

# SOME PROBLEMS OF MODEL-BUILDING IN SYN BIOLOGY

## PART 1

### SPATIAL DIVERSITY PROCESS OF THE BINARY TYPE IN A SIMPLE SITUATION

By

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

Department of Plant Taxonomy and Ecology of the Eötvös Loránd University,\*\* Budapest  
Hydrobiological Group, Department of Zoology of the L. Kossuth, L. University  
Debrecen\*

Received on the 10<sup>th</sup> September, 1970

### I. Introductory notes

I.1 Diversity (in the present sense of this concept) is a synbiological basic phenomenon, similar to other basic phenomena (e.g. similarity, preference, etc.). The intuitive meaning of the term "basic phenomenon" should first be investigated.

I.2 Let us make the following, very general "mental experiment". Let  $Z$  be the set of all populations which ever lived and live, a set of finite cardinality  $z$ , i.e. a set of  $z$  points. Let, furthermore, be given an  $n$ -dimensional abstract space,  $X_n$ , whose coordinates are: (a) the 3 coordinates of the real (topographical) space, (b) the time axis (as the real number line), (c) the coordinates characterizing the possible quantitative representations of the populations (e.g. density, abundance, etc.)

Now, an absurd hypothesis (say, "existential indifference hypothesis") that "whatever population may be found in wherever, whenever, and in whichever quantity" is equivalent with the statement that all elements (points) of  $Z$  may be characterized by any  $n$ -dimensional vector of  $X_n$ . Of course, if this statement were true, it would mean that Nature as "experimenter" randomizes  $z$  points in  $X_n$ .

We know that this hypothesis is false: this is a trivial non-sense. It is shown by a huge body of evidence that the  $z$  points of  $Z$  form very different sorts of aggregates in  $X_n$ . But, of course, it is almost always extremely difficult to say how the points form "groupings", how they become denser or rarer in  $X_n$ .

This is why the distinction of basic phenomena is necessary. It may be truly said that all research work in the field of synbiology may be regarded as special sort of negation of the central hypothesis mentioned above with respect to a certain subset of  $Z$  and to a certain subspace of  $X_n$ . But, of course, this negation is possible (even, with respect to the same object, i.e. to the same subset of points in the same subspace of  $X_n$ ) in a vast number of ways; it depends – *inter alia* – on the mathematical

model of a particular investigation. Now, model and basic phenomenon are clearly closely related.

By basic phenomenon the following on the whole is meant. If the different methodological types of the negation of the central hypothesis is approached so that they can be satisfactorily described by some "general model" (e.g. by some urn model, etc.), further, by some "working model" (e.g. by that of the statistical type), in such a way that their biological meaning is more or less clear, then a phenomenon (related to a logical type of negation) can be considered a basic phenomenon. There are many types of basic phenomena (e.g. similarity among biotas, association among populations, species-area-relations, and so on; cf. e.g. Preston 1948, Elton 1949, Black-Dobzhansky-Pavan 1950, Goodall 1952, Andrewartha-Birch 1954, Sokal 1961, Greig-Smith 1964, etc.)

I.3 The trouble lies naturally in the fact that several authors delimit the given basic phenomena by different means and investigate them by various methods, frequently without ever submitting or publishing an adequately precise description of their approach. (For instance, it is rarely recognised that "interspecific correlation" is not synonym with "association" but the former is a definite subcase of the latter, if and only if certain linearity and normality conditions are satisfied.) It is quite possible that more care on the "general model" and the relation between general and working models would improve the conceptual clarity. Take, for instance, the concept diversity. Most of the diversity types are distinguished according to their working models, only (cf. McIntosh 1967), but almost all consideration are related to one single general (theoretical) type of diversity, namely, "individual diversity" (where the distribution of numbered individuals is studied into categories, e.g. taxa). On the other hand, there are a number of other general types of diversity; in fact, all "natural" frequency (or probability) distributions (e.g. distribution of biotal combinations in this paper, distribution of "Raunkiaer frequency values", etc.) may be regarded from the same point of view. But, just for this reason, one has the feeling that the "general concept" of diversity is essentially undefined. Perhaps such a general definition would relate the possible and "realized" (empirical) vectors of some properly defined abstract spaces (a definite subspace of  $X_n$ ), and, in the same way, "similarity" may be regarded as a basic phenomenon due to the systematic comparisons of empirical vectors, and so on.

