# Measuring of Interspecific Association and Similarity between Communities

by

**Masaaki MORISITA**

*Reprinted from the Memoirs of the Faculty of Science,*
*Kyushu University, Series E (Biology),*
*Vol. 3, No. 1*

**FUKUOKA, JAPAN**

**1959**

# Measuring of Interspecific Association and Similarity between Communities[1)]

by

**Masaaki MORISITA**

(Department of Biology, Faculty of Science, Kyushu University Fukuoka, Japan)

*Received July 18, 1959*

## Introduction

Many indices based on the quadrat sampling have been deviced for determining the degree of interspecific association and also of similarity between communities. However, almost of them have a disadvantage that the values obtained are considerably influenced by the average number of individuals per quadrat as has been partly noticed by Nash (1950). As the consequence, the interpretations of the results obtained through the use of these indices dealing with the analysis of interspecific relationships, grouping of samples or species, ordination of communities and etc. might sometimes be erroneous, since it may happen that a difference between the index values is nothing but a reflection of the difference between densities per quadrat, not of the true relation between two species or between two samples.

In this paper, new indices for measuring the interspecific association and similarity between communities which are almost unaffected, though not perfectly, by the average density per quadrat shall be given, and they may be useful in applying to the studies of population and community ecology.

## Interspecific association

There are two standpoints for treating the interspecific association: one deals with the overlapping of distributions of two species regardless of the area on which neither of the two species is present, and the other treats the whole area examined considering the single or joint occurrence by chance of the two species. The association index by Dice (1945) and by Whittaker (1952) and the index of amplitudinal correspondence by Bray (1956) might be derived from the former and the coefficient

---

[1)] Contributions from the Department of Biology, Faculty of Science, Kyushu University, No. 83. Aided in part by the Scientific Research Fund from the Ministry of Education.

of association by Cole (1949) and the quantile correlation coefficient by De Vries (1954) from the latter. The former indices measure, say, the absolute amount of interspecific overlapping and the latter ones, the relative amount which is determined by both of the absolute overlapping and the area where both the species are absent.[1] In this paper, the indices from the two standpoints are described separately.

*1. Index of interspecific overlapping*

Let $n_{xi}$ and $n_{yi}$ be the numbers of individuals of species X and Y occurring in the $i$-th quadrat ($i=1, 2, \ldots, q$), and $N_x$ and $N_y$ be the total numbers of individuals of X and Y sampled, then

$$\delta_x = \frac{\sum_{i=1}^{q} n_{xi}(n_{xi}-1)}{N_x(N_x-1)} \qquad (1)$$

and

$$\delta_y = \frac{\sum_{i=1}^{q} n_{yi}(n_{yi}-1)}{N_y(N_y-1)} \qquad (2)$$

will be independent of the sizes of $N_x$ and $N_y$ if the individuals are distributed at random on each subarea from which each quadrat is taken (Morisita, 1959).

Putting

$$C_\delta = \frac{2\sum_{i=1}^{q} n_{xi}\, n_{yi}}{(\delta_x+\delta_y)N_x N_y}, \qquad (\delta_x+\delta_y>0) \qquad (3)$$

and letting the value of $C_\delta$ obtained when $N_x=N_y=N$ be $C_{\delta_0}$, we have

$$C_{\delta_0} = 2\left(2-\frac{1}{N}\right)\frac{\delta_{x+y}}{\delta_x+\delta_y}-\left(1-\frac{1}{N}\right), \qquad (4)$$

where

$$\delta_{x+y} = \frac{\sum_{i=1}^{q}(n_{xi}+n_{yi})(n_{xi}+n_{yi}-1)}{2N(2N-1)}. \qquad (5)$$

If $N$ is large,

$$C_{\delta_0} \doteq 4\frac{\delta_{x+y}}{\delta_x+\delta_y}-1. \qquad (6)$$

---

[1] This classification will somewhat differ from that of Bray (1956) who has claimed that the interspecific association and the amplitudinal correspondence are quite independent of each other, the one showing the influence of direct species reactions and the other, the degree of coincidence in ecologic amplitudes of two species. However, Bray's claim seems not to be fully reasonable as an interspecific association value, though it is obtained from a stand, will reflect not only the direct species reactions but also the effect of micro-habitat conditions in the stand, and, on the other hand, an amplitudinal correspondence value obtained from several stands may sometimes reflect the direct species reaction.

Since $\delta_{x+y}$ will also be uninfluenced by the size of $N$, $C_{\delta_0}$ can be used as an index of overlapping unless $q\delta_x$ and $q\delta_y$ are significantly smaller than unity (Morisita, 1959), taking the value of about 1[1] when the densities of both species are equal to each other on each subarea in spite of the difference among the subarea desities, and the value of zero when there is no quadrat of joint occurrence of the two species.

The next problem is whether or not $C_\delta$ is affected by the sizes of $N_x$ and $N_y$, when $N_x > N_y$.

