identifying opportunities for decision makers with different views to compromise on alternative(s) that are nonetheless highly ranked by all.

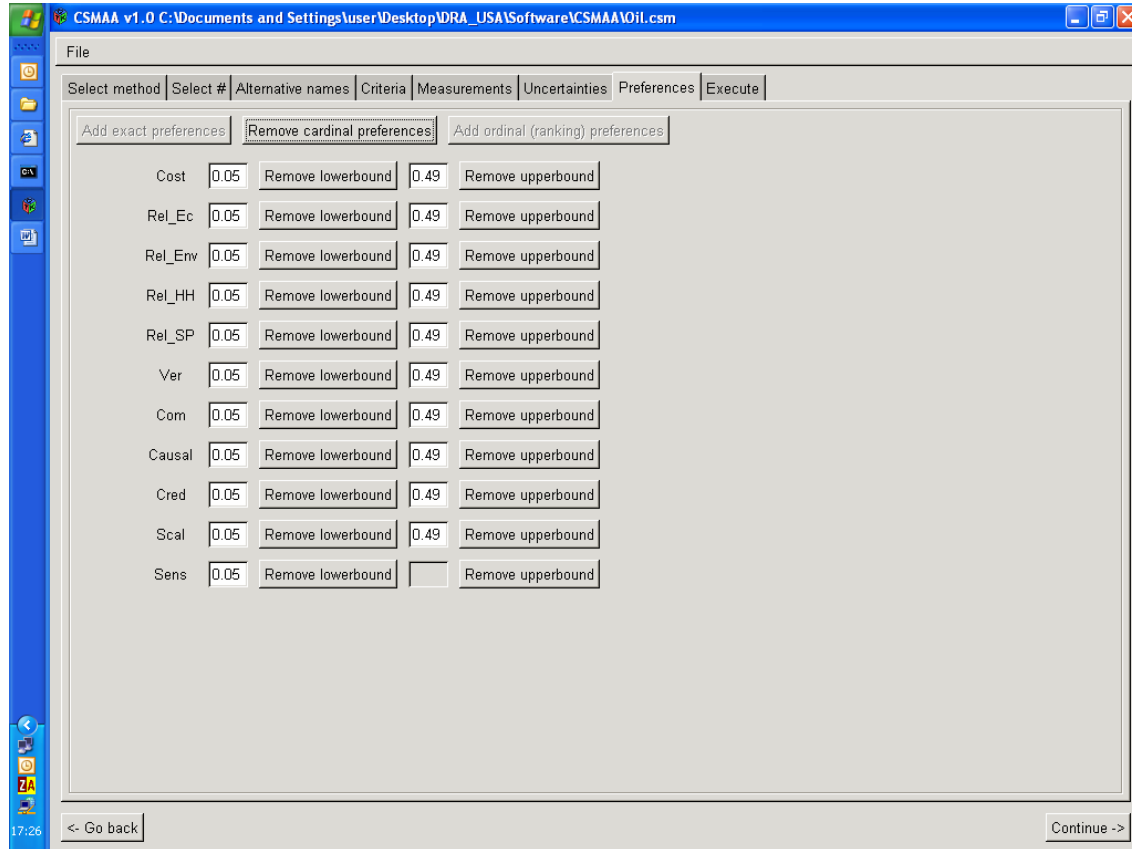


Figure 3: The range of weights tested for each assessment criteria range from a lower bound of 5% to an upper bound of 49%.

4.1 RANK ACCEPTABILITY ANALYSIS

The rank acceptability index measures the share of feasible weights that grant each metric specific rank in the order of importance by taking into account simultaneously uncertainty in all parameters and criterion evaluations. It represents the share of all feasible parameter combinations that make the alternative acceptable for a particular rank, and it is most conveniently expressed percentage-wise. The most acceptable (“best”) metrics are those with high acceptabilities for the best ranks. These are the metrics that are most likely to be valued by decision-makers with a random assignment of criteria weights between the lower and upper bounds. That is, they may be the

alternatives most likely to result in consensus. The results of the rank acceptability analysis for our case are presented at Figure 4. The left most-box represents the likelihood of the alternative metric being ranked as the most important, whereas the right-most indicates the likelihood of being ranked last. The top metrics include TP (threatened population), LOS (length of oiled shoreline), CLC (cleanup cost) and PD (property damage), which are likely to be ranked among the top four preferred metrics.

