

SOME PROBLEMS OF MODEL-BUILDING IN SYN BIOLOGY PART 2

Associatum process in a simple situation

By

P. JUHÁSZ-NAGY* - I. DÉVAI** - K. HORVÁTH***

Dept. of Pl. Taxonomy and Ecology of the L. Eötvös University, Budapest*;

Hydrological Works, Debrecen**; Agricultural School, Debrecen***

Received on the 30th, May, 1972

I. Introduction

I. 1 The present paper is a continuation of the first part of this series (Dévai - Horváth - Juhász - Nagy, 1971, quoted here in brief as "P. I."), where a new type of diversity ("faunal diversity") and its trend of spatial change has been described with respect to the *Cladocera* fauna (of 10 taxa) living in a fish-pond (located in the central part of the Great Hungarian Plain). This text refers to the same object, basic concepts, sampling procedure, data etc. as P. I. do. Thus an attempt will be made to reduce repetition to a minimum.

I. 2 Now we are here concerned with *association*, more correctly, association among populations ("interlocal similarity" in our terms; cf. P. I.: IV. 1). There is no need to enter here into details; many excellent references are easily available (e. g. Goodall 1952; Andrewartha - Birch 1954; Dagnelie 1960, 1965; Greig - Smith 1964; Kershaw 1964; Lambert - Dale 1964; Pielou 1969, etc.). It is fairly well-known that the vast majority of studies on association (in particular, in recent years) is motivated by the problems of ordination of living communities. Much work has been done mostly on sorting algorithmus based on independence, both on interlocal and interfloral (-faunal) similarity relations (cf. Williams - Dale 1965; Macnaughton - Smith 1965; McIntosh 1967; Aleksandrova 1969; etc.). In spite of this - or perhaps because of it - some rather important problems (being outside the realm of the present-day ordination practice) seem to be virtually neglected. Consider only two of them.

I. 2. 1 Our knowledge of *association process in space* (roughly, change of association in space) is extremely limited. Greig - Smith's warning (1964, p. 105: „The effect of sampling size on indications of association has been largely overlooked. .”) is still apt to make. Clearly, any statement of the sign, number or strength of association measures is

related to a point (or an interval) of a spatial process, if only one definite size of sampling units was used.

I. 2. 2 We know next to nothing about the spatial changes of *partial* or *multiple association measures* among 3, 4, . . . populations. Except a very limited number of papers (e. g. Cole 1957; J u h á s z — N a g y 1967) almost all studies are concerned with pairwise association which — one has the feeling — is somewhat “unnatural”. A rather “more natural” measure would be *total association* of a faunal or floral universe (cf. P. I.).

I. 3 Considerations such as those just mentioned suggest that it would be useful to know something about the trend of spatial changes of total association. This paper is aimed at showing a simple modeling of such a problem.

Associatum is defined to be the total association of a faunal universe (faunal set of interest). By *associatum process* a spatial process (of faunistic nature) is meant (by means of a discrete approximation), excl. temporal and grouping processes. *Associatum* is expressed in terms of information theory (more precisely, in terms of information *sensu* S h a n n o n). There is again no need to enter into details; the reader can consult some excellent references (e. g. W i l l i a m s — L a m b e r t 1966; P i e l o u 1969; O r l o c i 1971; D a l e 1971) and their bibliographies.

II. Modeling

II. 1 Let Q_T be a faunal universe (as a set universe of populations) of a given object (*Cladocera* fauna of a pond, in this case), $Q_T = \{q_1, q_2, \dots, q_e, \dots, q_s\}$. Let, further, V_x be a class of sampling units (volumes), according to an X „increasing algorithm”, $V_x = \{V_1, V_2, \dots, V_j, \dots, V_u\}$, where, in our case, V_j is a set of sampling units of 2^j ml. size, $V_j = \{v_{j1}, v_{j2}, \dots, v_{jg}, \dots, v_{jm}\}$; $u = 8$ (the largest sampling size is 256 ml.) and $m = 64$ (all V_x set elements consists of 64 volumes of a given size randomized over the surface of the pond).

