Pierre LEGENDRE and Andrzej CHODOROWSKI

Centre de Recherche en Sciences de l'Environnement, Université du Québec à Montréal, Montréal, Canada

A GENERALIZATION OF JACCARD'S ASSOCIATION COEFFICIENT FOR Q ANALYSIS OF MULTI-STATE ECOLOGICAL DATA MATRICES

ABSTRACT: A new association coefficient is described, for comparing localities described by species abundance data on a normalized or a relative scale. It uses a partial similarity function adjustable to the scale used. Computer program is available.

Contents

1. INTRODUCTION

Very many similarity coefficients have been described in the biological literature, and many of them have been discussed in recent textbooks, for instance in S n e a t h and S o k a l (1973) and in O r l o c i (1975). These coefficients have been classified under four headings by S n e a t h and S o k a l (1973): distance, association, correlation and probabilistic. Distance and association coefficients are both used for Q studies (comparison of objects over all descriptors) and usually correspond to one another, where the distance coefficient is to be used when a metric is needed, and the association coefficient is used for instance for hierarchical clustering. Correlation coefficients are used for R studies (comparison of descriptors through all the objects) and the choice of the appropriate correlation coefficient depends mainly on the type of descriptors to compare. Probabilistic coefficients form a small group of Q- or R-type

coefficients which tend to compare the behaviour of pairs of objects or descriptors, in relation to the behaviour of all the other objects or descriptors in the study.

Association coefficients are the most diversified, probably because they are more readily interpreted as to their meaning and results. They can be divided in two main categories, although some have been used for both purposes: coefficients developed by taxonomists, and coefficients directed towards ecology. The main difference between these categories resides in the interpretation of a double-nought: in taxonomy, where a nought usually (but not always) refers to the absence of the morphological descriptor under consideration, the fact for this structure to be absent in two of the (otherwise somewhat related) objects under study, means something as to the relationship existing between these two objects, as much as the presence of one of the forms of that structure would. Coefficients from this group are also adequate to handle, in ecology, physical and chemical characteristics of the milieu, where the "zero" for a variable is a valid basis for the comparison of two localities. However, when the nought refers to the absence of a species, the fact for a species to be absent from two localities does not mean much about the relationship of these two stations. The absence of blue whales from two of several pools, for instance, does not mean that these two pools are any more similar to one another, which illustrates that negative matches have to be excluded from such comparisons. On the other hand, replacement species may well be present. This is the main difference between the association coefficients developed by taxonomists and by ecologists, although authors have at times borrowed a coefficient from one field to apply it to the other.

Ecological association coefficients have been developed at first to deal with presence-absence type of data. J a c c a r d (1908) was the first to introduce such a coefficient, which compared two stations by counting the number of species present at the two locations, divided by the number of species present in the combination of the two stations. This coefficient has also been applied to abundance data (Fig. 1A), where it works nicely as long as there are not too many classes of abundance, and as long as the species are diversified enough.

In the case of a diversified environment, when one is trying to differentiate stations which are very close (similar) to one another, it may be helpful to use such a strong coefficient, in order to get a better spreading. But in a less diversified environment, where every piece of information has to be used, it is quite another story. In the search of taxocenes (C h o d o - r o w s k i 1959, 1960) we need a coefficient sensitive enough ·to account for partial similarities of two stations with respect to a given species. As an example, let us take two pools from a group of semi-permanent pools, in which zooplankton species can be found, and imagine that a single species, the same one, is found in the two pools, but in different abundances, on a scale from 0 to 5. The vectors representing the 21 potentially present species in the two pools would look as follows:

