Journal of Environmental Management (2001) **62**, 429–442
doi:10.1006/jema.2001.0451, available online at http://www.idealibrary.com on IDEAL®

A critique of EPA's index of watershed indicators

M. T. Schultz

Numerous indices have been developed to assess water quality and the impact of programs to improve quality. The Index of Watershed Indicators (IWI) is one such index created by the US Environmental Protection Agency to assess watershed vulnerability and condition in the United States. The credibility and applicability of subjective indices such as IWI depends upon their ability to withstand tests that challenge their internal consistency and interpretation. This paper critiques IWI on the basis of these tests and other considerations, and suggests that explicitly basing the index on multiattribute utility theory and methods could help resolve many of these difficulties.

Keywords: environmental quality indices, watersheds, water-quality management, multiattribute utility theory.

Introduction

Interest in aggregate measures of water and habitat quality arises from a desire for tools to assess progress toward national water quality goals. While some studies have tried to measure temporal trends in individual water-quality parameters, conflicting trends among parameters and across locations make it difficult to draw conclusions from study results (Smith *et al.*, 1987; Lettenmaier *et al.*, 1991; Knopman and Smith, 1993). Factors complicating assessments of water quality include the multidimensional nature of the water-quality concept, the use-dependency of perceptions regarding water quality and the lack of consistency in regional and national databases. Indices have been suggested as a means of aggregating dimensions of the water-quality concept to make inferences about trends in watershed environmental quality. An index aggregates information about water-quality parameters at different times and in different places and translates this information into a single statistic that is representative of the time period and spatial unit under consideration. There are no hard and fast rules for constructing an index. In each case, rules are derived from a specific understanding of how the index will be interpreted and how it will be used. Therefore, a water-quality index should be specific to a water use or a set of goals.

The US Environmental Protection Agency (EPA) is using an index number approach to evaluate water quality in individual watersheds. The purpose of this index, called the Index of Watershed Indicators (IWI), is to: (1) characterize the condition and vulnerability to pollution of the nation's watersheds; (2) inform water managers and citizens about their watersheds and work to protect them; and (3) measure progress towards EPA's goal that all watersheds be healthy and productive places (US EPA, 1997; Spooner and Lehmann, 1998). The form and components of the index have been evolving for 4 years and EPA is candid about the developmental stage of its work, which has also been the subject of an EPA Science Advisory Board (SAB) review. The SAB recommends development of a conceptual model to guide the choice of IWI indicators, and notes that the current index falls short of the goal of characterizing watershed condition and vulnerability (US EPA, 1999b). In contrast to the SAB review, this paper more narrowly addresses the index

Email of author: mschultz@andrew.cmu.edu

Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213-3890, USA

Received 29 November 1999; accepted 16 March 2001

0301–4797/01/080429+14 $35.00/0

structure and suggests MAUT as a possible framework for developing a conceptual model.

This paper provides a constructive critique of the index in its present form and describes issues that, if addressed, could improve the ability of this index to achieve the goals specified by EPA. The paper discusses bilateral indices, used by economists to calculate an aggregate growth rate in price or quantity between two time periods. This discussion suggests that the axiomatic properties of bilateral indices, and the assumptions surrounding their applicability and use, could also apply to subjective environmental indices developed using multiattribute utility theory (MAUT). MAUT has been widely used (National Research Council (NRC), 1994; McDaniels, 1996; Kim *et al.*, 1998) and it is consistent with the theory underlying other water-quality indices such as the National Sanitation Foundation Water Quality Index (Brown, 1970; Ott, 1978) and the Great Lakes Nearshore Index (Steinhart *et al.*, 1982; Schierow and Chesters, 1988). Although IWI was not specifically developed using MAUT, it is similar in structure. Regardless of the methods used to develop an index, the credibility and applicability of subjective indices such as IWI depends upon their ability to withstand tests that challenge their internal consistency and interpretation. This paper presents a critique based on these tests and other considerations, and suggests where an MAUT approach might help resolve these apparent difficulties.

Index theory

Economists have developed a substantial body of literature on indices (Fisher, 1922; Diewert, 1993). The basic building blocks are bilateral indices that measure the aggregate growth rate in prices or quantities between two periods or locations. Economic indices such as the consumer price index (CPI) are a visible and influential application that are one model for water-quality indices. A price index is a time series of the ratio of a value function in one time period to the value function in another time period. Meaningful units of time (month, year) and space (household, state or nation) are defined. A value function can take many forms but additive forms that sum the product of N prices and quantities sold within a spatial and temporal unit are most common:

$$V^{\mathrm{t}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} p_{\mathrm{i}}^{\mathrm{t}} q_{\mathrm{i}}^{\mathrm{t}} \quad i = \{1, 2, 3, \ldots, N\}$$

where:

V = economic value

p = price

q = quantity

The subscript i is an index of goods and the superscript t denotes a spatial and temporal unit. In the axiomatic approach to price indices described here, prices and quantities are assumed to be independent. The difficulty with taking a simple ratio of prices or quantities in one time period relative to a base period is that both commodity prices and quantities change. Numerous axiomatic tests have been devised to determine whether an index form can indeed serve as a metric (Fisher, 1922). Fisher's ideal price index is described here because it has been shown to meet 22 axiomatic validity tests for an aggregate growth index, more than any other method (Diewert, 1993). Fisher's ideal price index is the square root of the product of the Laspeyeres' and Paasche's price indices. Laspeyere's price index is calculated holding the quantity fixed at its value in the base period:

$$P_{\mathrm{L}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} p_{\mathrm{i}}^{2} q_{\mathrm{i}}^{1}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} p_{\mathrm{i}}^{1} q_{\mathrm{i}}^{1}} \quad i = \{1, 2, 3, \ldots, N\}$$

where:

