Comparative vs. Absolute Performance Assessment with Environmental Sustainability Metrics

Xun Jin and Karen A. High

School of Chemical Engineering
Oklahoma State University
423 Engineering North
Stillwater, Oklahoma 74078
U.S.A.

Phone: 405-744-5280, Fax: 405-744-6338
xun.jin@okstate.edu, high@okstate.edu

ABSTRACT

Different goals and potential audiences determine that two types of environmental performance assessments (comparative vs. absolute) can be distinguished. Whether site-specific environmental conditions are accounted for constitutes an underlying watershed between them. As a result, disparate metrics are needed to accommodate these different assessments, which essentially demand "realism" to different extents. In the context of sustainability, the dilemma of pursuing "relative" or "actual" environmental outcomes is getting more prominent. A four-class scheme is proposed for classifying metrics in such a manner that appropriate metrics for either comparative or absolute assessment can be identified, respectively. Discussion is also given to the metrics of each class, in terms of their different characteristics and utilization.

1 INTRODUCTION

Sustainability means different things to different people. Since 1992 when Agenda 21 [1] officially called upon developing indicators for sustainable development, enormous metrics have emerged with a title of "sustainability." Some argue that the title may have been abused, as many of today's sustainability metrics are actually "old wine in a new bottle." Then, what exactly makes the difference for sustainability metrics?

In the environmental context, disputes exist regarding the key features that a desired sustainability metric is supposed to gain. Sound discussions are given in [2] and Chapter 11 of [3]. Basically, previous environmental metrics address apparent, straightforward, clear-cut environmental problems (e.g. oil spill, belching smokestack), which, to date, have been understood fairly well. However, people's environmental concerns have recently switched to more complex, obscure, consequential, far-reaching and less recognized issues such as global warming, ozone layer, species annihilation, and so on, as those are believed to be more relevant to human's welfare in a longer time scale. This focus shift obviously requires to set a different and higher standard for the so-called "sustainability metrics."

On the other hand, environmental metrics can be partitioned into two camps. One suite of metrics aim to assess the environmental performance of a particular human activity, while the other gauges the condition of the ecosystem. These two classes of metrics were developed by two different cadres of professionals from their own perspectives. Unfortunately, little interaction took place between the two camps in the past. This scarcity of communications has severely impeded the progress of either side [3].

Assessors' different perspectives also give rise to another dilemma. This can be further explained as follows. For example, the environmental performance of a manufacturing plant to a large extent relies upon its inherent properties, including design conditions, pollution control and operation, etc. However, the specific environmental scenario (e.g. wind direction, hydrological conditions) in which the plant locates also plays an important role in the actual environmental deterioration. To this end, assessors have to determine ahead of time whether a site-specific assessment is intended. Accordingly, different metrics will be needed.

This study focuses on the metrics that help assess the environmental performance of a particular human activity in the context of sustainability. The authors, leaving out those ideologically debatable underpinnings of "sustainability metrics," adopted a rather loose criterion for metric screening. More specifically, as long as a given metric environmentally describes the performance of a human activity, regardless of its title, is considered a candidate.

In this paper, environmental performance assessments are distinguished into either "comparative" or "absolute" assessments. An assessment without considering specific environmental conditions is known as "comparative", which supposedly compare alternatives (e.g. processes, products, policies, etc.) in terms of their relative performance. Contrastively, an "absolute" assessment provides the prediction of "real" environmental outcomes by taking into account specific environmental information. The metrics that meet the different needs of both assessments are addressed in parallel.

2 COMPARATIVE VS. ABSOLUTE ASSESSMENTS

The ultimate goal of an environmental performance assessment is to predict or measure the extent to which negative outcomes will be or have been caused to the environment. In the past, assessors used simple and crude measures (e.g. mass flow), which do not reflect real environmental effects. Today progress has been made to persistently move closer to revealing actual environmental damages. In addition, growing inspiration for sustainability further stimulated people's curiosity of exploring what exactly is going to happen in the environment. However, the pursuit of realism is costly, because too many factors contribute to it.

