Are Species Association Coefficients Really Necessary ?

by William Knight
Albert V. Tyler

FISHERIES RESEARCH BOARD OF CANADA
TECHNICAL REPORT NO. 397
1973

FRB

FISHERIES RESEARCH BOARD OF CANADA

Technical Reports

FRB Technical Reports are research documents that are of sufficient importance to be preserved, but which for some reason are not appropriate for primary scientific publication. No restriction is placed on subject matter and the series should reflect the broad research interests of FRB.

These Reports can be cited in publications, but care should be taken to indicate their manuscript status. Some of the material in these Reports will eventually appear in the primary scientific literature.

Inquiries concerning any particular Report should be directed to the issuing FRB establishment which is indicated on the title page.

FISHERIES RESEARCH BOARD OF CANADA

TECHNICAL REPORT NO. 397

ARE SPECIES ASSOCIATION COEFFICIENTS REALLY NECESSARY?

by

William Knight

Department of Mathematics
University of New Brunswick
Fredericton, N. B.

and

Albert V. Tyler

Fisheries and Marine Service
Department of Environment
Biological Station
St. Andrews, N.B.

This is the sixty-first FRB Technical Report from the
Fisheries Research Board of Canada
Biological Station, St. Andrews, N.B.

1973

Are Species Association Coefficients Really Necessary?

This note is against species association coefficients. Rather than urge, probably in vain, that you calculate no more species association coefficients, a weaker thesis is defended. If you must calculate them, don't flaunt them. We will write first of reasons for avoiding such coefficients, and then of ways and means of doing so.

The objections are really obvious. The mass of coefficients generated by a study of moderate size is suffocating;a study of 50 species provides 1225 coefficients. Nobody can look at a mixed bag of 1225 numbers to much comprehension; the result is confusion rather than clarification, a situation the more ironic as one of the reasons for having the coefficients is simplification.

There are lots of papers inventing new coefficients and disputing old ones. [As a sample at hazard, Cole 1949, Hurlbert 1969, Fager 1957, Levandowski 1972, and their references!] It is by no means clear to the practical user which coefficient to use, upon what criteria to base a choice, nor having made a choice, what the coefficient means.

Finally, we hold the admittedly subjective belief that in fact attempts at using species association coefficients usually don't "come off".

Most of us have seen studies something like this. The data are gathered and the association coefficients are calculated. The data being voluminous, a considerable number of these coefficients are significantly different from zero. The larger coefficients are selected for some such mention as this and that are seen to be highly associated. Some speculations as to the reason(s) for these "associations" are proposed but not subjected to scientific (as opposed to statistical) test. We submit that the results and conclusions have more the flavor of armchair speculation than working science.

What lessons are there from these failings? First, the species association coefficient is not a fundamental concept in the sense that energy or instantaneous mortality rate are. The act of defining one creates no insight. It, like the median or the correlation, is a descriptive statistic; it has no life apart from the data from which it was calculated, and whose job it is to summarize.

Second, as a descriptive device, it does not go far enough. The ideal description is short enough and simple enough for easy mental assimilation. A large table of association coefficients is seldom simple and never short. At best the table of coefficients is a half-way house, a fact which would appear generally recognized, for the table is often the beginning of a cluster analysis or an ordination.

How is such a cluster analysis or ordination to be interpreted? The straightforward way, tracing the meaning back through the association coefficients and thence to the recorded data and ultimately the "real world" is unappetizing in view of the dubious interpretation of the coefficients, and indeed is not usually attempted. People try to interpret clusters or ordinate axes more directly.

The result is speculation on the meaning of the clusters or ordinate axes, not subject to objective test. There are also the studies wherein massive mathematical mauling of the data, reported in detail, produces nothing more exciting than a depth gradient.

At their best the interpretations may be clear, even useful. Such examples as a forgotten burn and old trails lying unnoticed until brought out by

clustering, [Williams and Lambert 1960] or an ordination insisting that topography is the dominant factor where it had been thought there was so little topography it could be ignored [Hatheway 1971], bring joy to the statistician though they are probably rarer than he would care to admit.

Such factors could be found without the machinery of association coefficients and ordination methods. In the clear light of hindsight this is easy to say; since in fact these were not noticed until the machinery was employed it is pointless to say it. Nevertheless there is a real moral here: however obvious such things as old burns and small scale topography may or not be in prospect, in retrospect they are quite obvious without reference to the mathematical machinery required, or in any event used, to find them. A simple, verbal statement of the result is sufficient; nay, it is better to stop there, for detailed reporting of the mathematical machinery adds no insight and is likely actually to confuse the issue. A detailed account of the mathematical apparatus is no more to be included than, say, the blueprints of the pH meter used!

Association coefficients, and their ensuing superstructure of cluster analysis or ordination, can, and should, be exorcised from the verbal part of a study. Moreover, even at the mathematical level, association coefficients can be superfluous. We have found this to be so in a collection of data from research trawls on the Nova Scotia continental shelf where the association coefficients added no information to the location of the trawl. This topic occupies the remainder of this paper.

A table of habitat preferences is a natural description for species association data. Any environment can be stratified in some way, by altitude, by depth, by temperature, by moisture gradient, etc. The table shows the frequency, or alternately density, of each species at each environmental stratum. Of course this is nothing but the distribution curves used in gradient analysis [Whittaker 1967]. Not only is the table of such habitat preferences smaller than that of the association coefficients (except in small studies), and easier to understand; it also relates species to environment which is supposed to be the aim of ecology.

Do the species association coefficients contain any more information than the table of habitat preferences? To rephrase, are species associated only insofar as they happen to live in the same kind of place, or is there association over and above that of cohabitation? For any given set of data this question, in principle at least, can be decided by statistical test: Can the association coefficients be calculated up to sampling error given only the habitat preferences? The details of the test will vary with the type of data; we have designed one for presence-absence data (i.e. frequency rather than density).

We tried this test upon a collection of research trawl data at the St. Andrews Biological Station of Environment Canada. Almost all information in the table of association coefficients was contained in the table of habitat preferences. Typically all but one or two out of the order of 500 coefficients would be reproduced up to sampling error. The conclusion should be emphasized: Anything that can be done with the association coefficients can be done with the habitat preferences, and that with more understanding and less work. (This, bye the bye, became evident when we tried to group the species; grouping from the habitat preferences

could be done by eye; grouping from the association coefficients was done by computer, it is hard to see how a hand calculation could be done.) Note that, once again, association coefficients are calculated only to be discarded, to be used in the test which shows them unnecessary.

LITERATURE CITED

Cole, LaMont C. 1949

The measurement of interspecific association
Ecology 30: 411-424

Fager, E. W. 1957

Determination and analysis of recurrent groups
Ecology 38: 586-595

Hatheway, W. H. 1971

Contingency-table analysis of rain forest vegetation.
In Statistical Ecology, Vol. 3, Penn. State Univ.
Press, pp. 271-314.

Hurlbert, Stuart H. 1969

A coefficient of interspecific association
Ecology 50: 1-9

Levandowsky, M. 1972

An ordination of phytoplankton populations in ponds
of varying salinity and temperature
Ecology 53: 398-407.

Whittaker, R. H. 1967

Gradient analysis in vegetation
Biol. Review 42:207-264

Williams, W. T., and J. M. Lambert. 1960

Multivariate methods in plant ecology
J. of Ecology 48: 689-710