P_{L} = Laspeyere's price index

Paasche's price index holds quantity constant at its value in the second period:

$$P_{\mathrm{P}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} p_{\mathrm{i}}^{2} q_{\mathrm{i}}^{2}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} p_{\mathrm{i}}^{1} q_{\mathrm{i}}^{2}} \quad i = \{1, 2, 3, \ldots, N\}$$

where:

$$P_P = \text{Paasche's price index}$$

Laspeyere's price index is regarded as a lower bound on the 'true' price index and Paasche's price index is regarded as an upper bound. Fisher's ideal price index is the geometric mean of the two price indices:

$$P_F = (P_L \cdot P_P)^{0.5}$$

where:

$$P_F = \text{Fisher's ideal price index}$$

The ratio of value in one period relative to a base period yields one plus the aggregate growth rate in prices:

$$P_F = 1 + r$$

where:

$$r = \text{aggregate growth rate in prices}$$

P_F is said to be representative of all component prices. The principles and calculations described for price indices are applicable to quantity indices, but first or base period prices are held constant while allowing quantities to vary in the temporal or spatial dimension (Diewert, 1993).

The application of bilateral indices from classical economics to environmental or water quality indices is not straightforward. Most notable is the absence of prices used as a measure of an environmental attribute's contribution to overall environmental quality. How can the relative importance of various attributes be evaluated and how can credible value scales be constructed? Also notable is the difficulty of defining a quantity vector. What environmental or water quality constituents should an index include? Once characteristics are identified, how can these be aggregated in a value function? Water quality and environmental characteristics are often representative of a point in time and space. What are natural, meaningful and convenient units of time and space? In the case of economic and price indices, a large data collection and reporting network is in place to obtain data. While environmental data collection systems exist, these are sparse. Data collection methods lack consistency and sampling tends to occur on an irregular basis. This makes it difficult to obtain a complete set of data from the right time and place.

MAUT provides an intuitive and theoretically sound framework for developing indices in the absence of prices and relating these to decision objectives or other organizational goals. The multiattribute utility (MAU) function converts quantitative and qualitative indicators of one or more attributes, chosen on the basis of organizational goals, to compare two or more tangible or intangible objects. Indicators, which characterize attributes, are converted to common values using carefully constructed value scales and then combined using a set of weights. A common and convenient form of the MAU function is a weighted linear sum:

$$MAU = \sum_{i=1}^{N} w_i v_i(x_i)$$

$$w_i \geq 0, \sum_{i=1}^{N} w_i = 1$$

The variable w is a weight that gives the relative contribution of the ith attribute to overall utility. The term $v_i(x_i)$ is a value function that is dependent upon the indicator x and that equates to a carefully constructed value or utility scale. Weights are analogous to prices and values to quantities in the economic value function. The choice of weights is often a source of controversy. Sometimes natural weighting strategies are suggested for technical, logical or other reasons and various elicitation protocols have been devised (von Winterfeldt and Edwards, 1986). As indicated, all weights sum to one. The use of elicitation procedures and mathematical operations to aggregate abstract value scales imposes axiomatic requirements on the structure of preferences and indicator scales (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986).

Calculating aggregate trends over time periods

A trend in environmental or water quality is the aggregate change in status of attributes relative to goals. Recall that Fisher's ideal price index method is used to calculate a price or quantity index between two time periods. Calculation of a trend using the MAU function described above is accomplished by holding the weights (prices) constant and allowing attributes to vary across time periods. If weights remain constant across time periods, the aggregate growth rate in utility, r, is calculated:

$$r=\left[\frac{\sum_{i=1}^{N} w_i^1 v_i(x_i)^2}{\sum_{i=1}^{N} w_i^1 v_i(x_i)^1}\cdot\frac{\sum_{i=1}^{N} w_i^2 v_i(x_i)^2}{\sum_{i=1}^{N} w_i^2 v_i(x_i)^1}\right]^{0\cdot 5}-1$$

$$=\frac{MAU^2}{MAU^1}-1$$

Index of watershed indicators

EPA's effort to develop an index of watershed indicators began in 1996, after the completion of a four-year study to define environmental indicators of water quality (US EPA, 1996). IWI utilizes the concepts, concerns and interests that emerged from the 1996 report, but uses a modified set of indicators. Indicators were revised and modified through internal and external review and made suitable for implementation on a national scale. IWI uses available data in existing databases to calculate an index that reflects condition, vulnerability and data sufficiency in 2262 watersheds defined by US Geological Survey hydrologic unit code (HUC) boundaries. EPA states its purposes as follows (US EPA, 1997, 1999a): (1) to characterize the condition and vulnerability to pollution of the watersheds of the United States; (2) to make this information available in a way that would inform and inspire Americans to learn more about their water resources, what affects those resources, and how to protect and restore them for our use and enjoyment and that of future generations; (3) to create a tool to help water quality management professionals make better decisions on strategies and priorities for environmental programs; (4) to measure progress toward EPA's goal that all watersheds be healthy and productive places. EPA released its original version of IWI in October 1997. The index has since undergone three revisions (Version 1.1 in July 1998, Version 1.2 in October 1998 and Version 1.3 in April 1999).

IWI assigns each watershed to one of seven categories based on the value of watershed condition and vulnerability subindices. Watershed condition is defined as the quality of the water. Vulnerability is defined as the susceptibility of the water to future declines in aquatic health given information about pollutant releases and other stressors. Watershed condition and vulnerability are each the weighted sum of component indicators (Table 1). A more detailed description of each indicator is provided in US EPA (1997, 1999a). Each indicator score takes a numeric value between zero and three that reflects the assignment of a watershed to one of three or four condition or vulnerability categories devised by EPA for that indicator. Lower indicator scores and lower values of the condition and vulnerability subindices indicate better conditions or lower vulnerability, respectively.