pool 1: 000 000 030 000 000 000 000

pool 2: 000 000 010 000 000 000 000

Considering only presence-absence, Jaccard's would give a similarity of 1 between these two pools, and 0 with multi-state abundance data. Obviously, the truth is somewhere in between, and Jaccard's coefficient shows here its incompleteness since it can take two opposite values with the same data, depending on how it is applied.
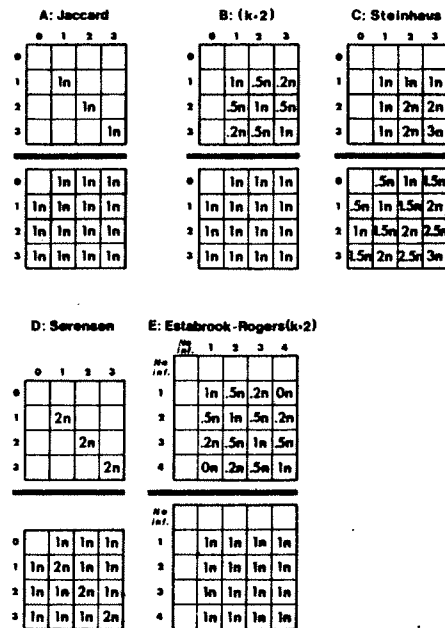
Fig. 1. Comparison of five association coefficients for pairs of objects

A — coefficient of J a c c a r d (1908), B — the generalization of Jaccard's coefficient presented in this paper, using the equation of partial similarity $f(d, k)$, with $k = 2$ in the example, C — coefficient attributed to Steinhaus by M o t y k a, D o b r z a ń s k i and Z a w a d z k i (1950), D — coefficient of S ø r e n s e n (1948) also used by D i c e (1945) as "coincidence index" for R studies, E — coefficient of E s t a b r o o k and R o g e r s (1966) for $k = 2$, it was developed for taxonomic purposes and uses the same partial similarity function $f(d, k)$. In this coefficient, state zero means "no information available".

In the upper part of the figure, the new coefficient presented in this paper (B) is compared to two other coefficients used on the same kind of data. The lower part of the figure presents mathematical variants of the coefficients just above: the coefficient of Sørensen is monotonic to the coefficient of Jaccard, and our new coefficient (B) uses the same partial similarity function as the coefficient of Estabrook and Rogers.

In examples A to D, the various descriptors have up to three states of abundance, plus state 0 for absence. For each coefficient, the upper matrix (numerator) and the lower matrix (denominator) are separated by a dark bar which symbolizes division. n represents the number of descriptors which code the two objects in the pair of states corresponding to the position in the matrix. A blank represents a structural zero. The numerator of the coefficient is obtained by summing the squares of the upper matrix, and the denominator likewise by summing the squares of the lower matrix

2. A PARTIAL SIMILARITY FUNCTION

The way around this difficulty, which we would like to suggest and illustrate in this paper, is to apply to the various abundances of each species, a function, f, from the family of decreasing functions, which gives the value of association for the various abundance combinations of a given species. This function should map the distance (d) between two abundances, together with a measure of the spreading (k) one wants to make of partial similarity, into the partial similarity relation (s) of the various pairs (i and j) of abundances of a species E:

$$s(E_i, E_j) = f(d, k)$$

Among all such functions available, the following one is here proposed because it has been proven useful in parallel cases in taxonomy, although it is clear that any other function with the same properties, which a worker would feel corresponds better to reality, could be used as well. The similarity value on species E, for a pair of sampling stations in which abundances i and j of species E have been respectively found, is thus a function of two parameters, d and k, such that

$$s(E_i, E_j) = f(d, k) = \frac{2(k + 1 - d)}{2k + 2 + dk} \text{ whenever } d \leqslant k \text{ (part 1)}$$
$$= 0 \text{ when } d > k \qquad \text{(part 2)}$$
$$= 0 \text{ when } E_i \text{ or } E_j = 0 \qquad \text{(part 3)}$$

where d is the "distance" between abundance E_i and abundance E_j, expressed by the absolute value $|i - j| = d$; k is a parameter set by the biologist for each species, which indicates the largest "distance" $|i - j|$ between abundances of the given species, for which the worker wishes to make a non-zero assignment (E s t a b r o o k and R o g e r s 1966). L e g e n d r e and R o g e r s (1972: 593) discuss in detail the criteria for fixing the value of k for each variable.

If for instance the various abundance states of a species are coded 0 to 7, if one established the maximum "distance" in which partial similarity is allowed, to be $k = 2$, then two sampling stations with abundances $E_1 = 2$ and $E_2 = 3$ would have, for that species, a similarity equal to

$$f(|2 - 3|, 2) = f(1, 2) = \frac{2(2 + 1 - 1)}{4 + 2 + 2} = .5 \text{ according to part 1 of the equation.}$$

But for another species, if $E_1 = 2$ and $E_2 = 5$ with $k = 2$:

$|2 - 5| = 3$ and $3 > k$, then $f(3, 2) = 0$ according to part 2 of the equation.