Let us consider the Cartesian product of Q_T and V_j , $Q_T \times V_j$, and a binary relation, r_{jeg} , called *intersection*, defined between each ordered pair $\langle q_e, v_{jg} \rangle$,

$$r_{jeg} = \begin{cases} 1, & \text{if } q'_e \cap v_{jg} \neq \emptyset \\ 0, & \text{if } q'_e \cap v_{jg} = \emptyset \end{cases} \quad (\text{II}; 1)$$

where q'_e stands for the set of body points of q_e . Let $Q_T \times V_j$ be transformed into a U_j binary contingency table (by permitting summation in marginals) so that

$$\sum_{g=1}^m r_{jeg} = r_{je1}, \quad \sum_{e=1}^s r_{jeg} = r_{1jg}, \quad \sum_{g=1}^m \sum_{e=1}^s r_{jeg} = N_j, \quad (\text{II}; 2)$$

and let

$$m - r_{jel} = r_{je0}, \quad s - r_{l jg} = r_{0jg}, \quad ms - N_j = n_j, \quad (\text{II}; 3)$$

where the following terminology (cf. P. I. in detail) is adopted:

$$\begin{array}{ll} r_{jel} \text{ (local valence)} & r_{l jg} \text{ (faunal valence)} \\ r_{je0} \text{ (local invalence)} & r_{0jg} \text{ (faunal invalence)} \\ N_j \text{ (total valence)} & n_j \text{ (total invalence).} \end{array}$$

Let II_Q be faunal power set of Q_T , $II_Q = \{\pi_0, \pi_1, \dots, \pi_k, \dots, \pi_w\}$, $w = 2^s - 1$, where π_k is a potential fauna (any subset of Q_T), and let f_{jk} be the frequency of π_k and F_j the frequency distribution of π_k 's in U_j . Faunal diversity is characterized in P. I. by the function

$$m\hat{H}_j = m \log m - \sum_{k=0}^w f_{jk} \log f_{jk} \quad (\text{II}; 4)$$

For the sake of convenience, let

$$\sum_{k=0}^w f_{jk} \log f_{jk} = G_j \quad (\text{II}; 5)$$

$$\sum_{e=1}^s r_{jel} \log r_{jel} = A_j \quad (\text{II}; 6)$$

$$\sum_{e=1}^s r_{je0} \log r_{je0} = a_j \quad (\text{II}; 7)$$

$$\sum_{g=1}^m r_{l jg} \log r_{l jg} = B_j \quad (\text{II}; 8)$$

$$\sum_{g=1}^m r_{0jg} \log r_{0jg} = b_j \quad (\text{II}; 9)$$

$$\left. \begin{array}{l} N_j \log N_j = C_j \\ n_j \log n_j = c_j \end{array} \right\} \quad (\text{II}; 10)$$

$$\left. \begin{array}{l} A_j + a_j = \alpha_j \\ B_j + b_j = \beta_j \\ C_j + c_j = \gamma_j \end{array} \right\} \quad (\text{II}; 11)$$

The simple concepts and estimators above are preliminary ones to the following reasoning.

II. 2 Association ("simple", pairwise association) is interpreted between *pairs of local vectors* of U_j (cf. P. I.) by a 2×2 scheme, whose molecular events are: ($<11>$, $<10>$, $<01>$, $<00>$). Comparing 3 local vectors begins with attaching frequencies to molecular events ($<111>$, $<110>$, $<101>$, ..., $<000>$). In the case if all the local vectors of U_j are to be compared, we have a $2 \times 2 \times \dots \times 2$ table with 2^s molecular events representing all ordered s -tets possible (all subsets of Q_T). Let us

call such a table K_j with respect to U_j . Frequency distribution of K_j clearly corresponds to F_j (because events of K_j are equivalent with that of Π_Q) and marginals of K_j correspond to local valence and invalence sets of U_j . Thus a weighted estimate of associatum expressed as multipel information has the form (cf. Kullback, 1959):

$$m\hat{I}_j[\lambda] = G_j - \alpha_j + (s-1)m \log m. \quad (\text{II}; 12)$$

With a slight change of quantities in (II; 12)

$$m\hat{I}_j[\lambda] = (sm \log m - \alpha_j) - (m \log m - G_j), \quad (\text{II}; 13)$$

where

– the first term,

$$ms\hat{I}_j(v, [q, i]) = sm \log m - \alpha_j, \quad (\text{II}; 14)$$

called *local distinction*, requires some further considerations, and

– the second term corresponds to $m\hat{H}_j$ of (II; 4),
faunal diversity.