First of all, the magnitude of an undesired chemical release to the environment has to be primarily considered. A general experience tells that "more release, more harm". A premise for this to hold valid is that the comparison is carried out with respect to two different quantities of a same chemical species in identical environmental conditions. Obviously, a comparison like this is of little meaning in practice. Therefore, more factors have to be taken into consideration.

Second, the properties of a chemical essentially affect its environmental behaviors. For example, both carbon dioxide and methane are identified as greenhouse gases. However, their ability to cause greenhouse effects differs. In other words, the same amount of methane and carbon dioxide will result in disparate effects of the so-called "global warming." Chemicals exhibit a wide range of environment relevant properties, such as toxicity, transport, persistency, reactivity, bioaccumulation, heat-trapping capacity and so on, varying with the environmental problem that is concerned. More importantly, derivation of these properties is closely related to specific environmental contexts in which they are addressed.

Third, environmental conditions also have significant influence on the potential environmental consequence. Before a chemical causes the damage of interest, it may transport, degrade, accumulate, transform or even react with others in the environment. All those behaviors rely on environmental conditions, which is site-specific.

In practice, a contradiction is always present between what should be measured and what can be measured. People are interested in gaining awareness as much as possible to the actual environmental effects resulting from a targeted activity. Sustainability, over the recent years, has fostered a remarkable raise in the attention given to more consequential and less discovered environmental impacts. However, sustainability, on the other hand, calls for proactive measurement of obscure environmental effects over an expanded time scale. In this case, chemicals spend more time in the environment. As a consequence, specific environmental conditions will likely contribute more to the final damage. As just mentioned, an assessment measuring real effects has to involve comprehensive considerations of three aspects of information, say, release quantity, chemical properties, and environmental conditions. This comprehensiveness usually leads to a significant increase in complexity, sophistication, and uncertainty of the assessment, which could exceed assessors' tolerance.

This contradiction necessitates the effort of distinguishing comparative and absolute assessments, because such a single assessment did not exist until today that could reconcile realism and conciseness to a satisfactory extent. Generally speaking, whether a particular assessment should be comparative or absolute depends on its goals and potential audience. Table 1 shows some possible cases where two kinds of assessments are encountered, respectively.

Table 1. Comparative vs. absolute assessments

		Assessment goals	Potential audience
Comparative assessment	1	Comparison of alternatives	Process, product, policy developers
	2	Performance progress over times	Regulators, stakeholders, managers and technology developers, etc.
Absolution assessment	1	Location comparison of alternatives	Process, product developers
	2	Performance reporting to certain groups	Victims, community, neighborhood environmentalists, and regulators

3 PERFORMANCE-ORIENTED METRIC CLASSIFICATION

Metrics are needed in any assessment. A useful dichotomy mentioned in the introduction splits the full spectrum of environmental metrics into two subsets. Extraordinarily high diversity exists in the first subset: performance-oriented metrics, which vary from simplest waste flows to sophisticated ecosystem effects. Further classification arises as an important task helping an interested audience who wants to make good use of these metrics. As stated in [4], classification not only “helps in identifying scope and limits of our current knowledge and in reviewing the available methods,” but also “helps select the most adequate format and methods of measurement.”

There are different ways to classify performance-oriented environmental metrics. In [3], metrics are grouped into three categories according to their utility in different areas. The three classes are:

- Operational metrics
- Management metrics
- Environmental condition metrics

These three classes are expressed differently in [5] as: lagging, leading and environmental condition metrics, respectively. This scheme is consistent with the ISO14031 standards developed by the International Organization for Standardization (ISO). According to [3], the third class is the least developed, but of the greatest interest to industry and external stakeholders. The challenge for developing or implementing the metrics of this kind lies in a defendable causal relationship linking a specific human activity to an environmental outcome, which, if conquered, could ideally express environmental performance in the units of ecosystem condition.

The authors in previous work [6, 7] proposed a “Stressor-Status-Effect-Integrality-Well-being” conceptual hierarchy for classifying environmental sustainability metrics. This scheme, along with its peers, essentially stresses on the difference in identifying the interested environmental attributes affected by the interaction between two distinguished systems, say, human system and the encompassing ecosystem. Other similar schemes include: “Pressure-State-Response” (PSR) by Organization for Economic Cooperation and Development (OECD); “Driving force-State-Response” (DSR) by the United Nation Commission on Sustainable Development (UNCSD); the “Pressure-State-Impact-Response” (PSIR) by the United National Environmental Programme (UNEP) and the Netherlands National Institute of Public Health and the Environment (RIVM); as well as the “Driving force-Pressure-State-Impact-Response” (DPSIR) by European Union (EU).

