Water Res. Vol. 16, pp. 945 to 948, 1982
Printed in Great Britain.

0043-1354/82/060945-04$03.00/0

A WATER QUALITY INDEX FOR LAKE BEACHES

NORMAND ST-LOUIS[1] and PIERRE LEGENDRE[2]

[1]Eco-Recherches Inc. (C.I.L.), 121 Hymus Blvd, Pointe-Claire, Québec, H9R 1E6 and
[2]Département de Sciences biologiques, Université de Montréal C.P. 6128, Succursale A, Montréal, Québec, H3C 3J7, Canada

(*Received September* 1981)

Abstract—Discriminant analysis of ten years of data from seven beaches located on three lakes in Gatineau Park (Québec) leads to a canonical vector accounting for most of the variability of the three microbial count variables recorded. This discriminant axis orders the beaches along a cline of increasing pollution. It is then transformed into a microbial water quality index calibrated for the given set of beaches, which is in turn used to study the medium-term variability and the long-term evolution of water quality of the beaches.

INTRODUCTION

There is a need for indices of potential ecological risk, to be used as diagnostic tools for water pollution monitoring. Håkanson (1980), for instance, has developed a rapid indicator system based on the dosage of chemical elements in a few sediment samples. Zamfir (1980) shows that bacteria are important indicators of surface water pollution, together with chemical, animal and vegetal indicators; Starzeka *et al.* (1979) use a series of 10 physical, chemical and bacterial indices, which are averaged to form a 4-class water pollution index. Tebbutt (1977) reports the details of the World Health Organization standards for drinking water, in terms of permissible levels of its physical and chemical properties, and of the number of total and fecal coliforms. In the same vein of thought, Mahloch (1974) used physical, chemical and bacterial variables together in various types of multivariate factor analyses, which are successful in reducing the original set of variables to a small number of synthetic factors explaining the major effects on water quality.

In long-term monitoring of a group of beaches used for water sports, the variability of bacterial descriptors is likely to be a much more important indicator of water quality than variations in other descriptors. This is why the government norms are formulated in terms of total and fecal coliforms, and of streptococci. This note describes a water quality index based upon these gross bacterial counts. Other such indices are designed to take into account only the quantitative aspect of the question, pooling all bacterial types; this is the case for instance with the Air Microbic Index of Pitzurra *et al.* (1980). Instead, the index we seek is a linear combination of the three bacterial variables, thus making it possible to weight them differentially.

The index described hereafter results from the comparison of a group of beaches of the same area and subjected to similar uses. It is thus calibrated for those beaches and makes it possible to follow their medium- (weeks) and long-term (years) evolution.

Key words: bacteria, coliform, discriminant analysis, index, pollution, water quality.

DATA BASE

Bacteriological data were gathered from 1971 to 1980 on 7 beaches located on lakes Philippe, Meach and Lapêche in Gatineau Park (Québec) by Environment-Canada and the National Capital Commission, Government of Canada.

5651 samples were obtained from 42 sampling stations on these seven beaches, and analyzed for bacterial concentration. The results were provided in the form of three variables, as is usual in such monitoring programs: total coliforms, fecal coliforms, and streptococci.

The data, as well as the detailed results, are fully presented in the report to the sponsor of this study (St-Louis, 1981).

METHODS

Fecal coliforms were included in the total coliform counts. We left them in this form, instead of reworking them into "fecal" and "non-fecal" groups, in order to make the results easier to use with data in the same form, taken in the future from the same beaches. On the other hand, since there is a difference of about an order of magnitude between these two variables, the discriminant function coefficients (below) should be little affected.

The method explained hereafter involves a discriminant analysis, which produces better results when the variables satisfy the condition of normality. Kolmogorov–Smirnov goodness of fit tests showed that the three variables were far from normal; various simple normalizing transformations were tested, and a log (base 10) transformation was retained as the one which brought the data closest to normal. More complicated transformations were not needed, since discriminant analysis is known to be sufficiently robust to permit a certain amount of departure from normality. Furthermore, the additional complexity could have made the water quality index more difficult to use for field biologists.

