

Typological Review of Environmental Performance Metrics (with Illustrative Examples for Oil Spill Response)



Thomas P Seager,[†] F Kyle Satterstrom,[‡] Igor Linkov,^{*‡} Seth P Tuler,[§] and Rebecca Kay[§]

[†]Civil Engineering, Purdue University, West Lafayette, Indiana 47907, USA

[‡]INTERTOX, 83 Winchester Street, Suite 1, Brookline, Massachusetts 02446, USA

[§]Social and Environmental Research Institute, 278 Main Street, Room 404, Greenfield, Massachusetts 01301, USA

(Received 6 March 2006; Accepted 16 July 2006)

ABSTRACT

An intensification of interest in environmental assessment during the last 2 decades has driven corporate efforts to better document environmental goals, improve environmental management systems, and increase awareness of the environmental and ecological effects of business operations. This trend has been motivated partly by regulatory requirements (such as the Toxics Release Inventory in the United States) and partly by the inclination of some large manufacturing firms to embrace a broader social and environmental mission characterized as “sustainability” or “eco-efficiency.” Moreover, the importance of measurable objectives in the US government has been recognized at least since the Government Performance Results Act of 1993, which was intended to both improve the efficiency of government and the confidence of the American public in government managers. However, in management of environmental crises—such as catastrophic oil or chemical spills—development of measurable performance standards has lagged. Consequently, government spill managers are unable to define success in terms that are easily communicated to public and other stakeholder groups, and they could be disadvantaged in their efforts to deploy response resources with maximum efficiency. In this paper, we present a typological review of environmental assessment measures and summarize some of the current practices and strategic goals among federal agencies with regard to oil and chemical spills. A general approach to organizing metrics for oil spill response, restoration, and recovery is also presented. The results could improve planning efforts and communication among different federal, state, and local agencies and public or stakeholder groups involved in spill management.

Keywords: Oil spills Performance metrics Crisis management Stakeholder communication

INTRODUCTION

A series of environment crises in the mid- to late 1980s significantly raised public and private awareness of the need to be able to assess the broader environmental and ecological effects of industrial activities. Incidents such as the catastrophic release of toxic chemicals from a Union Carbide plant in Bhopal, India, discovery of the stratospheric ozone hole above Antarctica, and the grounding of the *Exxon Valdez* in Prince William Sound, Alaska, USA, focused intense interest on developing new methods of managing environmental risks that emphasize corporate accountability and quantitative management.

The Bhopal case is credited with motivating passage of amendments to the Emergency Planning and Community Right-to-Know Act of 1986 (USEPA 2002) and the Pollution Prevention Act of 1990 that created the Toxics Release Inventory (USEPA 2005b), which requires publication of the total mass of reportable chemical releases by factories in the United States (Neumann 1998). Since its inception, the TRI has become a focal point of public attention that motivates bad actors found at the top of the list of worst polluters to reduce chemical releases as a way of deflecting negative attention (Hamilton 1995; Konar and Cohen 1997). Consequently, total reported releases have trended down for the last decade (Figure 1).

The discovery of the ozone hole resulted in a nearly global ban on manufacture of chlorofluorocarbons (and their initial substitutes, hydrochlorofluorocarbons). With respect to stratospheric ozone depletion, the environmental effects of these compounds are characterized by ozone depletion potential (ODP), a measure of the capacity of a chemical released at the surface of the earth to destroy ozone in the stratosphere relative to the destructive capacity of CFC-11. Negotiation of the international treaty (the Montreal Protocol) that promulgated the manufacturing ban relied heavily on the ODP measure (WMO 2003). The success of the ODP approach has led to a proliferation of other novel environmental assessment measures such as global warming potential (GWP; WMO 2003), tropospheric ozone formation potential, human toxicity potential, pollution potential, and others (e.g., Hertwich et al. 1997; Seager and Theis 2004). Each measure is designed to inform management, policy, or design decisions in relation to environmental or ecological dimensions that might have not been considered by previous generations. In the sense that stratospheric ozone levels have stabilized and likely increased in recent years (Figure 2), the ODP concept has contributed to the successful design of effective policy measures.

By contrast, the regulatory response to the 1989 *Exxon Valdez* catastrophe has emphasized better planning and coordination, increased accountability, investment in response equipment and technologies, and preventive shipping regulations, such as double-hulled tankers or crew work rules (USEPA 1990; Tannahill and Steen 2001). As part of ongoing

* To whom correspondence may be addressed:
ilinkov@intertox.com

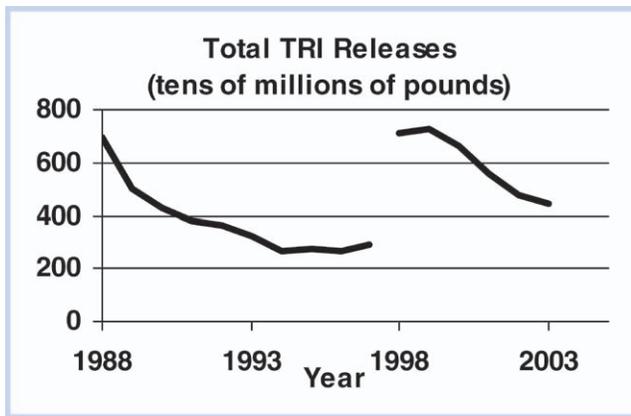


Figure 1. Total reportable toxic releases have trended downward since reporting requirements have taken effect. The total number of industries required to report expanded in 1998, accounting for the jump in that year (USEPA 2005b).

efforts to improve national, state, and local spill preparedness, extensive planning and coordinating efforts have been undertaken during the last 15 y, culminating in the creation of the National Response Plan (NRP; DHS 2004). In the case of oil spills, the NRP

... describes the lead coordination roles, the division and specification of responsibilities among Federal agencies (particularly the US Coast Guard, and US Environmental Protection Agency [USEPA]) under anticipated crisis scenarios and the national, regional, and onsite response organizations, personnel, and resources that may be used to support response actions (p ESF 10-1).

However, the NRP has not resulted in the advancement of quantitative strategic goals to track the success of efforts to minimize the damage caused by oil spills. Although the frequency of domestic spills has declined precipitously since passage of the Oil Pollution Act of 1990 (Figure 3), the effectiveness of response efforts is more difficult to gauge (partly because spills are rarer than ever; Kim 2002). Depending on the location, size, timing, and environmental conditions of the spill, the potential ecological and economic effects are highly variable.

According to NOAA (2001),

... oil spill response goals, in order of priority, are to 1) maintain safety of human life, 2) stabilize the situation to preclude it from worsening, and 3) minimize adverse environmental and socio-economic effects by coordinating all containment and removal activities to carry out a timely, effective response.

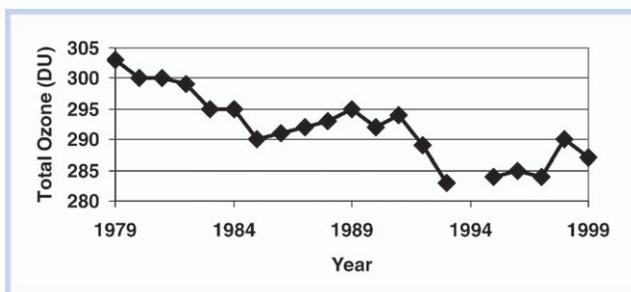


Figure 2. Global averaged total stratospheric ozone levels, measured in Dobson units, stabilized in the late 1990s as a result of the Montreal Protocol prohibition on manufacture of ozone-depleting substances (adapted from Gorova et al. 2001).