IWI is a function of calculated watershed condition, watershed vulnerability and data sufficiency:

$$IWI_k=f\{C_k, V_k, S_k\}$$

where:

IWI=index of watershed indicators

C=condition subindex

V=vulnerability subindex

S=data sufficiency

k=index of watersheds

Watersheds are assigned to IWI categories using condition and vulnerability subindices as described in Table 2. For example, a watershed assigned a condition score of 12 and a vulnerability score of 15 would be assigned to IWI category 4. If the calculated condition and vulnerability subindices would place the watershed in IWI categories 1 or 2 but fewer than four condition indicators or fewer than 5 vulnerability indicators are available, the watershed is placed in IWI category 7 to indicate insufficient information for classification.

According to EPA, 'these [seven IWI] categories array a spectrum of watershed health

Table 1. Indicators used in EPA's Index of Watershed Indicators (IWI)

Condition indicators	Vulnerability indicators
1. Fraction of assessed stream miles meeting all designated uses	8. Number of wetland aquatic species at risk
2. Limitations on fish consumption	9. Aggregate excess toxic pollutant loads as a percent over permitted toxic loads
3. Degree of source water impairment	10. Aggregate conventional pollutant loads as a percent over permitted conventional loads
4. Degree of concern over contaminated sediments	11. Percent of cover that is greater than or equal to 25% impervious
5. Percent of STORET observations in which a toxic pollutant concentration exceeds the national chronic criteria	12. Watershed rank in terms of potential pesticide and nitrogen runoff and potential in-stream sediment loads
6. Percent of STORET observations in which a conventional pollutant exceeds a reference level	13. Percent change in population between 1980 and 1990
7. Percent of historic wetlands lost since 1780	14. Volume of water impounded behind dams
	15. Relative ability of each estuary to concentrate dissolved and particulate pollutants
	16. Atmospheric deposition of nitrogen by unit area[a]

[a]Indicator 16, atmospheric deposition of nitrogen by unit area, is numbered Indicator 17 in EPA documents and the original Indicator 16 was dropped. The indicator is renumbered 16 here for clarity.

Table 2. Index of Watershed Indicators

Watershed condition score	$0 \leq V \leq 9$	Watershed vulnerability score $10 \leq V \leq 18$	V undetermined
$0 \leq C \leq 7$	1	2	7
$8 \leq C \leq 18$	3	4	—
$19 \leq C \leq 30$	5	6	—
C Undetermined	7	—	7

that suggest opportunities for different management responses (US EPA, 1999a).' Table 3 lists a verbal description for each IWI category. This description consists of a statement about the quality of water (condition) and the susceptibility of the water quality to degradation (vulnerability). An IWI category is interpreted using three qualitative statements, one describes condition, the second describes watershed vulnerability, and the third describes data sufficiency. There are three possible conditions: 'better water quality', 'less serious water-quality problems', and 'more serious water quality problems'. According to US EPA (1999a): (1) '*better water quality*' implies that data are sufficient to assert that the designated uses are largely met. The term 'designated uses' refers to the specific set of beneficial uses of water each state or other local authority assigns to the stream segment. Beneficial uses include swimming, agriculture, drinking water supply, freshwater life support and other uses that suggest appropriate numeric water quality criteria for pollutants in that stream segment. (2) '*less serious water quality problems*' implies watersheds with aquatic conditions below water-quality goals that have problems revealed by other indicators. (3) '*more serious water-quality problems*' implies aquatic conditions well below water-quality goals that have serious problems exposed by other indicators.

Table 3. IWI category descriptions

Category	Description
1	Watersheds with better water quality and lower vulnerability to stressors such as pollutant loadings
2	Watersheds with better water quality and higher vulnerability to stressors such as pollutant loadings
3	Watersheds with less serious water-quality problems and lower vulnerability to stressors such as pollutant loadings
4	Watersheds with less serious water-quality problems and higher vulnerability to stressors such as pollutant loadings
5	Watersheds with more serious water-quality problems and lower vulnerability to stressors such as pollutant loadings
6	Watersheds with more serious water-quality problems and higher vulnerability to stressors such as pollutant loadings
7	Watersheds for which insufficient data exist to make an assertion about condition and vulnerability

EPA defines two possible levels of vulnerability: (1) '*low vulnerability*' implies that pollutants or other stressors are low, and, therefore there exists a lower potential for future declines in aquatic health. (2) '*high vulnerability*' implies significant pollution and other stressors and, therefore, a higher vulnerability to declines in aquatic health.

EPA defines two possible levels of data sufficiency: (1) '*Data sufficient*' implies that four or more indicators of watershed condition could be calculated and that six or more indicators of watershed vulnerability could be calculated. (2) '*Data not sufficient*' implies that not enough data were available in this watershed to calculate either watershed condition or watershed vulnerability. (Approximately 27% of watersheds fall into this category).

Watershed condition subindex

Seven component indicators are aggregated to arrive at a watershed condition subindex between zero and 30. This subindex is the weighted sum of condition indicator scores that are assigned a value between zero and three based on Indicators 1 through 7 described in Table 1. A weighted linear sum combines the seven indicator scores to calculate a condition subindex:

$$C_{\mathrm{k}}=\sum_{\mathrm{i}} w_{\mathrm{i}} c_{\mathrm{ik}}$$

where:

w=weight

c=condition indicator score

C=condition subindex

i=index of indicator

A weight of six is used for Indicator 1 (designated use indicator) to reflect the relative importance of biennial 305(b) reports, state water quality assessments required under the Clean Water Act. A weight of one is used for all other condition indicators. This weighting mechanism ensures that if the designated use indicator receives a score of three because 20% or fewer assessed stream miles meet all designated uses, 'more serious water quality problems' are indicated regardless of other indicator scores. In the case that either fewer than 20% of stream miles are assessed or no estimate of the percentage of stream miles satisfying designated uses is available, the condition index is calculated by increasing the weight of remaining indicators from one to three. There must be enough data available to assign scores to at least four indicators or the condition subindex for that watershed is undetermined.