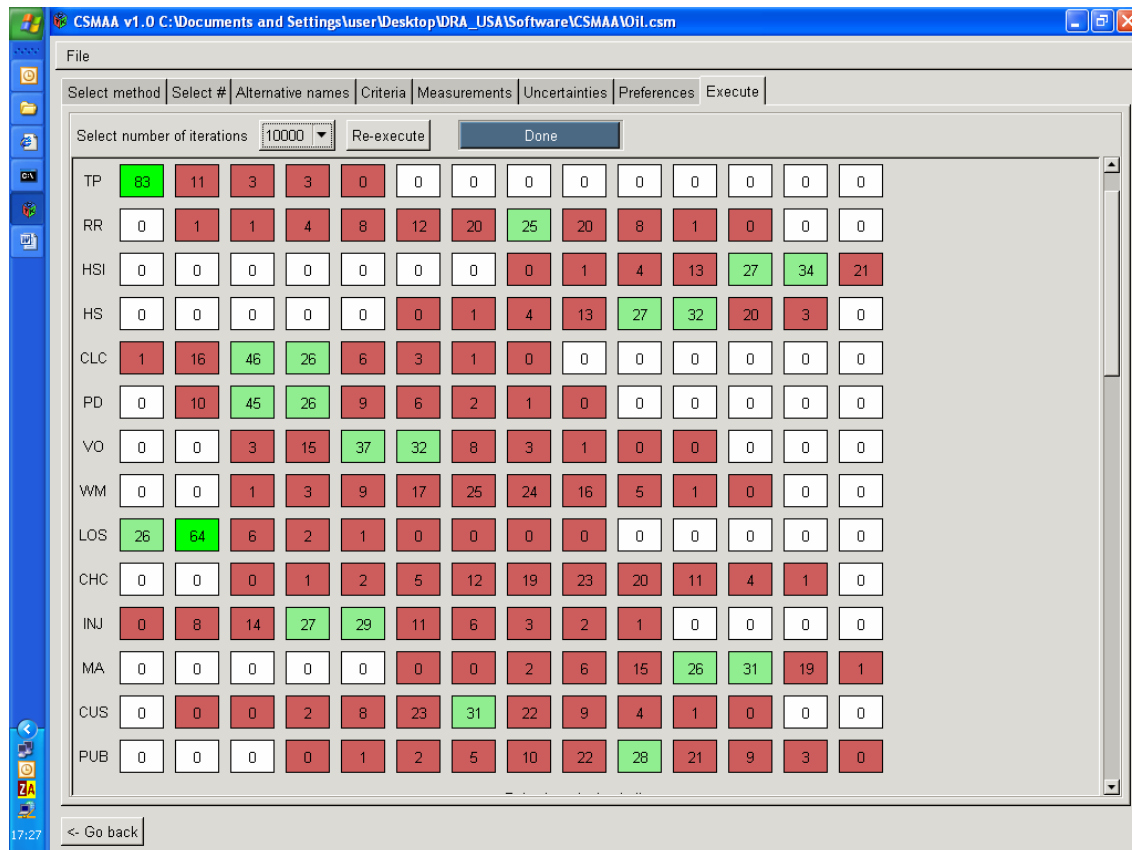


Figure 4: Rank acceptability analysis results show the likelihood of each alternative metric being ranked by decision makers as most preferred (left hand box) or least preferred (right hand box).

4.3 PAIR-WISE ALTERNATIVE COMPARISON

Alternatively, it is possible to compare the likelihood that one metric may be preferred to one other metric in a pair-wise comparison rather than an overall ranking. Figure 5 is a pair-wise comparison table. The alternative metrics are listed in the row titles at left. Although the column headings are omitted, the order of the alternatives is exactly the same. To determine the likelihood that a decision maker might prefer TP (threatened population) to length of oiled shoreline, it is necessary to follow the TP row to the right nine boxes, where we see the result is 74. (LOS is listed in the ninth row and therefore corresponds to the ninth column).