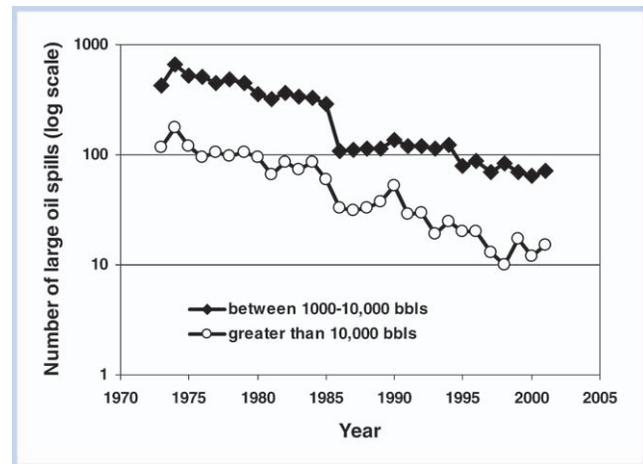


Figure 3. Number of oil spills greater than 1,000 bbl in US waters (from US Coast Guard Office of Investigations and Analysis, <http://www.uscg.mil/hq/g-m/nmc/response/stats/aa.htm>. Last updated August 2003).

However, establishing a baseline context to measure the effectiveness of a response is extremely challenging. Because incident-specific strategies must be identified early and on a case-by-case basis, the complexity and challenge faced by spill managers increases (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, setting objectives, tracking progress, and communicating or determining success is an ad hoc process depending 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 Harrald 1994; Kuchin and Hereth 1999; Chess et al. 2005).

On the other hand, in private organizations (such as manufacturing firms), recognition of the need to evaluate the effectiveness of environmental practices, remedial efforts, and performance goals and measures as part of a comprehensive environmental management strategy has been increasing (e.g., GEMI 1998; OECD 1998; CIEPM 1999; Schulze 1999; Olsthoorn et al. 2001). More stringent reporting requirements, tighter product or emission constraints, consumer demands, the availability of international management standards (i.e., ISO 14001) and international agreements (such as the Montreal and Kyoto Protocols) have driven industry attention toward more quantitative management practices and improved environmental assessment methods. In the last 15 y, environmental performance indicators and assessment tools, such as ecoefficiency, life-cycle assessment, ecological footprint analysis, and others, have expanded enormously (Hammond et al. 1995; Gray and Wiedemann 1999).

In this paper, we characterize current quantitative environmental assessment metric types, discuss the qualities that make for effective versus ineffective metrics, contrast government

and industry practices, and propose a new typology of oil spill performance metrics that could improve communication between different stakeholder and responder organizations and facilitate analysis of the different perspectives, objectives, or concerns of agency and nongovernment personnel.

THE EVOLUTION OF ENVIRONMENTAL ASSESSMENT

As the focus on environmental management has shifted from reactive or remedial to proactive and preventive, industrial firms have implemented more sophisticated environmental management systems to better document chemical releases, resource consumption, and potential environmental and ecological effects. Concurrent with these developments in industry has been an increased emphasis by researchers on the quantitative assessment of the environmental and ecological effects of industrial activities. Although historically assessment had been focused on the toxicological properties of specific chemicals, the success of new measures such as ODP and GWP in gauging nontoxic hazards has spurred a rapid expansion of measures designed to inform managers, policy makers, and designers of broader implications such as smog formation, acidification, biodiversity, ecosystem health, or resource depletion. In general, the boundaries of interest have expanded beyond the classic toxicological approach to ecotoxicology and even broader indicators of ecosystem health (e.g., Rapport 1999) or sustainability (e.g., Hammond et al. 1995; Sikdar et al. 2004). In ecotoxicology, the human toxicological model is extended to more complete ecosystems, including exposure through food webs and bioaccumulative effects (Schuurmann and Market 1998). In sustainability, the primary tool of quantitative analysis is life cycle assessment, in which the emphasis is on the industrial product or material chain and in aggregating all resources consumed (such as energy) and chemical releases concomitant to resource extraction, production, use, and disposal of a specific product (Seager and Theis 2004).

The expansion of interests and metrics has led to an increased level of sophistication in both the interpretation of assessments and the methods for conducting them. However, managers and researchers still have some considerable doubt as to which metrics are most important, whether metrics accurately capture the intended effect, and whether the metrics employed will have meaning to customers or stakeholder groups (e.g., Chess et al. 2005). Once a metric is implemented, it becomes a tool for prioritization, resource allocation, or intentional structuring of management efforts to shape a system in accordance with organizational objectives. Therefore, the selected metrics are an expression of the values that guide the activities of an organization and must be designed with the objectives of the organization in mind.

Three general approaches characterize the environmental metrics described in the literature. Metrics can be sorted by their mathematical properties, relation to organizational objectives, or position in a cause and effect chronological sequence. Each approach is briefly described in subsequent paragraphs, with an emphasis on examples relevant to oil spill response. Additionally, we introduce a new value-based approach to characterizing performance metrics in terms of whether they relate to economic, thermodynamic, environmental, ecological, human health, or sociopolitical values. Each of these characterization approaches is intended to be complementary rather than comprehensive.

Mathematical properties

Metrics can be quantitative (such as length), semiquantitative (such as an ordinal ranking), nonquantitative (such as a favorite color), or qualitative (better or worse), although without a context for comparison, any metric is meaningless. For simple systems, metrics can be easy to enumerate and interpret and inexpensive to gather data on. However, establishing metrics for complex environmental and ecological systems presents a significant challenge. Both natural and human systems are complicated and relate to one another in a myriad number of ways. Consequently, any set of metrics is incomplete and could at best be considered only representative of the decision factors that could be brought to bear on the situation. For this reason, metrics are often referred to as “indicators” to emphasize the representational relationship these measures have to the state of complex systems. They are indicative—but not definitive—gauges and consequently must be interpreted with their limitations in mind.

The total amount of information obtainable to describe the state of an ecological system can be infinite. As the quantity of information increases, the ease of interpretation of that information decreases. Therefore, it is essential to aggregate measures to provide a simpler assessment of progress with respect to a single dimension. Aggregation mathematically combines related measures, for example, by summing, averaging, or combining by more complex methods such as net present value computation. Data can be aggregated over a geographic area, over time, or over other independent variables such as species, habitat type, or demographic profile. Aggregated data are easier to work with but contain less information than the original data set from which aggregated measures are compiled. Moreover, the mathematical methods used to aggregate different measures can constrain or confuse the interpretation of those methods. Particular attention must be paid to aggregation of data that are expressed in different units. Methodologically unsound approaches to aggregation might render information meaningless or cause managers to reach unsound conclusions. For example, intensive measures characterize a system in terms of a ratio, fraction, or percentage, such as concentration (e.g., mg/L or ppm) or miles per gallon. These are independent of the total size of the system. Therefore, it is not proper to add or subtract intensive measures without first converting these to extensive measures (such as mass, volume, or miles) expressed in units that are identical.