Watershed vulnerability subindex

Nine component indicators are aggregated to determine a watershed vulnerability subindex between zero and 18. This subindex is the weighted sum of vulnerability indicator scores that are assigned a value between zero and two based on Indicators 8 through 16 described in Table 1. In contrast to condition indicators, these vulnerability indicators reflect 'the presence, absence, or trends in stressors that can cause degraded water or habitat quality' (US EPA, 1999a). A weighted linear sum calculates a vulnerability subindex that assesses the potential for future declines in aquatic health:

$$V_{\mathrm{k}}=\sum_{\mathrm{i}} w_{\mathrm{i}} v_{\mathrm{ik}}$$

where:

v=vulnerability indicator score

V=vulnerability subindex

Weights used in this aggregation of indicators are uniformly one, thus each indicator score contributes equally to the assessment of watershed vulnerability. Not all indicators are available in all watersheds. Less than 5% of watersheds have a full complement of vulnerability indicators. When some indicators are not available the method assigns a value of 0 to the missing indicator. There must be enough data available to assign scores to at least six vulnerability indicators or the vulnerability subindex for that watershed is undetermined.

IWI Version 1.3 Results

According to EPA, IWI categories array a spectrum of watershed health that suggest opportunities for different management responses (US EPA, 1999a). For example, developers suggest that watersheds assigned higher condition scores and lower vulnerability scores might be made targets of federal programs that fund watershed restoration projects. In a similar way, watersheds that are assigned lower condition scores and higher vulnerability scores might be made targets of federal programs that fund watershed protection programs. EPA does not intend that users interpret the numbers one through seven used to denote IWI categories on either an interval or an ordinal scale (Lehmann, pers. comm.; Spooner, pers. comm.).

IWI results are summarized in Table 4 for Version 1.3 completed in April 1999. Of the 2262 watersheds in the 50 US States and Puerto Rico, both condition and vulnerability scores can be calculated for 1649 watersheds. There are 303 watersheds (13%) with both better water-quality condition and low vulnerability to degradation. Less than 6% of watersheds have a vulnerability subindex that is greater than eight. Scores greater than eight suggest a high vulnerability to future degradation. Table 4 also shows that 516 watersheds (23%) are said to have 'more serious water-quality problems' and that no statements could be made about 613 watersheds (27%) because of insufficient information.

Table 5 lists by indicator the number of watersheds assigned each indicator score. Indicator 1 is a function of designated use attainment as reported by the states in biennial 305(b) reports. There are 437 watersheds that, according to the states, support designated uses in 20% or fewer assessed stream miles and therefore receive an indicator score of 3 and an IWI condition subindex of 18 or greater, indicating 'more serious water quality problems'. Comparing this result from Table 5 with results in Table 4 shows that these 437 watersheds account for 84% of the 516 watersheds falling in the 'more serious water-quality problems' category ($18 \leq C \leq 30$).

Table 4. Index of Watershed Indicator results (Version 1.3)

Watershed condition score	$0 \leq V \leq 8$	Watershed vulnerability score $9 \leq V \leq 18$	V undetermined
$0 \leq C \leq 7$	303	31	
$8 \leq C \leq 17$	739	60	
$18 \leq C \leq 30$	480	36	
C Undetermined			613

Table 5. Number of watersheds by indicator score (Version 1.3)

Indicator	Indicator score				
	0	1	2	3	S
1	486	454	414	437	471
2	95	203	500	—	1464
3	555	340	297	—	1070
4	1270	56	40	—	896
5	565	189	8	—	1500
6	487	583	288	—	904
7	31	1581	499	—	151
8	430	745	422	—	665
9	2082[a]	77	41	—	62
10	2144[a]	8	21	—	89
11	1645	218	115	—	284
12	526	1055	529	—	152
13	1223	322	657	—	60
14	491	960	481	—	330
15	8	81	60	—	2113
16	1173	777	161		153

Column S lists the number of watersheds in which data sufficiency conditions for this indicator were not met.
[a]These sums include watersheds in which either no discharge monitoring reports are required or no permitted dischargers are located. These data are from US EPA (1999a).

Results of the index are shown graphically in Figure 1. Darker shading indicates a larger number of watersheds by condition and vulnerability score. Lower condition scores are associated with watersheds in better condition and lower vulnerability scores are associated with watersheds less vulnerable to degradation. This graph shows that watersheds tend to be clustered around four points and much of the space in the vulnerability dimension is underutilized because few watersheds are assigned high vulnerability scores. There is also a negative correlation between the 1649 pairs of condition and vulnerability scores ($r = -0{\cdot}1951$, $P \leq 0{\cdot}0001$).