With a third species, if $E_1 = 2$ and $E_2 = 0$, there is not much sense recognizing a partial similarity between an abundance of 0 and any other abundance, even though d is not larger than k. This is why the limitation has been added in part 3 of the equation, that $f(d, k) = 0$ when E_i or $E_j = 0$.

3. THE SIMILARITY MEASURE FOR PAIRS OF LOCALITIES

After the two localities have been compared for each one species, it is now possible to combine these species similarity measures into a measure of the similarity $S(a, b)$ between two sampling stations a and b,\for their species composition.   It is calculated as the sum of the partial similarity values for all species, divided by the total number of species found in the combination of the two sampling stations (Fig. 1B).

$$S \; (\text{a, b}) = \frac{\Sigma \; s(E \; (\text{a}), E \; (\text{b}))}{\text{total n. of species present at one or the other station}}$$

In this equation, the theoretical abundances i and j of species E are replaced by their actual values obtained by sampling localities a and b, and are noted E(a) and E(b) respectively.

There are two main differences with the taxonomic coefficient of E s t a b r o o k and R o g e r s (1966) illustrated in Figure 1E: first, the description 0 means "no information available" in Estabrook-Rogers' coefficient. Secondly, and since it is a taxonomic coefficient, all the comparisons are included in the denominator, except in the case of "no information available": so, the denominator in Estabrook-Rogers' measure includes only the variables with information available for both objects.

To summarize, the properties of this coefficient are the following:

1. When d is larger than k, the two sampling stations are considered as not similar at all for that species, i.e., $s(E_i, E_j) = 0$ (equation, part 2).

2. When $d = 0$, the value $s(E_i, E_j)$ is equal to 1, as expected, except when i and j are equal to 0 (equation, part 1; Fig. 2).

3. $f(d, k)$ decreases as d increases for fixed k (Fig. 2).

4. $f(d, k)$ increases as k increases for fixed d (Fig. 2).

5. When E_i or E_j equals 0, the two sampling stations are considered as not similar for that species, i.e., $s(E_i, E_j) = 0$, even though d is not larger than k (equation, part 3; Fig. 1B, numerator).

6. Jaccard's coefficient is obtained by setting $k = 0$ (Figs. 1A, B).

Table I. The values of $f(d, k)$ for the most commonly used values of k. See equation in text. (Modified from L e g e n d r e and R o g e r s 1972)

k	d							
	0	1	2	3	4	5	6	7
0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	1	.40	0.00	0.00	0.00	0.00	0.00	0.00
2	1	.50	.20	0.00	0.00	0.00	0.00	0.00
3	1	.55	.28	.12	0.00	0.00	0.00	0.00
4	1	.57	.33	.18	.08	0.00	0.00	0.00
5	1	.59	.36	.22	.13	.05	0.00	0.00

Even though this formula is empirical, its properties reflect the judgments an ecologist can make about the problem of partial similarity. The values of $f(d, k)$ for the most commonly used values of k are given in Table I.

4. NEGATIVE VALUES OF PARTIAL SIMILARITY?

To help the user appreciate the rates of decrease associated with the various values of k, $f(d, k)$ has been plotted in Figure 2A. A direct consequence of this exercise is to make one wonder: why could not $f(d, k)$ take negative values? This could be done by removing from the definition above, the line which says that $f(d, k) = 0$ when $d > k$. This solution is plotted as Figure 2B, and it changes $f(d, k)$ from a monotone decreasing function into a strictly decreasing

function on interval $[0, \infty)$. It has the consequence that $f(d, k)$ takes negative values when $d > (k + 1)$. Whether this modification would be of any interest could only be assessed by trying it on real data. For sure, it would help to spread apart things which are not alike, and as such it could compensate for the overestimation of the similarity (compared to Jaccard's formula) which results from using $f(d, k)$ as defined above.

