Statistical Considerations in the Development of Environmental Indices: the Example of the Canadian Water Quality Index

Esterby, Sylvia R.
University of British Columbia Okanagan ; Mathematics, Statistics and Physics
ASC 347, 3333 University Way
Kelowna, Canada V1V 1V7
E-mail: sylvia.esterby@ubc.ca

El-Shaarawi, Abdel H.
McMaster University and National Water Research Institute
867 Lakeshore Road, P.O. Box 5050
Burlington, Canada L7R 4A6
E-mail: Abdel.El-Shaarawi@ec.gc.ca

Aggregated environmental indices are seen as a means of providing information to two groups: the public and decision makers. Thus, from the monitoring programs designed to track the health of the aquatic environment, results are reported in an understandable form to decision-makers and the public, in addition to dissemination through technical publications. Implicit in this method is the assumption that the complex environmental information can be reduced to a simple aggregate index. The question arises as to whether such a reduction can be justified and under what conditions, since even a simple message must be defensible and accurate in the most important parts of the complex picture. In this paper we review some cases of the use of environmental quality indicators, consider criteria such indicators should meet, and discuss the Canadian Council of Ministers of the Environment (CCME) Water Quality Index (WQI).

Examples of indicators and indices

Four examples which vary in some aspect are the following : 1) an indicator which measures quality relative to a specific use such as an indicator organism for safe drinking water, 2) a measurable variable used to trace the response of a water body to a regulation, where a decrease in ambient concentration was expected as the result of a decrease in loading (phosphorous), 3) a biological indicator as an integrator of the effects of pollution on the health of the water body, and 4) a water quality index which is an aggregated index designed to represent some empirical model of reality.

The first three examples draw on North American experience. Microbiological quality is determined by testing drinking water for *Escherichia coli* (Health Canada, 2010) which serves as the indicator organism for disease-causing bacteria, protozoa and viruses. Coliforms will be the first bacteria present in contaminated water and in larger concentrations than other types of pathogenic microbes, thus providing an indicator that is also readily detectable. Statistical properties of a drinking water regulation are given by Carbonez, El-Shaarawi and Teugels (1999). Eutrophication in the North American Great Lakes provides the second example. By 1964 extensive algal growth in the lakes prompted the International Joint Commission, through participating US and Canadian agencies, to seek a means of remedying the situation (Stevens and Neilson, 1987). Phosphorus was deemed the limiting nutrient and regulations were enacted to control the loading of total phosphorus (TP) to Lake Ontario. To follow the response of the lake to reduced loading, the 1972 and 1978 Great Lakes Water Quality Agreements called for monitoring of ambient TP concentrations in Lake Ontario. The third example reminds us that the search for a general indicator of water quality has been underway for sometime. An International Joint Commission Workshop was held in 1985 where scientists who specialized in the study of aquatic organisms debated the merits of using specific organisms as integrative indicators of pollution, where criteria such as broad occurrence of a species across the continent and sensitivity to pollutants were considered (Evans, 1988).

The fourth type of indicator can be exemplified by the eutrophication index, one of six indicators comprising a composite pollution index for the Netherlands (Hammond *et al.*, 1995). Phosphates and nitrogen compounds are the primary substances contributing to eutrophication (although as noted above, phosphorous compounds were thought to be limiting). Releases of phosphates and nitrogen compounds are expressed in eutrophication equivalents through an aggregated index which combines information on the eutrophication-causing properties of a compound as well as amounts. Thus the index quantifies a measure related directly to the problem identified, ie. eutrophication.

Perhaps the broadest environmental indicator is the EPI. The 2010 report ranked 163 countries on the basis of the EPI score, itself built up from sets of indicators in categories (Emerson *et al*., 2010). Therein, the water quality index is a proximity-to-target composite indicator based on 5 water quality variables and with a station density adjustment. The current development of sustainability indicators for Canada involves 6 indicators (Statistics Canada, 2007, and The Sustainability Report, 2011), one of which is the CCME WQI.

Criteria for aggregated indices

Environmental aggregated indices are thought to be one way of tracking policy performance beyond the well established economic indicators such as the Gross Domestic Product (GDP). Groups of indicators are being combined to measure sustainability or combined with socio-economic factors to provide indicators of well-being (eg. Emerson *et al.*, 2010; CSIN, 2010). Thus it seems reasonable to consider some criteria.