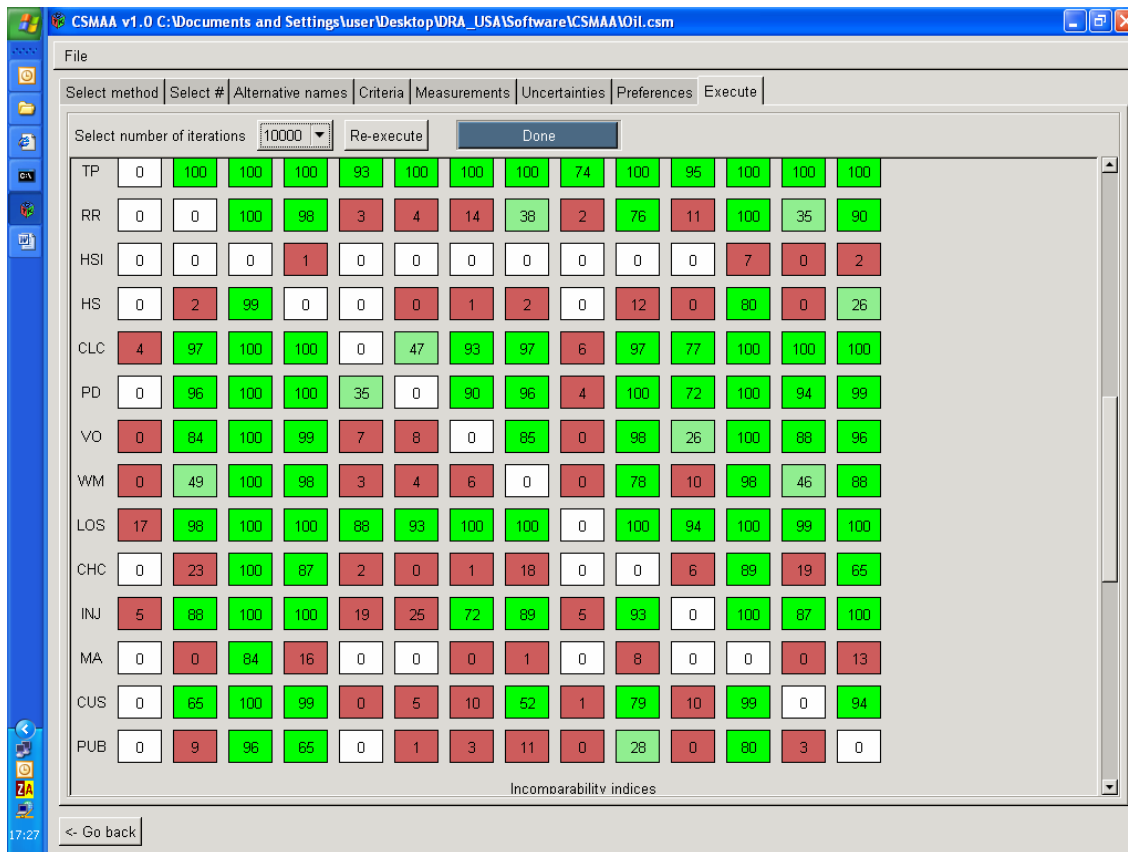


Figure 5: Pair-wise comparison of alternative metrics in a 14x14 matrix showing the likelihood of the row metric dominating the column metric. Note that column metrics are presented in the same order as the rows, although the headings are omitted.

4.4 SENSITIVITY TO WEIGHTING BOUNDS

In some cases, decision makers will express a stronger preference for some assessment criteria than others (compared with a randomly distributed, equally likely distribution of weights). In Figure 6 below we assign greater importance to cost by raising the upper bound on this weighting to 25% (without changing the upper bound of 49% or the bounds of any other criteria). For comparison, Figure 7 depicts the results of the rank acceptability analysis for over weighting of environmental considerations (Rel_Env) in the same way. It is interesting to note that in both cases, the results show that LOS (length of oiled shoreline) is now most likely the preferred metric, compared with TP (threatened population). While LOS is not the cheapest metric and it is not the most relevant metric to ecosystem health, its relatively low cost and good performance on all criteria make it acceptable for both budget-conscious and environmentally-conscious stakeholder communities when these priorities are given greater weight.

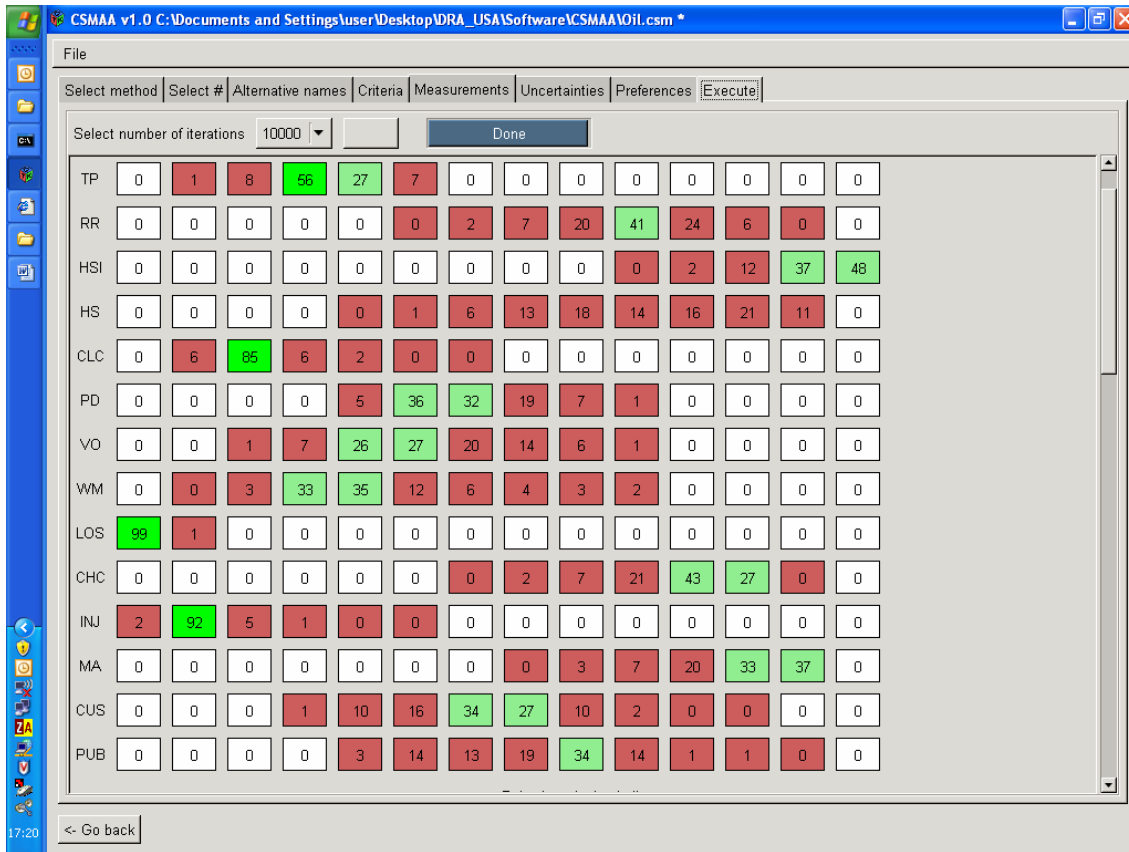


Figure 6: Rank acceptability analysis for potential over weighting of cost.

5. Discussion

The example presented here is a simplified, hypothetical case to illustrate the utility of a SMAA-III approach to analysis of a suite of candidate metrics for gauging the success of oil spill response. Selection and ranking of performance metrics for specific region/situation requires a more detailed elicitation of candidate metrics, criteria and assessments. Moreover, the number of performance metrics that may be relevant in an initial elicitation such as the *Bouchard* and Chalk Point cases is well above 100. (See Appendix). Prioritization and ranking of so many performance metrics may require an initial screening. Nevertheless, when populating the typology of metrics (e.g., Table 4), significant gaps may be found. The critical contribution of SMAA-III in this case is not that it facilitates gathering of information or generating the data required to meet the GPRA requirements, but that it provides a structured approach to *analysis* of the data.