Even more problematic is aggregating data that might relate to qualitatively different objectives. For example, it is common to assess the severity of an oil spill by estimating the volume of oil released into the environment. However, the ecotoxicity of oil varies according to the type of oil and particular hydrocarbon components. Where spills of different types are compared, an ecotoxicity-weighted approach might provide a different perspective than a simple mass-based approach. With regard to the TRI, a toxicity-weighted aggregation has been proposed as an improvement on conventional mass-based reporting (Horvath et al. 1995). Whenever qualitative characteristics can differentiate data, aggregation inevitably involves application of a value-based weighting scheme (such as a weighted average) that emphasizes some aspects more than others.

Organizational objective

Decision making in any organization typically includes 3 levels of thinking: Strategic, tactical, and operational. The

strategic level is the broadest level. It typically involves longer term planning and is intended to align all components of an organization toward realization of strategic objectives. Tactical decision making typically engages an intermediate time frame. At the tactical level, organizational units may select from several alternative approaches to implementing a strategy, especially with regard to the response of other groups or systems to the alternatives chosen. Last, operational decision making engages the shortest time frames. Operations typically are those specific actions that together form the tactical alternative.

In relation to oil spills, strategic thinking can involve prevention, preparedness, response, mitigation, or restoration. In the case of response planning, specifically, decision makers must be prepared for many different contingencies. Strategy can involve the purchase and pre-positioning of equipment, delineation of responsibilities or organizational authority, or dedication of other resources. Therefore, decisions made at the strategic level create and constrain alternatives at the tactical level (Wilhelm and Srinivasa 1997). In response to a specific spill, tactical decision making involves the deployment of resources and selection of alternatives specific to the circumstances. At the operational level, the effectiveness of the tactical decisions must be assessed in relation to specific equipment or other resources. For example, protection of an estuary in an oil spill might involve a strategy of containment to prevent incursion of a spill into estuarine tributaries. Containment booms must be purchased and pre-positioned to enable this strategy. In the event of a spill, the tactical response might be to deploy booms in the areas considered at risk, whereas at the operational level, the effectiveness of the booms in containing the slick must be assessed. If either the operational execution fails (e.g., the slick runs under the boom) or the tactical response is inadequate (e.g., booms are not deployed), then the overall strategy could fail. Similarly, if the strategy is flawed (e.g., oil becomes uncontainable by either sinking or emulsifying), the tactical and operational efforts based on that strategy could be pointless. To monitor the effectiveness of strategic, tactical, and operational efforts, it is essential to design metrics that inform all 3 levels of thinking.

Cause-effect relation

With regard to environmental risks, indicators can be characterized as descriptive of 3 different stages of hazard development: Pressure, state, or response (Gray and Wiedemann 1999). Pressure indicators relate to the level of stress placed on the environment by human systems, whereas state indicators relate to characterization of environmental-ecological systems. Response indicators relate to the changes in human systems that eventually result from the overall chain of cause and effect relationships. For example, oil spills put pressure on the environment, thereby effecting a change in the environmental or ecological state, such as the presence of a slick or reduced bird populations. The anthropogenic response might be mechanical removal, chemical dispersants, burning, bioremediation of the oil, or restoration of bird habitat. Ultimately, the human response could be political (such as increased regulation) rather than technological.

Many different types of cause-effect relational chains can be embedded within the pressure, state, and response stages. Moreover, different groups might perceive causal linkages differently, leading to misunderstandings (Webler et al. 1995). For example, in the case of oil spill response, it might be difficult

to link response operations to the economic and ecological effects that the response is intended to minimize, partly because the systems are complex and partly because damage assessment necessarily lags response decision making in time. Therefore, leading and intermediate indicators are necessary to provide early feedback regarding response effectiveness.

Value-based

The existing literature characterizing indicators and performance metrics emphasizes the way the metric is expressed (mathematical), the purpose of the metric (within the organization), and the cause-effect relationships between different indicators. However, more recent attention has shifted to indicators as an expression of the values of an organization and as a method of facilitating communication both within the organization and with outside or stakeholder groups (e.g., Chess et al. 2005). In this regard, it is helpful to create a taxonomy that classifies different indicators according to their qualitative, value-based characteristics (Seager and Theis 2004).

Virtually all metrics relevant to chemical release management can be characterized into 6 broad dimensions: Economic; thermodynamic; environmental; ecological; human health; and sociopolitical. However, no single tool or approach encompasses all of these dimensions. For example, in the life cycle assessment model developed by the USEPA (USEPA 2005b), effects are broken down into 6 categories: Ozone depletion, global warming, photochemical oxidation, eutrophication, acidification, and human and environmental health effects. Although each of these relates to a recognizable environmental problem, the purpose of the life cycle approach is narrowly focused on environmental assessment rather than holistic decision making. As they relate to oil spills, the broader categories are described next (and in further detail in Seager and Theis 2004).

Economic—In addition to direct and indirect costs, economic metrics convert nonmarket resources or effects into monetary values to allow comparison with monetary transactions or industrial accounts. Economic estimates of non-market effects are required by benefit-cost analysis for estimating the value of damages caused by an oil spill in terms of fish catch, property damage, clean up costs, or for prioritizing new investments. Broader economic analysis could include estimates of lost tourism revenues, decreased property values, or opportunity costs (Loureiro et al. 2006). In theory, proper pricing of environmental goods and services could allow market forces to optimally allocate resources between ecological and industrial activities. However, in practice, both the calculation methods and the validity of the concept of pricing the environment are recognized as controversial. Because most environmental goods, such as pollution attenuation, have no markets, external or social costs are highly uncertain, as are the methods and figures reported for the value of ecosystem services. Moreover, monetization could lead to the erroneous assumption that environmental exploitation can be revocable in a manner analogous to pecuniary transactions, although in some cases, ecological systems could be damaged beyond recovery.

Thermodynamic—Metrics such as total pollutant loading or release are indicative of environmental pressure (e.g., pollution to be attenuated), whereas measures such as energy use are more indicative of resource consumption or scarcity. Sometimes, thermodynamic metrics are normalized to intensive units (e.g., kg/person) or oil equivalents (energy/

product), which attempt to capture the ecoefficiency of a process. However, in the case of oil spills, extensive measures such as total barrels lost or recovered are appropriate. Usually thermodynamic metrics do not indicate the specific environmental response associated with resource consumption or loss. For example, the severity of an oil spill could be determined on the basis of total volume spilled. Nonetheless, ecological effects are dependent on a number of other factors, such as the type of oil and the location, mobility, and timing of the spill. On the basis of a thermodynamic measure called “energy,” which measures energy consumption in terms of the equivalent solar energy required to replace the consumption, Odum (1996) criticized the extensive clean up efforts that followed the grounding of the *Exxon Valdez* as an unproductive deployment of energy resources. His study claimed that more diesel fuel was expended on clean up efforts than barrels of oil were lost in the spill. Nonetheless, thermodynamic metrics are only indirectly related to the human and ecological health objectives that guide oil spill response—conservation of diesel fuel is not the primary objective of any large spill response.

Environmental—In measuring the extent of chemical change or hazard in the environment, environmental metrics often use physical or chemical units such as pH, temperature, or concentration. Concentration measures—especially for toxic oil components such as polycyclic aromatic hydrocarbons—are difficult to put in an appropriate context unless they are tied to some ecological or human manifestation, such as carcinogenicity, mutagenicity, or even nonhealth-based endpoints such as beach or fishery closures. Environmental metrics are generally measures of the residuals released by industrial processes that pressure the environment or are indicative of the environmental state (e.g., chemical contamination). Similarly, response (such as cleanup or remediation) actions are often motivated by measurable environmental objectives, such as reducing contaminant concentrations.