Let the whole area examined be composed of $Z$ subareas on each of which the individuals of the two species are distributed at random, and $t_l$ quadrats be taken from the $l$-th subarea ($l=1, 2, \ldots, Z$),[2] then we have

$$C_\delta = \frac{2\sum_{l=1}^{Z}\sum_{j=1}^{t_l} n_{xlj}\, n_{ylj}}{(\delta_x+\delta_y)\, N_x N_y}, \tag{7}$$

where $n_{xlj}$ and $n_{ylj}$ are the numbers of individuals of species X and Y occurring in the $l_j$-th quadrat ($j=1, 2, \ldots, t_l$).

If $t_l$ is large, $\sum_{l=1}^{Z}\sum_{j=1}^{t_l} n_{xlj}\, n_{ylj}$ will be nearly equal to $\sum_{l=1}^{Z} t_l\, \bar{n}_{xl}\, \bar{n}_{yl}$ since the two species are premised to be distributed independently of each other on each subarea.

Now consider that small quadrats of a fixed size in total of which $N_y$ individuals of species X are contained be sampled from the quadrats, each from each quadrat, and that the number of individuals ($n'_x$) of species X occurring in each small quadrat be compared with that ($n_y$) of species Y occurring in the corresponding large quadrat, then, putting $N_x = kN_y$, we have

$$C_{\delta_0} = \frac{2\sum_{l=1}^{Z}\sum_{j=1}^{t_l} n'_{xlj}\, n_{ylj}}{(\delta_x+\delta_y)\, N_y^2} \doteqdot \frac{2\sum_{l=1}^{Z} t_l\, \bar{n}'_{xl}\, \bar{n}_{yl}}{(\delta_x+\delta_y) N_y^2}$$

$$= \frac{2\sum_{l=1}^{Z} t_l\, k\bar{n}'_{xl}\, \bar{n}_{yl}}{(\delta_x+\delta_y)\, kN_y^2} = \frac{2\sum_{l=1}^{Z} t_l\, \bar{n}_{xl}\, \bar{n}_{yl}}{(\delta_x+\delta_y)\, N_x N_y} \doteqdot C_\delta . \tag{8}$$

[1] The variance of $\delta$ is approximately equal to $4\{\sum\pi^3-(\sum\pi^2)^2\}/N$ which tends to zero as $N$ becomes infinity, $\pi$ being the proportion of density in each subarea ($\sum\pi=1$) (Simpson, 1949; Morisita, 1959). Consequently the application of Tchebycheff's theorem will give us that for each assigned positive number $\varepsilon$, the probability that the values $\delta_x$, $\delta_y$ and $\delta_{x+y}$ lie in a common interval ($m$, $M$) is greater than $1-\varepsilon$ for sufficiently large $N$. Since $m$ and $M$ may be taken as near as we wish, we can conclude that the probability that the value of $C_{\delta_0}$ is much larger or smaller than unity will be very small for large $N$ in this case. (This note is due to Prof. Kitagawa's advice.)

[2] If the quadrat size is very small as compared with the size of each subarea, the probability that a quadrat randomly sampled from the whole area is taken from the border area between subareas and contain the fragments of two or more subareas in it will be very small (Morisita, 1959).

Therefore, it is known that $C_\delta$ takes nearly equal value to $C_{\delta_0}$ at least when $t_l$ is large, indicating that it is almost uninfluenced by the sizes of $N_x$ and $N_y$. And even if $t_l$ is small, it may be considered that $C_\delta$ does not take much different value from the $C_{\delta_0}$ value when $n_x$, $N_x$ and $N_y$ are large, because the $\delta_x$ value will be almost unchanged by substituting $n_x/k$ and $N_x/k$ for $n_x$ and $N_x$ in this case.

In Table 1, an example of $C_\delta$ values computed from the samples with different sizes of $N_x$ and $N_y$ taken from the two populations having equal density ratio in

**Table 1.** The values of $C_\delta$ and $R_\delta$ for different sizes of $N_x$ and $N_y$ sampled from the populations mapped in Fig. 1.

| $N_x$ | $N_y$ | Number of quadrats | $C_\delta$ | $R_\delta$ | Note |
|---|---|---|---|---|---|
| 210 | 105 | 125 | 1.081 | +0.566 | Whole area was subdivided into 125 quadrats. |
| 210 | 28 | 125 | 1.087 | +0.543 | The numbers of individuals of sp. X occurring in the areas with the size of 1/4 quadrat, each taken from each of 125 quadrats, were compared with those of sp. Y occurring in the corresponding quadrats. |
| | 27 | 125 | 0.886 | +0.475 | |
| | 30 | 125 | 1.285 | +0.699 | |
| Average | | | 1.086 | +0.572 | |
| 102 | 105 | 500 | 1.127 | +0.618 | Whole area was subdivided into 500 quadrats, and the numbers of individuals of sp. X occurring in the areas with the size of 1/2 quadrat area were compared with those of sp. Y occurring in the corresponding quadrats. |
| 108 | 105 | 500 | 0.991 | +0.567 | |
| Average | | | 1.059 | +0.593 | |