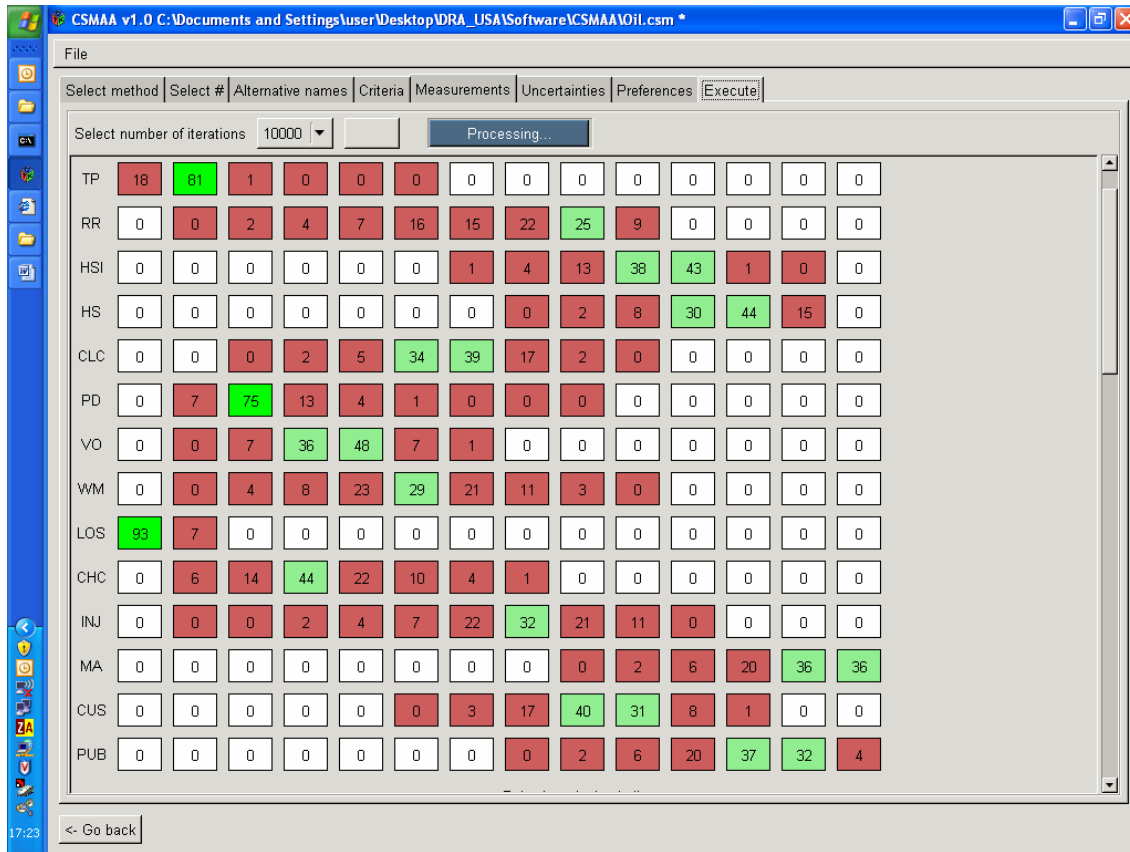


Figure 7: Rank acceptability analysis for potential over weighting of environmental relevance.

6. Conclusions

Effective oil spill performance metric prioritization requires an explicit structure for jointly considering the environmental, ecological, technological, economic, human health and socio-political factors relevant to evaluating alternatives and making a decision. Integrating this heterogeneous information with respect to human aspirations and technical applications demands a systematic and understandable framework to organize the people, processes, and tools for making a structured and defensible decision. Previous papers by the authors (Linkov et al. 2004; Kiker et al. 2005) introduced a general framework for management of environmental challenges based on multi-criteria decision analysis (MCDA). We have also tailored this framework specifically for application to management of contaminated sediments (Linkov et al. 2006a). In general, we recommend a four-stage process of *elicitation*, *characterization*, *assessment* and *analytic-deliberation*. The work presented in this paper relies upon the results of the first three steps to help facilitate the final, deliberative step in which a decision is actually made. The overall goal of any MCDA process is not to make a decision, but rather to *support* a decision making process. In the case of group decision-making processes, one of the strengths of the approach presented is the ability to characterize the sensitivity of the result to different weightings, assessments or perspectives – especially in recognition of the fact that these perspectives may change. As managers gain experience with new spills or policies evolve, it is possible to revisit the process by considering new metrics,

revising assessments or weighting boundaries and revisiting the suite of metrics considered most applicable to the next spill. An adaptive management process is called. (See Linkov et al. 2006b for an example charting adaptive MCDA management processes).

The result of the entire process is a comprehensive, structured process for selecting the GPRA metrics for in any given situation, and vetting these metrics with stakeholder groups in a way that incorporates their value judgments as well as scientific modeling and risk analysis. This process could be of great benefit to management of oil spill response, where there is currently no structured approach for making justifiable and transparent determinations about the success of any particular response. In particular, an MCDA framework allows visualization and quantification of trade-offs. The proposed framework can also be used to prioritize research and information gathering activities and thus can be useful for value of information analysis.