Thus, associatum (total association) is separable into two additive components, meaningful in themselves:

$$m\hat{I}_j[\lambda] = ms\hat{I}_j(v, [q, i]) - m\hat{H}_j \quad (\text{II}; 15)$$

II. 3 Let us examine *local distinction* more closely. Such quantity as $ms\hat{I}_j(v, [q, i])$ is best interpreted as one related to *preferential relations* (cf. Quastler, 1953; Juhász — Nagy, 1964). The positions of I 's in U_j show how certain populations prefer certain volumes (of j -size) to others, and *vice versa*, according to the attribute duality principle of the Southampton group (cf. e. g. Lambert — Dale, 1964). Preferential relations may be first approached by a conceptual triplet (choice — selection — preference).

a) A random *choice* of an ordered pair in U_j is characterized by (II; 1).

b) Random *selection* (as a repetitive choice) of a particular object is characterized by a corresponding pair of valence and invalence value.

c) *Preference* (of some kind) is characterized by some function defined on the frequency distributions of valences and invalences.

Let us define three abstract *partners* in our selective situation, an average *population* (q), an average sampling *volume* of a fixed size (v), and an average *intersection* (i) related to that size. All partners are endowed with selective property. Namely,

- q selects volumes from the elements of V_j ,
- v selects populations from the elements of Q_T ,
- i selects ordered pairs from the set of ms ordered pairs of a U_j .

Let us call a “positive random choice” in (II; 1) a *coincidence* ($r_{jeg} = 1$). Assume a classical probability field (cf. K h i n c h i n, 1957), equiprobability of coincidences, i. e. $p(q) = 1/s$, $p(v) = 1/m$, $p(i) = 1/ms$. Thus, in terms of information theory, we have

$$\left. \begin{aligned} H(q) &= \log m \\ H(v) &= \log s \\ \widehat{H}_j(i) &= \log ms - \gamma_j/ms, \end{aligned} \right\} \quad (\text{II}; 16)$$

some uncertainty measures of the Hartley type, and further, for joint entropy measures, we have

$$\left. \begin{aligned} \widehat{H}_j([q, i]) &= \log ms - \alpha_j/ms \\ \widehat{H}_j([v, i]) &= \log ms - \beta_j/ms \\ H([q, v]) &= \log ms \\ H([q, v, i]) &= \log ms \end{aligned} \right\} \quad (\text{II}; 17)$$

It is fairly easy to see that having (II; 16) and (II; 17) a whole system of information functions

$$\widehat{I}_j(q, i) = \log s + \alpha_j/ms - \gamma_j/ms \quad (\text{II}; 18)$$

$$\widehat{I}_j(v, i) = \log m + \beta_j/ms - \gamma_j/ms \quad (\text{II}; 19)$$

$$\vdots$$

$$\widehat{I}_j([q, v, i]) = \log ms - \alpha_j/ms - \beta_j/ms \quad (\text{II}; 20)$$

can be obtained, where each function has a definite bearing on a certain aspect of preferential relations. Unfortunately, owing to a lack of space, without referring to some important properties of these information quantities, we must confine ourselves to our special function of (II; 14).

III. 3.1 Now, from a formal point of view, this quantity can be expressed either as multiple information of partner v with respect to the joint finite scheme of the other partners, $[q, i]$,

$$\begin{aligned} \widehat{I}_j(v, [q, i]) &= H(v) - \widehat{H}_j(v|[q, i]) = H(v) - H([q, v, i]) + \widehat{H}_j([q, i]) = \\ &= \log m - \log ms + \log ms - \alpha_j/ms = \log m - \alpha_j/ms, \end{aligned}$$

or as conditional information of joint finite scheme $[v, i]$ with respect to partner q ,

$$\begin{aligned} \widehat{I}_j([v, i]|q) &= H([v, i]|q) - \widehat{H}_j(v|[q, i]) - \widehat{H}_j(i|[q, v]) = \\ &= \log ms - \log s - \log ms + \log ms - \alpha_j/ms - 0 = \log m - \alpha_j/ms. \end{aligned}$$

The equality $\widehat{I}_j(v, [q, i]) = \widehat{I}_j([v, i]|q)$, of course is, a consequence of the equiprobability condition for our model system. These two quantities ought to have been different if basic assumptions were different.