Critique

Use of an index to characterize watershed condition or watershed vulnerability is

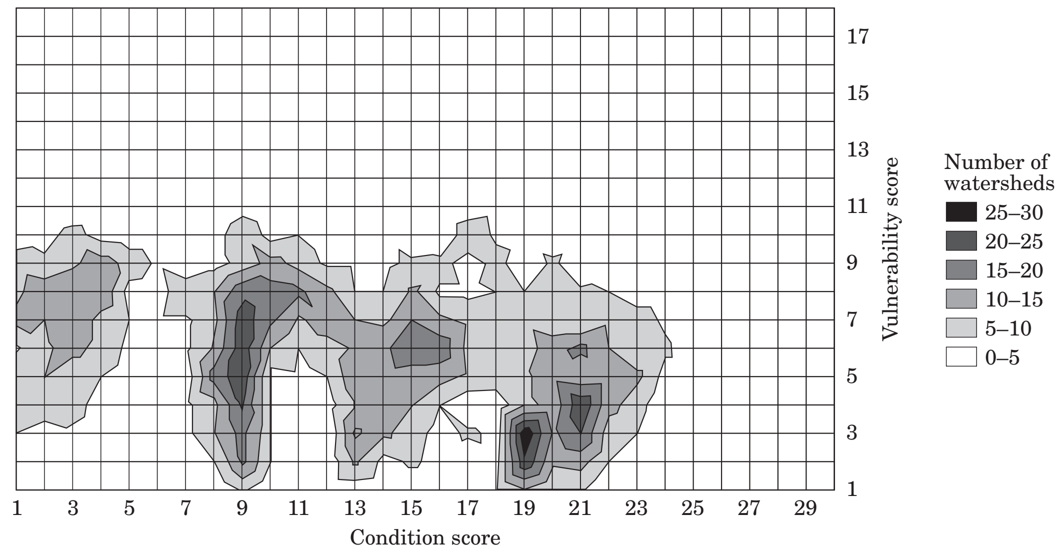

Figure 1. National results of the Index of Watershed Indicators showing 1649 watersheds by condition and vulnerability scores. Either a condition score or a vulnerability score could not be calculated for remaining watersheds.

justified because these are multi-dimensional concepts. A systematic simplification of data is needed to aggregate dimensions of watershed condition and vulnerability and to discern and evaluate differences in spatial and temporal dimensions. However, several characteristics of this index create problems that impede its interpretation. Other characteristics are inconsistent with assumptions implied by the use of an additive aggregation function. The following discussion highlights these and other difficulties that interfere with interpretation of IWI.

Relationship between indicators and water-quality objectives

An index of this type cannot be validated by comparison with any field data because it has no independent physical interpretation (NRC, 1994). Validity of an index of this type is demonstrated by a logical connection between agency goals or other objectives and the aggregation of indicators. IWI lacks such a connection to agency goals and its connection to the particular set of designated uses assigned in the watershed is rather weak. The absence of a connection between the chosen indicators and water-quality or watershed-management objectives raises questions about what kind of progress toward goals is measured by changes in the value of the index.

Operationality and transparency of indicators

In general, the credibility of the index will increase with greater transparency of calculation procedures. Transparency of IWI indicators is sometimes limited. For example, EPA uses an *adhoc* procedure to interpret three different national databases in order to calculate a single indicator of source water condition (Indicator 3). In a similar way, the indicator of agricultural runoff potential (Indicator 12) is calculated using results of four separate watershed ranking procedures. Operationally complex calculation procedures can obscure interpretation and make recalculation and verification difficult and this detracts from the credibility of the index. Because a correlation between the calculated indicator and actual source water condition or agricultural runoff potential cannot be verified, there is a need to explain why it is believed that such calculations yield indicators that are correlated with unobservable measures of watershed condition and vulnerability.

Combination of relative and absolute indicators

Some indicator scores are assigned on the basis of relative comparisons among watersheds (Agricultural runoff potential, Indicator 12). Indicators that are based on a relative comparison among watersheds are not descriptive of the watershed, but rather descriptive of the relationship among watersheds. Combining relative indicators with descriptive indicators in an aggregation function creates confusion about the meaning of the value scale. It also creates a situation in which the index for any one watershed is not independent of the index for other watersheds. This may complicate or prevent attempts to aggregate indices in the spatial dimension, although EPA has not recommended that IWI be used this way.

Definition of spatial and temporal units

An index of watershed condition or vulnerability should utilize a set of indicators that is descriptive of the entire area under consideration rather than isolated points within the area. Examples of such indicators include the percent of stream miles meeting all designated uses, and the percent cover that is greater than 25% impervious. However, it appears that some indicators are more descriptive of events or points within the watershed. For example, condition indicators for toxic and conventional pollutants are based on isolated events recorded in STORET, a water quality database EPA uses to compile results of water quality samples taken by independent agencies, organizations and individuals. These measurements are cataloged by HUC code, but tend to be representative of just a few points at a few times within a watershed. The ability to extrapolate from these points to the watershed as a whole depends upon many site-specific factors such as the frequency of sampling and the distribution of samples within the watershed. Such factors should be investigated before making conclusions as to the extent to which these data are meaningful in the context of a watershed index.

In a similar way, the ability to estimate trends over time requires that all data used to calculate the index be representative of a well-specified time period. For example, it is said that IWI will be used to chart progress towards national water-quality goals. Therefore, it will be necessary to update indicators periodically to recalculate the index. Although it would be best if all indicators were updated together on the basis of new information, this is likely to be costly and impractical. However, updating only some of the indicators periodically as data become available creates lag and interaction effects between time periods that will complicate trend estimation. Because this cost creates a disincentive for updating and maintenance of the index, minimizing data requirements could improve overall utility of IWI or any comparable index.

Additivity of condition and vulnerability indicators

Use of a weighted linear sum to aggregate value scales requires that the additivity axiom be satisfied (von Winterfeldt and Edwards, 1986). The additivity axiom is satisfied if each interval of the value scale represents an equivalent change in condition or vulnerability. For example, the difference between condition in those watersheds assigned an indicator score of zero and those watersheds assigned an indicator score of one must be the same as the difference in condition between watersheds assigned an indicator score of one and those watersheds assigned an indicator score of two. The same must be true across condition attributes. An indicator score of one for the designated use indicator should imply the same contribution to the multiattribute condition as an indicator score of one for the source water condition indicator. This requirement can be quite difficult to achieve in practice because the value scale has no physical interpretation.