7. References

- Belton V. and Stewart T. *Multiple Criteria Decision Analysis: An Integrated Approach*. Boston, MA: Kluwer Academic Publishers, 2002.
- Chess C, Johnson BB, Gibson G. 2005. Communicating about environmental indicators. *J Risk Research*. 8:63–75.
- Figueira J, Greco S, Ehrgott M, eds. 2005. *Multiple Criteria Decision Analysis: State of the Art Surveys*. New York: Springer.
- Gore A. 1993. *From Red Tape to Results: Creating a Government That Works Better & Costs Less*. Report of the National Performance Review. U.S. Government Printing Office: Washington DC.
- Grabowski MR, Harrald JH, Roberts KH. 1997. Decision support and organizational forms in a high velocity environment: Responses to catastrophic oil spills. In: Grabowski MR, Wallace WA, editors *Advances in Expert Systems for Management*, vol. 2. JAI Press: Greenwich CT.
- Harrald JR. 1994. Preparing for success: A systems approach to pollution response. Paper presented to the Five Years After the Exxon Valdez Oil Spill Conference, Anchorage Alaska.
- Harry MJ. 1994 Six Sigma: A Breakthrough Strategy for Profitability - Making strides along the road to defect-free work. *Quality Progress*. 31(5):60–66.
- Kiker G, Bridges T, Varghese AS, Seager TP, Linkov I. Application of multi-criteria decision analysis in environmental management. *Integr. Environ. Assess. Manage*. 2005, 1, 49–58.
- Klein G. 1999. *Sources of Power*. MIT Press: Cambridge MA.
- Kravchuck R and Schack R. 1996. Designing effective performance-measurement systems under the Government Performance and Results Act of 1993. *Public Administration Review*. 56(4):348–358.
- Lahdelma R, Salminen P. 2001. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. *Operations Research*. 49(3):444–454.
- Lahdelma R, Salminen P. 2002. Pseudo-criteria versus linear utility function in Stochastic Multicriteria Acceptability Analysis. *European Journal of Operational Research*. 141(2):454–469.
- Lahdelma R, Salminen P and Hokkanen J. 2000. Using multicriteria methods in environmental planning and management. *Environmental Management*. 26(6):595–605.
- Linkov I, Varghese A, Jamil S, Seager TP, Kiker G, Bridges T. 2004. Multi-criteria decision analysis: framework for applications in remedial planning for contaminated sites. In: Comparative Risk Assessment and Environmental Decision Making, Linkov I, Ramadan A, Ed. Kluwer, Amsterdam.
- Linkov I, Satterstrom FK, Seager TP, Kiker G, Bridges T, Belluck D, Meyer A. 2006a. Multi-criteria decision analysis: Comprehensive decision analysis tool for risk management of contaminated sediments. *Risk Analysis*. 26(1):61–78.
- Linkov I, Satterstrom FK, Kiker G, Bridges T, Benjamin S, Belluck D. 2006b. From optimization to adaptation: Shifting paradigms in environmental management and their application to remedial decisions. *Integrated Environmental Assessment and Management*. 2(1):92–98.

- NOAA. 2001. *Characteristics Of Response Strategies: A Guide For Spill Response Planning In Marine Environments*. National Oceanic and Atmospheric Association: Washington DC. http://response.restoration.noaa.gov/book_shelf/910_response.pdf.
- Seager TP. 2004. Understanding industrial ecology and the multiple dimensions of sustainability. In *Strategic Environmental Management* by O'Brien and Gere Engineers. New York: John Wiley & Sons, 2004.
- Tervonen T, Figueira JR, Lahdelma R, Salminen P. 2007. Towards robust ELECTRE III with simulation: Theory and software of SMAA-III. *Decision Support Systems*. Under review.
- Tuler SP, Seager TP, Kay R, Linkov I. 2007. Objectives and performance metrics in oil spill response: The *Bouchard-120* and Chalk Point spill responses. In *Environmental Security in Harbors and Coastal Areas* edited by Linkov I, Kiker G, Wenning R. Springer: Amsterdam, The Netherlands. In press. pp. 209-228.
- Tuler SP 1988. Individual, group and organizational decision-making in technological emergencies. *Indust. Crisis Quarterly*. 2:109-138.
- Yoe C. *Trade-Off Analysis Planning and Procedures Guidebook*. Prepared for Institute for Water Resources, U.S. Army Corps of Engineers. April 2002.