III. 3.2 Coming to a more interpretative view, let us consider first the following additive relation:

$$\begin{aligned}
 (1) \quad ms\hat{I}_j(q, i) &= ms \log s + \alpha_j - \gamma_j \\
 (2) + ms\hat{I}_j(v, [q, i]) &= ms \log m - \alpha_j \\
 (3) \quad ms\hat{I}_j(i, [q, v]) &= ms \log ms - \gamma_j
 \end{aligned}
 \tag{II; 21}$$

(1) in (II; 21), being contingency information of a $2 \times s$ table related to U_j of local valences and invalences, reflects upon the difference among populations in their *local preferences*. On the one hand, if we deal with an equiprobable local valence set (i.e. all local valence values of a U_j is equal to N_j/s , *local density*), then $\hat{I}_j(q, i)$ is zero, and $\hat{I}_j(v, [q, i])$ is of maximum value. $\hat{I}_j(v, [q, i]) = \hat{I}_j(i, [q, v])$. On the other hand, if local valence set has an "unequal" frequency distribution of the highest degree (i.e. N_j/m values are equal to m , and the rest is zero), then $\hat{I}_j(q, i) = \hat{I}_j(i, [q, v])$, and $\hat{I}_j(v, [q, i]) = 0$. These boundary relations have different bearings on a number of synbiological problems. At the moment we are interested here in the following intuitive meaning.

II. 4. Let

$$\begin{aligned}
 M_q^{(j)} &= \{r_{j11}, r_{j21}, \dots\} \\
 M_v^{(j)} &= \{r_{1j1}, r_{1j2}, \dots\} \\
 m_q^{(j)} &= \{r_{j10}, r_{j20}, \dots\} \\
 m_v^{(j)} &= \{r_{0j1}, r_{0j2}, \dots\}
 \end{aligned}
 \tag{II; 22}$$

called *faunistic marginals* (in brief, *f-marginals*) with respect to U_j . We can think of *f-marginals* in pairs, calling them local, faunal, valence and invence marginals, resp. Valence marginals of U_j can be characterized by an ordered triplet of *faunistic mean density values*, $\langle \bar{m}_j, \bar{s}_j, \bar{N}_j \rangle$, where

$$\begin{aligned}
 -\bar{m}_j &\text{ is a mean local density, } \bar{m}_j = N_j/s, \\
 -\bar{s}_j &\text{ is a mean faunal density, } \bar{s}_j = N_j/m, \\
 -\bar{N}_j &\text{ is a mean total density, } \bar{N}_j = N_j/ms,
 \end{aligned}$$

and where "faunistic density" means a "rate of scoring", instead of "number of individuals".

II. 4.1 Let us define the following weighted entropy values (on the *f-marginals* of U_j), referring to marginal uncertainties:

$$\begin{aligned}
 N_j \widehat{H}_j(M_q) &= C_j - A_j \mid \\
 N_j \widehat{H}_j(M_v) &= C_j - B_j \mid
 \end{aligned}
 \tag{II; 23}$$

$$\begin{aligned}
 n_j \widehat{H}_j(m_q) &= c_j - a_j \mid \\
 n_j \widehat{H}_j(m_v) &= c_j - b_j \mid
 \end{aligned}
 \tag{II; 24}$$

