DEVELOPMENT OF WATER QUALITY INDEX

In this era of technological development, man has collected vast quantities of data and information about himself, his society, and the physical world around him. This large body of data has grown so rapidly that it challenges man's ability to understand and assimilate it. The same technology which made it possible to create this large data base also has produced the automatic computers which makes the task of storing, analyzing and processing the data more reliable and efficient. The computer, however, is just a tool, a slave to the programmers will, and there still remains the task of extracting from the data the pertinent information required to answer questions of importance. Not only must the data be manipulated and reformulated in a way that is understandable to the user, but exactly the right information must be extracted that is relevant to the questions that are being asked.

In the environmental field, an interested member of the public, a representative of a citizens group, or a governmental official typically may seek to determine whether a particular environmental problem is becoming better or worse. The questioners usually will seek answers in the simplest form. The environmental scientists or professional working in the field may feel, on the other hand, that the answer to the question is complex, requiring the interpretation of hundreds of thousand of measurements of different pollutant concentrations and other variables, some times compounded by missing data, inconsistencies, and quality control problems and often giving vague or uncertain results. Unfortunately, however, the questioner usually will not be satisfied by a 500-page telephone book full of raw data, time series plots, statistical analyses of pollutant concentrations at different locations, and other complex findings. He wants a simpler answer.

The questioner could, of course, hire a consultant already familiar with the data to go through the book of numbers to determine a simple answer to the question. This sometimes happens. Another common but unfortunate is for the questioner to be told that the problem is "too complex", that his question can not be answered unless he is willing to learn more about the technical details of the problems. Usually, the fault does not lie with the person asking the question but with those in the technical and scientific communities who may be unwilling or unable to take the trouble to express the answer in terms that the lay man will understand. One reason, of course, is that technical specialists often do not feel comfortable with simple answers to complex questions, they see many nuances of the questions and possible areas for misunderstanding. They prefer to give no answer rather than an imperfect answer that could lead to misunderstanding. Yet the layman usually prefers an imperfect to no answer at all.

Here is where "indices" can play a potentially important communications role. Ideally, an index or an indicator is a means devised to reduce a large quantity of data down to its simplest form, retaining essential meaning of the questions that are being asked of the data. In short, an index is designed to simplify. In the process of simplification, of course, some information is lost. Hopefully, if the index is designed properly, the lost information will not seriously distort the answer to the question. Unfortunately, however, one may not know in advance which question will be asked. This situation creates the hazards that the index yard stick to judge the effectiveness of regulatory programs in improving environmental quality. From a purely conceptual point of view, environmental monitoring data serve as a feed back loop to evaluate the effectiveness of regulatory activities. Once the environmental monitoring data are collected, there is a further need to translate it into a form that is easily understood. Once the indices are developed and applied, they should serve as a 'tools' to examine trends, to highlight specific environmental conditions, and to help governmental decision-makers in evaluating the effectiveness of regulatory programme.

Environmental indices, of course, are not the only source of information that is brought to bear on environmental decisions. Decision-making will be based on many other considerations besides indices and the monitoring data on which they are based. Ott (1978) identified six basic uses of environmental indices -

i) Resource allocation

Indices may be applied to environmental decisions to assist managers in allocating funds and determining priorities.

ii) Ranking of allocations

Indices may be applied to assist in comparing environmental conditions at different locations or geographical areas.

iii) Enforcement of standards

Indices may be applied to specific locations to determine the extent to which legislative standards and existing criteria are being met or exceeded.

iv) Trend analysis

Indices may be applied to environmental data at different points in time to determine the changes in environmental quality (degradation or improvement) which have occurred over the period.

v) Public information

Indices may be used to inform the public about environmental conditions.

vi) Scientific research

Indices may be applied as a means for reducing a large quantity of data to a form that gives insights to the researchers conducting a study of some environmental phenomenon.

In each of these applications, the index helps in conveying information about the state-of the-environmental phenomenon. Because the questions being asked are different in each application, however the index may differ in terms of the variables included, the basic structure, and the manner inwhich it is applied. Because different users have different data-reporting needs, identification of the users should be critical part of the development and application of any environmental indices.

STRUCTURE OF ENVIRONMENTAL INDICES

The environmental indices can be formulated in two general environmental index forms: (1) those in which the index numbers increase with the degree of pollution (increasing scale indices), and (2) those in which the index numbers decrease with the degree of pollution (decreasing scale indices). Some specialists in the field refer to the former as "environmental pollution indices and the later as "environmental quality" indices. This framework is better suited to representing absolute indices than relative indices.

Mathematical Structure

In this general framework, calculation of an index consists of two fundamental steps:

i) calculation of subindices for the pollutant variables used in the index, and
ii) aggregation of the subindices into the overall index.

If we consider a set of n pollutant variables denoted as (x_1, x_2, x_3 ----- x_i, x_n), then for each pollutant variable x_i, a subindex I_i is computed using subindex function f_i (x_i):

$$I_i = f_i (x_i) \tag{1}$$

In most environmental index, a different mathematical function is used to compute each pollutant variable, giving the subindex functions f_1 (x_1), f_2 (x_2) .... f_n (x_n). Each subindex function is intended to represent the environmental characteristics of the particular pollutant variable. It may consists of simple multiplier, or the pollutant variable raised to a power, or some other functional relationship.

Once the subindices are calculated, they usually are aggregated together in a second mathematical step to form the final index:

$$I = g(I_1 ,I_2, \ldots.. I_n) \tag{2}$$

The aggregation function, Eq. (2), usually consists either of a summation operation, in which individual subindices are added together, or a multiplication operation, in which a product is formed of some or all the subindices, or a maximum operation, in which just the maximum subindex is reported.

The overall process-calculation of subindices and aggregation of subindices to form the index can be illustrated in a flow diagram (Fig. 1). In this process, the information contained in the raw data (environmental measurements) flow from left to right and is reduced to a more parsimonious form. Some information may be lost; however, in a properly designed index, the information loss should be of such a nature that it does not cause the results to be distorted or ultimately misinterpreted.

INFORMATION FLOW

Subindices

Subindices can be classifed as one of four general types:

i) Linear
ii) Non linear
iii) Segmented linear
iv) Segmented nonlinear

Linear function

The simplest subindex function is the linear equation:

$$I = \alpha x + \beta \tag{3}$$

where, I = subindex
x = pollutant variable
α, β = constants

with this function, a direct proportion exists between the subindex and the pollutant variable. The linear indices have the advantages that they are simple to compute and easy to understand. The disadvantage with linear system is that they provides little flexibility. As an example see Fig. 2&3.

Segmented linear function

A segmented linear function consists of two or more straight line segments joined at break points (threshold level). It offers more flexibility. It is especially useful for incorporatiing administratively recommended limits, such as indian standards limits, WHO limits etc. An important segmented linear function is the step function, which exhibits just two states and therefore is called a dichotomous function. Subindices also may consist of a staircase of steps, giving a multiple-state function. For example, Horton's index (1965) uses subindex functions containing three, four, and five steps. In Horton's dissolved oxygen subindex, I=0 for x less than 10% saturation, while I=30 for x between 10% and 30% saturation, and I=100 for x above 70% saturation.

Mathematically, the general form of segmented linear function can be formulated as:

Suppose x and I coordinates of the break points are represented by (a_1, b_2), (a_2, b_2), ... (a_j, b_j) as depicted in Fig. 4. Any segmented linear function with m segments can be presented by the following general equation

$$I = \frac{b_{j+1} - b_j}{a_{j+1} - a_j}(x - a_j) + b_j, \quad a_j \le x \le a_{j+1} \tag{4}$$

where J = 1, 2, 3, ..., m.

Although segmented linear functions are flexible, they are not ideally suited to some situations, particularly those in which the slope changes very gradually with increasing levels of environmental pollution. In these instances, a non linear function usually is more appropriate. Fig. 5 to Fig. 7 are the examples of various segmented linear functions.

Non-linear function

A non-linear function is any relationship which exhibit curvature when plotted on linear paper. The non-linear functions can be further divided in two basic types:

i) an implicit function, which can be plotted on a graph but for which no equation is given such as subindex for pH proposed by Brown et al. (1970) (Fig. 8)

ii) an explicit function, for which a mathematical equation is given (see Fig. 9, 10 & 11)

Implicit functions usually arise when some empirical curve has been obtained from a process under study. For example, Brown et al (1970) proposed an implicit nonlinear subindex function for pH.