5. CONTINUOUS DESCRIPTORS

In the present paper, only integer values of d are used, since abundance classes are discontinuous ordered descriptors. The curves of Figure 2 suggest however that the same partial similarity function could be used with continuous data, since the domain of the function is the set of natural numbers.
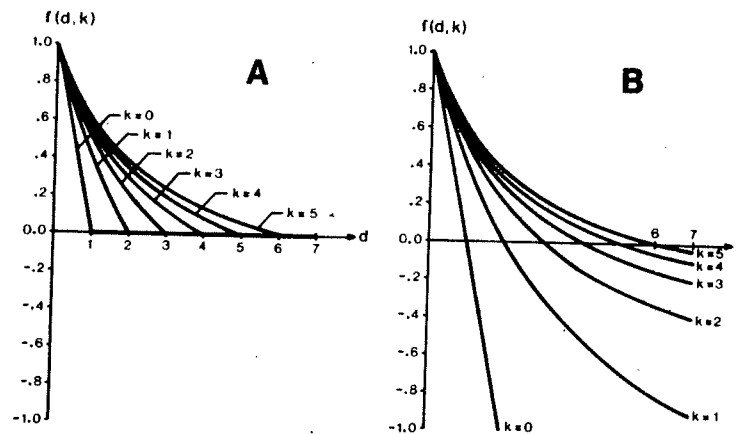

Fig. 2. $f(d, k)$ is plotted as a function of d for the various of k. A — with the restriction that $f(d, k) = 0$ when $d > k$. B — without this restriction

6. COMPUTER PROGRAM

A program written in Fortran—IV and optimized for Control Data series 6,000 computers is available from the authors[1]. This is actually a modified version of a clustering model in graph theory which has been described and published before (L e g e n d r e and R o g e r s 1972; E s t a b r o c k et al. 1972), in which the similarity measure has been replaced with the one described here.

The procedure starts by calculating, in Q mode, the said measure of association between all pairs of sampling stations in the study. Then these similarities are placed in order of decreasing value, and the clusters are formed following this list of ordered similarities. The procedure is

[1]Please specify which of the three following listings is required: (a) the measure of association only, or (b) the measure of association together with the clustering program it is a part of, or (c) the graph-clustering package which includes 7 measures of association designed for problems of taxonomy and ecology, in Q and R modes. Listings will be sent free of charge. If cards or tape are needed, please write for further instructions.

agglomerative and of the single-linkage type. An histogram-type visual display summarizes the clustering activity.

In the program, a sophistication added to the calculation of the measure of association makes it possible for a worker to enter for any one species a new set of partial similarity values (between 0 and 1) replace those produced by function $f(d, k)$, together with the input data, if he so wishes, without modifying the program. This facility may be useful in the following cases: (a) if the worker does not like the slope of function $f(d, k)$ for the particular problem at hand; (b) if some of the states do not follow the other ordered states of abundance: for example, in a study of pools, if the various states of abundance are followed by the state "dry pool"; (c) if some of the states are built on a different scale of abundance than the others: for example, 0–10, 10–20, 20–30, 30–40, 40–50, then 50–100, 100–1,000, 1,000 and more; (d) if, instead of the coefficient described hereinbefore, another measure of similarity is to be used, which differs from this one only by the numerator (see section 3 and Fig. 1); (e) in comparisons implying ecological variables, other than abundances, which are not logically ordered on a scale, when some pairs of stations in non-identical states of a variable are still judged by the worker to be more similar to each other, for this variable, than some other pairs of stations in another pair of states.

7. EXAMPLE: PLANKTONIC SIMILARITY OF POOLS

To illustrate the use of this generalization of Jaccard's association coefficient, we present hereafter data from a problem which has been treated with this method, in which various astatic pools located on islands of the St. Lawrence river southeast of Montreal (Quebec) were characterized for their content in zooplankton. Abundance of the 38 species found was coded on an estimative scale from 0 (absent) to 5 (very abundant) as used in the description of taxocenes (C h o d o r o w s k i 1960). This scale, quite helpful when results are needed rapidly, corresponds roughly to a log transformation of actual counts.

Figure 3 illustrates the results when the planktonic similarity was calculated with the generalized Jaccard's coefficient presented here ($k = 2$). The pools are positioned in a space formed by the first (abscissa) and the second (ordinate) principal coordinates (G o w e r 1966), calculated from the association matrix of the 20 pools. The program used for principal coordinates analysis was basically the one listed in B l a c k i t h and R e y m e n t (1971). The lines between the pools, called primary connections, or dendrites (Ł u k a s z e w i c z 1951, adapted to taxocenes by M o r a c z e w s k i 1962), stand for the highest similarity values which link a pool to another, until all the pools are linked by a single chain of connections. In other words, it is the chain of primary connections in a single linkage clustering strategy. This projection of single linkage clusters onto a space of reduced dimensionality obtained by principal coordinates analysis, is documented elsewhere (L e g e n d r e 1976).