Functions of (II; 23) are monotonously increasing, those of (II; 24) are monotonously decreasing with increasing of sampling unit sizes. (We believe that these functions are optimum substituents of "species area functions" used so far.) The best property of these functions is that at each V_j they are bounded by the following inequalities:

$$\begin{aligned} N_j \log \bar{s}_j &\leq N_j \widehat{H}_j(M_q) \leq N_j \log s \\ N_j \log \bar{m}_j &\leq N_j \widehat{H}_j(M_v) \leq N_j \log m \\ n_j \log (s - \bar{s}_j) &\leq n_j \widehat{H}_j(m_q) \leq n_j \log s \\ n_j \log (m - \bar{m}_j) &\leq n_j \widehat{H}_j(m_v) \leq n_j \log m \end{aligned} \quad (\text{II; 25})$$

where *maxima* are due to *monovalence state* (if the proper valence values are all equal to one another, i.e. to the corresponding mean density), and *minima* are due to *oligovalence state* (if a number of valence values reach their maximum and others remain zero). This relation was already commented upon in III. 3.2.

III. 4.2 According to inequalities of (II; 25), either *upper deviates*,

$$\begin{aligned} \max N_j \widehat{H}_j(M_q) - N_j \widehat{H}_j(M_q) &= \Delta N_j \widehat{H}_j(M_q) = N_j \log s - C_j + A_j, \\ \Delta n_j \widehat{H}_j(m_q) &= n_j \log s - c_j + a_j, \\ \Delta N_j \widehat{H}_j(M_v) &= N_j \log m - C_j + B_j, \\ \Delta n_j \widehat{H}_j(m_v) &= n_j \log m - c_j + b_j \end{aligned} \quad (\text{II; 26})$$

or, *lower deviates*,

$$\begin{aligned} N_j \widehat{H}_j(M_q) - \min N_j \widehat{H}_j(M_q) &= \nabla N_j \widehat{H}_j(M_q) = N_j \log m - A_j, \\ \nabla n_j \widehat{H}_j(m_q) &= n_j \log m - a_j, \\ \nabla N_j \widehat{H}_j(M_v) &= N_j \log s - B_j, \\ \nabla n_j \widehat{H}_j(m_v) &= n_j \log s - b_j, \end{aligned} \quad (\text{II; 27})$$

as empirical *gain of information* measures at V_j indicate how *diverse* empirical f -marginals are. In a monovalence state, when there is no marginal diversity, the proper upper deviates become zero. On the other hand, in the same case, proper lower deviates have maximum value. This shows that *marginal diversity (sensu generali)* should be estimated from both extremities at the same time.

III. 4.3 Now, it is rather easy to see that

$$\Delta N_j \widehat{H}_j(M_q) + \Delta n_j \widehat{H}_j(m_q) = ms \log s + \alpha_j - \gamma_j = ms \hat{I}_j(q, i), \quad (\text{II; 28})$$

and

$$\nabla N_j H_j(M_q) + \nabla n_j H_j(m_q) = ms \log m - \alpha_j = ms \hat{I}_j(u, [q, i]) \quad (\text{II}; 29)$$

i.e. the pairwise sums of upper and lower deviates, resp. correspond to the quantities in (II; 21).

III. 4.4 This relation allows us to reformulate (II; 15) by partitioning associatum into three diversity components

$$m \hat{I}_j[\lambda] = \nabla N_j \hat{H}_j(M_q) + \nabla n_j \hat{H}_j(m_q) - m \hat{H}_j, \quad (\text{II}; 30)$$

where the meaning of each component is interpretable and related to a framework of concepts.

III. Illustration and interpretation

III. 1 *Data.* In our case, where (a) Q_T , faunal universe is a set of 10 *Cladocera* taxa ($s = 10$) denoted as (A, B, \dots, K) in *P. I.*, (b) V_j is a set of cubic shaped sampling units ($m = 64$) of 2_j ml. size, (c) $j = 1, 2, \dots, 8$ ($u = 8$), i. e. V_x is a series according to the increasing powers of 2, we have 8 binary contingency tables (U_j) of 10×64 size whose *valence sets* are given by tables 1 and 2.