I.4 Within the scope of the present paper, however, we cannot undertake this task even with regard to a single basic phenomenon. On the other hand, our endeavour is to assure a proper unequivocality of our very simple considerations by the construction of a mathematical model system presented herewith.

I.5 The investigation of the chosen basic phenomenon (diversity) had the following stages in recent times. First, E.H. Simpson (1949) published a simple "diversity index". Beside its simplicity, this index (intended

to serve mainly large-scale evolutionary comparisons) has a great number of drawbacks. Their discussion is given by C.B. Williams (1964, p. 147—).

Following Simpson, several authors published diversity indices, mostly of mere empiric nature. A significant breakthrough was given only in the works of Margalef (1957), Hutchinson (1953, 1958), and R.H. MacArthur—J.W. MacArthur (1961) as well as their disciples and followers (e.g. Hairston 1959, 1954, Klopfer 1959, Klopfer—MacArthur 1960, Connell—Orias 1964, Grice—Hart 1962, Lloyd—Ghelardi 1964, etc.); an excellent review is due to McIntosh (1967). It is they who recognize the close relationship of the synbiological diversity with certain physical occupancy problems, and the importance of the diversity conception in connection with niche-segregation.

## II. Material and method

II.1 In relation to such a general consideration of diversity, a number of formulations and also a variety of their approaches are naturally possible. Under the "simple situation" indicated in the title of this paper we refer to the following restrictions.

We have delimited our investigations (*a*) to a *single* animal group (*b*) of a *single* geographical object, and (*c*) to a *faunistic-binary* approach (*d*) connected only with *spatial* (and no temporal) diversity processes (*e*) based enterily on one simple model system of *information theory* (neglecting other types of statistical working models; cf. McIntosh 1967).

The basic data of this type of study were obtained from a survey (the 26th of July — the 5th of August, 1968) of *Cladocera fauna*, present in Pond N° 10 of the fish-pond system at *Bánhalma* (in the central part of the Great Hungarian Plain; near the village *Kenderes*), a property of a state estate, that is, by counting 66 565 individuals of 10 taxa in  $8 \cdot 64 = 512$  sampling units (cf. II.3). The hydrological, chemical and hydrobiological properties of this shallow pond (cca. 90 cm., in average; cca. 51 cadastral yoke, in size) are well-known, but, owing to a lack of space, these important data must be omitted.

II.2 The of taxa of *Cladocera fauna* are as follows:

1. *Bosmina longirostris* C.F. Müller  
var. *pellucida* Stingelin (henceforth: A)
2. *Bosmina longirostris* O.F. Müller  
var. *similis* Liljeborg (B)
3. *Bosmina longirostris* O.F. Müller  
var. *cornuta* Jurine (C)
4. *Bosmina longirostris* O.F. Müller  
var. *brevicornis* Hellich (D)
5. *Daphnia longispina* O.F. Müller (E)
6. *Chydorus sphaericus* O.F. Müller (F)
7. *Ceriodaphnia pulchella* Sars (G)

8. *Ceriodaphnia quadrangula* O.F. Müller (H)
9. *Moina brachiata* Jurine (I)
10. *Moina rectoris* Leydig (K)

II.3 Concerning the plan and execution of the *sampling* the following should be mentioned. The tree-dimensional shape of the sampling units was approximately a cube. The sizes of the sampling units were selected according to the increasing powers of 2, from 1 to 8, thus, we took 2, 4, 8, 16, 32, 64, 128 and 256 milliliter (ml.) water for different sampling sizes (volumes). The cardinality (number) of sampling units was 64 for all sampling volumes. The laying-out of sampling points (for locating the sampling units) was of the random type. The randomizing instructions of sampling were obtained by constructing a table of pairs of random numbers so that its first column represent random values for a compass (0–360), and in its second column the corresponding random values show the number of oar-strokes to be made in the given (random) direction. In such a way the sampling is similar to a "Brownian motion" in the surface of the pond, i.e. the  $g$ th sampling point is located by a pair of a random angle and a random distance with respect to the  $(g-1)$ th point. It could thus be assumed that the several sampling points may be more or less independent of one another, that is, all points of the pond may be selected equiprobably.