8. Appendix: Objectives and metrics elicited for *Bouchard* and Chalk Point oil spill response

Bouchard-120

Type of metric	End-Point	Process	Resource
Protect worker and public health and safety	<ul style="list-style-type: none"> Number people killed or injured Number of mishaps during hours worked 	NA	<ul style="list-style-type: none"> Number of IRAC team members OSHA/HAZMAT trained
Protect environment/ ecological systems	<ul style="list-style-type: none"> No re-oiling or residual oil causing chronic toxicity to something Number of estuaries “shut off” (i.e., areas prevented from oiling from via booming) Number of inlets that were able to boom off Number of miles of shoreline impacted Amount of oil or globules on shore Number of areas where prevented oil contamination Oil contamination is on sandy beach (easier to clean-up than rocky inter-tidal zone) Shellfish beds protected Number of miles that have been cleaned to an acceptable level How many pounds of oil contaminated debris has been recovered and sent off for disposal? How many gallons of oil have been recovered? Number of birds oiled? Number birds released and survived Mortality/survival of wildlife (birds) Number roseate terns lost Number dead birds per unit search area Number dead birds. Percent birds rehabilitated Number of birds rehabilitated Have clean-up standards for shoreline contamination been met? 	<ul style="list-style-type: none"> Did getting required permits delay response action? Booming deployed around critical habitat (Roseate tern nesting habitat, salt marshes)? Attempt to direct oil to “sacrificial areas”? Good choices made about where to deploy boom? Oil being captured in open water before it hits the beach? Proper placement of skimmers? Resources placed in the proper locations How fast [# of days/ # of hours] rehabilitation efforts brought in and set up operations Is rehabilitation center handling all live birds coming in? Speed at which able to handle birds coming in for rehabilitation? Amount of area covered in search and recovery Volunteers able to get plugged in well? Time it takes to implement tasks, such as boom deployment? 	<ul style="list-style-type: none"> Number sandbags deployed? Number people on cleanup crews to deal with oiled beaches? Amount of boom deployed Number of floating resources to pick up oil in open water oil? Number skimming units available and operable Number of volunteers Number of bodies to manage different aspects of response, including organizing SCATs? Is there a ‘bird searcher’ on each team?
Establish a coordinated and effective response framework	<ul style="list-style-type: none"> Modelers able to get 24 hours ahead of spill with accurate projections Are pre-identified areas potentially being affected? Did we have information to keep Governor’s office and other state Senators and Reps abreast of what was going on Understanding of whether oil is still stored offshore re-contaminating cleaned up beaches? Is oil coming on shore several days later? Accurate accounting of volume oil spilled? 	<ul style="list-style-type: none"> How quickly decisions made? Chain of command clear? Resources placed in the proper locations Are decisions correct (in hindsight)? Location of command center convenient and easily accessible? Are crews visiting hard hit areas every day? Shoreline assessment teams record information in a unified way 1) where is oil, 2) how much, and 3) what does shoreline look like; want minimal variability Organized, systematic way of recording where oil has come 	<ul style="list-style-type: none"> Number of teams of trained observers walk coastline and make observations of extent and coverage area of oil Number crews trained Type of oil Phones available and working Number of hours worked Number supervisors out in field

	<ul style="list-style-type: none"> • Accurate accounting of amount of oil on the shore? • How many areas have been cleaned as of today? • Number of miles of shoreline impacted right now? • Bad feelings amongst locals responders toward the unified command staff? • Level of conflict about plans and activities? 	<p>ashore and impacted is used</p> <ul style="list-style-type: none"> • Systematic, ‘non-political’ approach used to deploy clean-up crews • Presence of watchdog to see what’s going on? • Is there conflict or chaos in command center? • Are players familiar with each other? and reliable communication technology available (e.g., cell phones working)? • System established to track progress? • Experts consulted for input on response strategies? • Command center set-up and people mobilized? • Plans communicated day in advance? • Ability to effectively and timely monitor what happening in field? • Are there clear protocols? • Are there clear schedules? • Time it takes to implement tasks, such as boom deployment? • Is all pertinent data gathered and recorded? • Local responders notified quickly? • Conflicts and heated emotions among responders addressed? • All goals worked on together? • Attend to short, medium, and long term needs simultaneously? • Follow ‘best response’ protocol and integrated command system? • Revise objectives and activities based on monitoring effort? • Coordination of volunteers performed? • Volunteers integrated into clean-up appropriately? • Volunteers integrated into clean-up quickly? • Equipment and personnel demobilized when no longer needed? 	
Implement an effective and timely response	<ul style="list-style-type: none"> • Volume spilled compared to what’s removed from the shoreline • What it looks like a year later • Evidence of oil on shoreline • How much oil was recovered off the water? • How much oil was recovered off the shore? • Endpoints in plan achieved? • Are desired endpoints achieved in reasonable amount of time? • Number of days until endpoints achieved? • Are response actions having the desired effect? • Members of the public voice support? • Number of public meetings organized? 	<ul style="list-style-type: none"> • Cleanup of impacted areas organized to be manageable and able to monitor? • Access established for recovery and clean-up crews through private property? • Plans communicated day in advance? • Location of command center convenient and easily accessible? • Ability to effectively and timely monitor what happening in field? • Muster all forces in town? • Local officials set up task force for how to handle spill response? • Clear and reliable communication technology available (e.g., cell phones working)? • Do efforts correspond to tides? 	<ul style="list-style-type: none"> • Number of people in the field? • A ‘bird searcher’ on each SCAT? • Hours worked? • Supervisors out in field? • Number of monitors in field to give direction and warning to clean up crews • Resources adequate for planned tasks? • Amount of boom deployed? • How much money spent? • Types of skills represented on team? • Number of volunteers? • Number of packets of baby oil for oil