III. 1.1 Local valence sets are shown by the columns of table 1, where double row symbols (as alternative notation) stand for the elements of Q_T (cf. *P. I.*, II. 2)

Table 1.

		v ₂	v ₂	v ₃	v ₄	v ₅	v ₆	v ₇	v ₈
A	q ₁	6	11	21	51	53	62	64	64
B	q ₂	15	24	36	63	63	64	64	64
C	q ₃	5	10	21	51	55	62	64	64
D	q ₄	1	1	2	3	6	10	42	64
E	q ₅	1	1	3	12	14	26	50	64
F	q ₆	21	39	57	64	64	64	64	64
G	q ₇	4	9	17	21	32	60	64	64
H	q ₈	4	6	24	17	27	60	64	64
I	q ₉	0	0	0	1	7	11	38	64
K	q ₁₀	1	0	1	3	5	12	33	64
		58	101	182	286	326	431	547	640

III. 1.2 In table II. a variable ψ is introduced which has possible values of faunal valences ($\psi = 0, 1, \dots, 10$). Thus, an entry, $k_{j\psi}$, of the first part of table 2 represents frequency of a given value of faunal valences in U_j . Naturally, weighted frequencies of that sort sum up to total valence,

$$\sum_{\psi=0}^8 \psi k_{j\psi} = N_j$$

as it is shown in the second part of table 2, where series of total valences, $N_1 < N_2 < \dots < N_8$, corresponds to that of table I.

Table II.

	V ₁	V ₂	V ₃	V ₄	V ₅	V ₆	V ₇	V ₈										
$\psi = 0$	17	7	0	0	0	0	0	0	V ₁	V ₂	V ₃	V ₄	V ₅	V ₆	V ₇	V ₈		
1	36	29	9	0	1	0	0	0	36	29	9	0	1	0	0	0	0	
2	11	16	17	2	2	0	0	0	22	32	34	4	4	0	0	0	0	
3	0	8	21	12	3	0	0	0	0	24	63	36	9	0	0	0	0	
4	0	4	11	22	15	1	0	0	0	16	44	88	60	4	0	0	0	
5	0	0	4	15	17	5	0	0	0	0	20	75	85	25	0	0	0	
6	0	0	2	8	16	23	3	0	0	0	12	48	96	138	18	0	0	
7	0	0	0	5	9	23	12	0	0	0	0	35	63	161	84	0	0	
8	0	0	0	0	1	6	15	0	0	0	0	0	8	48	120	0	0	
9	0	0	0	0	0	5	15	0	0	0	0	0	0	45	135	0	0	
10	0	0	0	0	0	1	19	64	0	0	0	0	0	10	190	640	0	
	64	64	64	64	64	64	64	64	58	101	182	286	326	431	547	640		

III. 1.3 *Results* based on data of tables I. and II. in relation to our basic function (I; 15) is shown by table III. All quantities are given in weighted *nits* (natural units), i. e. *log. nat.* was used.

We believe that at least "the qualitative trend" of these functions in table III. are clear and interpretable. This view is supported by some considerations presented below and by experience of one of us (*P. J.-N.*) with respect to terrestrial communities; all functions studied so far are similar to that of table III.

III. 2 *Trends*. Examining the trends of functions in table III., our question is how change associatum in value with increasing size of sampling units.

Table III.

	$m\hat{I}_j[\lambda]$	$= msI_j(v, [q, i])$	$m\hat{H}_j$
$j = 1$	20.30	158.26	137.96
2	25.52	198.54	173.02
3	50.35	252.58	202.23
4	30.23	207.63	177.40
5	47.62	241.69	194.07
6	22.53	179.02	156.49
7	16.50	162.44	145.94
8	0.00	0.00	0.00

III. 2.1 It is observable that $m\hat{I}_j[\lambda]$, $j = 1, 2, \dots$ has two "peaks" such that the first one (cca. at 8 ml.) is somewhat greater than the second one (somewhere about 32 ml.). We believe that these two peaks correspond two volume sizes characterized by "the dominance of negative and positive association", resp. We intend to show in a further part of this series (or elsewhere) that this is the case. At the moment we confine our attention to the additive relation of table III., i. e. to the components of associatum.

III. 2.2 We know already (from *P. I.*) the trend of $m\hat{H}_j$ which function has maximum values (somewhere about 8 ml. and 32 ml.), called *maximum volumes*, V_{\max} , where faunal diversity (with respect to Π_Q) or impredecability of F_j is of the highest degree. V_{\max} is contrasted with *minimum volume*, V_{\min} , where (at a size less than or equal to 256 ml.) there is no faunal diversity whatever and faunal predicability is perfect.