In Figure 3, the astatic pools surrounded by a circle are dry at least part of the year. They are filled during the high water season and/or by the rain, and they are referred to as "periodic". Figure 3 shows that the association coefficient is doing its job as expected: periodic and permanent or semi-permanent pools form two well-defined groups. On the other hand, the tendency is for a pool to be close to, and to cluster first with, pools of the same area.

Figure 4 illustrates very similar results, obtained with the association coefficient attributed to the Polish mathematician Steinhaus by M o t y k a, D o b r z a ń s k i and Z a w a d z k i (1950) instead of the generalized Jaccard's coefficient.
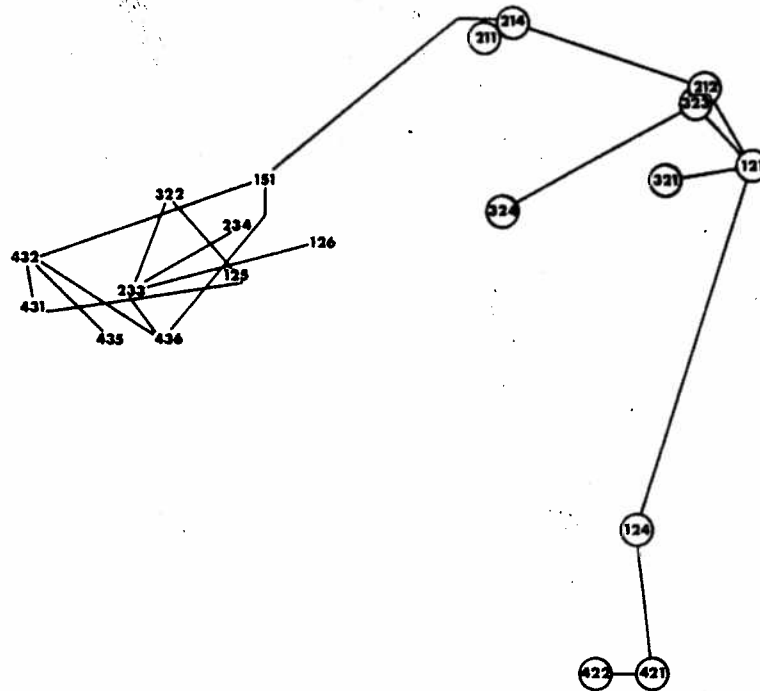
Fig. 3. Twenty astatic pools are compared for their planktonic similarity, through the use of the generalized association coefficient introduced in the present paper (with $k = 2$). From the association coefficient, principal coordinates were calculated (see text), and the pools are presented here in the space formed by the first (abscissa) and second (ordinate) principal coordinates. The highest similarity values are also drawn, which unite all the pools through a single chain of primary connections. Pools in a circle are periodic. Numbering system as in Figure 4

8. WHEN TO USE THIS GENERALIZED COEFFICIENT

This generalization of Jaccard's association coefficient has been designed for Q analysis (comparison of sampling stations) of ecological data matrices in which the abundance of the species characterizing the samples is known. So it pertains to the small group of quantitative Q-type ecological coefficients, the most famous of which is the one attributed to Steinhaus by Motyka, Dobrzański and Zawadzki (1950), which compares the smallest number to the average abundance of each species at the two stations:

$$S = \frac{W}{(A + B)/2}$$

where W is the sum (for all the species) of the lesser values at two stations; A and B are the sum of abundances at each of the two stations compared[2].

A related coefficient is that of K u l c z y ń s k i (1928), which compares the sum of minima to the sum total at each station, then averages the two values obtained[3]:

$$K = \frac{1}{2}\left(\frac{W}{A} + \frac{W}{B}\right)$$

In his study K u l c z y ń s k i (1928), multiplied this value by 100 to obtain percentages. Since Kulczyński's method of averaging two comparisons is somewhat more arbitrary than comparing directly minima to averages as does Steinhaus, the latter should be preferred, although the results should be almost monotonic. M c C o n n a u g h e y (1964) has used a truncated linear transformation of this coefficient on presence-absence data in an R study.