In explicit nonlinear functions, curvature is achieved automatically. An important general non-linear function is one in which the pollutant variable is raised to a power other than one, the power subindex function:

$$I = x^c \tag{5}$$

where, c = 1

Walski and Parker (1974) used the following general parabolic form in evolving the subindices for temperature and pH.

$$I = -\frac{b}{a^2}(x-a)^2 + b, 0 \leq x \leq 2a \tag{6}$$

Another common nonlinear function is the exponential function, in which pollutant variable x is the exponent of a constant:

$$I = c^x \tag{7}$$

The constant usually selected is either 10 or e, the base of the natural logarithm. If a and b are constants, the general form of an exponential function is written as follows:

$$I = a\, e^{bx} \tag{8}$$

Segmented Nonlinear Function

Segmented nonlinear functions consist of line segments similar to the segmented linear function; however, at least one segment is nonlinear. Usually, each segment is represented by a different equation which applies over a specific range of the pollutant variable. Segmented nonlinear function being more flexible, has been used in a number of water quality indices. Prati et. al (1971) used a segmented nonlinear function for the pH subindex [Fig. 12] in their water quality index. The pH subindex function contains four segments as given in Table 1:

Table 1 : Parti's Subindex Functions used for p^H

Segment	Limits	Function
1	$0 \leq x \leq 5$	$I = -0.4x^2 + 14$
2	$5 \leq x \leq 7$	$I = -2.0x + 14$
3	$7 \leq x \leq 9$	$I = x^2 - 14x + 49$
4	$9 \leq x \leq 14$	$I = -0.4x^2 + 11.2x - 64.4$

Aggregation of Subindices

The aggregation process is one of the most important steps in calculating any environmental index. In this step most of the simplification (reduction of information) and distortion takes place. In general, four types of aggregation functions are available as described below:

Additive forms

The simplest aggregation functions are the additive forms which can be further divided in to following three forms:

Linear sum

Linear sum is the addition of unweighted subindices, in which no subindex is raised to a power other than 1.

$$I = \sum_{i=1}^{n} I_i \tag{9}$$

where, I_i = Subindex for pollutant variable i
n = number of pollutant variables

In an increasing scale index, the linear sum unfortunately exhibits an ambiguous region; that is, the overall index can report poor enviromental quality when no subindex exhibits poor environmental quality as explained below:

Suppose that a linear sum water pollution index is formed consisting of just two subindices, I_1 and I_2

$$I = I_1 + I_2 \tag{10}$$

In this simple index, we shall assume that I_1 and I_2 are dichotomous subindices in which $I_1=0$ and $I_2=0$ represent zero water pollutant concentrations for pollutant variables x_1 and x_2, and $I_1 \geq 100$ or $I_2 \geq 100$ represent concentration at or above the permissible level. Most users will expect I above 100 to mean unequivocally that permissible level is violated for at least one subindex, and it is unfortunately possible for I to exceed 100 without a permissible limit being violated. For example, if moderate pollution levels occur for both pollutant variables, giving I = 50, and I = 50 then I = 100. Similarly if I = 60 and I = 70 then I = 130. The index conveys the impression that a permissible level has been violated when it has not been, giving an exaggerated an ambiguous reading. This problem is called as ambiguity problem. Graphically, it is described in Fig. 13.

Weighted sum

The weighted linear sum has the following general form:

$$I = \sum_{i=1}^{n} w_i \, I_i \tag{11}$$

where I_i = subindex for i^{th} variable
w_i = weight for i^{th} variable

$$\sum_{i=1}^{n} w_i = 1 \tag{12}$$

The weighted linear sum avoids the ambiguity problem but introduces a more serious problem called 'eclipsing'. Eclipsing occurs when at least one subindex exhibits poor environmental quality, but the overall index does not exhibit poor environmental quality as explained below:

For the two variable case,

$$I = w_1 I_1 + w_2 I_2 \tag{13}$$

$$w_1 + w_2 = 1 \tag{14}$$

Equation 13 and 14 can be written in a single equation as:

$$I = w_1 I_1 + (1 - w_2) I_2 \tag{15}$$

from Eq.(15) it is clear that $I=0$ when both I_1 & I_2 = o i.e. (15) report the zero pollution properly. Further, I will not be 100 until and unless one of the subindex is more or equal to 100. Hence the problem of ambiguity is also removed.

Now putting $I_1=50$ and $I_2=110$ with $w_1=0.5$, gives $I=80$. Because the overall index is less than 100, violation of the permissible level for variable x_2 ($I_2 > 100$) is eclipsed. Graphically, it has been described in Fig. 14.

Root Sum Power

To alleviate the eclipsing problem, a somewhat more complex additive form is available. The root-sum-power is a nonlinear aggregation function of the following form:

$$I = \left[\sum_{i=1}^{n} I_i^p\right]^{1/p} \tag{16}$$

where, p = is a positive real number, greater than 1. As p becomes larger, the ambiguous region becomes increasingly smaller. Thus, for large value of p, the ambiguous region is almost entirely eliminated. For the limiting case in which p approaches infinity, the root-sum-power has desirable properties for aggregating subindices. It possesses neither an eclipsing region nor an ambiguous region. However, because it is a limiting function, it is somewhat unwidely to write and use. As an example refer Fig. 15.

Maximum Operator

The maximum operator can be viewed as the limiting case of the root-sum-power as p approaches infinity. The general form of the maximum operator is as follows:

$$I = \max \{I_1, I_2, \ldots\ldots, I_n\} \qquad (17)$$

In the maximum operator, I takes on the largest of any of the subindices, and I=0 if and only if I=0 for all i. It is ideally suited to determine if a permissible value is violated and by how much.

The limitation of the maximum operator becomes apparent when fine gradations of environmental quality, rather than discrete events, are to be reported and a number of subindices are to be aggregated.

The maximum operator is ideally suited to applications in which an index must report if at least one recommended limit is violated and by how much. Of course, if several subindices violate a recommended limit, the maximum operator will report the worst subindex. The suitability of the maximum operator for use in water pollution indices has not been investigated, however, and none of the published water quality indices have employed this aggregation function.

Multiplicative Forms

The multiplicative forms have found use primarily in indices that have decreasing scales. Most of the water quality indices are based on decreasing scale forms. The water quality index proposed by Brown et. al (1970) originally used an additive aggregation function, the weighted linear sum. Later Landwehr (1974) evaluated multiplicative aggregation functions that could be substituted for the additive form, and the multiplicative form has become the most popular version of this index.

Like increasing scale indices, many decreasing scale indices exhibit both the ambiguity and eclipsing problems. In general, the additive forms do not appear well suited for aggregating decreasing scale subindices.

To avoid such problems, the multiplicative forms have been proposed. The most common multiplicative aggregation function is the weighted product, which has the following general form:

$$I = \prod_{i=1}^{n} I_i^{w_i} \qquad (18)$$

where

$$\sum_{i=1}^{n} w_i = 1 \qquad (19)$$

In this aggregation function, as with all multiplicative forms, the index is zero if any one subindex is zero. This characteristic eliminates the eclipsing problem, because, if any one subindex exhibits poor environmental quality, the overall index will exhibit poor environmental quality. Conversely, I=0 if and only if at least one subindex is zero, and this characteristic eliminates the ambiguity problem.

If the weights in equation (19) are set equal, $w_i=w$ for all i, then Eq. (19) can be written as follows:

$$\sum_{i=1}^{n} w_i = n\,w = 1 \qquad (20)$$

For this situation, w=1/n, and Eq. (18) becomes the geometric mean of subindices:

$$I = \left[\prod_{i=1}^{n} I_i\right]^{w} = \left[\prod_{i=1}^{n} I_i\right]^{1/n} \qquad (21)$$

Thus, the geometric mean is a special case of the weighted product aggregation function. A common version of the weighted product is the geometric aggregation function.

$$I = \left[\prod_{i=1}^{n} I_i^{g_i}\right]^{1/\gamma} \qquad (22)$$

where

$$\gamma = \sum_{i=1}^{n} g_i \qquad (23)$$

Graphically, it has been explained in Fig. 16.

Minimum Operator

The minimum operator, when applied to decreasing scale subindices, performs in a fashion similar to the increasing scale maximum operator. The general form of the minimum operator is as follows:

$$I = \min \left\{I_1, I_2, \ldots, I_n\right\} \qquad (24)$$

Like the weighted product in the minimum operator functions, eclipsing can not occure, and no ambiguous region exists. Consequently, the minimum operator appears to be a good candidate for aggregating decreasing scale subindices. However, none of the publised environmental indices employ the minimum operator, and its potential apparently remains unexplored.

SUMMARY & NEED FOR FURTHER WORK

- WQI frequently uses implicit or segmented function which can not be readily agregated.
- a number of explicit function have been also developed by various researchers but most of them can not be used over the full range of pollutant variation.
- the other problems are - dimensionally inconsistent, indeterminate forms at zero pollutant concentration.
- most of the water quality indices use the weighted linear sum agregation function which has serious eclipsing problem.
- to circumvent this problem the weighted product aggregation function was used which reduces the problem of eclipsing to a certain extent if the number of water quality variables are small (say 2 or 3).
- but in general a WQI has atleast 9 or 10 parameters which has eclipsing problem in both the weighted product and weighted sum forms.