Ecological—Ecological metrics attempt to estimate the effects of human intervention on natural systems in ways that are related to living things and ecosystem functions. The rates of species extinction and loss of biodiversity are good examples, and they are incorporated into the concept of ecosystem health (Rapport et al. 1998; Rapport 1999). Oiled bird counts, marine mammal death counts, and time to ecological recovery are all examples of ecological metrics that are typically applied to oil spill damage assessments and restoration efforts. Although agreement among experts on the importance and relevancy of ecological outcomes might be good, considerable disagreement could still be present about the response alternatives (such as mechanical removal, in situ burning, dispersion, or natural attenuation) that will best achieve the ecological objectives. When ecological systems are unable to recover naturally from the effects of a spill, responsible parties are typically required to sponsor restoration efforts that return ecological systems to an approximation of their undamaged state.

Human health metrics—Human health metrics are indicative of the state of the human population just as ecological metrics indicate the state of natural systems. Human health includes worker and public safety. For example, worker injuries per total response hours worked is representative of one type of anthropocentric oil spill effect. Increased health risks could arise from inhalation exposure to toxic chemicals released to the environment. These risks are difficult to

aggregate because of the breadth of different endpoints entailed (Hofstetter and Hammit 2002). Protection of human health is a primary goal for federal agencies engaged in oil spill response, reflecting the primacy of anthropocentric concerns. Nevertheless, the measures devised (e.g., injuries, deaths, treated patients) could be directly analogous, if not identical, to those used to track animal effects.

Sociopolitical—These metrics evaluate whether industrial activities are consistent with political goals like energy independence or ecojustice or whether collaborative relationships exist that foster social solutions to shared problems. Major oil spills undoubtedly have far-reaching social and political effects (e.g., Shaw 1992). However, these are difficult to gauge quantitatively. In some cases, the political and social dimensions are translated or communicated primarily through the media. That is, although spill responders might understand the importance of public perception, they might have no basis for measuring improvement or deterioration of public sentiment, except through the tone of media coverage, which they could feel powerless to influence (Harrold 1994).

As a general rule, aggregating data across 2 different dimensions, such as economic and ecological, risks obfuscating the meaning of each dimension. Metrics that are designed to capture qualitatively different characteristics are incomparable. However, to gain an overall sense of the state of a system, it might be necessary to evaluate trade-offs between different dimensions. For example, how much money and energy should be spent cleaning beaches to improve environmental measures such as oil concentration or appearance if few (if any) ecological benefits result? Assessing such interdimensional trade-offs is a value-laden problem suitable for multicriteria decision analysis (Linkov, Satterstrom, Kiker, et al. 2006; Linkov, Satterstrom, Seager, et al. 2006), and it suggests a need for broad stakeholder involvement in setting of objectives to ensure that a wide range of values are represented (National Research Council 1996; Dietz and Stern 1998).

Application of the life cycle assessment methods developed for industrial processes is not likely to be directly applicable to oil spills. Although the metrics themselves (such as GWP or human and ecological toxicity) might be accurate, the methods for gathering and analyzing data will likely be very different when the focus is on a catastrophic, unplanned release instead of on product life cycle. Moreover, the focus in oil spill management is on minimizing acute impact categories (such as ecotoxicity) rather than on chronic effects such as global warming or eutrophication.

Nonetheless, meaningful decision processes must inevitably rely on some credible assessment measures that are accessible or explainable to the public. To date, planning efforts have focused more on defining resource availability, agency responsibility, and coordination rather than definitions of success and feedback measures. Therefore, it is essential to consider what information is available to spill responders, when it is available, the quality of the information, and its relevance to the purpose of the response. An ideal metric would have several characteristics (Graedel and Allenby 2002; Seager and Theis 2004):

- It would be scientifically verifiable. Two independent assessments would yield similar results.
- It would be cost-effective. It would use technology that is economically feasible and does not require an intensive deployment of labor to track.

Table 1. Comprehensive scheme for characterizing indicators and performance metrics

Mathematical		Organizational objective		Cause–effect relation	
Quantitative		Strategic		Pressure	Resource
Semiquantitative		Tactical		State	Process
Nonquantitative		Operational		Response	Endpoint
Qualitative					
Value-based					
Economic	Thermodynamic	Environmental	Ecological	Human health	Sociopolitical

- It would be easy to communicate to a wide audience. The public would understand the scale and context and be able to interpret the metric with little additional explanation.
- It could be changed by human intervention. The metric would have a causal relationship between the state of the system and the variables that are under a decisionmaker's control. Metrics that are independent of human action do not inform a management, policy-making, or design process.
- It would be credible. It would be perceived by most of the stakeholders as accurately measuring what it is intended to measure.
- It would be scalable over an appropriate time period and geographic region. It would be indicative of short-, medium-, and long-term effects as appropriate. For example, it would not be meaningful to attempt to measure the effects of chronic low-level toxic dosages over a period of weeks or months, just as it would not be appropriate to average local environmental conditions over a widely varying region.
- It would be relevant. It would reflect the priorities of the public and other stakeholders and enhance the ability of spill managers, regulators, or both to faithfully execute their stewardship responsibilities. There is no point assembling a metric no one cares about.
- It would be sensitive enough to capture the minimum meaningful level of change or make the smallest distinctions that are still significant, and it would have uncertainty bounds that are easy to communicate.

It might be difficult, if not impossible, to find metrics that satisfy all of these conditions for all stakeholders. Nevertheless, a suite of metrics that have several of these characteristics could prove to be especially useful to spill managers. In press accounts, damage assessment reports, and conversations with oil spill experts, it is clear that many performance metrics are in play during any given spill. Although new metrics might enhance the information available to decision makers, what is most likely called for 1st is an explicit approach to organizing the metrics already available—that is, assessing the quality of the metric with respect to the ideal characteristics listed above, understanding what the metric is intended to describe and how, relating the metric to the proper level of organizational thinking, understanding the cause and effect relations between metrics, and knowing to whom the metrics are important.

Table 1 summarizes the previous sections to create a comprehensive scheme for characterizing indicators and performance metrics. Table 2 classifies a number of example oil spill metrics with regard to the multiple value-based dimensions described in the previous section.

ENVIRONMENTAL PERFORMANCE METRICS IN FEDERAL AGENCIES

Within the US government, the primary regulatory driver of performance measurement has been the Government Performance and Results Act (GPRA) of 1993. GPRA requires federal agencies to prepare performance reports that are then reviewed by the Office of Management and Budget (OMB). GPRA was enacted to “provide for the establishment of strategic planning and performance measurement in the Federal Government” (OMB 2005a). It embodied a push for better planning, greater accountability, and straightforward performance evaluation in government by requiring a federal program to have an overall strategic plan and to prepare annual performance plans and reports (OMB 2005b).

Many US federal agencies use performance measures to assess their progress toward achieving ecological or environmental goals. Different agencies are in differing stages of performance metric adoption. For example, the National Oceanic and Atmospheric Administration (NOAA) is explicit in articulating performance measures for increasing the number of fish stocks managed at sustainable levels. NOAA defines performance measures as “indicators of conditions in natural and human systems that have been selected as targets for restoration. Collectively, a well-selected set of performance measures provides a quantitative representation of the overall environmental health of these systems” (NOAA 2005b). Some agencies have explicit goals but leave performance measures implicit in their strategic plan documents. For example, the US Fish and Wildlife Service states the specific goal of delisting 15 species from the Endangered Species Act between 2000 and 2005. The number of listed endangered species is the metric implied.