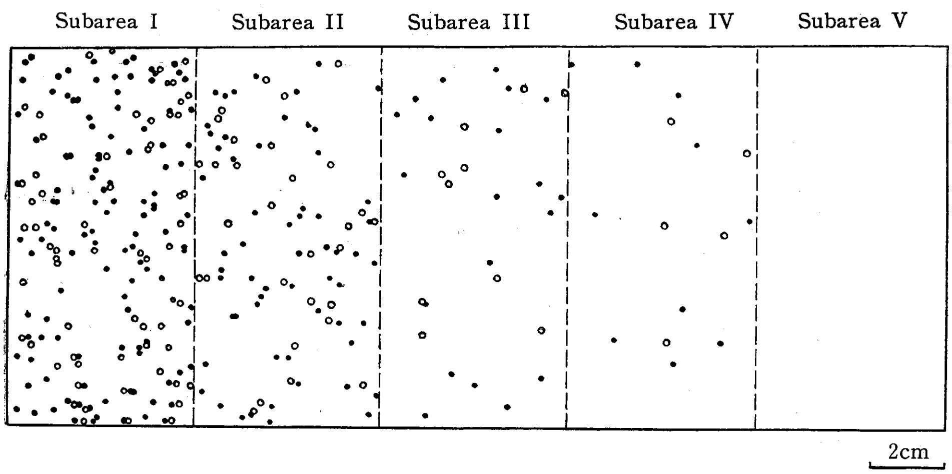


Fig. 1. Artificial population map (A)

every subarea mapped in Fig. 1 is shown. Another example of $C_\delta$ values computed from the samples of only five quadrats, each of which was taken from each subarea in Fig. 1 is given in Table 2. In these examples, it can be seen that $C_\delta$ is affected neither by the sizes of $N_x$ and $N_y$ nor by the number of quadrats.

Then it may be said that $C_\delta$ is an appropriate index of interspecific overlapping which takes the value of about 1 when the ratios of the density of species X to that of species Y are not different for different subareas and the value of zero when no overlapping is found between the two species.

**Table 2.** The values of $C_\delta$ and $R_\delta$ from the samples of five quadrats (2.5×2.5 cm), each taken from each subarea, for the populations mapped in Fig. 1.

| $N_x$ | $N_y$ | Number of quadrats | Maximum number of individuals occurring in a quadrat | | $C_\delta$ | $R_\delta$ |
|---|---|---|---|---|---|---|
| | | | $n_x$ | $n_y$ | | |
| 29 | 10 | 5 | 15 | 5 | 1.024 | +0.453 |
| 28 | 13 | 5 | 18 | 9 | 1.049 | +0.617 |
| 22 | 16 | 5 | 15 | 7 | 0.956 | +0.504 |
| 30 | 15 | 5 | 17 | 7 | 1 015 | +0.538 |
| 19 | 11 | 5 | 8 | 7 | 0.965 | +0.522 |
| 27 | 13 | 5 | 14 | 7 | 1.068 | +0.615 |
| 29 | 12 | 5 | 18 | 9 | 1.025 | +0.628 |
| 26 | 15 | 5 | 15 | 10 | 0.908 | +0.497 |
| Average | | | | | 1.001 | +0.547 |

2. *Index of interspecific correlation*

If the individuals of two species are distributed independently to each other, the following relation may be expected at least when $q$ is large, irrespective as to whether each distribution is at random or contagious:

$$\sum_{i=1}^{q} n_{xi}\, n_{yi} \doteqdot q\, \bar{n}_x \bar{n}_y = \frac{1}{q} N_x N_y. \tag{9}$$

Putting the $C_\delta$ value of this case as $W_\delta$, we have

$$W_\delta = \frac{2}{(\delta_x+\delta_y)\, q}\,. \tag{10}$$

Then, if we put

$$R'_\delta = C_\delta - W_\delta = \frac{2}{(\delta_x+\delta_y)} \left( \frac{\sum_{i=1}^{q} n_{xi}\, n_{yi}}{N_x N_y} - \frac{1}{q} \right) \tag{11}$$

$R'_\delta$ will express the degree of interspecific correlation taking positive values when the distributions of both species are positively correlated and negative values for the negative correlations. If the two species are distributed independently of each other, $R'_\delta$ will take the value of zero. These relations will be clearly indicated by the following formula:

$$R'_\delta=\frac{2(q\sum n_x n_y-N_x N_y)}{q(\delta_x+\delta_y)N_x N_y}=\frac{2\sqrt{\sum(n_x-\bar{n}_x)^2\sum(n_y-\bar{n}_y)^2}}{(\delta_x+\delta_y)N_x N_y}r \tag{12}$$

where $r$ is the correlation coefficient.

When the examined area is enlarged infinitively beyond the range within which both of the two species are distributed, $W_\delta$ will approximate zero, because $q$ will get infinitively large value notwithstanding $\delta_x$ and $\delta_y$ remain constant and accordingly $R'_\delta$ approximates $C_\delta$.