	<ul style="list-style-type: none"> • Good working relationships with all parties involved? • command center and command personnel located in accessible location 	<ul style="list-style-type: none"> • Are there clear protocols? • Are there clear schedules? • Data sheets available and sufficient? • Are clean-up crews assigned effectively to do a good job? • Recovery or rehabilitation of wildlife conducted? • How soon have experts been called and set up triage and rehab centers? • Time it takes to implement tasks, such as boom deployment? • Is entire shoreline being inspected – both oiled and non-oiled? • How often is beach searched, all shoreline searched? • Is all pertinent data gathered and recorded? • Basic training for volunteers on bird collection and rehabilitation conducted? • Equipment and personnel demobilized when no longer needed? 	removal distributed
Mitigate economic impacts	<ul style="list-style-type: none"> • Lost rental income? • Lost tourism dollars? • Loss of summer rentals? • When beaches put back in use? • Recreational opportunities that were lost that are now back to what they were? • Acres of shellfish areas closed? • How many acres of closures reopened? • Duration of shell fish bed closures? • Acres of shellfish beds lost? • Costs of laboratory work? • Money spent on response? 		
Meet legal and regulatory requirements	<ul style="list-style-type: none"> • Achieve termination endpoints? • No oil should come off to the touch • Shoreline back to conditions prior to spill? • evaluate response with respect to endpoints achieved • For sandy beach no visible oil • For groin (jetties between properties), riprap no sheen or no oil available when touched • For marshes no sheen <p>No odor of oil</p>	<ul style="list-style-type: none"> • Number of days until endpoints achieved. 	
Mitigate social nuisance impacts	<ul style="list-style-type: none"> • Presence of stained rocks that lead people to say that it's not successful because you can still see evidence of that? • Complaints about stained rocks? • Oil on the beach? 		
Address needs and concerns of the affected	<ul style="list-style-type: none"> • Level of staining and other aesthetic damage? • Complaints about stained rocks? 	<ul style="list-style-type: none"> • A timeframe for ending the cleanup is established • Ongoing monitoring and addressing of issues post-spill? 	<ul style="list-style-type: none"> • Number of pamphlets distributed to inform public of hazards

public/communities	<ul style="list-style-type: none"> • Residual oil on shore (tar balls)? • Number of phone calls to complain? • Spirit of state regulations for public involvement was met 	<ul style="list-style-type: none"> • People have a place or someone to go to with concerns? • Public receives assurances that beaches will be cleaned up to the level of their expectation? • Information fairs and public meetings set-up? • A forum is provided to public so they can hear what's going on in response action and give their feedback • Number of public meetings • Level of conflict/anger • People are happy • Number of visuals for media • People given examples of what was impacted and what kinds of cleanup was going on • Public provided the kinds of information it wants? • Establish and keep up to date website for public information 	<ul style="list-style-type: none"> • Number of stakeholders involved in setting clean-up standards • Number of dispatch teams arranged to reach-out to various stakeholders • Frequency of information postings on BBP website • Number of flyers passed out about educating public on risks
Gain public support for the response	<ul style="list-style-type: none"> • Members of the public voice support? • Level of staining and other aesthetic damage? • Complaints about stained rocks? • Residual oil on shore (tar balls)? • Number of phone calls to complain? • Was there trust from the public? • Public comments of support from critics and local residents (about response effort, not about outcomes)? • Spirit of state regulations for public involvement was met 	<ul style="list-style-type: none"> • A timeframe for ending the cleanup is established • Ongoing monitoring and addressing of issues post-spill? • People have a place or someone to go to with concerns? • Public receives assurances that beaches will be cleaned up to the level of their expectation? • Information fairs and public meetings set-up? • Relationships and trust with local officials developed? • A forum is provided to public so they can hear what's going on in response action and give their feedback • Number of public meetings • Level of conflict/anger • People are happy • Number of visuals for media • People given examples of what was impacted and what kinds of cleanup was going on • Public provided the kinds of information it wants? • Quality of questions from media • Establish and keep up to date website for public information • Able to 'stay on message' during public meetings, press conferences, etc.? • Unified Command accessible for public questions and comments? 	<ul style="list-style-type: none"> • Number of pamphlets distributed to inform public of hazards • Number of stakeholders involved in setting clean-up standards • Number of dispatch teams arranged to reach-out to various stakeholders • Frequency of information postings on BBP website • Number of flyers passed out about educating public on risks

Chalk Point

Type of metric	End-Point	Process	Resource
Protect worker and public health and safety	<ul style="list-style-type: none"> • Presence of PHCs in water samples? • Concentrations of oil in fish samples • Level of PHCs in clam/oyster tissue • Number life threatening situations to human health • Toxins in smoke plume if do in situ burning 		<ul style="list-style-type: none"> •
Protect environment and mitigate environmental impacts	<ul style="list-style-type: none"> • Number of oiled swans • Number of fish killed • Damage from clean-up actions remediated? • Number of animals saved • “Appropriate” wildlife saved (e.g., not exotics or ‘varmints’) • Oiling of osprey eggs • Number of dead ducks • Number Injured birds around marsh area • Oiling on Diamondback terrapin eggs • Mortality of wildlife • How far sheen at surface extended out [miles]? • How long oil stayed? • Presence of odors of oil • How much grass was destroyed? • How many acres of marsh were impacted? • How many oiled birds were observed? • Dead and stressed organisms found (rather than estimated)? • How long it takes to reach background levels/concentrations • Observe water blowing over booms • How long does it take for area to recover from clean-up related damage? • Degree of change to beaches and sandbars from clean-up actions • Types of animals and vegetation present after spill cleanup • Areas that were impacted cleaned up to conditions before the spill happened 	<ul style="list-style-type: none"> • Accurate cataloguing and enumeration of what found • monitoring stations established • Deploy booming – and double-booming in sensitive areas • Immediacy of rehabilitator organization’s response to call for assistance • Change of helicopter flight patterns in response to requests from biologists to not disturb nesting birds 	<ul style="list-style-type: none"> • Presence of booms • Amount of oil containment boom deployed
Protect cultural resources	<ul style="list-style-type: none"> • Number of critical sites protected • Residual presence of oil on artifacts • Smell of oil • Soil concentrations of oil 	<ul style="list-style-type: none"> • Did trench digging affect sites? • Experts contacted early for input about sites potentially at risk? • Command responsive to requests for protection of sites (e.g., deployment of booms) • Were less destructive response actions chosen (e.g., sorbents and booms rather than burning) 	<ul style="list-style-type: none"> • Number of GIS & hard maps • Number of laptops • Amount of boom deployed • Accurate spatial information about location of sites and artifacts
Meet legal and regulatory requirements		<ul style="list-style-type: none"> • All procedures followed (e.g., NIMS) 	
Mitigate economic impacts	<ul style="list-style-type: none"> • Devaluation of property? • Number dead fish 	<ul style="list-style-type: none"> • Funds spent on important and appropriate projects 	<ul style="list-style-type: none"> • Amount of money spent on studies