The development of performance metrics is not always straightforward. Problems commonly arise when attempting to measure the performance of programs that

- have outcomes that are extremely difficult to measure,
- have many contributors to a desired outcome,
- have results that will not be achieved for many years,
- are characterized by causal relationships or feedbacks are not well understood,
- relate to inherently uncertain or stochastic systems,
- operate at multiple temporal and spatial scales,
- relate to deterrence or prevention of specific behaviors,
- have multiple purposes and funding that can be used for a range of activities, and
- are administrative or process oriented, relating to bureaucratic effectiveness rather than outcomes (OMB 2005c).

Table 2. Classification of a number of example oil spill metrics with regard to multiple value-based dimensions

Economic	Thermodynamic	Environmental	Ecological	Human health	Sociopolitical
Cleanup costs Property and 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 cleanup wastes generated Volume of cleaning agent deployed	Chemical concentration and toxicity Habitat suitability (e.g., acres of shellfish beds) Length of oiled shoreline Degradation rates	Wildlife deaths, populations, fecundity, and recovery rates Biodiversity Catch sizes Plantings, seedlings	Quality-adjusted life-years (QALYS) Disability-adjusted life-years (DALYS) Life expectancy Lost time injuries	Newspaper column inches, minutes TV coverage, Web hits Volunteerism Public meeting attendance Direct messages (e-mail, letters, phone calls)

With regard to tactical metrics in particular, written information describing Federal agency ecological and environmental metrics is scant. For example, measuring fish populations over time is an effective metric in the sense that it has many of the ideal qualities listed above. It is easy to communicate, important to a wide range of stakeholders, likely responsive to human intervention, credible, scalable, and relevant. However, it could be difficult to verify when a species is being managed “sustainably.” The point at which the transition from unsustainable to sustainable management occurs can be contested on the basis of whether sustainability is defined narrowly as an ecological measure (e.g., wild populations), more broadly to include environmental or human health-based considerations (e.g., mercury contaminant levels), or in economic terms (e.g., price per pound, total pounds catch, or availability of farm-raised alternatives). As seen in Table 3, the number of sustainable fish stocks can be described as quantitative, strategic, and descriptive of the state of the ecological system. However, to achieve the goal of improving wild fish stocks, it is essential to understand the cause-effect relational aspects that describe both the pressures on the ecological resources of concern and the human response, as well as what sort of tactical measures (i.e., alternatives such as catch limits or habitat restoration) the responsible federal agency can deploy and the effectiveness of the operations involved (such as dam removal, fish ladder installation, or conservation areas).

OIL SPILL PERFORMANCE METRICS

Although many US federal agencies have developed ecological performance metrics, few specifically address oil spill response or effects (Table 4). For example, the USEPA Strategic Plan 2003–2008 performance measures related to oil spills are predicated on the number of spill responses and the percentage of oil storage facilities inspected, rather than the severity of the spills or damage that results (USEPA 2003). The US Department of Transportation has targeted a 20% reduction in the volume of oil spilled by maritime shipping sources and by piped sources between the years 2001 and 2006 (DOT 2002). Although a laudable goal, this performance metric is unaffected by spill response and is independent of the ecological damage caused.

However, the State of Rhode Island has established an Incident Severity Scale to better gauge the potential effect of spills with regard to human safety and environmental threats (RIDEM 2004). Although the size of the spill is important, the Rhode Island approach also accounts for other factors that might mitigate or exacerbate the effects of the oil. The scale ranges from minor (category 1) to severe (category 4) and is assessed on the basis of 4 general criteria: 1) proximity to or danger to critical areas, 2) level of public concern, 3) association with enforcement actions, and 4) threat to public health or welfare.

An effective response presumably results in reduction of the severity of the spill by addressing each of these 4 general areas. Although the Rhode Island approach is semiquantitative (i.e., an ordinal ranking rather than a cardinal scale) and the assessment might rely heavily on subjective or expert judgment, the approach holds promise for organizing response efforts and focusing communications with the public on mitigation of the potential effects of the spill. For example, the Rhode Island system is reminiscent of the Saffir–Simpson Hurricane Scale (NOAA 2005a). The public readily under-

Table 3. Environmental performance metrics in US federal agencies

Agency	Topic	Strategic/intermediate objective	Performance goal/measure
NOAA (2004)	Healthy and productive coastal and marine ecosystems that benefit society	Increase number of fish stocks managed at sustainable levels	Number of overfished major stocks of fish
			Number of major stocks with an “unknown” stock status
			Percentage of plans to rebuild overfished major stocks to sustainable levels
USFS (2004)	Reduce the risk from catastrophic wildland fire	Improve the health of National Forest Service lands that have the greatest potential for catastrophic wildland fire	Number of acres of hazardous fuels treated in the wildland–urban interface and percent identified as high priority
			Number of acres in the wildland–urban interface treated per \$1 million gross investment
USFWS (2000)	Sustainability of fish and wildlife populations	Lessen number of imperiled species	Through 2005, 371 species listed under the Endangered Species Act as endangered or threatened for a decade or more either stable or improving, 15 species delisted because of recovery, and the listing of 12 species at risk is made unnecessary because of conservation agreements
USEPA (2003)	Clean air and global climate change	Healthier outdoor air	Reduce stationary source emissions by 2008 of nitrogen oxides by 3 million tons from the 2000 level of 5.1 million tons
			Reduce mobile source emissions by 2010 of nitrogen oxides by 3.4 million tons from the 2000 level of 11.8 million tons
DOE (2003)	Protect the environment by providing a responsible resolution to the environmental legacy of the Cold War	Accelerate cleanup of nuclear weapons manufacturing and testing sites	Complete cleanup of 108 of 114 sites by 2025

stands the Saffir–Simpson scale, the basis for establishing it (maximum sustained wind speed), and the relationship of hurricane strength to risk. Preparations and response, including evacuation in extreme circumstances, are facilitated by the fact that the Saffir–Simpson scale has gained broad acceptance. An analogous system of assessing and communicating oil spill severity such as the Rhode Island system could be similarly advantageous, although it likely needs to be more sophisticated. At the tactical and operational level, it might be possible to study alternatives or technologies that are effective in reducing the strategic goal of reduced spill severity.

Kuchin and Hereth (1999, p. 347) note that despite the necessity of critical success factors and measures of outcomes, they “found no comprehensive system, agreed upon by the

response community, that systematically evaluates the success of the response effort.” They discuss an evaluation model that includes essentially the same criteria identified by the Rhode Island approach, with the additional consideration of economics and organization: 1) human health and safety, 2) natural environment, 3) economic impact, 4) public communication, 5) stakeholder service and support, and 6) quality and effectiveness of the response organization.

After each criterion is evaluated, they are integrated as part of a balanced response scorecard for measuring the success of spill response.