Tabulation of each of the three bacterial count variables by beach and by sampling station showed that the

Table 1. The use of the unstandardized canonical discriminant function coefficients in the computation of a sample's discriminant score on the first canonical axis; the example shown uses as data the government norms for minimum acceptable water quality. The figures at the right are the product of a log-transformed number of bacteria by a discriminant coefficient. Adding a constant to the sum of these products, one obtains the score of this sample on the discriminant axis

Variables	Minimum water quality (government norms)	Log_{10} transformation	Unstandardized discriminant coefficients
Total coliforms	1000	3.00000	x −0.16471 = −0.49413
Fecal coliforms	200	2.30103	x −0.21449 = −0.49355
Streptococci	100	2.00000	x −1.18271 = −2.36542
(Constant)			1.42791 = 1.42791
			Discriminant score = −1.92519

between-beach variability was the most important. It was therefore hypothesized that the study of the between-beach variance, through a discriminant analysis model, could give a basis from which the evolution of water quality in time could be studied.

The primary aim of discriminant analysis is to produce a weighting of variables that best distinguishes between predefined groups. Here, the groups are the various beaches and all the samples from the same beach will be considered as forming a group; the log-transformed bacterial counts are the discriminating variables. Since some variables may perform this task better than others, the problem solved by discriminant analysis is to assemble these variables into linear (that is: first degree) functions, giving to each an appropriate weight. A single function can differentiate two groups, but two or more may be needed to separate three or more groups, inasmuch as the variables allow. Each function corresponds to a unique weighting of all the variables, different from that of the other functions. Furthermore, each such function gives rise to an axis, called discriminant axis, on which it is possible to plot the position of the samples forming the groups. Finally, the importance of each of these axes can be measured by the amount of between-group variability that it takes into account (the so-called eigenvalues). These properties (Nie *et al.*, 1975; Legendre & Legendre, 1979) will be used in the following results and discussion.

The first axis obtained from such factor analyses usually acts as a general size factor (Blackith & Reyment, 1971); so it was hoped that this axis would order the beaches along a general water quality gradient, and would also account for a large fraction of the variability exhibited by the three bacterial count variables. These attributes would enable us to use the first discriminant function as a basis for the development of a water quality index, valid for the 7 beaches during the decade of study. The index is described in the Results section.

The index was then used to examine the time variations of water quality on the various beaches, graphically for weekly fluctuations and by linear regression for the long-term variability.

All computations were performed with the aid of the widely available SPSS software package (Nie *et al.*, 1975) on the CDC CYBER 171 and 173 of the Multiple Access Computer Group (Montreal) and the Université de Montréal, respectively.

RESULTS

The first discriminant axis obtained from this analysis accounts for 83% of the total between-group variance. The standardized canonical discriminant function coefficients give the following weights to the three variables in the formation of this first axis:

0.12618 to log-transformed total coliforms
0.14697 to log-transformed fecal coliforms
0.81478 to log-transformed streptococci.

Since these coefficients are computed from variables which are standardized (expressed in terms of their respective standard deviations), they indicate that the streptococci have a predominant weight in the formation of the first discriminant axis, about 6 times as much as the total coliforms or the entric coliforms. On the other hand, the three log-transformed variables have very similar standard deviations (0.78, 0.70 and 0.71 respectively); so, these figures are, with relation to one another, not very different from the actual unstandardized canonical discriminant function coefficients (computed from the log-transformed but unstandardized variables), also known as identification functions, presented in Table 1: all three standardized-to-unstandardized-coefficient ratios are about equal to 0.70. These make it possible to compute each sample's discriminant score from the raw, log-transformed data. The constant is computed so as to center the mean of the sample's distribution on the origin (point zero) of the axis. The negative weights merely make it possible to represent the worst quality samples in the negative part of the axis and the best in the positive part.

Before presenting the ordination of the beaches thus realized, the discriminant axis will be transformed into a water quality index I, with a 0–1 scale of variation, which is easier to use:

$$I = \frac{\text{(discriminant score)} - \text{min}}{\text{max} - \text{min}}.$$

Min is the minimum value obtained for a sample (−4.478) and max is the maximum one (1.428 in the present study). Consequently, the government norms

Table 2. Comparison of the average I (second column) obtained from the mean value of the three bacterial counts, log transformed (three "bacterial" columns), for each beach, along 10 years of sampling. In parenthesis: $\log_{10}^{-1}$ (mean log count), that is, the bacterial count corresponding to the mean of the log counts. The government norms are added for reference

Beaches	Mean index value	Total coliforms	Fecal coliforms	Streptococci
La Pêche	0.8222	1.46 (29.0)	0.51 (3.2)	0.59 (3.9)
Smith	0.7932	1.54 (34.4)	0.63 (4.3)	0.70 (5.1)
Raby	0.7908	1.44 (27.4)	0.62 (4.1)	0.73 (5.4)
Breton	0.7658	1.72 (52.4)	0.77 (5.9)	0.79 (6.2)
Parent	0.7645	1.66 (46.1)	0.71 (5.2)	0.81 (6.5)
Blanchette	0.7531	1.61 (41.2)	0.76 (5.7)	0.87 (7.4)
O'Brien	0.6745	1.92 (83.4)	0.98 (9.5)	1.18 (15.1)
Government norms	0.432	3.00 (1000)	2.30 (200)	2.00 (100)

(Table 1) correspond to an index value of

$$I = \frac{-1.925 - (-4.478)}{1.428 - (-4.478)} = 0.432$$

for the beaches studied in the present note. The position of the beaches' mean point on the scale of this water quality index is shown in the second column of Table 2.