The report of the World Resources Institute (Hammond *et al*., 1995) considers what is required of an indicator which is to be used to communicate information to decision makers. The authors contend that we must have an underlying model of reality, the indicator must provide a metric against which some aspects of public policy can be measured, it must be analytically sound and have a fixed methodology of measurement, and we must be explicit about the model and the metric.

It is clear that we must understand just what is being measured with an aggregated index. The WRI report (Hammond *et al*., 1995) cites the example of the Dutch indicator for environmental eutrophication, mentioned above. Similarly, the Dutch indicator for climate change is derived from a weighted sum of the discharge of gases based on gases which increase the atmosphere's warming potential. In both cases a phenomenon is identified and a metric based on measurable variables is defined. It is interesting to compare these indicators with the GDP. Statistics Canada defines the GDP as the total unduplicated value of the goods and services produced in a specified economic territory over a specified period (Statistics Canada, 2011). From this definition, methods of estimating the GDP are discussed.

Arguments about the role of single-variable indicators for regulation versus indices for communication and broad comparisons could also be developed. When the action taken on the basis of an indicator is immediate and involves, for example, whether we will get sick if the organism is present or the contaminant has been dumped, we will tend to emphasize the technical validity of the indicator. When we must communicate environmental quality and make comparisons between diverse regions or countries, cases where there are a many variables which don't match from region to region or country to country, then we may have no choice but to use a broader brush, and aggregate and approximate more. The tool to use and the confidence we can attach will differ in these different scenarios. The case of diversity indices and progress towards the Convention on Biological Diversity also provides insight on aggregated indices (Walpole *et al., 2009).*

Canadian Water Quality Index

A logical first step in discussing the idea of a water quality index is to define the term, water quality. The USGS primer on water quality (U.S. Geological Survey, 2001), designed for the general public, states: "Water quality can be thought of as a measure of the suitability of water for a particular use based on selected physical, chemical and biological characteristics" . The OECD Glossary of Statistical Terms (1997) gives the definition: "Water quality refers to the physical, chemical, biological and organoleptic (taste-related) properties of water". Other definitions of water quality are given with respective to some use, usually a use of benefit to human beings, but it could also be relative to conditions for a healthy aquatic ecosystem.

The Canadian Council of Ministers of the Environment (CCME) endorsed what is know as the Water Quality Index (WQI) in 2001 (CCME National Water Quality Index Workshop, 2004). It is the result of a joint effort of the Canadian federal and provincial governments. The index is basically the index developed in the 1990s by the province of British Columbia and modified to match the expectations of the Alberta index such that high values of the index correspond to good quality.

Two types of data are required to calculate the WQI: first, measurements on several water quality variables made over the period and region of interest, and second, numerical values representing water quality objectives associated with each water quality variable. Three measures of the violation of objectives are computed, called factors, based on comparing the data with their objectives, and the factors are combined to provide the value of the index. The factors are:

F1, the Scope, equal to the percentage of variables that failed to meet their objectives (limits) at least once during the time period under consideration.

F2, the Frequency, equal to the percentage of individual tests that failed to meet objectives.

F3, the Amplitude, which measures the amount by which each failed observation differs from its objective.

The ideal data set can be represented as a matrix **X** with p columns and n rows where the entry in ith row and jth column x_{ij} represents the measurement on jth water quality variable (j =1, 2, …, p) for ith case (i =1, 2, …, n). For example, if the index is being computed for a sampling station, where monthly samples were taken for one year, then i=1,2,...,12. There is one or more water quality limit(s), u_j , associated with the jth variable (column). It is more common to have the objective stated as a single limit representing an upper or lower concentration limit for the water quality variable. An example of an upper limit objective is bacterial level (*E. coli* in drinking or recreational waters) and an example of a lower limit is dissolved oxygen concentration in the water. For some variables, an allowable range is used, for example, pH or water level where both droughts and floods are of concern.

The index is defined as (Canadian Council of Ministers of the Environment, 2001):

$$WQI = 100 - \frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732}$$

where:

$$F_1 = \frac{Number\ of\ \ failed\ \mathrm{var}iables}{Total\ number\ of\ \ \mathrm{var}iables} \times 100$$

$$F_2 = \frac{Number\ of\ \ failed\ tests}{Total\ number\ of\ \ tests} \times 100$$

$$F_3 = \frac{nse}{0.01 nse + 0.01}$$

$$nse = \frac{\sum_{i=1}^{n} excursion_i}{\#of\ Tests}$$, and excursions are defined as

$$excursion_i = \left[\frac{Failed\ \ test\ value_i}{objective_j} \right] - 1$$, when the failing observed value exceeds the objective and

$$excursion_i = \left[\frac{objective_j}{Failed\ \ test\ value_i}\right] - 1$$, when the failing observed value is less than the objective.