Figure 2 shows how category boundaries can be adjusted so two indicators have equivalent value scale increments. The x-axis is the initial indicator dimension, for example, limitations on fish consumption (Indicator 2) or degree of source water impairment (Indicator 3). The y-axis is divided into three categories, the numeric labels of which correspond to the indicator score assigned to watersheds falling in the first, second or third interval of the x-axis. Boundaries on the x-axis are adjusted for each indicator so the decision maker's value assessment

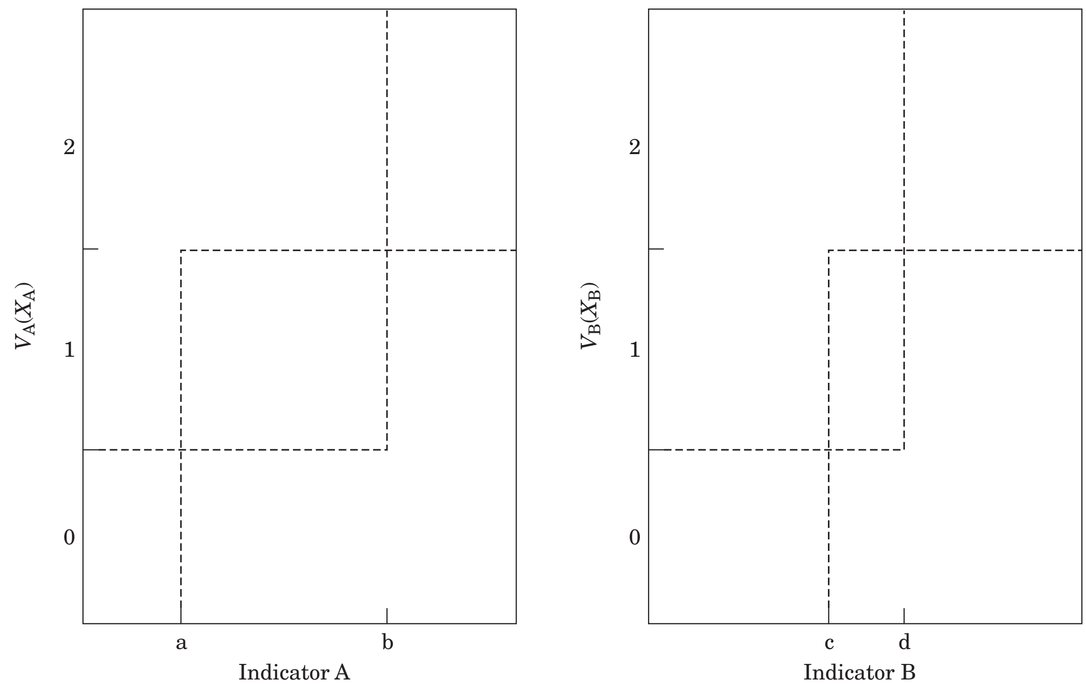

Figure 2. Adjustment of category boundaries. Category boundaries a, b, c and d can be adjusted so that value scale increments are equivalent for each indicator.

is consistent with $v_A(a)=v_B(c)$ and $v_A(b)=v_B(d)$. In practice, category boundaries can be adjusted in the indicator dimension by elicitation procedures (von Winterfeldt and Edwards, 1986). Another approach that has been used for indices of water quality is to equate the value scale with an environmental damage function. This leads to consistency among indicators and is the approach used in the National Sanitation Foundation's Water Quality Index (Brown *et al.*, 1970; Ott, 1978), and the Great Lakes Nearshore Index (Steinhart *et al.*, 1982; Schierow and Chesters, 1988). Unlike these water-quality indices that are based on a few selected water-quality parameters, IWI indicators are more diverse and a watershed environmental damage function is harder to define. Therefore, IWI category boundaries should be carefully justified.

Another situation under which an additive aggregation function may not be appropriate occurs when independence assumptions are violated. Independence is violated when a change in the value of one indicator produces a change in another indicator. In such cases, one indicator can be expressed as the function of the other. This is the reason that condition and vulnerability subindices cannot be easily combined in a single additive index. The condition subindex reflects ambient conditions that are the direct result of environmental stressors assessed in the vulnerability subindex. Such dependencies invalidate the index because some changes in watershed condition or vulnerability may be double counted. IWI Indicators 1, 5 and 6 illustrate the problem of redundant indicators. Indicator 1 is a measure of designated use attainment, and Indicators 5 and 6 compare ambient water-quality data with national chronic and reference level criteria for conventional and toxic pollutants, respectively. The indicators may be redundant because a reduction in conventional or toxic pollutant loadings could improve water quality and consequently change the value of two indicator scores. If so and depending upon the level of each change, the index would appear to double count a single improvement in water quality. In this case, a multiplicative aggregation function is needed to resolve dependencies (von Winterfeldt and Edwards, 1986). However, it should also be possible to devise indices that avoid the need for multiplicative aggregation functions through the initial choice of indicators.

Bias in condition and vulnerability subindices

Condition and vulnerability subindices tend to underestimate condition and vulnerability. This bias arises because available information for many watersheds is insufficient to calculate all of the indicators in each subindex. If an indicator cannot be calculated, EPA assigns it the score of zero, the lowest score on the value scale. The effect of this bias can be seen in Figure 3. The average vulnerability score increases from 1·8 to 7·8 as the number of vulnerability indicators increases from 5 to 9. The change in average condition score is less pronounced, the average condition score increases from 11·1 to 12·6 as the number of condition subindices increases from 4 to 7. EPA recognizes this problem and uses data sufficiency criteria to help reduce distortion caused by missing data. No condition subindex is calculated if fewer than four of the seven indicators are available and no vulnerability subindex is calculated if fewer than five of the nine indicators are available. However, differences in mean values understate the effect of missing data because these differences are large enough to alter the assignment of individual watersheds to IWI categories.