III. 2.3 V_{\min} of course, is a concept of general importance. It can be said as axiomatic that all *characteristic functions* of faunistic nature must reach zero value at V_{\min} (since all objects — faunas, loci, etc. — are equivalent faunistically to one another); for instance, all information functions of II. 3 have this property. (Naturally, the phrase "characteristic function" is used here in a synbiological sense of the word, not to be confounded with the well-known term of probability theory.) V_{\max} and V_{\min} are *characteristic volumes*. For a more general definition we can put forward the following one. A characteristic volume is defined by the extreme values (*maxima, minima*) of some characteristic functions with respect to the interval $[V_0, V_{\min}]$. We can guess that there are other important characteristic volumes than V_{\max} , V_{\min} .

Indeed, a new type of characteristic volume, called *compensatory volume*, V_{comp} , is one, where $N_j = n_j = ms/2$, i. e. where total valence and invalence values (or, number of coincidences and incoincidences; cf. II. 3) are — at least, approximately — equal to each other. In our case V_{comp} is at V_5 , where $N_5 = 326$, incidentally quite close to $ms/2 = 320$.

Clearly, V_{comp} need not be equal to V_{max} . (In terrestrial plant communities for maximum area, A_{max} , and compensatory area, A_{comp} , the relation $A_{\text{max}} < A_{\text{comp}}$ is generally true.) Unfortunately, our data are insufficient to make such a distinction, if there is any. But we can infer even from our "weak data" of table III. that a number of characteristic volumes (e. g. *maxima* and *minimum* of associatum) are more or less close in size to V_{comp} . This is so as each characteristic function has a "rising part" (when sampling units are "small") and a "falling part" (toward V_{min} , A_{min} , etc.), and, between these two, at an interval, a "part", where one or several characteristic volumes occur. We may call this interval a *compensatory interval*.

III. 2.4 Function $ms\hat{I}_j(i, [q, v])$ in (II; 21), for instance, has always a unique maximum at V_{comp} , $\max ms\hat{I}_j(i, [q, v]) = ms \log 2$, because its only variable, γ_j , has its own minimum, $\min \gamma_j = ms \log ms - ms \log 2$, at that size of volume. It is worth considering the two types of partition of this quantity mentioned in II. 3.2 - .

III. 2.4.1 First, table IV. shows the double partition of $ms\hat{I}_j(i, [q, v])$ in (II; 21).

Table IV.

	$ms\hat{I}_j(i, [q, v])$	$= ms\hat{I}_j(q, i)$	$+ ms\hat{I}_j(v, [q, i])$
$j = 1$	194.888	36.626	158.262
2	279.396	80.856	198.540
3	382.444	129.865	252.579
4	440.335	232.703	207.632
5	443.842	202.149	241.693
6	404.639	225.621	179.018
7	265.615	103.176	162.439
8	0.000	0.000	0.000

III. 2.4.2 Second, table V. shows the quadruple partition of the same quantity according to III. 4.3. Note that upper local deviates sum up to $ms\hat{I}_j(q, i)$, lower local deviates sum up to $ms\hat{I}_j(v, [q, i])$, and the sum of the four quantities involved is equal to the corresponding value of $ms\hat{I}_j(i, [q, v])$.

III. 2.5 Let us examine in brief table V. It is noteworthy that lower deviates have common *maxima* (at V_3 and V_5) contrary to the different trends of upper deviates. This is resulted in the difference between the trends of $ms\hat{I}_j(q, i)$ and $ms\hat{I}_j(v, [q, i])$ in table IV. All quantities (all the deviates of table V.) are some diversity measures but lower deviates show how empirical marginal diversity differs from a minimum marginal uncertainty.

Table V.

	$\Delta N_j \hat{H}_j (M_q)$	$\Delta n_j \hat{H}_j (m_q)$	$\nabla N_j \hat{H}_j (M_q)$	$\nabla n_j \hat{H}_j (m_q)$
$j = 1$	32.512	4.414	106.778	51.484
2	65.557	15.299	120.978	77.562
3	88.467	41.398	140.487	112.092
4	117.786	114.917	112.732	94.900
5	91.521	110.628	128.562	113.131
6	77.455	148.166	93.174	85.844
7	14.415	88.761	71.766	90.673
8	0.000	0.000	0.000	0.000

III. 3 We gratefully acknowledge the cordial help of staff people at the computer station of the L. Kossuth Univ., Debrecen.