Limitations of the WQI Formulation

There are a number of difficulties with the index formulation which can be seen by examination of the definitions above.

1. It concentrates on the violations of the criteria and thus ignores the data below the threshold. The ignored data contain information about trends and not including such information in the index means that the index will have low trend detection power. It is possible that the values are in compliance with the objectives but they are progressing toward non-compliance and such an index should reflect this trend tendency.

2. It is not easy to assess the statistical uncertainty in the index since the three factors F_1, F_2 and F_3 are not independent and indeed they are positively dependent.

3. The distribution of non-compliant values among the variables will have a major impact on the value of F_1. This can be seen when the number of non-compliant values are equal to the number of variables p. In this case the possible values of F_1 will range between 100/p and 100 with the minimum value of F_1 occurring when all the p values occur in a single variable while its maximum occurs when a single non-compliance occurs in each of p variables.

4. The index does not allow for weights to reflect the importance of the variables. For example, would we give a persistent toxic contaminant and turbidity the same weight?

5. The division by the observed value in the second excursion equation above, ie. the observation fails because it is below the limit, is inappropriate. The proportional exceedance should always be relative to the objective. To calculate it relative to the observed value means that the comparison changes with each different observed value and this makes F_3 unstable.

6. The Index as a statistic does not have a measure of uncertainty.

Points 3 and 4 above raise problems with the assumption of the exchangeability of variables inherent in the definition of the index. Also, the same numerical value could be obtained for the index in two quite different situations: (i) one water quality variable being responsible for all the failures and (ii) more than one variable failing at least once during the period, of course, under suitable values for the excursions. This loss of information is problematic for many purposes. For example, if the WQI is found to change over time, the index alone provides no information about the variable or variables responsible for the change.

Further considerations

There seems to be one major oversight in the use of the WQI. Returning to the criteria set out by the WRI report, particularly the requirement that a model be conceived at the outset, the question arises as to just what it is the WQI is measuring. The conceptual model of the British Columbia index (Canadian Council of Ministers, 2001) and hence of the CCME WQI involves the idea that the index measures a conceptual exceedance space. But in addition, it seems important to decide why it is being used. What is the environmental problem being tracked. For example, is eutrophication the problem? Or is it a problem of persistent toxic contaminants in the water, and so on. Once this is done, the variables of importance can be determined and taken as a fixed set. As long as the guidelines used are stated and either used in all regions to be compared or it is understood that quality is considered relative to local guidelines, and, of course held fixed if to be followed over time, then some of the problems of comparison are eliminated. At least one of the reasons for the qualifiers on its use will be removed. The question of how well the index estimates the conceptual exceedance space is another topic. The original ideas of considering the number of failures and how far the failing observations are from the objectives are in keeping with our ideas about comparison with a guideline or standard. Further examination of how the factors F1, F2 and F3 can be used to characterize water quality are worth considering.

Numerous reports are now available on application of the WQI within Canada (eg. CCME National Water Quality Index Workshop, 2004; Khan, Paterson and Khan, 2004; Mercier *et al*, 2005; Gartner Lee Limited, 2006; Kilgour and Associates, 2009). In all of these publications, the authors are reporting on applications of the WQI to existing monitoring data sets, often with the objective of determining how well it will work in a particular region. The 2004 Workshop is particularly informative since the participants, many having been involved with calculating the WQI from monitoring data in a region or across regions, list strengths, weaknesses and suggestions for improvement. Some of these echo the observations that were made above from looking at the formulae, without actually calculating values. In addition, several suggestions support the contention in the previous paragraph that it is important to be more explicit about what is being measured by the index. For example, suggested weaknesses included the lack of a clear definition of water quality or statement of the questions being asked and a need for more direction to be given for those charged with state-of-the-environment reporting .

REFERENCES

Canadian Council of Ministers, 2001. Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index 1.0, Technical Report. 13pp.