The assumption that is implicit in this approach is that, in the absence of information, the best possible condition and the lowest possible vulnerability are indicated. One justification for this assumption is that EPA has historically sought out and documented information about environmental problems (Spooner, pers. comm.). In contrast, data to calculate condition and vulnerability indicators may have been less frequently collected where few environmental problems have been noted. It is true that relatively little data may have been collected in areas where few problems exist or in remote areas. However, it is a mistake to infer the absence of environmental problems from the absence of data. Other means of treating missing indicators exist. For example, it is possible to divide a condition or vulnerability score by the maximum possible score to account for differences in the availability of information. This approach is similar to that used by the British Columbia Ministry of Environment, Lands and Parks in its watershed-level index of water quality (Zandbergen and Hall, 1998).

Choice of weights

EPA uses a weight of six for its designated use indicator and a weight of one for all other indicators. The use of equal weights on all but one indicator is arbitrary in the absence of a logical weighting mechanism. For example, EPA justifies a weight of six on the designated use indicator of the condition subindex by saying that any watershed in which at least 20% of stream miles do not satisfy water uses designated by the states must have, by definition, 'more serious water quality problems'. However, no rationale is given for assigning equal weights to remaining indicators. A logical and defensible weighting scheme could improve the interpretability and credibility of the index.

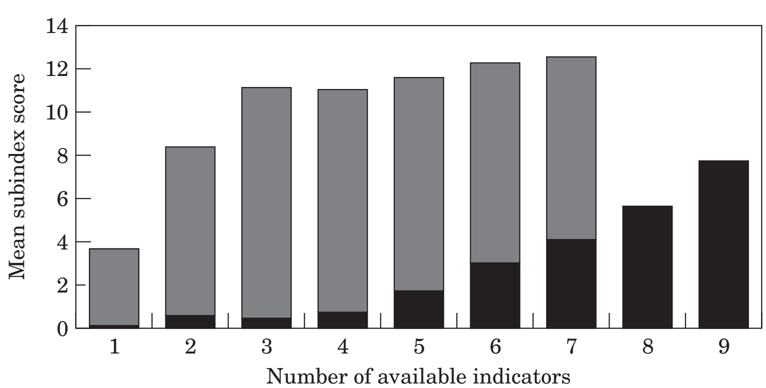

Figure 3. Bias in condition and vulnerability subindices. To reduce the effect of bias on IWI results, no subindex scores are calculated for watersheds with fewer than four of the seven condition indicators or fewer than five of the nine vulnerability indicators. Condition subindex score; ▩; vulnerability subindex score; ■.

Negative correlation between condition and vulnerability subindices

One means of evaluating an index is to consider whether results are consistent with available information. In this case, the pattern in condition and vulnerability subindices is observed for consistency with expectations about an assumed relationship between ambient conditions and environmental stressors. *A priori*, one might expect a positive correlation between the two subindices. Such a relationship would suggest that ambient water-quality conditions tend to be worse in watersheds where indicators of higher pollutant load and level of environmental stress are greatest. However, Figure 1 shows a negative Pearson correlation between these two subindices ($r=-0{\cdot}1951, P\leq 0{\cdot}0001, \mathrm{N}=1649$). This implies that watersheds with higher levels of pollutant loads and other environmental stressors have better ambient conditions. Although this correlation coefficient is not large, the pattern is statistically significant and not easy to explain. The pattern can be attributed to correlation between the condition subindex and two components of the vulnerability subindex, the population change code (Indicator 13, $r=-0{\cdot}1812, P\leq 0{\cdot}0001, \mathrm{N}=1637$) and the indicator of modifications caused by dams (Indicator 14, $r=-0{\cdot}1692, P\leq 0{\cdot}0001, \mathrm{N}=1560$). Pearson correlation coefficients between the condition subindex and remaining components of the vulnerability subindex are statistically insignificant.

Minimal contribution of many indicators

Simplicity and interpretability are desirable qualities of an index, and many individuals might argue that those indicators that contribute little to the meaning of an index should be excluded because they can obfuscate its meaning. This discussion illustrates that raw values of Indicator 1 of the IWI condition subindex can communicate much the same information as the entire condition subindex itself. Indicator 1 is the most heavily weighted variable in the condition subindex, and receives a score between zero and three depending upon the percentage of assessed stream miles in that watershed that satisfy their designated uses. This categorical classification and numeric coding reflects but is distinct from the raw values of the indicator scale, which are the actual percentage estimates in EPA's National Assessment Database (NAD) (US EPA, 1997).

A Spearman rank correlation (r_s) is used to help determine how well the simpler more easily interpreted index of actual percentage estimates agrees with the results of the IWI condition subindex. This method has been used by other authors to assess the level of agreement between indices (Spencer *et al.*, 1998). The Spearman rank correlation coefficient is $r_s=-0{\cdot}7589$ ($P\leq 0{\cdot}0001, \mathrm{N}=1226$), indicating that about three quarters of the information contained in the ranking of watersheds using the IWI condition subindex is also contained in the raw percentage estimates used to determine the indicator score. The well-known principle of Occam's razor would suggest that the simpler, more easily interpreted version might be preferred. However, the two indices differ in other respects. Since IWI allows the aggregation of indicators even if some indicator scores are missing, the condition subindex can be applied to a larger number of watersheds than the simplified index, which is available in only about 80% of all watersheds.

Conclusion

Subjective environmental indices are similar to axiomatic indices devised by economists to help measure changes in an economy over time and differences in economies across locations. They are similar because both measure changes that cannot be quantified using generally understood and accepted units based on physical concepts. Economic indices are accepted as measuring devices because they satisfy a set of axioms describing the desirable properties of such devices. It is reasonable to require that subjective environmental indices satisfy similar tests. If the index is similar in form to Fisher's ideal index, as is IWI, there seems no reason these expectations cannot be met. However, the calculation of IWI in practice requires inputs that are not easily quantified. MAUT provides a theoretically sound conceptual framework to overcome this obstacle that has been a notorious impediment to the development of environmental indices.