:: :::: ::

Table 1 : Key Water Quality Variables Classified by Sayers and Ott According to Various Water Uses

Public Water Supply	Industrial Water Supply	Agricultural Water Supply	Aquatic Life and Wildlife Maintenance	Recreation And Aesthetics
Coliform Bacteria Turbidity Color Taste-Odor Trace Metals Dissolved Solids Trace Organics Chlorides Fluorides Sulfates Nitrates Cyanides Radioactivity	Processing (Except foods) pH Tubidity Color Hardness Alkalinity/Acidity Dissolved Solids Suspended Solids Trace Metals Trace Organics Cooling pH Temperature Silica Aluminium Iron Manganese Hardness Alkalinity/Acidity Sulfates Dissolved Solids Suspended Solids Sanitary (same as for public supply)	Farmstead (same as for public supply) Livestock (similar to that for public supply) Irrigation Dissolved Solids Specific Conductance Sodium Calcium Magnesium Potassium Boron Chlorides Trace Metals	Temperature DO pH Alkalinity/Acidity Dissolved Solids Salinity Carbon Dioxide Turbidity Color Settleable Materials Floating Materials Tainting Substances Toxic Substances Nutrients	Recreation Coliforms Turbidity Color pH Odor Floating Materials Settleable Materials Nutrients Temperature Aesthetics Tubidity Color Odor Floating Materials Settleable Materials Nutrients Temperature Substances Adversely Affecting Wildlife

Table 2 : Mathematical Characteristics of General and Specific-Use Water Quality Indices Published in the Literature

Index	Subindices	Aggregation Function	Comments
General Water Quality indices			
Horton[5]	Segmented Linear (Step Functions)	Weighted Sum Multiplied by Two Dichotomous Terms	Eclipsing Region
Brown et al. (NSF WQI_a)	Implicit Nonlinear[b]	Weighted Sum	Eclipsing Region
Landwehr[6] (NSF WQI_m)	Implicit Nonlinear[b]	Weighted Product	Nonlinear
Prati et al.[14]	Segmented Nonlinear[b]	Weighted Sum (Arithmetic Mean)	Eclipsing Region
McDuffie and Haney[15]	Linear	Weighted Sum	Eclipsing Region
Dinius[16](1972)	Nonlinear(Linear, Power)	Weighted Sum	Eclipsing Region
Dee et al.[27]	Implicit Nonlinear[b]	Weighted Sum	Eclipsing Region
Dinius (1987)	Non-linear (power)	Geometric Mean	Eclipsing Region
Specific-Use Water Quality Indices			
O'Connor[18](FAWL, PWS)	Implicit Nonlinear[b]	Weighted Sum	Eclipsing Region
Deininger and Landwehr[19](PWS)	Implicit Nonlinear[b]	Weighted Sum/Weighted Product	Eclipsing/Nonlinear
Walski and Parker[17]	Nonlinear	Weighted Product (Geometric Mean)	Nonlinear
Stoner[20]	Nonlinear	Weighted Sum	Negative Values
Nemerow and Sumitomo[21]	Segmented Linear	Root-Mean-Square of Maximum and Arithmetic Mean	Minimizes Eclipsing
Bhargava (1985)	Nonlinear (Exponential)	Geometric mean	Eclipsing Region

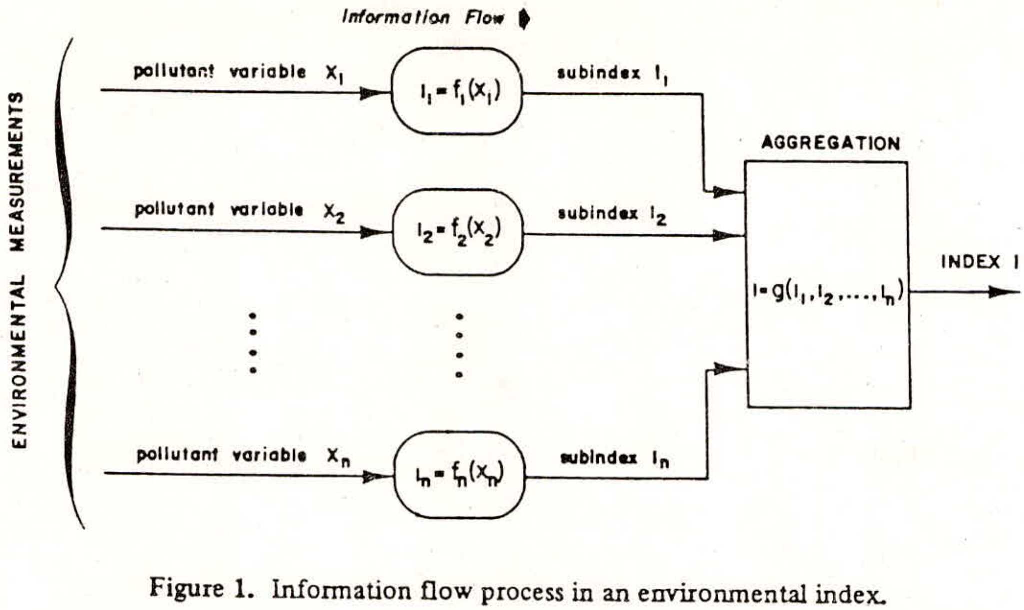

Figure 1. Information flow process in an environmental index.

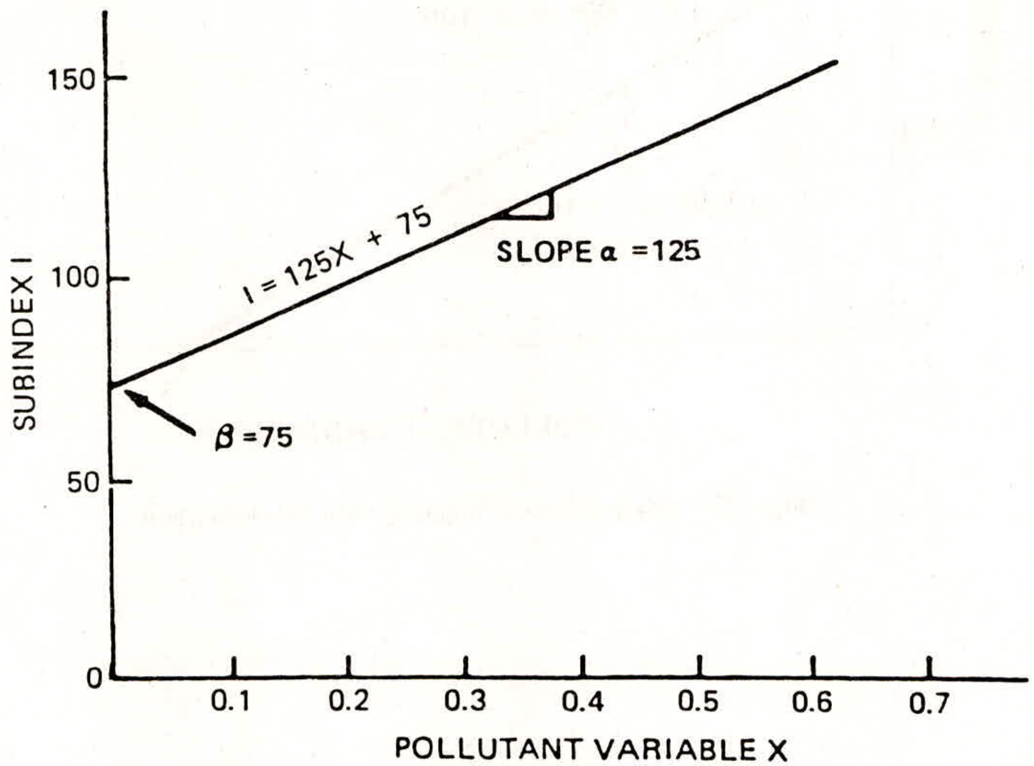

Figure 2. Simple linear (increasing scale) subindex function which does not pass through the origin.

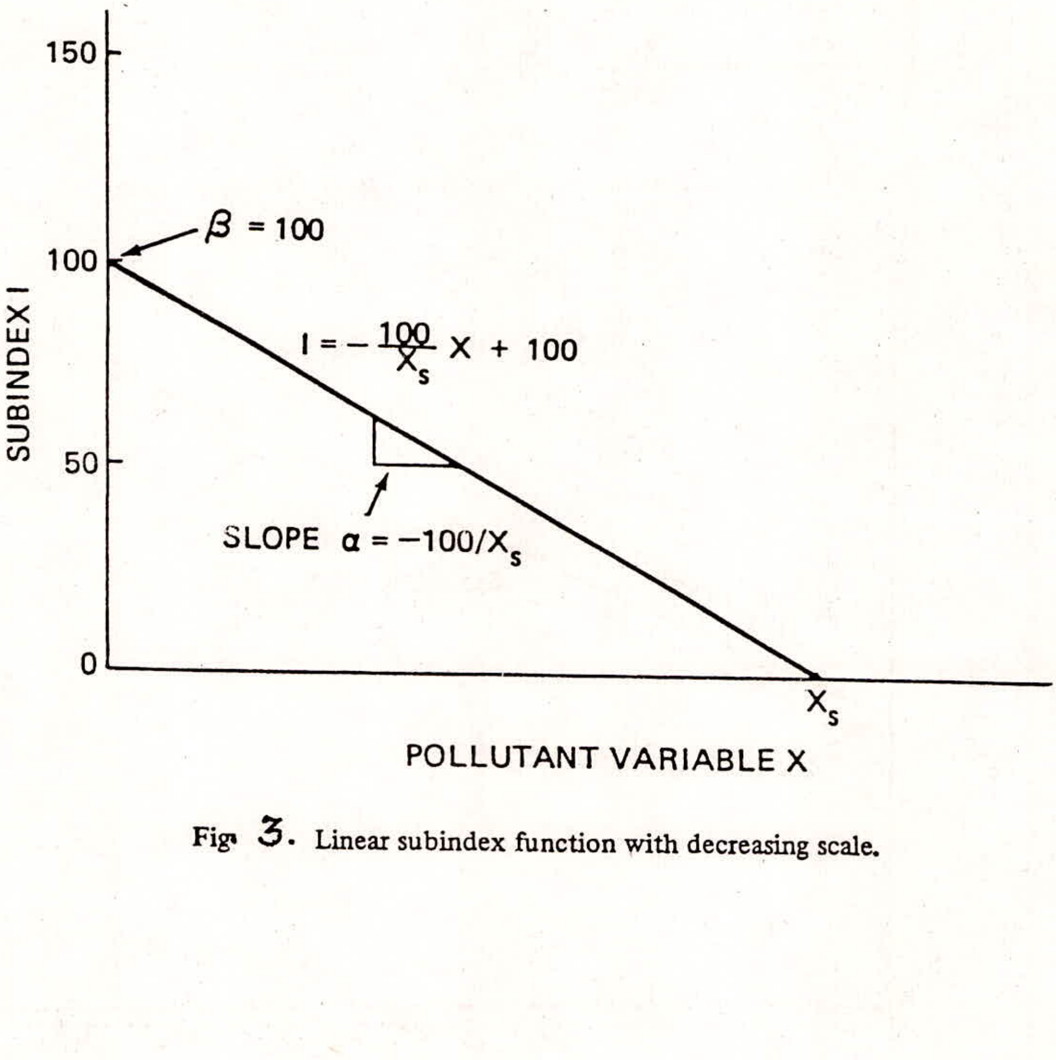

Fig. 3. Linear subindex function with decreasing scale.