	<ul style="list-style-type: none"> • Number dead ducks & geese • Lost wages to watermen • Number of lost fishing days • Dollar value for days of beach closures • Change in earnings relative to previous years • Cost of response actions compared to value of lost resources (e.g., artifacts) • Duration of recreational fishing closures • Duration of beach closures • Presence of PHCs in water samples • Increase in crime rate [in southern MD due to influx of people from cleanup crews] • Monetary compensations • PEPCO made restitution to counties for employee overtime and personal equipment used in helping response efforts 	<ul style="list-style-type: none"> • RP cleaned boats that were soiled with oil from spill 	<ul style="list-style-type: none"> • Total cost of clean-up
Establish a coordinated and effective response framework	<ul style="list-style-type: none"> • Clear chain of command established? • Incident Command System used? 	<ul style="list-style-type: none"> • Number of hours to set-up incident command center • Number of daily meetings • Daily reports for morning meetings held by UC? • Informed of meetings in advance (i.e., lead time) • RP is responsive • Key people became involved early? • Is Incident Commander able to reach key people, does their phone number work, did they respond? • Clear communication protocols? • Stable command posts established? • Time taken to re-staffing response people after contractor fired • Chauvinistic behavior among staff in command center? • Frequency of resource and personnel changes • All federal responders notified quickly? • Pick-up and shipping schedule for waste generated by clean-up organized? • Response organized by discrete (river) segments • Frequency of meetings • Accurate information obtained from the wildlife surveys and SCAT teams • Clear standards for sign-off established • People in field used to verify accuracy of information 	<ul style="list-style-type: none"> • Number of supervisors assigned per section • Enough Spanish speaking crew supervisors? • Number of radios available • Availability of GIS and computers

		<ul style="list-style-type: none"> • Did Unified Command resist information that did not conform to their expectations?) • Arguments and lack of cooperation among Unified Command • Clarity to all parties about stages of response effort • Clear understanding of rights the state trustees have as a state agency 	
Implement an effective and timely response	<ul style="list-style-type: none"> • Number feet of shoreline impacted • Presence/absence of sheen • Number gallons oil recovered • Number of days company (PEPCO) was shut down • No oil in water • How long takes to reach endpoints • Oil present in sediments? • Areas that were impacted to be cleaned up to the ecological state the environment was before the spill happened • Patchiness of oil • Solid sheet of heavy oil • Presence/absence of tarballs • Is oil being contained? • Number of areas to be signed-off compared to Number actually signed-off • Number of areas that were identified as sensitive areas ordered to be boomed off but order was not followed through and area impacted 	<ul style="list-style-type: none"> • Digging holes to look for oil on shoreline? • Quarterly checks to see if oil is present or not on beaches • Breach of water over boom? • Protection equipment put in the right place at the right time? • Number of times a single order was given but still not done • Accurate reporting and counting of crews in field • Number of newsletters published by RP • Number of public meetings held by RP • Time it takes to get response in order • Number/frequency of fly-overs for real time aerial photography • Time taken to set-up boom • Clear chain of command established? • Lead resources mobilized? • Communication funneled out to the appropriate people? • Quality of contractor work (e.g., perform required tasks?) • How quickly SCATs out in field? • Time it takes to resolve bad cell phone coverage, radio issues? 	<ul style="list-style-type: none"> • Cost of response • Number of cleanup teams and people on crews • Amount of money spent by RP • Number of laptops • Number of GIS & hard maps • Amount of equipment brought in • Pounds of sorbent material used • Amount of oil containment boom deployed • Number of people working at one time • Number of people per SCAT
Address needs and concerns of the affected public/communities	<ul style="list-style-type: none"> • Public reimbursements for private property losses • Number calls from public • Number of fliers and informational packets delivered door to door • Amount of oil removed manually from shoreline 	<ul style="list-style-type: none"> • Immediacy of public meetings (number days after spill occurs) • Incorrect information disclosed? 	<ul style="list-style-type: none"> • Number hours agencies spent on public outreach (meetings)
Gain support of public	<ul style="list-style-type: none"> • Public reimbursements for private property losses • Number calls from public • Number of fliers and informational packets delivered door to door 	<ul style="list-style-type: none"> • Immediacy of public meetings (number days after spill occurs) • Incorrect information disclosed? 	<ul style="list-style-type: none"> • Number hours agencies spent on public outreach (meetings)