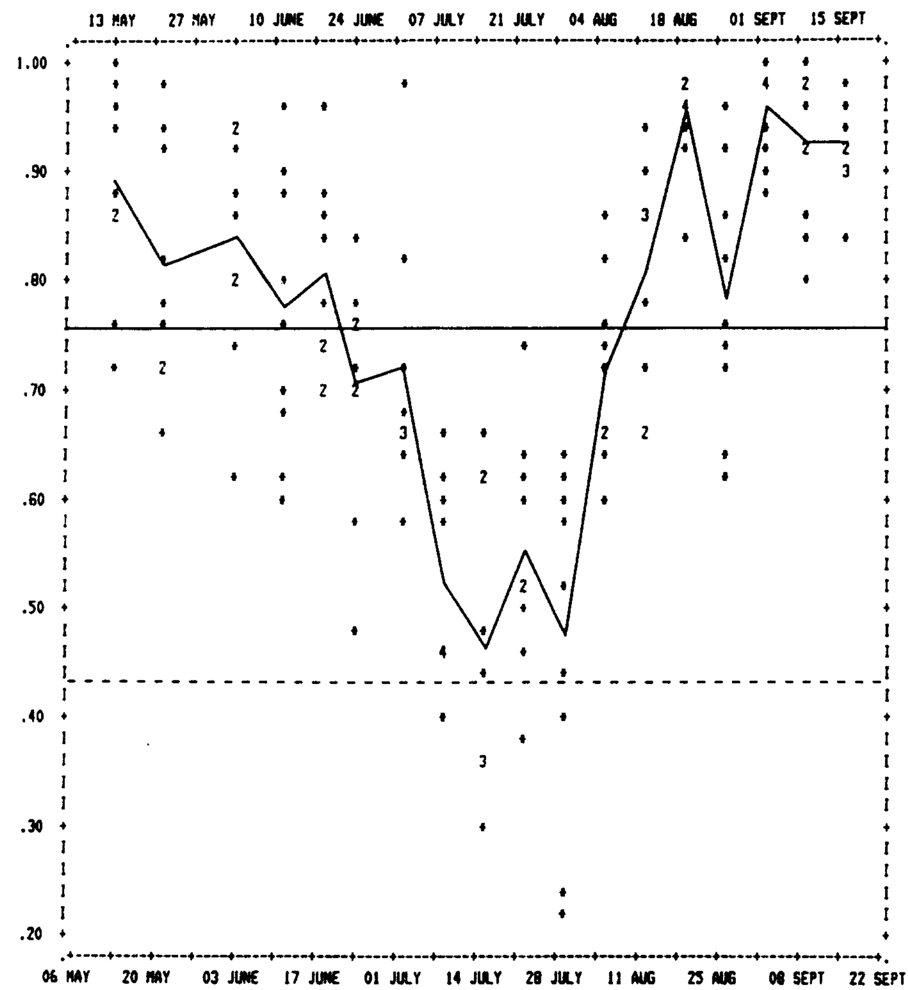

Fig. 1. Weekly variations in the microbial water quality index at nine sampling stations of the Breton beach, along 18 weeks of sampling from 13 May to 23 September, 1979. The index value corresponding to the government norms is plotted (dotted line) for reference. The solid straight line is the overall mean while the broken line joins each date's mean point. An integer means a multiple occurrence of the same I value, given the line printer's discriminating capacity.

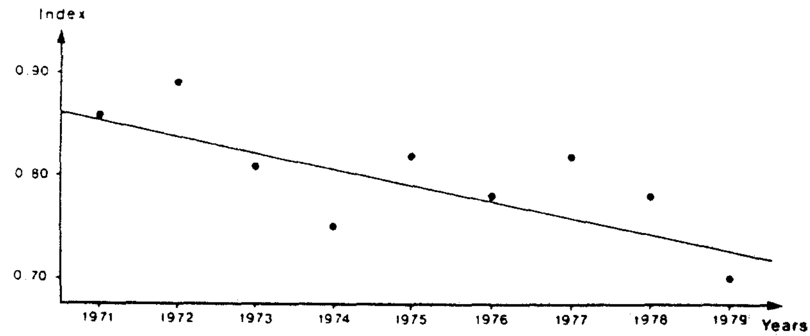

Fig. 2. Yearly averages of the microbial water quality index at the Smith beach along 9 years of sampling. The regression line shown was computed from the raw data. No data were available for this beach in 1980.