IV. Summary

Associatum is defined to be total association of a biotal universe. Associatum (expressed here as total information of some kind) can be partitioned into additive components (preference and diversity functions) whose change in space (with increasing unit sizes of sampling) can be related to associatum process in space.

REFERENCES

- Aleksandrova, V. D. 1969. Classification of vegetation (in Russian). Nauka, Leningrad.
- Andrewartha, H. — L. Birch, 1954. The distribution and abundance of animals. Univ. of Chicago Press, Illinois.
- Cole, L. C. 1957. The measurement of partial interspecific correlation. *Ecol.* 38: 226 — 233.
- Dagnelie, P. 1960. Contribution à l'étude des communautés végétales par l'analyse factorielle. *Bull. Serv. Carte Phytogéogr. Sér. B.* 5: 7 — 71; 93 — 195.
- Dagnelie, P. 1965: L'étude des communautés végétales par l'analyse statistique; des liaisons entre les espèces et les variables écologiques. *Biometrics*, 21: 345 — 361 890 — 907.
- Dale, M. B. 1971. Information analysis of quantitative data IN: „Statistical Ecology” vol. 3 (cf. Patil — Pielou — Waters): 133 — 148.
- Dévai, I. — K. Horváth — P. Juhász-Nagy, 1971. Some problems of model-building in synbiology. Part I. Spatial diversity process of the binary type in a simple situation. *Annales Univ. Budapestinensis, Sect. Biol.*, 13: 19 — 32.
- Goodall, D. W. 1952. Quantitative aspects of plant distribution. *Biol. Rev.*, 27: 194 — 245.
- Greig-Smith, P. 1964. Quantitative plant ecology. Butterworths, London.
- Juhász-Nagy, P. 1964. Some theoretical models of cenological fidelity. *Acta Biol. Debrecina*, III: 33 — 43.

- J u h á s z - N a g y, P. 1967. On association among plant populations. Part 1. Multiple and partial association: a new approach. *Ibidem*, V: 43 - 56.
- K e r s h a w, K. A. 1964. Quantitative and dynamic ecology. E. Arnold, London.
- K h i n c h i n, A. J. 1957. Mathematical foundations of information theory. Dover, New York.
- K u l l b a c k, S. 1959. Information theory and statistics. J. Wiley, New York.
- L a m b e r t, G. M. - M. B. Dale, 1965: The use of statistics in phytosociology. *Adv. ecol. Res.* 2: 59 - 99.
- M a c n a u g h t o n - S m i t h, P. 1965. Some statistical and other numerical techniques for classifying individuals. H. M. S. O., London
- M c I n t o s h, R. P. 1967. The continuum concept of vegetation. *Bot. Rev.*, 33: 130 - 187 (Responses. *Ibidem*, 34: 253 - 322).
- O r l o c i, L. 1971. Information theory techniques for classifying plant communities. IN: „Statistical Ecology”. 3 vols. (cf. Patil - Pielou - Waters): 259 - 270.
- P a t i l, G. P. - E. C. Pielou - W. E. Waters (ed.-s) 1971. Statistical Ecology. I - III. Pennsylv. St. Univ. Pr., Univ. Park.
- P i e l o u, E. C. 1969. An introduction to mathematical ecology. J. Wiley, New York.
- Q u a s t l e r, H. 1953. The measure of specificity. IN: „Information theory in biology” (ed. by H. Quastler; Univ. of Illinois Pr., Urbana) : 41 - 71.
- W i l l i a m s, W. T. - M. B. Dale, 1965. Fundamental problems in numerical taxonomy. *Adv. bot. Res.* 2: 35 - 68.
- W i l l i a m s, W. T. - J. M. Labert, 1966. Multivariate methods in plant ecology. V. Similarity analyses and information-analysis. *J. Ecol.* 54: 427 - 445.