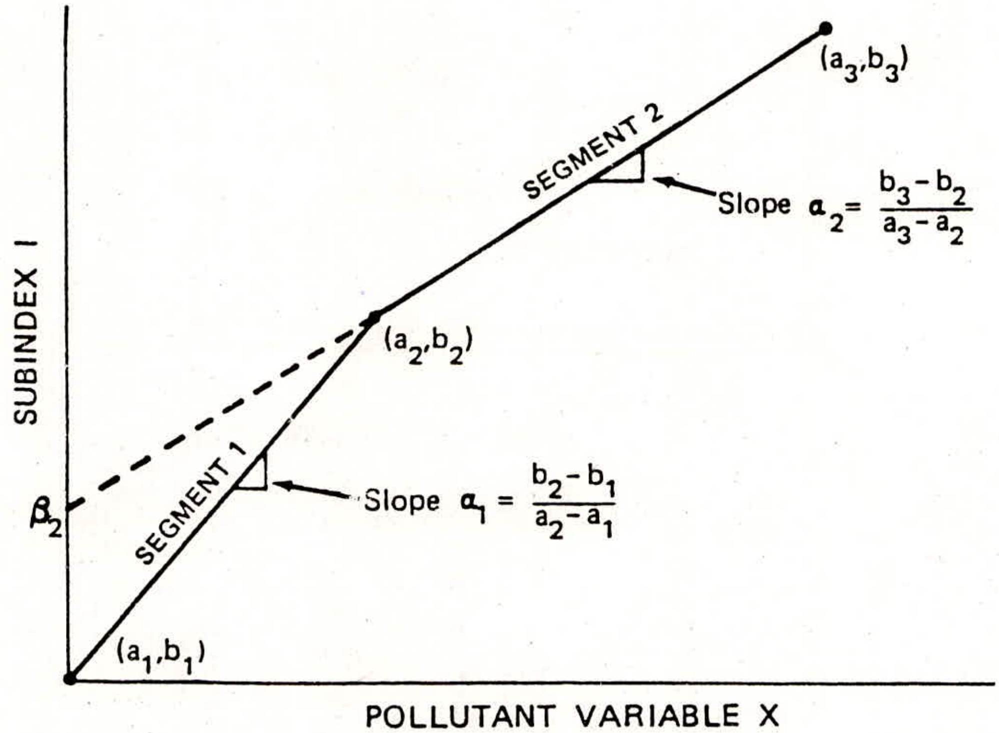

Figure 4. General form of segmented linear function.

$$I = \frac{b_{j+1} - b_j}{a_{j+1} - a_j}(X - a_j) + b_j$$

$$\text{for } a_j \leqslant X \leqslant a_{j+1}$$

where $j = 1, 2, 3, \ldots, m.$

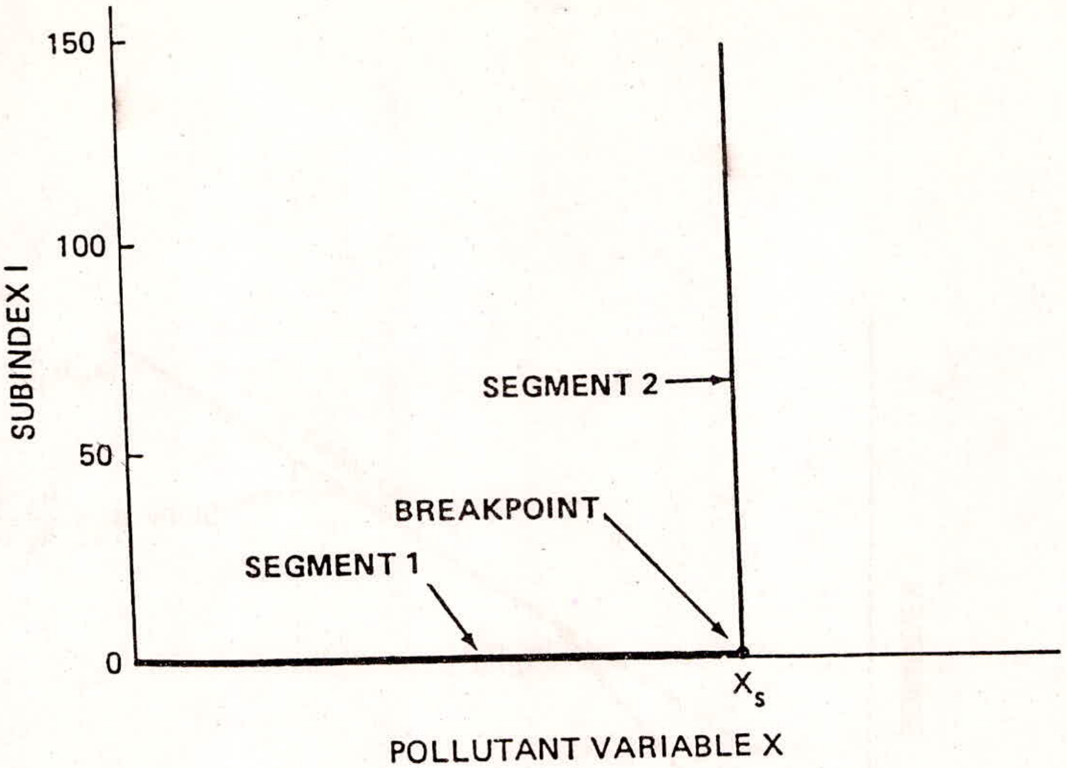

Figure 5. Example of a segmented linear (hockey stick) function.

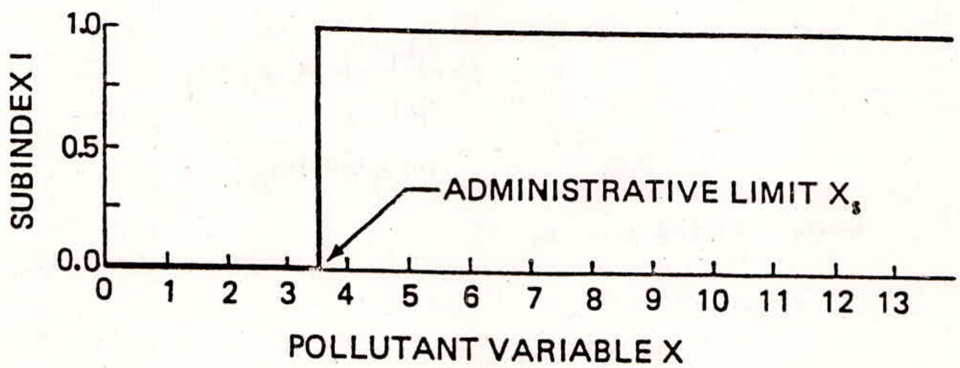

Figure 6. Example of a dichotomous step function.

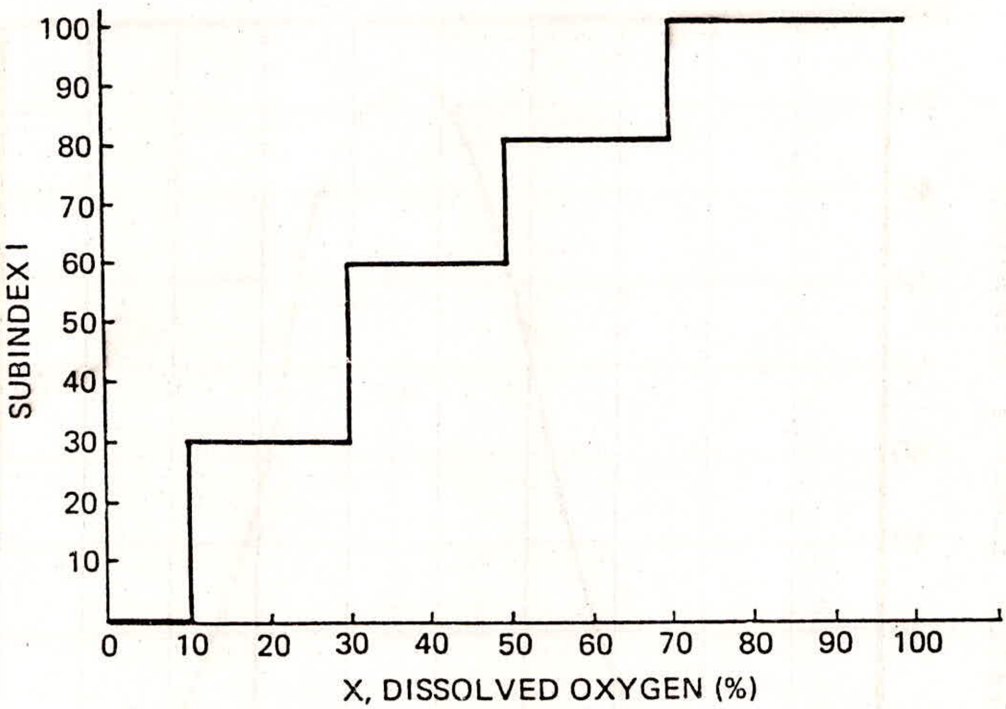

Figure 7. Staircase step function for dissolved oxygen from a water quality index proposed by Horton.[2]