In literature, various metric classification schemes exist, such as 5-level indicator hierarchy developed by the Lowell Center for Sustainable Production at the University of Massachusetts Lowell [8]; Two metric sets (midpoint vs. endpoint) mostly adopted in Life Cycle Assessment (LCA) based methods [9]; Three categories of environmental measurements (process measures, results measures and customer satisfaction) described by Wells et al. [10]; four-tiered (or five) metric hierarchies featuring disparate realism and complexity [11, 12, 13], and many others.

To specifically help identify different metrics needed by comparative and absolute environmental performance assessments. This paper presents a classification scheme based on different involvements of the factors that influence actual environmental outcomes. The scheme consists of four classes, each of which is described in Table 2 and Table 3.

Table 2. 4-class scheme of classifying environmental performance metrics

Class 1:	The metrics that only use quantity of releases.
Class 2:	The metrics that reflect the relative differences among chemicals, but without involving any effort to account for environmental conditions.
Class 3:	The metrics that measure the chemical-specific environmental properties using a "generic" or "standard" environmental scenario.
Class 4:	The metrics that measure the actual environmental effects by taking into account "real" environmental conditions.

Table 3. Characteristics and examples of the metrics in the 4-class scheme

Class	Characteristics	Suited to	Metric Examples
Class 1	Not chemical-specific Not site-specific	Comparative assessments	Toxical Release Inventory (TRI)
Class 2	Chemical-specific No environmental information	Comparative assessments	Human Toxicity Potential (HTP)
Class 3	Chemical-specific Generic environmental condition	Comparative assessments	Photochemical Ozone Creation Potential (POCP)
Class 4	Chemical-specific Site-specific	Absolution assessments	Human health and ecological risks

4 METRIC UTILIZATION AND EXAMPLES

The metrics in classes 1-4 basically covered various efforts in history that people have typically made to measure environmental effects associated with chemicals. Since the 1970s, the environmental performance metrics have evolved quite a bit from simplicity to sophistication, from universality to specificity, and from irreality to realism. The involvement of environmental conditions also received ever-growing attentions. As a consequence, the metrics are getting more sophisticated and complex. It becomes more difficult for an average metric user to establish sufficient insight so as to identify the metrics suited to their applications.

In this section, the detailed discussion is given to the four classes of metrics defined above, respectively, with respect to different characteristics and utilization of each class. Example metrics for each class are described to help readers understand their underlying distinctions.

Class 1- The metrics of Class 1 use direct inventory data or in variant forms (e.g. relative, indexed, aggregated, etc.). For instance, a waste emission can be expressed as annual emission, emission flowrate, emission vs. baseline value, emission per unit raw material, emission per unit product, emission per unit profit, etc. Historically, this kind of metrics dominated most applications in regulatory, business and industrial areas. However, due to their inherent deficiencies, they are confined to the assessments without ambition to measure environmental outcomes.

Toxic Release Inventory (TRI) is one most successful Class 1 metric in the United States, which was mandated by Emergency Planning and Community Right-to-Know Act (EPCRA) of 1986. Companies are required to annually report the quantity of each of their releases of over 600 listed chemicals. The information of TRI is maintained by EPA and publicly accessible. A typical TRI record contains information like "3,957 lb/year ethylene glycol emission from the point sources at Mercury Mercruiser facility, Stillwater, Oklahoma. [14]"

Class 2- Modifying inventory data by a factor in whatever titles (e.g. potentials, equivalency, characterization factors, potency, etc.) has become mainstream practice in the area of environmental performance assessment. This factor is used to account for chemical specificity via comparing relative significance of potential environmental effects caused by different chemicals. Metrics in both Classes 2 and 3 fall in this group.