The relation between Steinhaus coefficient and the generalized coefficient introduced in this paper is one of interest. In the example above (Figs. 3, 4), it can be seen that both do the job quite similarly, that is, in both cases the same pools lie close to and cluster (primary connection) mainly with each other, although the two measures are not algebraically monotonic. Indeed, both measures manage to account for partial similarity, but not in the same manner: our generalized coefficient uses function $f(d, k)$, while Steinhaus uses the lesser value of each pair of species abundances. Figure 1 (B and C) illustrates this difference in the case of a 3-states abundance scale.

A comparison has been made between our generalized coefficient and Steinhaus' coefficient, for the data of the example above. Association has been calculated between all pairs of pools with the generalized coefficient (for $k = 0$ to $k = 5$), and with Steinhaus' coefficient. Comparison of the five association matrices was made using Pearson's r linear correlation coefficient, thus controlling for scale differences (Steinhaus' values are consistently larger than those obtained with the generalized coefficient).

These comparisons are listed in Table II. It shows that the correlation with Steinhaus' coefficient increases as the value of k, the maximum "distance" on which partial similarity is

---

[2]Following the symbolism introduced above, the coefficient of Steinhaus could be written:

$$S(a, b) = \frac{2 \sum\limits_{E=1}^{n} \min\left[E(a), E(b)\right]}{\sum\limits_{E=1}^{n} \left[E(a), E(b)\right]} \quad \text{for the } n \text{ species } E.$$

[3]Likewise, the coefficient of Kulczyński could be written as follows:

$$K(a, b) = \frac{\sum\limits_{E=1}^{n} \min\left[E(a), E(b)\right]}{2} \cdot \left[\frac{1}{\sum\limits_{E=1}^{n} E(a)} + \frac{1}{\sum\limits_{E=1}^{n} E(b)}\right]$$

Table II. Correlation between the generalized coefficient introduced in this paper and Steinhaus' coefficient, on the planktonic data of the example. The generalized coefficient has been calculated for six values of k (descriptors were coded 0 to 5)

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$k = 1$	.8961					
$k = 2$	.8537	.9896				
$k = 3$	.8282	.9767	.9955			
$k = 4$	.8113	.9662	.9894	.9986		
$k = 5$	.7995	.9583	.9842	.9962	.9994	
Steinhaus	.6326	.8209	.8714	.9001	.9129	.9195

allowed, increases. On the other hand, and as expected, the similarities calculated with each value of k are in higher correlation with those of the closest values of k.

Looking into the equations, and referring to Figure 1 (B and C), the difference between the coefficient of Steinhaus (or the one of Kulczyński) and the one introduced here becomes obvious. With our coefficient, the partial similarity between equally distant states remains the
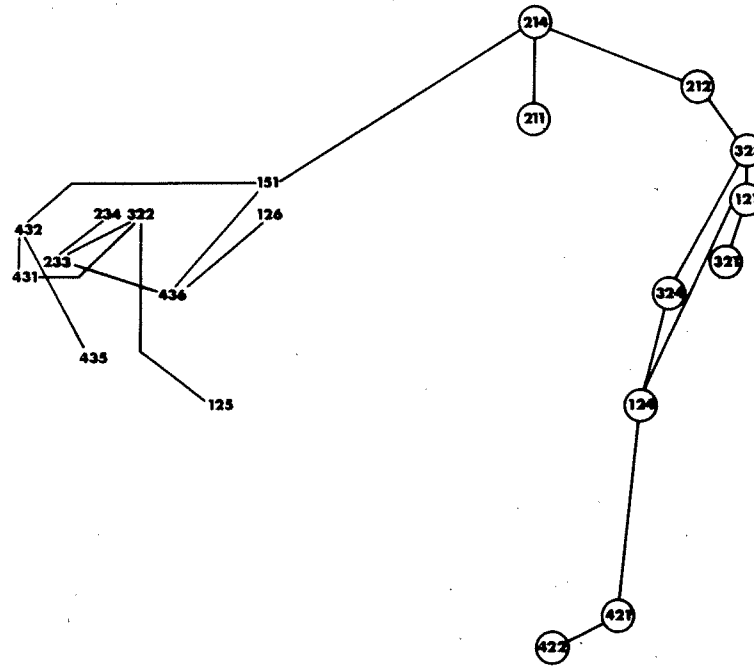