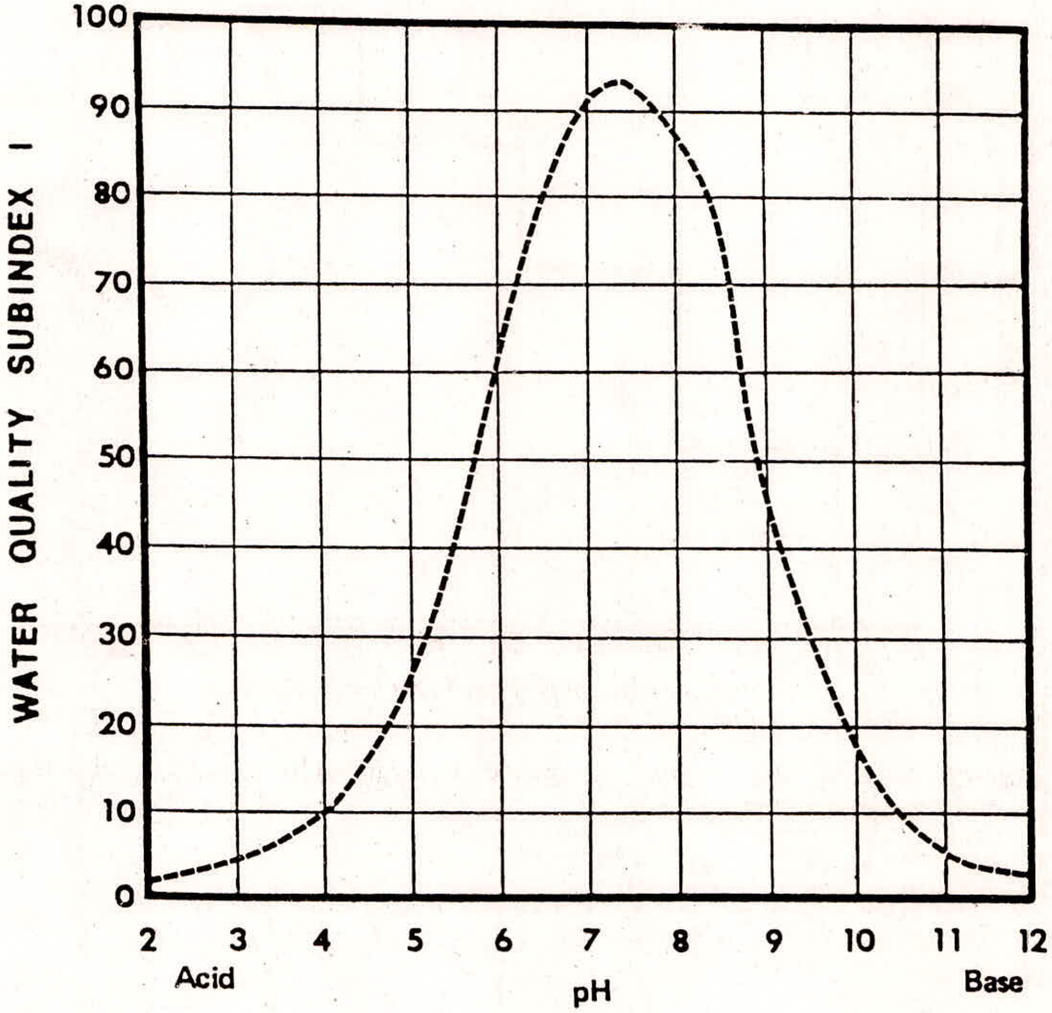

Fig. 8 : Implicit non-linear subindex function for pH proposed by Brown et al. (1970)

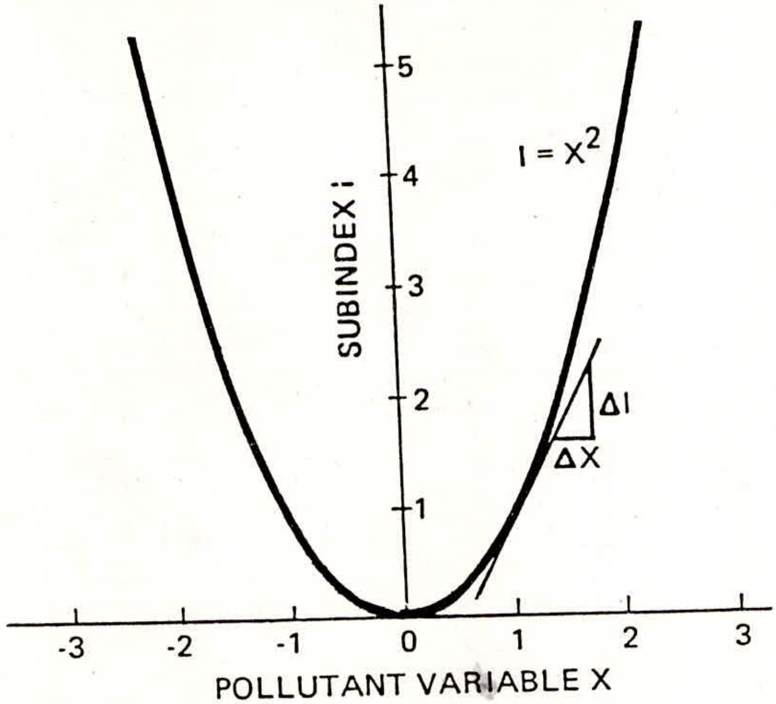

Figure 9: Example of explicit nonlinear subindex function, the parabola $I = X^2$.

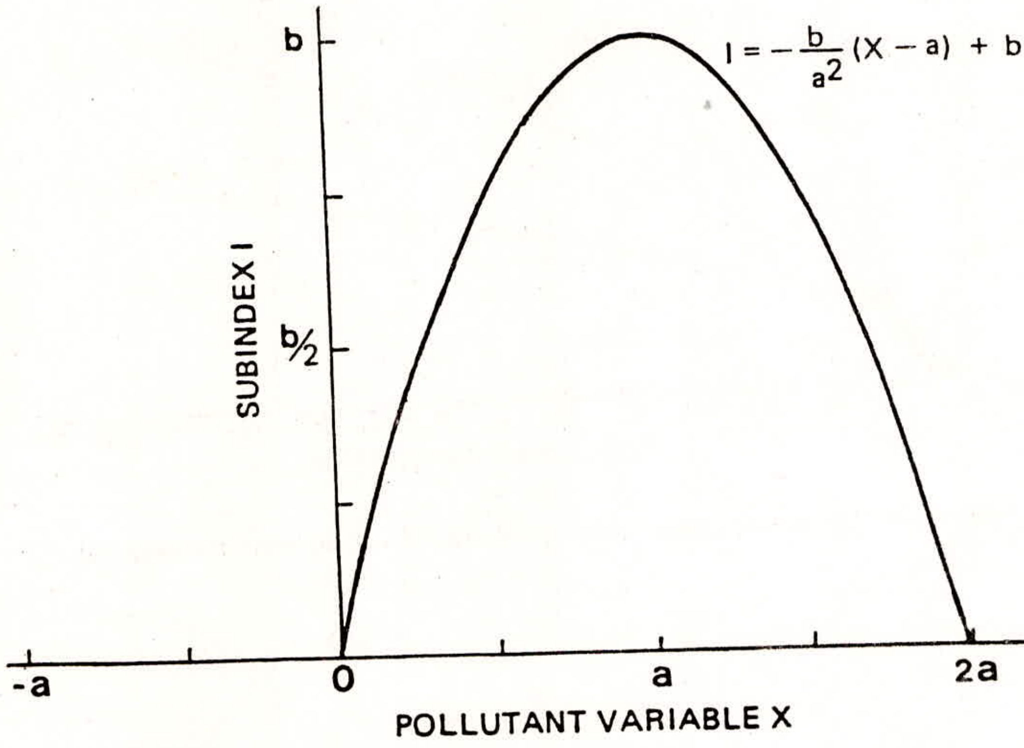

Fig. 10 : Example of a parabolic subindex function which was translated from the origin and inverted, based on a water quality index by Walski and Parker for Temperature and pH subindices