In general, chemical-specific properties need to be derived in certain environmental conditions. Characterization of specific environmental conditions is often conducted by performing a series of analysis. Metrics differ in their specific techniques to carry out different analysis, which may include:

- Fate analysis (e.g. degradation, accumulation, persistency, transformation etc.);
- Transport analysis (within a media or across medias);
- Exposure analysis (e.g. magnitude, frequency, duration, route of exposure)
- Effect analysis

In many cases, metric developers did not intentionally devise environmental conditions to be applied in their metric derivation, or the "default" environmental conditions therein are unspecified. This ignorance leads to difficulties in analyzing the extent to which the assessment results will deviate from actual environmental outcomes, which supposedly originates from the underlying deviation between "actual" and "applied" environmental condition. Therefore, Class 2 & 3 metrics are separated in the proposed scheme, just in order that for in a particular metric whether environmental conditions are either specified or not can be distinguished.

For Class 2 metrics, chemical-specificity is addressed usually via assigning scores for different chemicals. These scores are derived from experiments and/or model-based simulation in such a manner that the chemical's possible behaviors in the environment are not accounted for or specified.

Examples of Class 2 metrics can be found in many human toxicity metrics, such as Threshold Limit Value (TLV) by American Conference of Governmental Industrial Hygienists (ACGIH) and Permissible Exposure Limit (PEL) by Occupational Safety and Health Administration (OSHA). These metrics focusing on toxicity effects assume that chemicals are exposed to human receptors through direct oral, inhalation, or dermal contact. Therefore, they do not incorporate any indication of the effects associated with chemical's environmental behaviors. These metrics, in their original form, though have been useful in safety and health assessments, they are not suited for environmental performance assessments, especially when sustainability is concerned.

Class 3- Similar to the metrics in Class 2, Class 3 metrics reflect chemical-specific properties, ordinarily in the form of a scoring system. Nevertheless, Class 3 metrics contain readily identified environmental conditions that were devised or specified in the metrics' original derivation. This has given a big advantage to the metrics of Class 3, as opposed to Class 2 metrics, because the transparency of this background information, to some extent, allows users to be more convinced about metrics' utilization as well as the degree to which the obtained results should represent actual environmental impacts.

The embedded set of environmental conditions in a metric is known as "generic" or "standard" conditions. Unfortunately, "actual" environmental conditions always differentiate more or less from the "generic" conditions of the metric to be applied. Therefore, Class 3 metrics still cannot reflect actual effects. However, as the disparity between actual and generic environmental conditions are known, the eventual discrepancy from realism is almost predictable, though implicitly and qualitatively. It is a daunting task to explicitly state how a metric would perform in terms of its closeness to realism, because in most cases people's perception of actual environmental effects solely relies upon the measurements that they conduct.

The examples selected for Class 3 are Human Toxicity Potential (HTP) and Photochemical Ozone Creation Potential (POCP). HTP was developed in the University of California, Berkeley and the Lawrence Berkeley National Laboratory [15]. The generic environmental conditions are simulated by a multimedia, multiple pathway fate and exposure model, CalTOX. CalTOX determines pollutant concentrations in uniformly mixed environmental compartments from intercompartmental mass transfer equations. It models exposure pathways using partitioning and biotransfer relationships, and both cancer and noncancer health impacts are considered. POCP was developed in 1990s by European researchers in order to identify hydrocarbons that most significantly contribute to forming tropospheric ozone. A trajectory model is applied to describe multi-day photochemical behaviors of hydrocarbons during long range transport in air parcels across north west Europe towards the British Isles [16]. Users should be noted that POCP was made as realistic as possible to the conditions in northwest Europe. If it is applied elsewhere, deviations in geophysical quantities as well as environmental variables will reduce its credibility.

Class 4- Table look-up may constitute the only job for an average metric user to apply the metrics of Class 2 or 3, since those metrics simply modify inventory data by a score accounting for the interested chemical-specific properties. However, implementing the metrics of Class 4 turns out to be much more complicated, because site-specific environmental conditions need to be involved.

Class 4 metrics may differ widely from each other in answering a series of questions; what site-specific information is available? how this information is used, and how is the final measure devised? Usually it is difficult to account for widely variant environmental behaviors (e.g. fate, transport) with a same environmental model just via switching parameters. Therefore, models in a Class 4 metric sometimes need to be identified or even developed by assessors. This could impose an unsolvable burden on assessors without expertise. A metric, in this case, could possess similar degree of sophistication and complexity as a full assessment.