DISCUSSION

Comparison of the index with the mean of the log bacterial counts, for each of the three variables (Table 2), shows that I is indeed a water quality index, since it positions the beaches on a general bacterial gradient. The mean of the log-transformed data is used for this comparison instead of the mean of the raw data, since the arithmetic mean of a highly skewed distribution is likely to be an imprecise estimate of the distribution's central tendency.

The index can now be used to study the variations of water quality with time. A first example is provided by Fig. 1, where the weekly variations of the index, at the nine stations sampled in beach Breton, are represented for 1979. From this plot, one can appreciate the importance of the drop in bacterial water quality in July, corresponding both to the peak of the holiday season and to the highest water temperature period. The graph also shows that during this period the bacterial water quality may, at some stations on that beach, fall below the governmental minimal norms, even though the water quality is acceptable during most of the year.

In Fig. 2, the bacterial quality index of the Smith beach is plotted over a nine year period, revealing a long-term evolution of this beach towards worse average quality. For clarity, only the yearly averages of the index are plotted on this graph; however, the 363 individual data available from that beach (3 stations) over the same time period were used to compute the linear regression equation

$$Index = 0.87197 - 0.00004618\ (date)$$

the date being expressed in Julian days from the first of January 1971. Even though the coefficient of determination is relatively low ($r^2 = 0.08$ so that the date alone is a poor linear predictor of I, given the sampling time intervals), the negative slope coefficient is significantly different from zero at the 0.00001 level, thus showing that the degradation of the microbial water quality along the years is real.

These examples show that such an index, computed for a group of beaches from the same area and subjected to the same kind of use, can be a valuable synthetic tool in monitoring studies.

Acknowledgements—The data for this work was obtained and analyzed as part of a larger study undertaken by the Gatineau Park, National Capital Commission, Government of Canada; special thanks are due to Mr P. Kyer and R. Gélinas who are largely responsible for initiating this project. Our thanks are due to Dr Philippe Ross, Département de Sciences biologiques, Université de Montréal, for his critical review of the manuscript.

REFERENCES

Blackith R. E. & Reyment R. A. (1971) *Multivariate Morphometrics*. Academic Press, London.

Håkanson L. (1980) An ecological risk index for aquatic pollution control. A sedimentological approach. *Water Res.* **14**, 975–1002.

Legendre L. & Legendre P. (1979) *Ecologie Numérique*. Tome 1: Le traitement multiple des données écologiques. Tome 2: La structure des données écologiques. Collection d'Ecologie, Nos 12 et 13. Masson, Paris et Les Presses de l'Université du Québec.

Mahloch J. L. (1974) Multivariate techniques for water quality analysis. *J. envir. Engng Div., Proc. Am. Soc. civ. Engrs* **100** (EE5), 1119–1132.

Nie N. H., Hull C. H., Jenkins J. G., Steinbrenner K. & Bent D. H. (1975) *SPSS—Statistical Package for the Social Sciences*, Second Edition. McGraw-Hill, New York.

Pitzurra M., Iandoli M., Morlunghi P., Caroli G. & Scrucca F. (1980) L'indice microbico aria (I.M.A.) in camere di degenza di un reparto di medicina interna. *Ig. mod.* **73**, 857–872.

St-Louis N. (1981) Analyse statistique des résultats de dix ans d'échantillonnage des eaux des plages du parc de la Gatineau. Rapport de recherche présenté à la Comission de la Capitale nationale. Eco-Recherches Inc., Pointe-Claire.

Starzeka A., Pasternak K. & Ostrowski M. (1979) Essay in water purity classification on the basis of chosen biological and chemical properties. *Acta Hydrobiol.* **21**, 397–421.

Tebbutt T. H. Y. (1977) *Principles of Water Quality Control*, Second Edition. Pergamon Press, Oxford.

Zamfir G. (1980) Indicatori de apreciere comparativa si in dinamica a calitatii apelor de suprafata poluate cu substante organice biodegradabile. *Revtă Ig. Bacteriol. Virusol. Parazitol. Epidemiol. Pneumoftiziol., Ser. Igiena* **29**, 179–186.