Carbonez, A., El-Shaarawi, A. H. and Teugels, J. L. 1999. Maximum microbiological contaminant levels. Environmetrics, 10, 79-86.

CCME National Water Quality Index Workshop. 2004. A path forward for consistent implementation and reporting, Nov 24-25, 2003, Halifax, Nova Scotia, Workshop Proceedings.

CSIN 2010. Canadian Sustainability Indicators Network Conference 2010, http://www.csin-rcid.ca/conference_2010/

Emerson, J., Esty, D., Levy, M.A., Kim, C.H., Mara, V., de Sherbinin, A. and Srebotnjak, T. 2010. Environmental Performance Index. New Haven: Yale Center for Environmental Law and Policy.

Evans, M.S. 1988. Toxic contaminants and ecosystem health: a Great Lakes focus. Wiley series in Advances in environmental science and technology, v. 21, Wiley, New York.

Gartner Lee Limited. 2006. A sensitivity analysis of the Canadian Water Quality Index. PN 1355. Report to Canadian Council of Ministers of the Environment. 36pp.

Government of Canada. 2007. Canadian Environmental Sustainability Indicators. Environment Canada, Statistics Canada and Health Canada. Cat no. 16-251-XIE.

Government of Canada. 2011. The Sustainability Report, Canada's Sustainability Indicators Initiative. http://www.sustreport.org/indicators/nrtee_esdi.html

Hammond, A., Adriaanse, A., Rodenburg, R. Bryant, D. And Woodward, R. 1995. Environmental Indicators: A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development. World Resources Institute. ISBN 1-56973-026-1.

Health Canada. 2010. http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/2010-sum_guide-res_recom/index-eng.php

Khan, A.A., Paterson, R. And Khan, K. 2004. Modification and Application of the CCME WQI for the communication of drinking water quality in Newfoundland and Labrador. Water Qual. Res. J. Canada, 39, 285-293.

Kilgour & Associates Ltd. 2009. Reducing the sensitivity of the water quality index to episodic events. PN1435. Report to Canadian Council of Ministers of the Environment. 23pp.

Mercier, V. and 6 co-authors, 2005. Application and testing of the water quality index in Atlantic Canada, Report Summary. CCME web page. http://www.ccme.ca/assets/pdf/awi_en_posting.pdf

Statistics Canada. 2011. http://www.statcan.gc.ca/nea-cen/gloss/iea-crd-eng.htm

Stevens, R. J. J., and Neilson, M. A. 1987. Response of Lake Ontario to reductions in phosphorus load, 1967-82. Can. J. Fish. Aquati.Sci., 44, 2059-2068.

The OECD Glossary of Statistical Terms. 1997. http://stats.oecd.org/glossary/detail.asp?ID=2907

The Sustainability Report. 2011. Guidelines for Canadian Drinking Water Quality, http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/2010-sum_guide-res_recom/index-eng.php#a2

U.S. Geological Survey. 2001. http://ga.water.usgs.gov/edu/waterquality.html

Walpole, M., and 24 co-authors. 2009. Tracking progress toward the 2010 biodiversity target and beyond. Science, 325, 1503-1504.

RÉSUMÉ

It is now widely accepted that human health and well-being are highly dependent on the quality of the environment, yet the health of the environment plays a much less prominent role in public policy than do economic considerations. The development of environmental indicators and aggregated indices is seen as a way to facilitate the integration of environmental information with social and economic information. Indicators constructed for public policy purposes are expected to provide information in a form readily understood by decision makers and the public. An indicator should also be quantitative and based upon an underlying model which relates the indicator to the more complex phenomenon being summarized. The Canadian Water Quality Index (WQI) provides an example of the development of an index where 1) it was recognized that simpler summaries were needed for the public and decision makers, 2) monitoring data were already being collected, and 3) the scientific understanding of water quality was used to guide the development of the index. During the development of an index and subsequent evaluation, statistical thinking and methodology can contribute in various ways. Since the WQI is an aggregated index, investigation of the value of the index, generated by various scenarios of underlying water quality variable values, needed to be assessed to ensure that low or high water quality according to the index agrees with what a more multivariate assessment would have given. A measure of variability of the index is also necessary. A number of issues, including these two, will be considered for the WQI using data from several Canadian provinces. The role and form of the WQI in the context of a suite of sustainability indicators will also be considered.

Keywords: water quality index; sustainability; statistical consideration; public policy.