The NOAA Office of Response and Restoration (ORR) also recognizes the value of a multifaceted approach in discussion of the critical success factors (G. Ott, NOAA, Yorktown, VA,

Table 4. Ecological performance metrics in US federal agencies that address oil spill response or effects

Agency	Example goal/topic	Strategic/intermediate objective	Performance goal/measure
USEPA (2003)	Land preservation and restoration: restore land	Prepare for and respond to accidental and intentional releases	Each year through 2008, respond to 350 hazardous substance releases and 300 oil spills
			Each year through 2008, minimize effects of potential oil spills by inspecting or conducting exercises or drills at 6% of approximately 6,000 oil storage facilities required to have Facility Response Plans
DOT (2002)	Oil and pipeline spills	Reduce amount of oil spilled by 20% by 2006	Gallons spilled per million gallons shipped by maritime sources
			Tons of hazardous liquid materials spilled per million ton-miles shipped by pipelines
NOAA (2004)	Environmentally sound development and use of the US transportation system	Reduce human risk, environmental and economic consequences resulting from natural or human-induced emergencies	Listed as "To be determined"

USA, unpublished data). The factors over which ORR can exercise some influence are 1) human health, 2) natural environment, 3) economy, and 4) stakeholder support.

Compared with Kuchin and Hereth (1999), the ORR approach segregates the importance of an effective response organization as an administrative rather than outcome-based metric, and it lumps communication and stakeholder service and support into a single, broader category. The ORR approach also recognizes the importance of leading indicator metrics (such as hours worked) that are indicative of progress toward the eventual goal. Although they are not direct measures of the goal itself, such as safeguarding human or ecological health, the leading indicator metrics can be used to gauge the intensity of efforts or forecast potential outcomes. ORR identifies 4 key elements to a successful response: objectives, organization, resources, and mobilization.

Most performance measurements reported in the peer-reviewed literature describing spill response are operational metrics developed decades ago to evaluate the technical efficiency of slick skimmers deployed to recover floating oil (Lichte 1979; Schwartz 1979). Similar evaluations are still performed (Drieu and Hansen 2003), and they could play an important role in assessing the effectiveness of mechanical removal efforts, but they do not capture the broader ecological consequences of the spills. For example, the only performance metric that Caicedo et al. (2003) report with regard to a terrestrial oil spill response in Brazil is the efficiency (87%) of spilled oil recovered in 1 y of operation.

Dix and Hutley (2005) focus on measures of spill response velocity. They suggest the time for equipment to arrive on scene, or the time required for the incident command to order

initial resources through the incident command system, as a metric for initial response velocity. Additionally, they propose to monitor the tapering rate of funds expenditure as a metric for deceleration or demobilization velocity. Like the evaluations described above, they also use equipment efficiency and project cost as metrics for the oil removal phase of the response. This approach, however, is most effective for gauging the effectiveness of organizational or administrative efforts. It does not describe the effectiveness of those efforts with regard to ecological or economic outcomes. Moreover, response time is a resource-based operational measure rather than an endpoint measure related to a strategic goal. Consequently, response velocity might be most meaningful within the response organization itself (e.g., to assess the effectiveness of planning efforts) and less effective for communicating with public or stakeholder groups.

INTERPRETATION: A TYPOLOGY OF OIL SPILL PERFORMANCE METRICS

Oil spill response requires a rapid integration of several different organizations and government agencies with different areas of expertise, resources, and concerns. To reach maximum effectiveness, such a complex system must be focused on a set of shared goals that can clearly be articulated to outside groups. To date, oil spill contingency planning has focused primarily on the rules for establishing the organization itself (such as command and communications structures) or resource and technology development, rather than measures of effectiveness that are meaningful to stakeholder or public communities. It has not focused on developing measures for clearly defined strategic goals. Consequently,

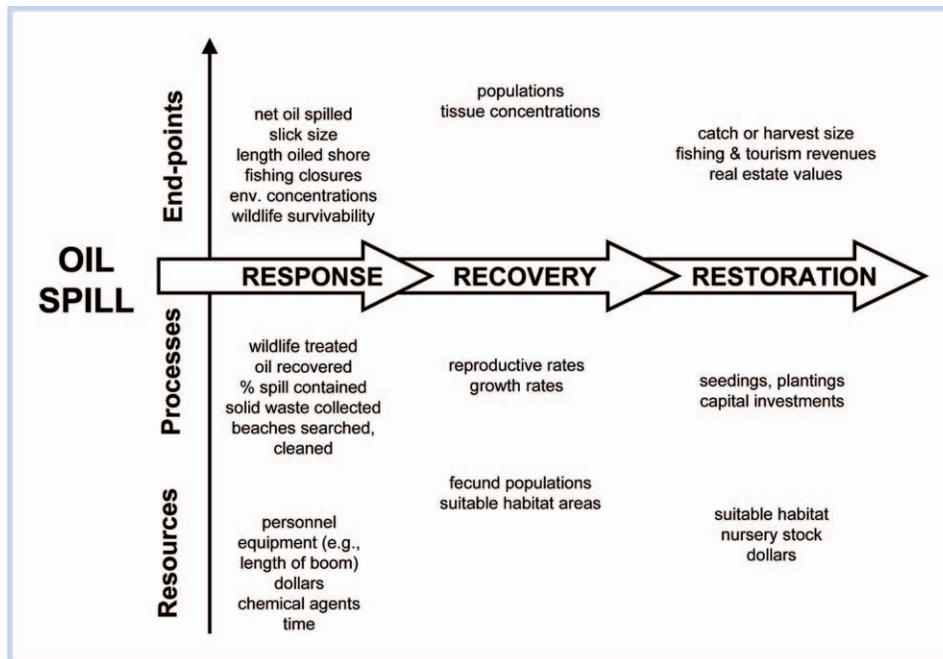


Figure 4. Typology of Oil Spill Response and Assessment Metrics. Management of oil spills can be characterized in 3 stages that generally blend together: response, recovery, and restoration. Each stage is overlapping in time, read from left to right in the direction of the block arrows. Within each stage, progress is from the lower edge of the figure toward the top as resources are applied in processes that are intended to improve endpoints. Decisions made in early stages can have a direct effect on the resources, alternatives, and endpoints in other stages. However, different stakeholders and government agencies are likely to have different areas of concern or different priorities within the same area. Mapping these concerns provides a visual depiction of potential opportunities for better communication and coordination, as well as potential conflicts.

spill managers could be unprepared to define the essential criteria for success, filter the wealth of imperfect and rapidly changing information available to focus the organization on the salient objectives, or disentangle what their actions achieve versus those that occur outside of their control (e.g., the vagaries of the weather and currents). Moreover, without clearly defined goals, it is difficult to respond to feedback, adapt to changing circumstances, or assess the effectiveness of response strategies. The result could be that personnel from different agencies feel disconnected from one another, the public at large, or both. At worst, they could be working toward cross-purposes.

For example, in a typical oil spill, response and damage assessment are conducted separately. The former entails establishing an ad hoc command structure, deploying resources, and making rapid judgments regarding initial efforts. The latter requires careful sampling and data collection, calibration of mathematical models, and scientific review. Different skill sets, thought patterns, and time frames are required. Although close cooperation of 1st responders with damage assessors might be advantageous, it is unlikely to occur spontaneously in the midst of a crisis. Adding to the complexity of the problem is that multiple stakeholders and public groups could hold mutually exclusive views about the best possible response. In this case, no response alternative is likely to satisfy all parties.

To establish a set of metrics acceptable or communicable to all groups that can be used to effectively assess the success of efforts to prevent and mitigate the effects of a spill, it is necessary to understand the concerns of all the parties involved. In addition to being related to the 6 broad value-based criteria

outlined above, different metrics could be relevant to different stages in the evolution of an oil spill response. For example:

- Response includes all efforts to contain and clean up the oil.
- Recovery is the period after initial cleanup during which both natural and human systems begin to stabilize and recover (e.g., marsh productivity and local tourism) and during which long-term effects are assessed, including natural resource damage assessment.
- Restoration includes all activities to restore natural and human systems to their state before the spill, including efforts to replace or offset damaged resources and effects.