Fig. 4. As in Figure 3, but the similarity was calculated through the use of the association coefficient of Steinhaus. Each pool was given a 3-digits name built as follows: the first digit for the area, the second digit for the insland; the third digit designates the pool on the insland. Areas 1 and 2 are located before Montreal on the St. Lawrence, and areas 3 and 4 are located in front of Montreal. 12X = Maple Island; 15X = Ste--Timothée Island; 21X = Claude Island; 23X = Pointe-des-Cascades; 32X = Des Soeurs (Nuns) Island; 42X = Grosbois Island; 43X = de la commune Island

same, independently from the position on the scale. With $k = 2$ for instance, a partial similarity of 0.5 will always be attributed to neighbouring states: states 1 and 2 or states 7 and 8 alike. With Steinhaus on the other hand, and with only one descriptor, the partial similarity between states 1 and 2 is $2 \min(1, 2)/(1 + 2) = .6667$, while between states 7 and 8, it is $2 \min(7, 8)/(7 + 8) = .9333$. One should then use coefficients like Steinhaus' or Kulczyński's on real abundance data (actual number of individuals) where one wants partial similarity to increase as the value increases, and the coefficient presented here on normalized abundance data (log transformation and the like), or corresponding estimative scales, like the scale used in our example, where the difference between two successive values corresponds approximately to an equal amount of dissimilarity. Depending upon the scale used (0 to 5 or 0 to 50, for instance), $f(d, k)$ makes it possible for the worker to establish how much he wants to depart from the case of no partial similarity, by adjusting the value of k, for each species if necessary.

9. SUMMARY

A new association coefficient is described, which has been designed for comparing sampling stations described by species abundance data.

The coefficient uses, for each species on which the two stations are compared, an empirical decreasing partial similarity function (Table I, Fig. 2) of two parameters, d and k, where d is the "distance" between the abundances of the given species at the two stations, and k is a parameter set by the biologist for each species, which indicates the largest "distance" on which he wishes to allow partial similarity.

This function is derived from that used in the taxonomic association coefficient of Estabrook and Rogers. The similarity between two sampling stations, for their species composition, is given by the sum of the value of partial similarity on each species, divided by the total number of species present at the two localities. A computer program is available from the authors, which contains an extra facility to input some matrix of predetermined values the worker wishes to use instead of the values given by the function of partial similarity. This is particularly useful when the abundance states are not well ordered. The new coefficient is compared with some other related measures of association (Fig. 1). Comparison with the coefficients of Steinhaus (Figs. 3, 4, Table II) and of Kulczyński indicates that these are best used with raw abundance data, while this new coefficient has been designed for transformed species abundance data: log or other normalizing transformation, or relative abundances.

10. POLISH SUMMARY (STRESZCZENIE)

Autorzy opisują nowy wskaźnik podobieństwa służący do porównania liczebności gatunków w badanych stanowiskach. Do wyliczania wskaźnika (dla każdego z gatunków obecnych w dwóch porównywanych stanowiskach) używa się empirycznie zmniejszanej funkcji częściowego podobieństwa (tab. I, fig. 2) dwóch parametrów: d i k. Parametr d jest „odległością" między liczebnościami danych gatunków w obu stanowiskach, a k jest parametrem wyznaczonym przez biologa dla każdego gatunku wskazującym największą „odległość" żądaną, dla znalezienia częściowego podobieństwa. Funkcja ta jest wyprowadzona ze wskaźnika pokrewieństwa taksonomicznego Estabrooka i Rogersa.

Podobieństwo składu gatunkowego dwóch badanych stacji jest wykazane przez sumę wartości częściowych podobieństw podzieloną przez ogólną liczbę gatunków obecnych w obu stacjach.

Autorzy dysponują programem komputerowym, który umożliwia ponadto wprowadzenie tablic z przygotowanymi uprzednio przez użytkownika danymi, na miejsce wartości otrzymanych przez zastosowanie funkcji częściowego podobieństwa. Jest to szczególnie użyteczne, gdy dane dotyczące liczebności nie są dobrze uporządkowane.