When no quadrat of joint occurrence of the two species is found, the value of $R'_\delta$ is $-W_\delta$ which is usually larger than $-0.5$. However, it may be desirable that the index which measures the interspecific correlation takes the value of $-1$ in such a case. Then, as the index of interspecific correlation, the following $R_\delta$ may be appropriate:

When $R'_\delta \geqq 0$,

$$R_\delta = R'_\delta, \tag{13}$$

and when $R'_\delta < 0$

$$R_\delta=\frac{R'_\delta}{W_\delta}=\frac{q\sum_{i=1}^{q} n_{xi}\, n_{yi}}{N_x N_y}-1. \tag{14}$$

It is evident that $R_\delta$, as well as $C_\delta$, is almost free from the sizes of $N_x$ and $N_y$ (Table 1, 2), and it can measure the correlation between two species irrespective of their distribution types whether they are normal or not, excepting the case that either or both of the two species are distributed uniformly over the whole area ($q\delta\,(=I_\delta)<1$). Furthermore, $R_\delta$ reflects the absolute differences of density ratios of the two species among the subareas which are not measured by the correlation coefficient. For example, though the density ratios of the hypothetical two populations shown in Table 3 are not much different among three quadrats ($P_{\chi^2}>0.3$), each population being distributed almost randomly as is indicated by the $I_{\delta x}$ and $I_{\delta y}$ values, the correlation coefficient takes the value of $-1$, while the $R_\delta$ value is ap proximately zero reflecting exactly the density ratio relationship between the two species.

**Table 3.** The values of $R_\delta$ and the correlation coefficient ($r$) applied to two hypothetical populations.

| Species | Number of individuals | | | | $I_\delta$ * | $C_\delta$ ** | $R_\delta$ *** | $r$ |
|---|---|---|---|---|---|---|---|---|
| | Quadrat | | | Total | | | | |
| | i | ii | iii | | | | | |
| X | 50 | 55 | 45 | 150 | 0.9915 | 1.003 | −0.006 | −1.000 |
| Y | 40 | 36 | 44 | 120 | 0.9900 | | | |
| $P_{\chi^2}$ | > 0.3 | | | | | | | |

*, **, ***

$$\delta_x=\frac{50\times 49+55\times 54+45\times 44}{150\times 149}=0.3305$$

$$\delta_y=\frac{40\times 39+36\times 35+44\times 43}{120\times 119}=0.3300$$

$$C_\delta=\frac{2\times(50\times 40+55\times 36+45\times 44)}{(0.3305+0.3300)\times 150\times 120}=1.003$$

$$W_\delta=\frac{2}{(0.3305+0.3300)\times 3}=1.009$$

$$R'_\delta=C_\delta-W_\delta=-0.006$$

$$R_\delta=\frac{-0.006}{1.009}=-0.006$$

$$I_{\delta_x}=3\times 0.3305=0.9915$$

$$I_{\delta_y}=3\times 0.3300=0.9900$$

*3. Comparison of $C_\delta$ and $R_\delta$ with other indices.*

For the tests of reliability of indices, an artificial population map of $10\times 25$ cm$^2$ having five subareas in it was made (Fig. 1), and on each of the subareas two kinds of points representing the individuals of two species with the density ratio 2:1 were plotted at random utilizing the random numbers table, the number of individuals contained in each subarea being as:

| | subarea | | | | |
|---|---|---|---|---|---|
| | I | II | III | IV | V |
| Species X | 120 | 60 | 20 | 10 | 0 |
| Species X | 60 | 30 | 10 | 5 | 0 . |

Then the whole area was subdivided into 0.25, 1, 2, 6.25 and 50 (=one subarea) cm$^2$ quadrats, and the values of indices hitherto proposed by several authors and of $C_\delta$ and $R_\delta$ were computed for each size of quadrats, utilizing either the frequencies of presence and absence of the two species or the numbers of individuals of both species occurring in each quadrat. The results are given in Table 4 (A), indicating that the indices which have been commonly used by the ecologists are much influenced by the average density per quadrat, taking quite different values at high and low densities, while the values of $C_\delta$ and $R_\delta$ are almost fixed in spite of much difference among the densities per quadrat.

**Table 4.** The values of $C_\delta$, $R_\delta$ and several other indices for the populations mapped in Fig. 1 and Fig. 2

| | Size of quadrat (cm²) | Number of quadrat | Average number of individuals per quadrat | | Degree of interspecific overlapping | | | | | Degree of interspecific correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | sp. X | sp. Y | Dice (1945) X/Y | Dice (1945) X/Y | Whittaker (1952) | Bray (1956) | $C_\delta$ | Forbes (1907) | Cole (1949) | Nash (1950) | De Vries (1954) | $R_\delta$ |
| | 0.25 | 1,000 | 0.210 | 0.105 | 0.344 | 0.185 | 0.200 | 0.241 | 1.095 | 1.981 | +0.204 | +0.145 | +0 269 | +0.592 |
| A | 1.00 | 250 | 0.840 | 0.420 | 0.731 | 0.462 | 0.429 | 0.566 | 0.897 | 1.723 | +0.534 | +0.372 | +0.468 | +0.464 |
| (Fig. 1) | 2.00 | 125 | 1.680 | 0.840 | 0.863 | 0.667 | 0.633 | 0.752 | 1.081 | 1.634 | +0.709 | +0.557 | +0.703 | +0.565 |
| | 6.25 | 40 | 5.250 | 2.625 | 1.000 | 0.839 | 0.833 | 0.912 | 1.069 | 1.290 | +1.000 | +0.734 | +0.914 | +0.566 |
| | 50.00 | 5 | 42.000 | 21.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.010 | 1.000 | +1.000 | +1.000 | +1.000 | +0.528 |
| | 1 | 192 | 0.568 | 0.406 | 0.210 | 0.169 | 0.126 | 0.187 | 0.296 | 0.384 | −0.477 | −0.428 | −0.623 | −0.661 |
| B | 2 | 96 | 1.114 | 0.813 | 0.422 | 0.352 | 0.193 | 0.384 | 0.301 | 0.751 | −0.283 | −0.266 | −0.405 | −0.616 |
| (Fig. 2) | 4 | 48 | 2.227 | 1.625 | 0.727 | 0.667 | 0.275 | 0.696 | 0.252 | 0.970 | −0.200 | −0.124 | −0.078 | −0.627 |
| | 16 | 12 | 9.083 | 6.500 | 1.000 | 0.917 | 0.296 | 0.957 | 0.232 | 1.000 | 0.000 | 0.000 | 0.000 | −0.609 |
| Possible range | | | | | 0~1 | 0~1 | | 0~1 | 0~1(±) | 0~ | −1~+1 | −1~+1 | −1~+1 | −1~+1(±) |