### III. Concepts and notation

III.1 There is given the  $V$  set class of the compound existential units that is,

$$V = \{V_1, V_2, \dots, V_i, \dots, V_8\} \quad (1)$$

where  $V$  is a set class of 8 elements in which  $V_i$  is a set of sampling units of  $2^i$  ml. size, and

$$V_i = \{v_1^{(i)}, v_2^{(i)}, \dots, v_k^{(i)}, \dots, v_m^{(i)}\} \quad (2)$$

where, again,  $V_i$  is a set of 64 elements ( $m = 64$ ) in which  $v_k^{(i)}$  is the  $k$ th sampling unit of  $2^i$  ml. size.

There is given, furthermore, the  $Q$  set of taxa in II.2,

$$Q = \{A, B, \dots, E, \dots, K\} \quad (3)$$

called *faunal universe* (i.e. a set universe, a set of all basic objects of a particular study).

There is given a random variable,  $R_{Ek}^{(.)}$ , showing the degree of representation of taxon  $E$  in  $v_k^{(.)}$ , binary,  $R_{Ek}^{(b)}$ , in the simplest case. Let  $R_{Ek}^{(b)} = r_{Ek}$ , and let

$$r_{iEk} = \{1, 0\} \quad (4)$$

that is,

$$\begin{aligned} r_{iEk} &= 1, \text{ if taxon } E \text{ is present in } v_k^{(i)}, \\ r_{iEk} &= 0, \text{ if taxon } E \text{ is absent in } v_k^{(i)}. \end{aligned}$$

III.2 Let us construct a set of binary matrices of size  $10 \times 64$ , one for each element of  $V[(1)]$ , where the rows and columns represent the elements of sets  $Q$  and  $V_i[(3), (2)]$ , respectively, and where the empirical value of  $r_{iEk}[(4)]$  is the  $\langle E, k \rangle$ th entry of the  $i$ th matrix. Let, further,

$$\sum_{k=1}^{64} r_{iEk} = n_{iE}, \quad \sum_{E=1}^{10} r_{iEk} = n_{\cdot ik}, \quad \sum_{k=1}^{64} \sum_{E=1}^{10} r_{iEk} = N_i \quad (5)$$

and, in such a way (permitting summation in the marginals), we obtain the  $i$ th binary contingency table of a set of such tables, called *faunistic composition* of order  $i$

$Q \backslash V_i$	$v_1^{(i)}$ .....	$v_k^{(i)}$	.....	$v_m^{(i)}$	
$A$					
⋮					
$E$	.....	$r_{iEk}$	.....	$n_{iE}$	(6)
⋮					
$K$					
.....		$n_{\cdot ik}$	.....	$N_i$	

abbreviated as  $U_i$ .

Consider the following definitions with respect to  $U_i$ :

$n_{iE}$ , called *local valence*, of taxon  $E$  in  $U_i$ , shows the number of sampling units of size  $2^i$ , where taxon  $E$  is present,

$n_{\cdot ik}$ , called *faunal valence* of  $v_k^{(i)}$  in  $U_i$ , shows the number of taxa which are present in  $v_k^{(i)}$ ,

$N_i$ , called *total valence* of  $U_i$ , shows the number of all presences in  $U_i$ ,

$q_{iE}$ , is the binary row vector as *local vector* belonging to taxon  $E$ ,

$q_{ik}$ , is the binary columnal vector as *faunal vector* belonging to  $v_k^{(i)}$

$\lambda_{iE}$ , the subset of those elements of set  $V_i$  whose  $r_{iE}$  values are "1", is called *loci* of taxon  $E$ ,

$\varphi_{ik}$ , the subset of those elements of set  $Q$  whose  $r_{\cdot ik}$  values are "1" is called *fauna* of  $v_k^{(i)}$ ,

$\Pi_i$ , the set of all empirical loci in  $U_i$ , is called *local class* of  $U_i$ ,

$\Phi_i$ , the set of all empirical faunas of  $U_i$ , is called *faunal class* of  $U_i$ .

It should be noted that the above definition of fauna is in agreement with the traditional interpretation of this concept, with the difference, that our outline is a result of a set theory approach. Fauna is regarded to be a subset of a set (called faunal universe), or, equivalently, an element of a power set (called faunal power set; cf. IV.). Note, further, that a concept ("attribute dual" to fauna) was missing and is needed; this is the concept called loci. Consider the difference between three adjectives:

- "faunal" (related to fauna)
- "local" (related to loci)
- "faunistic" (simultaneously related to both fauna and loci,  $U_i$  being a joint function of faunal and local binary relations).