As described in the critique, IWI has several limitations. These limitations interfere with its four stated purposes of characterization, communication, decision-making and measurement. IWI is not a reliable characterization tool because, among other things, there is ambiguity in the meaning of indicator value scales and inconsistency in the aggregation of these scales. To the extent that this index has limited ability as a characterization tool, its use as a communication tool is also limited. For example, IWI is implemented in conjunction with an internet-accessible information retrieval system that makes available to stakeholders a wide-variety of selected federal data at the watershed level. For the reasons discussed in this critique, IWI cannot meaningfully synthesize these data. Therefore, it cannot serve as a tool for communication with the public or as a tool for management decision-making. IWI is also described as a tool for measuring trends or progress towards EPA's goals. Structurally, IWI appears to satisfy axioms that qualify it as a metric in general. However, IWI cannot serve as a metric toward agency goals in the absence of a clear connection with these goals. A MAU approach could provide the basis for a more formal conceptual framework in which to consider agency objectives and assist EPA construct a theoretically sound and defensible index of watershed attributes.

MAU methods involve techniques that link an index to organizational or other objectives and specifically address some of the issues that are not now well addressed by IWI. For example, in MAUT the chosen indicator set consists of a minimum number of indicators that completely define and are uniquely assigned to organizational objectives. This facilitates interpretation of the index and helps avoid the kind of redundancies that may lead to double counting of improvements in watershed condition or vulnerability. MAU methods also provide means of weighting indicators so they are consistent with organizational objectives. If these or other methods can be employed effectively to address issues raised in this paper, IWI could become a more sound and interpretable index.

Acknowledgements

This work was completed at Carnegie Mellon University and supported by funding from the Center for the Study and Improvement of Regulation at Carnegie Mellon University. Scott Farrow, Paul Fischbeck, Sarah Lehmann, Mitchell Small and Charles Spooner provided helpful discussions and three anonymous reviewers provided comments and suggestions that helped improve the manuscript. However, the author is responsible for any errors, omissions or interpretations.

References

Brown, R. M., McClelland, N. I., Deininger, R. A. and Tozer, R. G. (1970). A water quality index: do we dare? *Water Sewage Works* **117**, 339–343.

Carson, R. T. and Mitchell, R. C. (1993). The value of clean water: the public's willingness to pay for boatable, fishable and swimmable quality water. *Water Resources Research* **29**, 2445–2454.

Diewert, W. E. (1993). Indices. In *Essays in Index Number Theory, Volume 1* (W. E. Diewert and A. O. Nakamura, eds), pp. 71–104. Amsterdam: North Holland Press.

Fisher, I. (1922). *The Making of Index Numbers*. Boston, MA: Houghton Mifflin.

Keeney, R. and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, 569 pp. New York: John Wiley and Sons.

Kim, T., Dwak, S. and Yoo, S. (1998). Applying multi-attribute utility theory to decision making in environmental planning: a case study of the electric utility in Korea. *Journal of Environmental Planning and Management* **41**, 597–609.

Knopman, D. and Smith, R. A. (1993). Twenty years of the clean water act. *Environment* **35**, 17–20.

Lettenmaier, D., Hooper, E., Wagoner, C. and Faris, K. (1991). Trends in stream quality in the continental United States, 1978–87. *Water Resources Research* **27**, 327–339.

McDaniels, T. L. (1996). A multiattribute index for evaluating environmental impacts of electric utilities. *Journal of Environmental Management* **46**, 57–66.

National Research Council (NRC) (1994). *Ranking Hazardous Waste Sites for Remedial Action*. Washington, DC: National Academy Press.

Ott, W. R. (1978). *Environmental Indices: Theory and Practice*. Ann Arbor, MI: Ann Arbor Science Publishers, Inc.

Schierow, L. and Chesters, G. (1988). Evaluation of the great lakes nearshore index. *Water Research* **22**, 269–277.

Smith, R., Alexander, R. and Wolman, M. G. (1991). Water quality trends in the nation's rivers. *Science* **235**, 1607–1615.

Spencer, C., Robertson, A. and Curtis, A. (1998). Development and testing of a rapid appraisal

wetland condition index of south-eastern Australia. *Journal of Environmental Management* **54**, 143–159.
Spooner, C. and Lehmann, S. (1998). Index of watershed indicators. *Sea Technology* **39**, 51–53.
Steinhart, C., Schierow L. and Chesters, G. (1982). An environmental quality index for the great lakes. *Water Resources Bulletin* **18**, 1025–1031.
US Environmental Protection Agency (US EPA) (1996). *Environmental Indicators of Water Quality in the United States (EPA 841-R-96-002)*. Office of Water, Washington DC.
US Environmental Protection Agency (US EPA) (1997). *The Index of Watershed Indicators (EPA-841-R-97-010)*. Office of Water, Washington, DC.
US Environmental Protection Agency (US EPA) (1999a). *Index of Watershed Indicators*. http://www.epa.gov /iwi.
US Environmental Protection Agency (US EPA) (1999b). *An SAB Report: Review of the Index of Watershed Indicators (EPA-SAB-EPEC-99-014)*, 27 pp. Ecological Processes and Effects Committee, Science Advisory Board, Washington DC.
von Winterfeldt, D. and Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge: Cambridge University Press.
Zandbergen, P. A. and Hall, K. J. (1998). Analysis of the British Columbia water quality index for watershed management. *Water Quality Research Journal of Canada* **33**, 519–549.