Dice (1945), Association index

X/Y .................... $a/(a+b)$

Y/X .................... $a/(a+c)$

Whittaker (1950), Association of species

$$I_a=\sum \min.\left(\frac{n_x}{N_x},\ \frac{n_y}{N_y}\right)$$

Bray (1956), Index of amplitudinal correspondence

$$C=\frac{2a}{(a+b)+(a+c)}$$

Forbes (1907), Coefficient of association

$$=\frac{a(a+b+c+d)}{(a+b)(a+c)}$$

Cole (1949), Index of interspecific association

$ad \geqq bc$ $\quad C=\dfrac{ad-bc}{(a+b)(b+d)}$

$bc > ad,\ d \geqq a$ $\quad C=\dfrac{ad-bc}{(a+b)(a+c)}$

$a > d$ $\quad C=\dfrac{ad-bc}{(b+d)(c+d)}$

Nash (1950), Pearson's ***phi*** coefficient

$$\varphi = \frac{ad-bc}{\sqrt{(a+b)(b+c)(c+d)(d+a)}}$$

De Vries (1954), Quantile correlation coefficient

$$r = -0.6 \log \frac{bc}{ad} \quad (r < 0.75)$$

or $r = \sin (90^\circ \times \varphi)$,

where the numbers of quadrats of presence and absence of Sp. X and Sp. Y are represented as:*

| | | Sp. X + | Sp. X − |
|---|---|---|---|
| Sp. Y | + | $a$ | $b$ |
| | − | $c$ | $d$ |

$(a+c > a+b)$

---

* The theoretical values of a, b, c, d are given as follows:

| | | Sp. X + | Sp. X − | Total |
|---|---|---|---|---|
| Sp. Y | + | $q\sum p_i(1-e^{-m_{xi}})(1-e^{-m_{yi}})$ | $q\sum p_i e^{-m_{xi}}(1-e^{-m_{yi}})$ | $q(1-\sum p_i e^{-m_{yi}})$ |
| | − | $q\sum p_i(1-e^{-m_{xi}})e^{-m_{xi}}$ | $q\sum p_i e^{-m_{xi}} e^{-m_{yi}}$ | $q\sum p_i e^{-m_{yi}}$ |
| Total | | $q(1-\sum p_i e^{-m_{xi}})$ | $q\sum p_i e^{-m_{xi}}$ | $q$ |

where $p_i$ = the proportion of the size of the $i$-th subarea

$m_{xi}$ = the average number of individuals of sp. X per quadrat in the $i$-th subarea

$m_{yi}$ = the average number of individuals of sp. Y per quadrat in the $i$-th subarea

$q$ = the total number of quadrats

For example, the value of Cole's index for positive association is given as:

$$C = \frac{\sum p_i e^{-m_{xi}} e^{-m_{yi}} - \sum p_i e^{-m_{xi}} \sum p_i e^{-m_{yi}}}{(1-\sum p_i e^{-m_{yi}}) \sum p_i e^{-m_{xi}}} .$$

It is evident that the value is influenced by the sizes of $m_{xi}$ and $m_{yi}$ though $m_{xi}/m_{yi}$ is constant. For instance, the theoretical $C$ value for 0.25 cm² quadrats in Fig. 1 is +0.194 while the value for 2 cm² quadrats is +0.638.

Another artificial population map in which two populations are negatively associated is shown in Fig. 2. As the two populations change their densities gradually from one side of the map to opposite side, the individuals found in any part which is not small in size though not very large of the map, may be considered to be distributed rather randomly, and in such a case $C_\delta$ and $R_\delta$ are expected to take almost fixed values for the quadrat samples of different sizes unless very large quadrats are used. The values of indices for 1, 2, 4 and 16 cm² quadrats are given in Table 4 (B) in which satisfactory results are seen in $C_\delta$ and $R_\delta$ values while the values of other indices are much different for the quadrats of different sizes being influenced by the differences of average number of individuals per quadrat.

Therefore, $C_\delta$ and $R_\delta$ may be used as the most reliable indices at present at least when density measurements can be utilized for measuring the interspecific association.