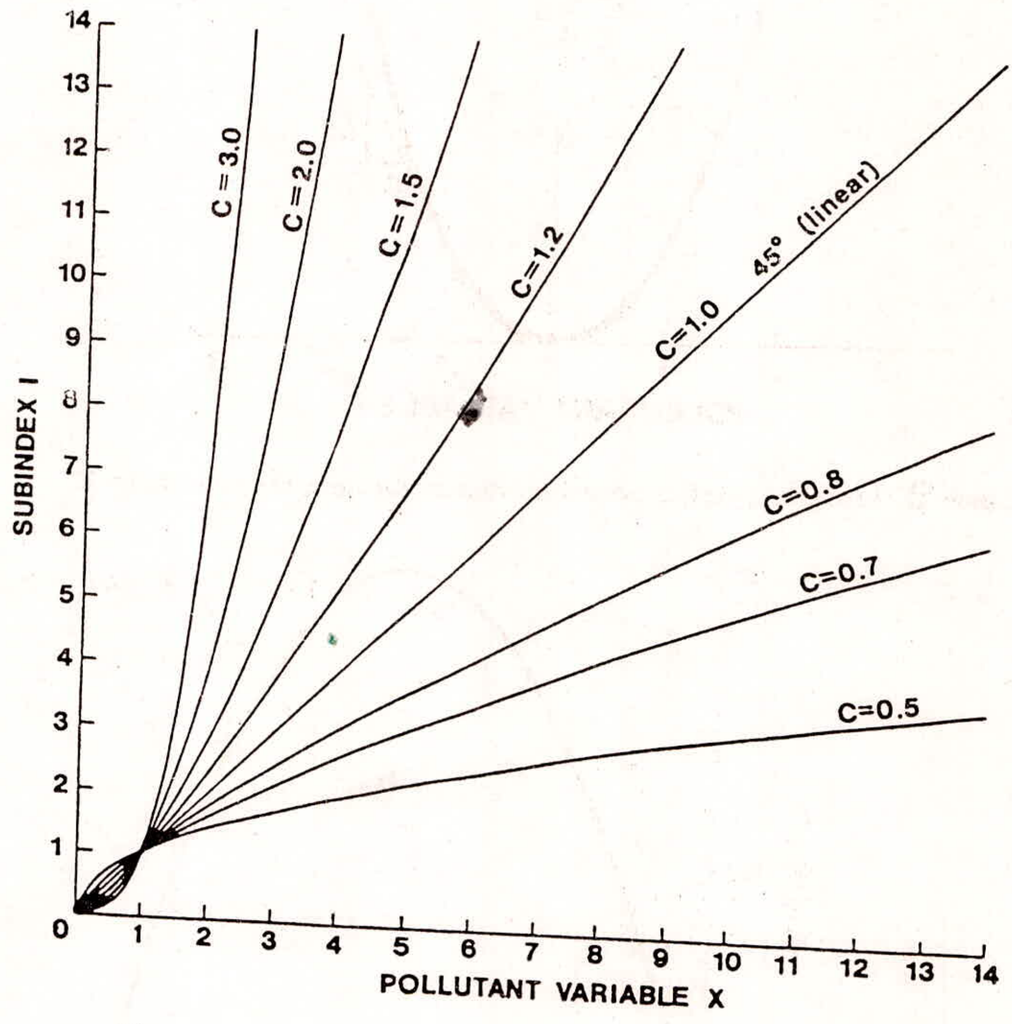

Figure 11. Plot of the power subindex function $I = X^c$ for selected values of c.

Segment 1 (AB)	$0 \leqslant X \leqslant 5$	$I = -0.4X^2 + 14$
Segment 2 (BC)	$5 \leqslant X \leqslant 7$	$I = -2X + 14$
Segment 3 (CD)	$7 \leqslant X \leqslant 9$	$I = X^2 - 14X + 49$
Segment 4 (DE)	$9 \leqslant X \leqslant 14$	$I = -0.4X^2 + 11.2X - 64.4$

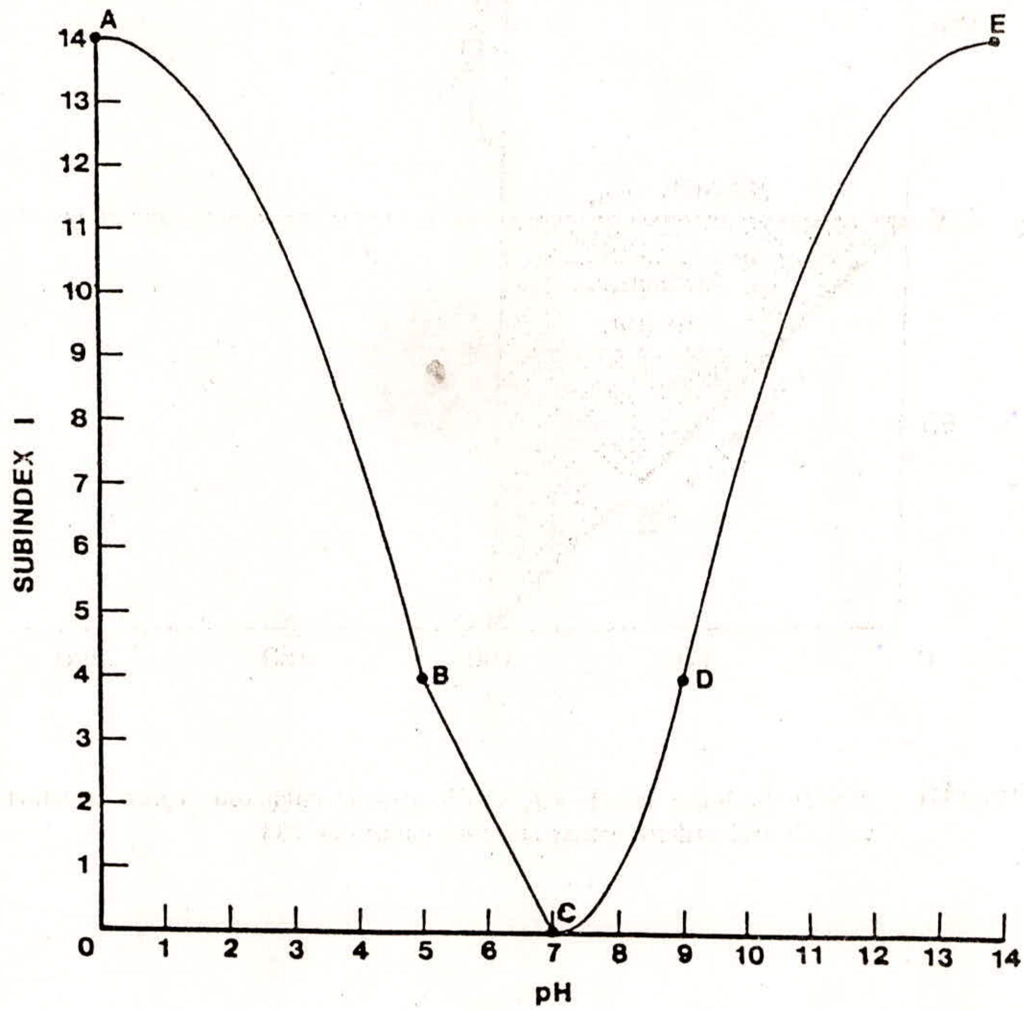

Figure 12. Example of a segmented nonlinear function for pH, from the water quality index of Prati, Paranello and Pesarin.[6]

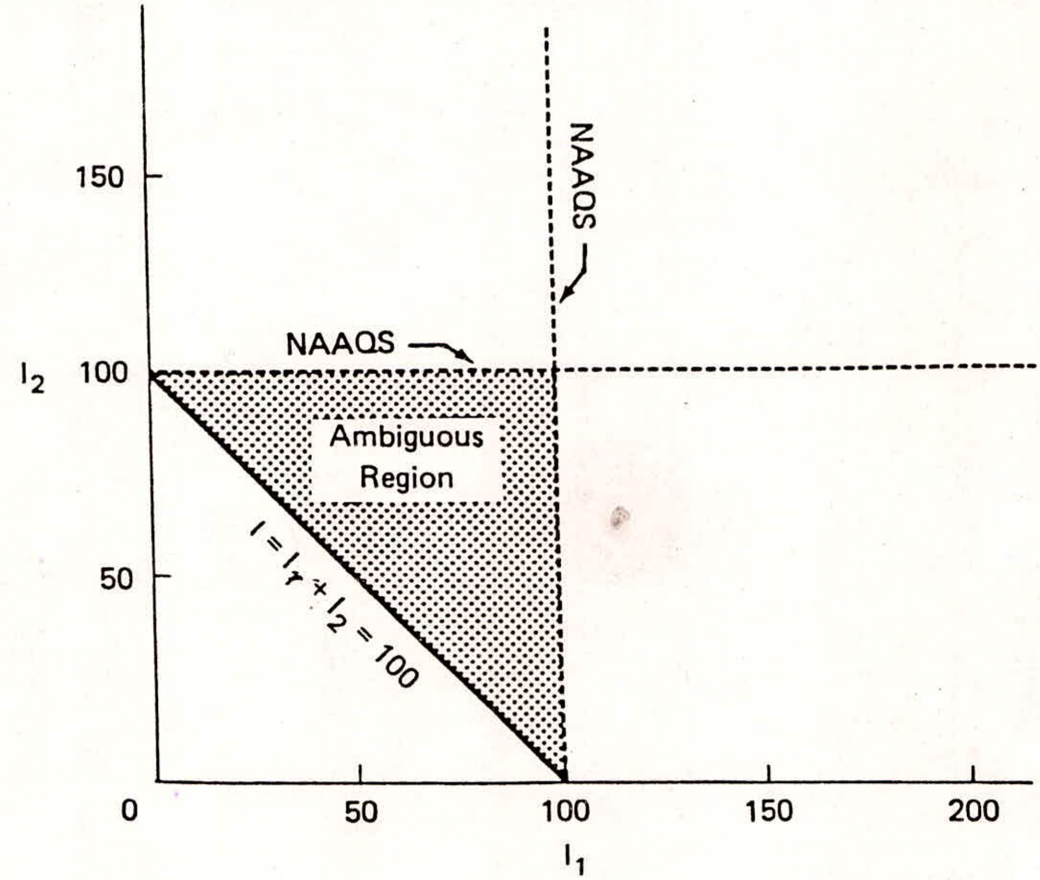

Figure 13. Plot of the linear sum $I_1 + I_2 = 100$ showing ambiguous region for which I exceeds 100 without either subindex exceeding 100.