Nowy wskaźnik jest porównywany z kilkoma innymi współczynnikami o zbliżonym charakterze (fig. 1). Porównania ze wskaźnikami Steinhausa (fig. 3, 4, tab. II) i Kulczyńskiego wskazują, że te ostatnie służą lepiej do porównywania surowych danych liczebności, podczas gdy nowy wskaźnik jest przeznaczony głównie do przekształconych danych liczebności gatunków (przekształcenia logarytmiczne i normalizacyjne albo liczebności względne).

11. REFERENCES

1. B l a c k i t h R. E., R e y m e n t R. A. 1971 – Multivariate morphometrics – Academic Press, London, New York, 412 pp.
2. C h o d o r o w s k i A. 1959 – Ecological differentiation of turbellarians in Harsz-Lake – Pol. Arch. Hydrobiol. 6: 33–73.
3. C h o d o r o w s k i A: 1960 – Taxoceny wirków (*Turbellaria*) i metodyka ich badania (Les taxocènes des Turbellariés et la méthode de leur étude) – Ekol. pol. B.6: 95–114.
4. D i c e L. R. 1945 – Measures of the amount of ecologic association between species – Ecology, 26: 297–302.
5. E s t a b r o o k C. F., F l e m i n g H. S., L e g e n d r e P., R o g e r s D. J., W i r t h |M. 1972 – Cluster analysis program using graph theory. User's manual, assembled by H. S. Fleming and S. G. Appan, Gulf Universities Research Corp., NASA/Mississippi Test Facility, Bay St-Luis – Centre de Recherche écologique, Université du Québec à Montréal, Montreal, 101 pp.
6. E s t a b r o o k G. F., R o g e r s D. J. 1966 – A general method of taxonomic description for a computed similarity measure – Bio Sci. 16: 789–/93.
7. G o w e r J. C. 1966 – Some distance properties of latent root and vector methods used in multivariate analysis – Biometrika, 53: 325–338.
8. J a c c a r d P. 1908 – Nouvelles recherches sur la distributior florale – Bull. Soc. vaud. Sci. nat. 44: 223–270.
9. K u l c z y ń s k i S. 1928 – Die Pflanzenassoziatonen der Pieninen – Bull. int. Acad. pol. Sci. Lett. Ser. B, Suppl. II (1927): 57–203.
10. L e g e n d r e P. 1976 – An appropriate space for clustering selected groups of Western North American *Salmo* – Syst. Zool. 25.
11. L e g e n d r e P., R o g e r s D. J. 1972 – Characters and clustering in taxonomy: a synthesis of two taximetric procedures – Taxon 21: 567–606.
12. Ł u k a s z e w i c z J. 1951 – Sur la liaison et la division des points d'un ensemble fini – Colloquium math. 2: 282–285.
13. M c C o n n a u g h e y B. H. 1964 – The determination and analysis of plankton communities – Mar. Res. Indonesia, special number: 1–40
14. M o r a c z e w s k i J. 1962 – Différentiation écologique de la faune des Testacés du littoral peu profond du lac Mamry – Pol. Arch. Hydrobiol. 10: 333–353.
15. M o t y k a J., D o b r z a ń s k i B., Ż a w a d z k i S. 1950 – Wstępne badania nad łąkami południowo-wschodniej Lubelszczyzny (Preliminary studies on meadows in the south-east of province Lublin) – Annls Univ. Mariae Curie-Skłodowska, Sect. E, 5: 367–447.
16. O r l o c i L. 1975 – Multivariate analysis in vegetation research – Dr W. Junk B. V. Publishers, The Hague, 276 pp.
17. S n e a t h P. H. A., S o k a l R. R. 1973 – Numerical taxonomy – The principles and practice of numerical classification – W. H. Freeman and Co., San Francisco, 573 pp.
18. S ø r e n s e n T. 1948 – A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analysis of the vegetation on Danish commons – Biol. Skr. 5: 1–34.

AUTHORS' ADDRESS:
Dr. Pierre Legendre
Dr. Andrzej Chodorowski
Centre de Recherche en Sciences de l'Environnement
Université du Québec à Montréal
C.P. 8888
Montréal, Québec H3C 3P8
Canada.