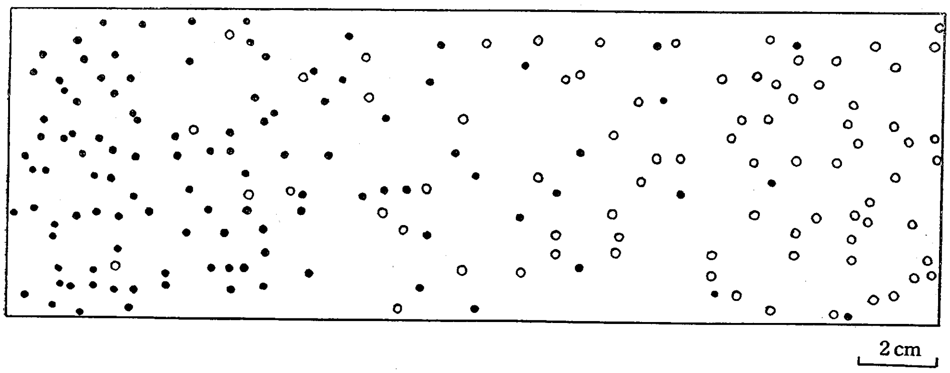


Fig. 2. Artificial population map (B)

## Similarity between communities

Substituting the number of individuals of each species found in a sample taken from a community for the number of individuals in each quadrat, we have, instead of $\delta$, the following Simpson's measure of diversity (Simpson, 1949):

$$\lambda = \frac{\sum n(n-1)}{N(N-1)} \tag{15}$$

where $N$ is the total number of individuals sampled.

Putting the $\lambda$ values of two samples as

$$\lambda_1 = \frac{\sum_{i=1}^{\infty} n_{1i}(n_{1i}-1)}{N_1(N_1-1)}$$

and

$$\lambda_2 = \frac{\sum_{i=1}^{\infty} n_{2i}(n_{2i}-1)}{N_2(N_2-1)},$$

where $\lambda_1$ and $\lambda_2$, $N_1$ and $N_2$ are $\lambda$ and $N$ of the sample I and II, and $n_{1i}$ and $n_{2i}$ are the numbers of individuals of species $i$ found in the sample I and II respectively, we may have

$$C_\lambda = \frac{2\sum_{i=1}^{\infty} n_{1i}\, n_{2i}}{(\lambda_1+\lambda_2)\, N_1\, N_2} \tag{16}$$

as an index of similarity between samples or between communities which, as well as $C_\delta$, is almost uninfluenced by the sizes of $N_1$ and $N_2$ unless either or both of $N_1$ and $N_2$ are small. The value of $C_\lambda$ will be about 1 when the two samples belong to same community and will be zero when no common species is found between them.

For the test of reliability of the index values, an artificial community including 499 species with the total number of 51,012 individuals having a Preston type (log-normal) distribution was constructed with paper cards, and sets of two samples of equal size were randomly taken from this community. The results of applying the $C_\lambda$ index and several other indices of community similarity to these sets of samples are given in Table 5.

As is seen in Table 5, the indices hitherto proposed are much affected by the sample size,[1] while $C_\lambda$ keeps almost fixed value for different sizes of $N$ indicating that it may be used as an appropriate index for measuring the similarity between samples or communities.

## Comparison of many species or many samples

When $N_x = N_y = N_z = \ldots = N$, $C_\delta$ and $R_\delta$ for species X, Y, Z, ... are computed as follows:

$$C_\delta = \frac{\sum_{i=1}^{q} T_i^2 - \sum_{i=1}^{q}(n^2_{xi}+n^2_{yi}+n^2_{zi}+\ldots)}{(s-1)(\delta_x+\delta_y+\delta_z+\ldots)\, N^2} \tag{17}$$

$$R_\delta = R'_\delta \qquad (R'_\delta \geqq 0) \tag{18}$$

$$R_\delta = R'_\delta / W_\delta \quad (R'_\delta < 0), \tag{19}$$

where

[1] The only one index which is free from the effect of sample size may be that of Williams (1947). However the use of his index is confined to the case that the distributions of the number of individuals in both samples follow the logarithmic series with equal parameter value.

**Table 5.** The values of $C_{\lambda}$ and several other indices for the samples taken from an artificial community.

| $N_1 : N_2$ | Number of sampling | Average number of species | | Average index value* | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | in a set of samples | common to both samples | Jaccard (1901) | S$\phi$rensen (1948) | Odum (1950) | Whittaker (1952) | Correlation coefficient (Motomura, 1935) | $C_{\lambda}$ |
| 50 : 50 | 3 | 82.0 | 7.0 | 0.086 | 0.158 | 0.153 | 0.153 | −0.550 | 1.000 |
| 100 : 100 | 3 | 127.7 | 23.3 | 0.184 | 0.310 | 0.297 | 0.297 | −0.225 | 0.906 |
| 500 : 500 | 1 | 316 | 144 | 0.456 | 0.626 | 0.554 | 0.554 | +0.321 | 1.046 |
| Expected value | | | | 1.000 | 1.000 | 1.000 | 1.000 | +1.000 | 1.000 |
| Possible range | | | | 0~1 | 0~1 | 0~1 | 0~1 | −1~+1 | 0~1(±) |

Jaccard (1901), Coefficient de communauté

$$=\frac{c}{a+b-c}$$

S$\phi$rensen (1948), Quotient of similarity between two populations

$$qs=\frac{2c}{a+b}$$

Odum (1950), Percentage similarity

$$=\frac{2\sum \min.(n_1, n_2)}{N_1+N_2}$$

Whittaker (1952), Association of samples

$$=\sum \min.\left(\frac{n_1}{N_1}, \frac{n_2}{N_2}\right),$$

where

$a$, $b$=number of species in the population I, II.