Certain methods of risk assessment involving site-specific data can be regarded as typical Class 4 metrics. As Class 4 metrics inherently need to be handled in a case-by-case fashion, due to its site-specificity. Also, risk assessments usually come to play as a methodological framework, instead of metrics. No specific class 4 metric is named here.

5 CONCLUSION

Site-specific environmental conditions of a target human activity have non-negligible influence on its actual environmental outcomes. Most today's environmental performance assessments do not take this site-specificity into consideration. Part of the ignorance can be explained as that only relative performance is interested, which has been referred as "comparative" assessments. However, admittedly, a so-called "absolute" assessment often makes assessors suffer from undiscovered issues, data shortage, complexity, and uncertainty. The classification scheme presented in this paper is specifically designed to help assessors to customize their environmental performance metrics, in order to meet the different needs of comparative and absolute assessments that are intended.

REFERENCE

1. United Nations Commission on Sustainable Development. (1992). Agenda 21. http://www.un.org/esa/sustdev/documents/agenda21/index.htm
2. Schulze, P. and Frosch, R. A. (1999). Overview: Measures of Environmental Performance and Ecosystem Condition. In "Measures of Environmental Performance and Ecosystem Condition", National Academy Press, Washington D.C.
3. National Academy of Engineering & National Research Council. (1999). Industrial Environmental Performance Metrics: Challenges and Opportunities, National Academy Press, Washington D.C.
4. Hardi, P. (2001). Trendsetters, Followers, and Skeptics: The State of Sustainability Development Indicators: A Review Essay. Journal of Industrial Ecology, 4 (4), 149-162.
5. Global Environmental Management Initiative. (1998). Measuring Environmental Performance: A Primer and Survey of Metrics in Use.
6. Jin, X. and High, K. (2004). A New Conceptual Hierarchy for Identifying Environmental Sustainability Metrics. Environmental Progress, In press.
7. Jin, X. and High, K. (2003). Toward Best Practice Environmental Sustainability Metrics for Chemical Engineering: Using a Hierarchical Life Cycle Impact Assessment. Annual Meeting of the American Institute of Chemical Engineers, November, San Francisco, California.
8. Veleva, V., Hart, M., Greiner, T. and Crumbley, C. (2003). Indicators for Measuring Environmental Sustainability: A Case Study of the Pharmaceutical Industrial. Benchmarking, 10 (2), 107-119.
9. Bare, J. C. et al. (2000). Life Cycle Impact Assessment Workshop Summary; Midpoints versus Endpoints: The Sacrifices and Benefits. International Journal of Life Cycle Assessment, 5(6), 319-326.
10. Wells, R. H., Calkins, P., and Balikov, H. (1994). Measuring Environmental Performance. International Symposium on Electronics and the Environment, San Francisco, CA.
11. Toffel, M. W. and Marshall, J. D. (2004). Improving Environmental Performance Assessment: A Comparative Analysis of Weighting Methods Used to Evaluate Chemical Release Inventories. Journal of Industrial Ecology, 8 (1-2), 143-172.
12. Pennington, D. W., Norris, G., Hoagland, T. and Bare, J. C. (2000). Environmental Comparison Metrics for Life Cycle Impact Assessment and Process Design. Environmental Progress, 19 (2), 83-91.

13. Pennington, D. W. and Yue, P. L. (2000). Options for Comparison of Process Alternatives in terms of Regional Environmental Impacts. Journal of Cleaner Production, 8, 1-9.

14. http://www.epa.gov/tri

15. Hertwich, E. G., Mateles, S. F., Pease, W. S. and McKone T. E. (2001). Human Toxicity Potentials for Life Cycle Assessment and Toxics Release Inventory Risk Screening. Environmental Toxicology and Chemistry, 20 (4), 928-939.

16. Derwent, R. G., Jenkin, M. E. and Saunders, S. M. (1996). Photochemical Ozone Creation Potentials for a Large Number of Reactive Hydrocarbons under European Conditions. Atmospheric Environment, 30 (2), 181-199.