Figure 4 depicts a typology of oil spill metrics that shows both how oil spill management progresses in time (from left to right) and how metrics can be characterized as a resource, process, or endpoint. An effective response will mitigate the damaging economic, thermodynamic, environmental, ecological, human health, and sociopolitical effects. Because these are manifested on different timescales, relational leading indicator metrics are essential to provide feedback to the response organization and allow adaptation to changing or unexpected events. As response efforts progress, the locus of concern moves from the lower left hand corner of the graph toward the upper right; that is, resources are required to drive processes that are directed at endpoints. However, response phase endpoints soon become recovery phase resources, and so on.

As different organizations become engaged in managing the spill, the metrics relevant to those organizations' efforts can be plotted in the metrics map depicted by Figure 4. For example, during the response stage, mechanical removal

contractors will be primarily concerned with resource availability and operational effectiveness in removing oil and emulsion from the environment. The metrics relevant to their work will be found in the lower left-hand region. A wildlife biologist, on the other hand, might be more concerned with habitat restoration and population recovery—concerns better plotted closer to the upper right-hand corner. Considering the distance between these 2 perspectives, it could be challenging to communicate to contractors how their efforts or decisions influence the measures of concern to the biologist. For example, response efforts that focus solely on washing or mechanical removal might engender perverse ecological outcomes such as destruction of sensitive ecological habitat, despite improvement of environmental endpoints (i.e., concentration reductions) that could be regarded as response priorities. Opportunities for closer coordination can be identified in which 2 different agencies identify measurable objectives that are plotted at some distance apart on the map. These agencies (or stakeholder or public groups) might disagree entirely on the perceived success of spill management efforts simply because they are examining different aspects of those efforts.

Alternatively, where concerns are overlapping, they are not necessarily in agreement. For example, disagreement on the appropriate response to a spill, such as whether to burn, chemically disperse, or attempt mechanical removal of the oil from the environment, could occur in the very early stages. Although the measures that define success (such as slick area, water column concentrations, total uncontained oil, or time oil is on the water) could plot on overlapping areas of the map, groups might place different emphasis on the different metrics relevant to that stage. Conflict will result, but this could likely be identified ahead of time if the contrasting value systems can be elicited or revealed.

CONCLUSIONS

Both industry and government have moved toward more quantitative approaches to environmental management in the years since the environmental crises of the 1980s. In areas that involve chronic environmental effects from continuous activities, such as release of toxic or ozone-depleting chemicals, rapid progress has been made in establishing indicators, and performance metrics to guide policy and decision making. However, in crisis response, such as the immediate aftermath of oil spills, measures for assessing the effectiveness of management actions have developed more slowly. Consequently, crisis managers are challenged to define success in terms that are easily verifiable and communicable to diverse stakeholder groups.

To develop more effective environmental performance metrics and to better understand the role of existing metrics, metrics can be characterized in several different ways: Mathematical, cause–effect relational, organizational, and value-based. Effective metrics of all types share qualities such as verifiability, cost effectiveness, communicability, importance, credibility, scalability, control, relevance, and sensitivity. Also, oil spill management metrics in particular can be identified within the 3 major phases of the spill: Response, recovery, and restoration. Creating effective metrics for oil spill response and understanding which metrics are important to the different people and organizations involved could result in more closely coordinated response efforts that are more satisfying of and responsive to stakeholder concerns.

Acknowledgment—The authors express their gratitude to Jim Clark and 2 anonymous reviewers for comments and suggestions. This study was funded by NOAA through the Coastal Response Research Center at the University of New Hampshire (NOAA grant NA04NOS4190063, project 05–983).

Disclaimer—The work presented by the authors was performed in their private capacities and does not reflect the policies or views of their parent institutions.

REFERENCES

- Caicedo NL, Finotti A, Bastos Neto A. 2003. Evaluation of volume and performance of a free phase recovery system in an oil spill—Case study of an accident in Araucaria, Brazil. In: Proceedings of the 2003 International Oil Spill Conference; 2003 April 6–10; Vancouver, Canada. Washington, DC: American Petroleum Institute.
- Chess C, Johnson BB, Gibson G. 2005. Communicating about environmental indicators. *J Risk Res* 8:63–75.
- [CIEPM] Committee on Industrial Environmental Performance Metrics. 1999. Industrial environmental performance metrics: Opportunities and challenges. Washington DC: National Academy Press. 252 p.
- [DHS] Department of Homeland Security. 2004. National response plan. http://www.dhs.gov/interweb/assetlibrary/NRP_FullText.pdf. Accessed 20 January 2006.
- Dietz T, Stern PC. 1998. Science, values, and biodiversity. *Bioscience* 48:441–444.
- Dix M, Hutley T. 2005. Response velocity. In: Proceedings of the 2005 International Oil Spill Conference; 2005 May 15–19; Miami Beach, FL. Washington, DC: American Petroleum Institute.
- [DOE] Department of Education. 2003. The Department of Energy strategic plan. <http://strategicplan.doe.gov/hires.pdf>. Accessed 20 May 2005.
- [DOT] Department of Transportation. 2002. DOT performance plan—FY 2003 and performance report—FY 2001 (2002). <http://www.dot.gov/performance/pdf/environment.pdf>. Accessed 24 June 2005.
- Drieu MD, Hansen K. 2003. Fast-water response equipment performance in two inland rivers. In: Proceedings of the 2003 International Oil Spill Conference; 2003 April 6–10; Vancouver, Canada. Washington, DC: American Petroleum Institute.
- Egorova TA, Rozanov EV, Schlesinger ME, Andronova NG, Malyshev SL, Karol IL, Zubov VA. 2001. Assessment of the effect of the Montreal Protocol on atmospheric ozone. *Geophys Res Lett* 28:2389–2392.
- [GEMI] Global Environmental Management Initiative. 1998. Measuring environmental performance: A primer and survey of metrics in use. http://www.gemi.org/MET_101.pdf. Accessed 23 January 2006.
- 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. Greenwich (CT): JAI.
- Graedel TE, Allenby BR. 2002. Hierarchical metrics for sustainability. *Environ Qual Manag* 12:21–30.
- Gray PCR, Wiedemann PM. 1999. Risk management and sustainable development: Mutual lessons from approaches to the use of indicators. *J Risk Res* 2:201–218.
- Hamilton JT. 1995. Pollution as news: Media and stock market reactions to the Toxics Release Inventory data. *J Environ Econ Manag* 28:98–113.
- Hammond A, Adriaanse A, Rodenburg E, Bryant D, Woodward R. 1995. Environmental indicators: A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development. Washington DC: World Resources Institute.
- 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. Washington, DC: Department of Engineering Management, George Washington University.
- Hertwich EG, Pease WS, Koshland CP. 1997. Evaluating the environmental impact of products and production processes: A comparison of six methods. *Sci Tot Environ* 196:13–29.
- Hofstetter P, Hammit JK. 2002. Selecting human health metrics for environmental decision-support tools. *Risk Anal* 22:965–983.