$c$=number of species common to the two populations.

* Indices of Gleason (1920), Kulczynski (1927), Raabe (1952), Clausen (1957) and Barkman (1958) are not compared in this table as they are weighted by frequency, cover or by constancy and their values for this sampling data can not be computed.

$$R'_\delta = C_\delta - W_\delta$$

$$W_\delta = \frac{s}{(\delta_x + \delta_y + \delta_z + \dots)\,q}$$

$s$ = number of species compared

$T_i$ = total number of individuals found in the $i$-th quadrat

$q$ = number of quadrats

$N$ = total number of individuals of each species.

Similarly, when $N_1 = N_2 = N_3 = \dots = N$, $C_\lambda$ is computed as follows:

$$C_\lambda = \frac{\sum_{i=1}^{\infty} V_i^2 - \sum_{m=1}^{h} \sum_{i=1}^{\infty} n^2_{mi}}{(h-1)(\lambda_1 + \lambda_2 + \dots + \lambda_h)}, \tag{20}$$

where

$h$ = number of samples compared

$V_i$ = total number of individuals of species $i$

$n_{mi}$ = number of individuals of species $i$ found in the $m$-th sample

$N$ = total number of individuals in each sample.

The formulae in the cases of $N_x \neq N_y \neq N_z \neq \dots$ and $N_1 \neq N_2 \neq N_3 \neq \dots$ are not yet obtained.

## Utilization of the quantities other than density

$C_\delta$, $R_\delta$ and $C_\lambda$ mentioned above are based on the density measurements. However, it may be desirable that the degree of interspecific association or of similarity between communities utilizing the quantities other than density is measured by similar method. In this respect, the following formulae may be useful for the ecological works though critical studies on these formulae will still be needed from the mathematical point of view.

### 1. *Weight and basal area*

If a quantity, the total of which changes in proportion to the change of the number of individuals, such as weight or basal area, is utilized, the index of similarity between community may be given as follows:

$$C_{\lambda(W)} = \frac{2\sum_{i=1}^{\infty} \omega_i^2\, n_{1i}\, n_{2i}}{(\lambda_{(W)1} + \lambda_{(W)2})\, W_1 W_2}, \tag{21}$$

where

$$\lambda_{(W)1} = \frac{\sum_{i=1}^{\infty} \omega_i^2 n_{1i}(n_{1i}-1)}{W_1^2 - \sum_{i=1}^{\infty} \omega_i^2 n_{1i}}, \quad \lambda_{(W)2} = \frac{\sum_{i=1}^{\infty} \omega_i^2 n_{2i}(n_{2i}-1)}{W_2^2 - \sum_{i=1}^{\infty} \omega_i^2 n_{2i}}$$

$\omega_i$ = average weight (or basal area) per individual of species $i$

$W_1$ = total weight (or basal area) of the sample I

$W_2$ = total weight (or basal area) of the sample II

$n_{1i}$ = number of individuals of species $i$ found in the sample I

$n_{2i}$ = number of individuals of species $i$ found in the sample II.

If the weight of every individual is very small, and the number of individuals is very large, putting $w$ as the total weight of each species and

$$\lambda_{(W)1} = \frac{\sum_{i=1}^{\infty} w_{1i}^2}{W_1^2} \qquad \lambda_{(W)2} = \frac{\sum_{i=1}^{\infty} w_{2i}^2}{W_2^2},$$

we have

$$C_{\lambda(W)} = \frac{2\sum_{i=1}^{\infty} w_{1i} w_{2i}}{(\lambda_{(W)1} + \lambda_{(W)2}) W_1 W_2}. \tag{22}$$

*2. Coverage*

If the quantity to be compared is the coverage, the indices of interspecific association, $C_{\delta(p)}$ and $R'_{\delta(p)}$ will be given as follows:

$$C_{\delta(p)} = \frac{2\sum_{i=1}^{q} p_{xi} p_{yi}}{(\delta_{(p)x} + \delta_{(p)y}) q^2 \bar{p}_x \bar{p}_y}, \tag{23}$$

$$R'_{\delta(p)} = C_{\delta(p)} - \frac{2}{(\delta_{(p)x} + \delta_{(p)y}) q} \tag{24}$$

where

$$\delta_{(p)x} = \frac{\sum_{i=1}^{q} p^2_{xi}}{q^2 (\bar{p}_x)^2}, \quad \delta_{(p)y} = \frac{\sum_{i=1}^{q} p^2_{yi}}{q^2 (\bar{p}_y)^2}$$

$p_{xi}$, $p_{yi}$ = coverage in per cent of species X, Y in the $i$-th quadrat

$\bar{p}_x$, $\bar{p}_y$ = mean coverage of species X, Y

$q$ = number of quadrat.