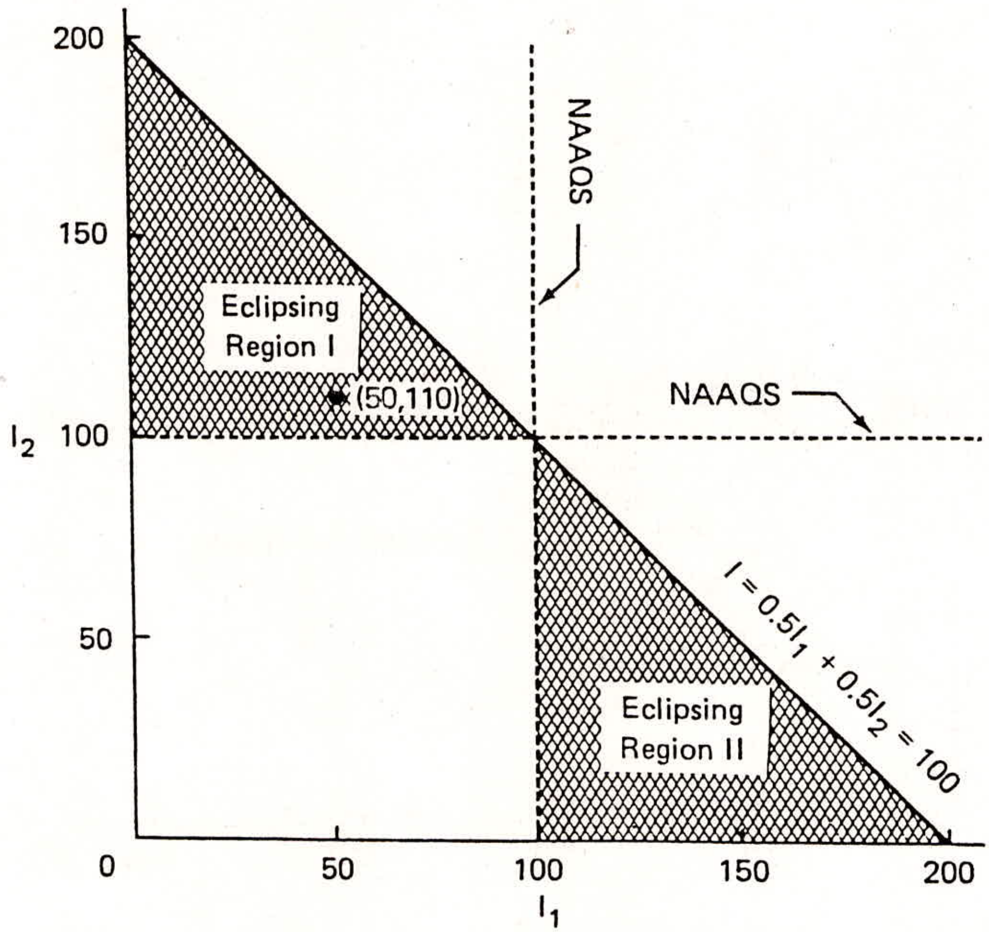

Figure 14 Plot of the weighted linear sum $0.5I_1 + 0.5I_2 = 100$ showing eclipsing regions for which a subindex exceeds 100 without the index exceeding 100.

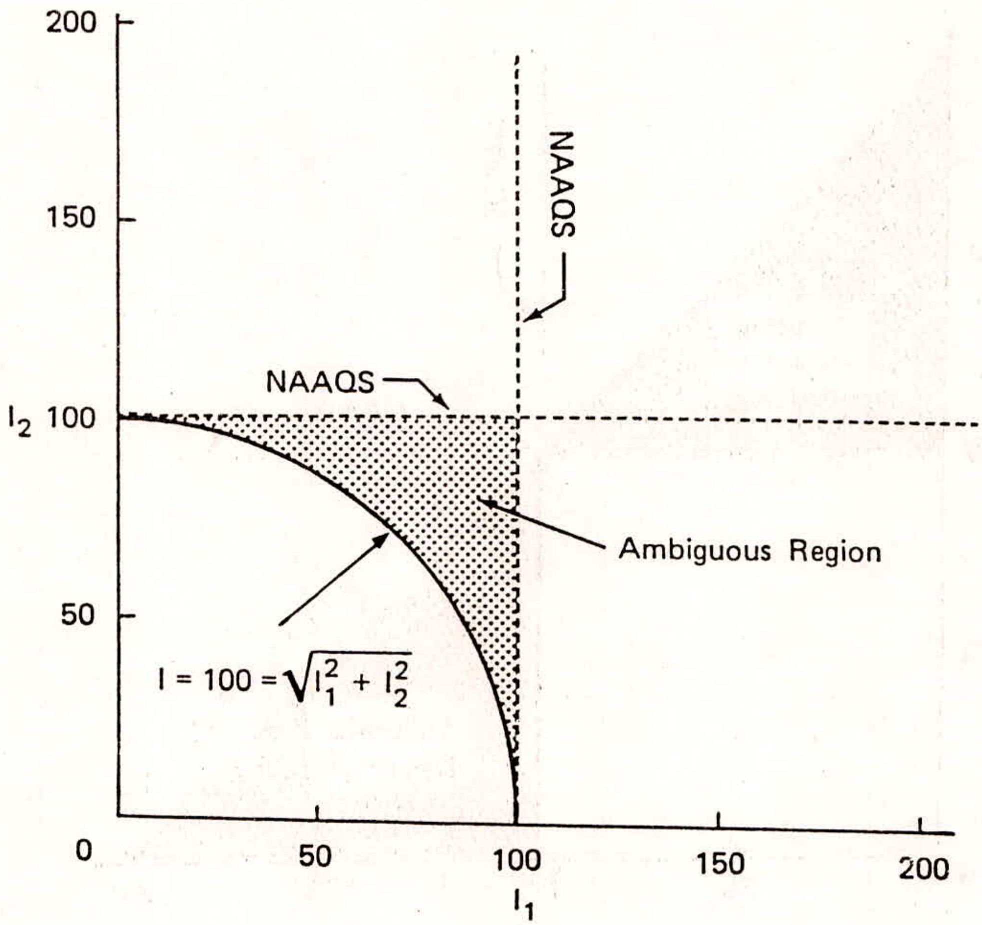

Figure 15. Plot of the root-sum-square aggregation function in the (I_1, I_2)-plane.

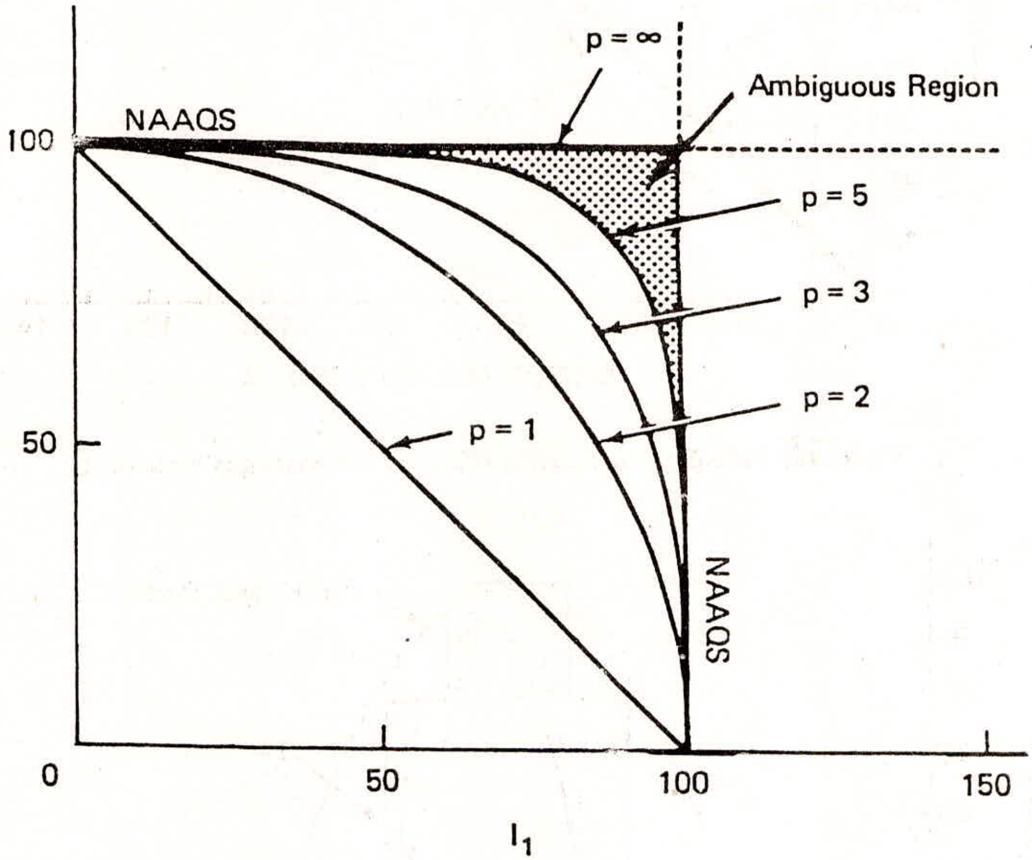

Figure 16 Plot of $I = (I_1^p + I_2^p)^{1/p}$ for selected values of p.