- Horvath A, Hendrickson CT, Lave LB, McMichael FC, Wu TS. 1995. Toxic emissions indexes for green design and inventory. *Environ Sci Technol* 29:A86–A90.
- Kim I. 2002. Ten years after the enactment of the Oil Pollution Act of 1990: A success or a failure. *Mar Pol* 26:197–207.
- Konar S, Cohen MA. 1997. Information as regulation: The effect of community right to know laws on toxic emissions. *J Environ Econ Manag* 32:109–124.
- Kuchin JT, Hereth L. 1999. Measuring response: A balanced response scorecard for evaluating success. #319 International Oil Spill Conference. *Korea Observer* 30:647–670.
- Lichte HW. 1979. Skimming barrier performance evaluation: Offshore version and harbor version. In: Proceedings of the 1979 Oil Spill Conference; 1979; Los Angeles, CA. Washington, DC: American Petroleum Institute.
- Linkov I, Satterstrom FK, Kiker G, Bridges T, Benjamin S, Belluck D. 2006. From optimization to adaptation: Shifting paradigms in environmental management and their application to remedial decisions. *Integr Environ Assess Manag* 2:92–98.
- Linkov I, Satterstrom FK, Seager TP, Kiker G, Bridges T, Belluck D, Meyer A. 2006. Multi-criteria decision analysis: Comprehensive decision analysis tool for risk management of contaminated sediments. *Risk Anal* 26:61–78.
- Loureiro ML, Ribas A, López E, Ojea E. 2006. Estimated costs and admissible claims linked to the Prestige oil spill. *Ecol Econ* 1:48–63.
- National Research Council. 1996. Understanding risk: Informing decisions in a democratic society. Washington DC: National Academy Press. 250 p.
- Neumann CM. 1998. Improving the US EPA toxic release inventory database for environmental health research. *J Toxicol Environ Health B Crit Rev* 1:259–270.
- [NOAA] National Oceanic and Atmospheric Association. 2001. Characteristics of response strategies: A guide for spill response planning in marine environments. http://response.restoration.noaa.gov/book_shelf/910_response.pdf. Accessed 6 October 2006.
- [NOAA] National Oceanic and Atmospheric Association. 2004. New priorities for the 21st century: NOAA's strategic plan, Updated for FY 2005–2010. http://www.spo.noaa.gov/pdfs/NOAA%20Strategic%20Plan_pm.pdf. Accessed 24 June 2005.
- [NOAA] National Oceanic and Atmospheric Association. 2005a. The Saffir–Simpson hurricane scale. <http://www.nhc.noaa.gov/aboutsshs.shtml>. Accessed 20 January 2006.
- [NOAA] National Oceanic and Atmospheric Association. 2005b. The science program for Florida Bay and adjacent marine wcosystems. http://www.aoml.noaa.gov/ocd/sferpm/scicon_synthesis/scicon_synthesis_ortner/synthintro.ppt. Accessed 17 May 2005.
- Odum HT. 1996. Environmental accounting. New York (NY): John Wiley. p 127–132.
- [OECD] Organization for Economic Cooperation and Development. 1998. Towards sustainable development: Environmental indicators. Paris (FR): OECD. 244 p.
- Olsthoorn X, Tyteca D, Wehrmeyer W, Wagner M. 2001. Environmental indicators for business: A review of the literature and standardization methods. *J Cleaner Prod* 9:453–463.
- [OMB] Office of Management and Budget. 2005a. Management/GPRA—Government Performance Results Act of 1993. <http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.html>. Accessed 17 May 2005.
- [OMB] Office of Management and Budget. 2005b. OMB circular A-11. http://www.whitehouse.gov/omb/circulars/a11/current_year/s26.pdf. Accessed 17 May 2005.
- [OMB] Office of Management and Budget. 2005c. Performance measurement challenges and strategies—Draft. http://www.whitehouse.gov/omb/part/challenges_strategies.html. Accessed 17 May 2005.
- Rapport DJ. 1999. Gaining respectability: Development of quantitative methods in ecosystem health. *Ecosyst Health* 5:1–2.
- Rapport DJ, Costanza R, McMichael AJ. 1998. Assessing ecosystem health. *Trends Ecol Evol* 13:397–402.
- [RIDEM] Rhode Island Department of Environmental Management. 2004. RIDEM—Emergency Response Plan. <http://www.dem.ri.gov/topics/erp.htm>. Accessed 20 January 2006.
- Schulze P, editors. 1999. Measures of environmental performance and ecosystem condition. Washington DC: National Academy Press.
- Schuermann G, Market B. 1998. Ecotoxicology: Ecological fundamentals, chemical exposure and biological effects. New York (NY): John Wiley. 936 p.
- Schwartz SH. 1979. Performance tests of four selected oil spill skimmers. In: Proceedings of the 1979 Oil Spill Conference 1979; Los Angeles, CA. Washington, DC: American Petroleum Institute.
- Seager TP, Theis TL. 2004. A taxonomy of metrics for testing the industrial ecology hypotheses and application to design of freezer insulation. *J Cleaner Prod* 12:865–875.
- Shaw DG. 1992. The Exxon-Valdez oil spill—Ecological and social consequences. *Environ Conserv* 19:253–258.
- Sikdar S, Glavic P, Jain R. 2004. Technological choices for sustainability. New York (NY): Heidelberg. 441 p.
- Tannahill G, Steen A. 2001. OPA effectiveness: Past, present, and future. In: Proceedings of 2001 International Oil Spill Conference. Washington, DC: American Petroleum Institute.
- [USEPA] US Environmental Protection Agency. 1990. Oil Pollution Act of 1990. Title 33 US Code 2702–2762. <http://www.epa.gov/Region5/defs/html/opa.htm>. Accessed 6 October 2006.
- [USEPA] US Environmental Protection Agency. 2002. EPCRA overview. <http://yosemite.epa.gov/oswer/Ceppoweb.nsf/content/epcraOverview.htm>. Accessed 17 January 2006.
- [USEPA] US Environmental Protection Agency. 2003. EPA strategic plan 2003–2008. <http://www.epa.gov/ocfo/plan/2003sp.pdf>. Accessed 23 June 2005.
- [USEPA] US Environmental Protection Agency. 2005a. Toxics release inventory (TRI) program. <http://www.epa.gov/tri/>. Accessed 17 January 2006.
- [USEPA] US Environmental Protection Agency. 2005b. Tool for the reduction and assessment of chemical and other environmental impacts. <http://www.epa.gov/ORD/NRMRL/std/sab/traci/>. Accessed 20 January 2006.
- [USFS] US Forest Service. 2004. USDA Forest Service strategic plan for fiscal years 2004–08. <http://www.fs.fed.us/publications/strategic/fs-sp-fy04–08.pdf>. Accessed 23 June 2005.
- [USFWS] US Fish and Wildlife Service. 2000. Strategic plan. <http://www.fws.gov/planning/USFWStrategicPlanv3.pdf>. Accessed 20 May 2005.
- Webler T, Rakel H, Renn O, Johnson B. 1995. Eliciting and classifying concerns: A methodological critique. *Risk Anal* 15:421–436.
- Wilhelm WE, Srinivasa AV. 1997. Prescribing tactical response for oil spill clean up operations. *Manag Sci* 43:386–402.
- [WMO] World Meteorological Organization. 2003. WMO scientific assessment of ozone depletion: 2002, Rep. 47, global ozone restoration and monitoring project, Geneva, Switzerland. http://www.wmo.ch/web/arep/reports/o3_assess_rep_2002_front_page.html. Accessed 25 January 2006.