The index of similarity between communities, $C_{\lambda(p)}$, will be

$$C_{\lambda(p)} = \frac{2\sum_{i=1}^{\infty} p_{1i} p_{2i}}{(\lambda_{1(p)} + \lambda_{2(p)}) \sum_{i=1}^{\infty} p_{1i} \sum_{i=1}^{\infty} p_{2i}}, \tag{25}$$

where

$$\lambda_{1(p)} = \frac{\sum_{i=1}^{\infty} p_{1i}^2}{(\sum_{i=1}^{\infty} p_{1i})^2}, \quad \lambda_{2(p)} = \frac{\sum_{i=1}^{\infty} p_{2i}^2}{(\sum_{i=1}^{\infty} p_{2i})^2}$$

$p_{1i}$, $p_{2i}$ = coverage in per cent of species $i$ in the sample I, II.

## Conclusion and Summary

For measuring the interspecific association or similarity between communities, finding out of such index that is free from the effects of sample size or of average density per quadrat might be said having been a pending problem in ecology. The indices, $C_\delta$, $R_\delta$ and $C_\lambda$, described in this paper may settle this question at least to some extent. As these indices are almost uninfluenced either by the average number of individuals per quadrat or by the number of quadrats and are applicable to any type of contagious distributions of individuals, one may safely analyse the relation between environmental factors and interspecific association using the quadrats of different sizes (Greig-Smith, 1957: p. 96) or may correctly determine the relative position of stands in the ordination study of communities (Bray and Curtis, 1957) through the use of these indices.

Though the indices were deviced on the basis of density measurement, utilization of other quantities, weight, basal area, coverage and etc., is also available. However, there are some difficulties in direct use of frequency measurement for the indices, and finding of the suitable method for treating the frequency measurement, and of the method for the comparison of many species or many samples with different number of total individuals are left in the future investigations.

The author is deeply indebted to Prof. T. Kitagawa, Department of Mathematics, Kyushu University, for his kind advice and through reading of the manuscript. Much appreciation is also owed to Miss. K. Motomura for her assistance in making and sampling of the artificial community.

### Literature Cited

**Barkman, J. J.** 1958. On the ecology of cryptogamic epiphytes. Leiden.

**Bray, J. R.** 1956. A study of mutual occurrence of plant species. Ecol. **37**: 21-28.

**Bray, J. R.** and **Curtis, J. T.** 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. **27**: 325-349.

**Clausen, J. J.** 1957. A phytosociological ordination of the conifer swamps of Wisconsin. Ecol. **38**: 638-645.

**Cole, LaMont C.** 1949. The measurement of interspecific association. Ecol. **30**: 411-424.

**De Vries, D. M.** 1954. Constellation of frequent herbage plants based on their correlation in occurrence. Vegetatio **5**: 105-111.

**Dice, L. R.** 1945. Measures of the amount of ecologic association between species. Ecol. **26**: 297-302.

**Forbes, S. A.** 1907. On the local distribution of certain Illinois fishes: an essay in statistical ecology. Bull. Ill. Lab. Nat. Hist. **7**: 273-303.

**Gleason, H. A.** 1920. Some applications of the quadrat method. Torry Bot. Club Bull. **47**: 21-33.

**Greig-Smith, P.** 1957. Quantitative plant ecology. London.

**Jaccard, P.** 1901. Étude comparative de la distribution florale dans une portion des Alpes et du Jura. Bull. Soc. Vaudoise des Sc. Nat. **37**: 547-579.

**Kulczyński, S.** 1927. Zespoly roślin w Pieninach. Internat. Acad. Polon. Sci. Lettr. Bull., Classe Sci. Math. et Nat., ser. B. Sci. Nat. Suppl. **2**: 1927: 57-203.

**Morisita, M.** 1959. Measuring of the dispersion of individuals and analysis of the distributional patterns. Mem. Fac. Sci. Kyushu Univ., Ser. E (Biol.) **2**: 215-235.

**Motomura, I.** 1935. Use of correlation coefficient in the statistics of animal community (in Japanese.) Ecol. Rev. (Sendai) **1**: 339-342.

**Nash C. B.** 1950. Associations between fish species in tributaries and shore waters of western Lake Erie. Ecol. **31**: 561-566.

**Odum, E. P.** 1950. Bird populations of the highlands (North Carolina) plateau in relation to plant succession and avian invasion. Ecol. **31**: 587-605.

**Raabe, E. W.** 1952. Über den "Affinitätswert" in der Pflanzensoziologie. Vegetatio **4**: 53-68.

**Simpson, E. H.** 1949. Measurement of diversity. Nature **163**: 688.

**Sφrensen, Th.** 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. D. Kgl. Dansk. Vidensk. Selsk. Biol. Skr. **5**: 1-34.

**Whittaker, R. H.** 1952. A study of summer foliage insect communities in the Great Smoky Mountains. Ecol. Monogr. **22**: 1-44.

**Williams, C. B.** 1947. The logarithmic series and the comparison of island floras. Proc. Linn. Soc. London, Session 158 (1945-46), Pt. **2**: 104-110.