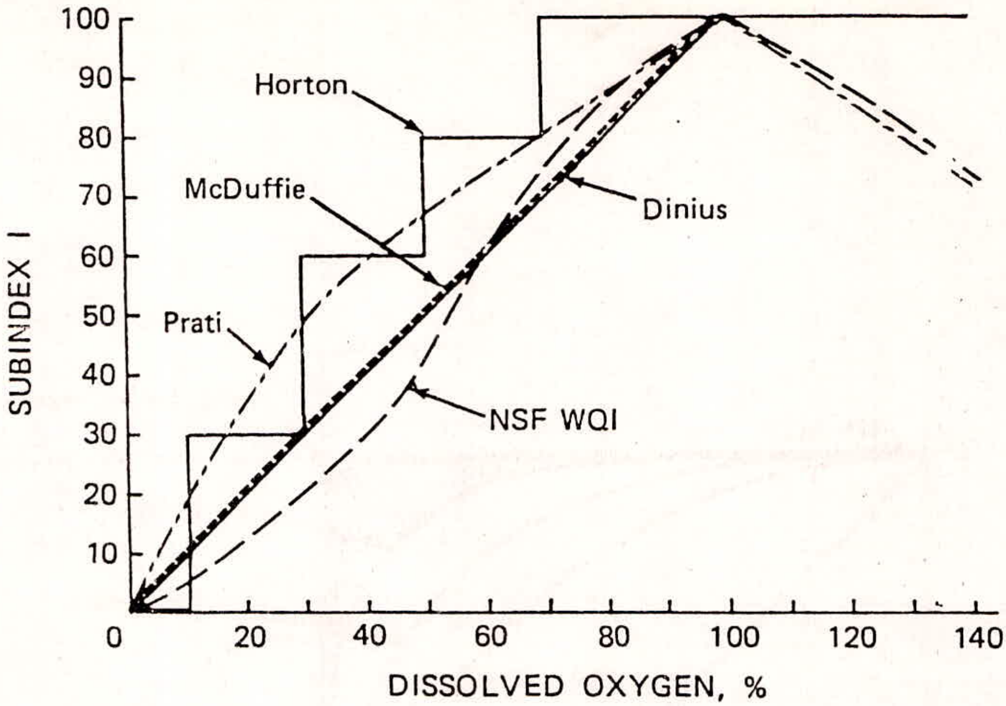

Figure 17. Subindex functions for DO from five water quality indices.

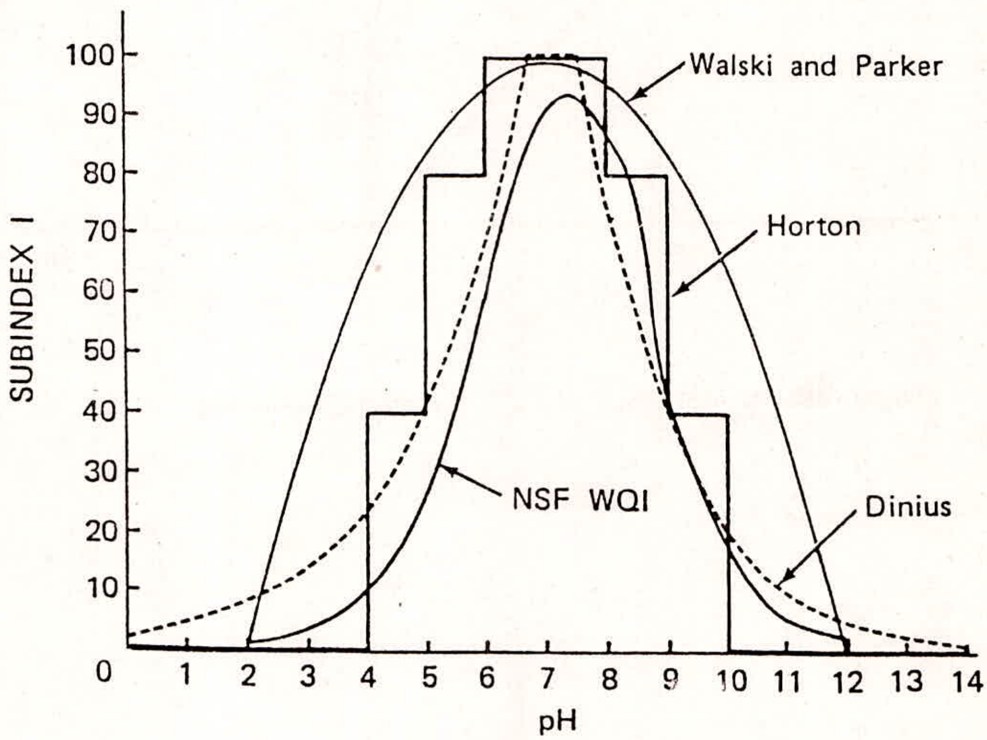

Figure 18. Subindex functions for pH from four water quality indices.

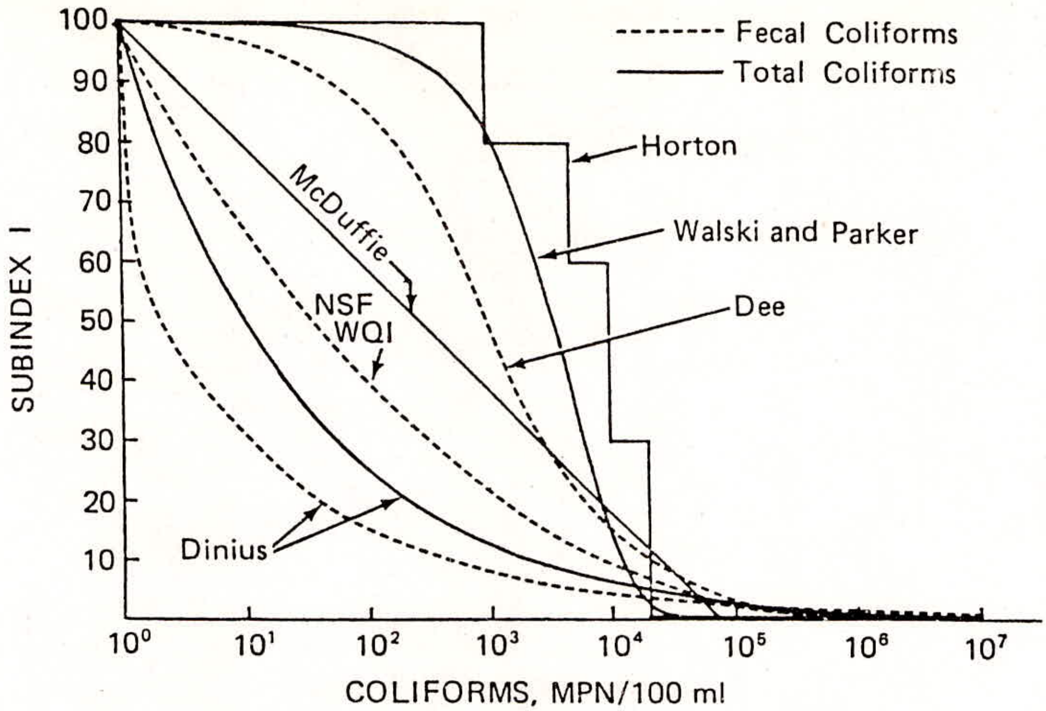

Figure 19. Subindex functions for coliform organisms from six water quality indices.

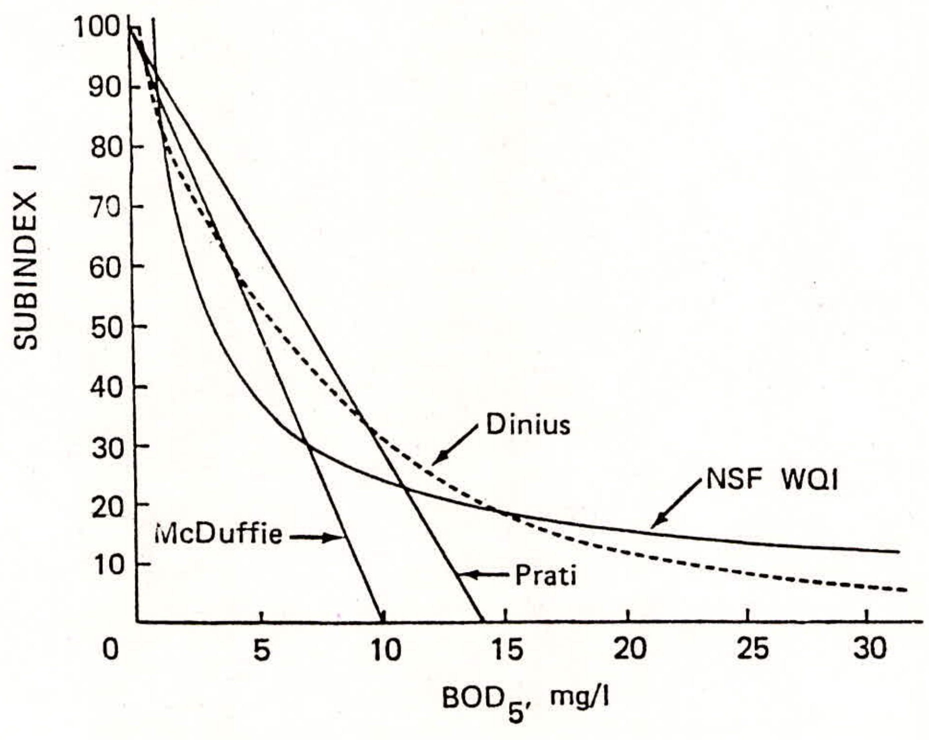

Figure 20. Subindex functions for BOD_5 from four water quality indices.