Note, still further, that "local valence" is synonymous — *mutatis mutandis* — with "frequency of Raunkiaer" (commonly used in plant sociology), and, "faunal valence" is synonymous with the common expression "number of species." Beside some drawbacks of these traditional terms (e.g. talking of frequency distribution of the former, one should think of "frequencies of frequencies"), it seems to be desirable to use a common terminology for both marginal distributions which are again "attribute duals" to each other.

III.3 All faunistic functions are related, of course, to Bernoulli model (as a "general model", in the sense of I.2) with "yes-no" type of elementary decisions.

Our sampling was made in such a way that, as practically possible, all individuals of every taxon within all sampling units have been registered (for further use), but in the present paper we consider only the above faunistic "presence-absence" data ( $8 \cdot 10 \cdot 64 = 5120$  elementary data in all).

#### IV. Working model, results, and interpretation

IV.1 The empirical contingency tables (see e.g. table I. as one table of 8 such tables) obtained after this (faunistic) type of elaboration of our empirical data can be evaluated in several ways with respect to the aims of investigation. Namely,

- (a) the frequency of the empirically obtained identical binary columnal vectors (faunal vectors) can be compared with that of the set of possible columnal vectors,
- (b) the frequency of the empirically obtained identical binary row vectors (local vectors) can be compared with that of the set of the possible row vectors,
- (c) the empirically obtained binary columnal vectors can be compared (per two, three, . . .) according to the frequency of agreeing elements,

- (d) the empirically obtained binary row vectors can be compared (per two, three, ...) according to the frequency of agreeing elements,
- (e) marginal distributions can be compared (in different ways).

Points (a) and (b) can be interpreted as combinational problems, since, for example, in the case of point (a), the question concerns the frequency distribution of the set of all possible subsets of  $Q$ , and elements (symbols) are not ordered. In the cases of (c) and (d,) the problem is related to the frequency distributions of ordered elements of vectors (for instance, in a pairwise comparison, the frequency distribution of a set of four elements, more precisely, ordered pairs,  $\{\langle 11 \rangle, \langle 10 \rangle, \langle 01 \rangle, \langle 00 \rangle\}$ , is considered), thus, roughly speaking, the model-building is interconnected with the construction of new contingency tables of  $2^2, 2^3, \dots$  cells. In the case of point e the marginal distributions (or, the frequency distribution of the valential values) are to be compared.

The possibilities of evaluation outlined above is a starting point to study to following basic phenomena:

- |  |   |              |
|--|---|--------------|
| (a) faunal diversity                       | } | { faunistic  |
| (b) local diversity                        | } | { diversity  |
| (c) interfaunal similarity                 | † | { faunistic  |
| (d) interlocal similarity                  | † | { similarity |
| (e) faunistic preference of different type |   |              |
| (e.g. fidelity, etc.)                      |   |              |

It is well-known that many types of further evaluation are used to the different basic phenomena, in particular (c) and (d), in "quantitative plant ecology" (cf. Greig-Smith, 1964), where, of course, "faunal" and "faunistic" are to be substituted by "floral" and "floristic", resp. (d) is commonly known as "association". According to our restrictions in II.1, we are concerned here only point (a), faunal diversity, a new type of diversity. The basic question of (a) is — to speak figuratively — how Nature as "experimentator" utilizes the "available possibilities" in distributing the set of available "alphabetic stock" (i.e. set  $Q$ ) as basic (elementary) objects into combined objects (namely, faunas), or "combinational words", and these combined objects into the existence units. It is rather easy to see that this problems corresponds to the "vectorial interpretation" of diversity, mentioned at the end of I.3.

IV.2 The task is, of course, to find a function with certain optimum properties characterizing faunal diversity. The following postulates are obvious:

- (1) the value of such a function is minimum (zero), if and only if every fauna is identical in every  $v_k^{(i)}$ ,  $k = 1, 2, \dots, m$ ,
- (2) the value of such a function reaches a (theoretical) maximum, if and only if every fauna is different in every  $v_k^i$ ,

- (3) the values of such a function are monotonically increasing between the minimum and maximum values indicated above.

Of course, there are an infinite number of functions satisfying these "weak conditions". But by a further postulate, or requirement,

- (4) let our function be additive,

although not too precise in itself, we can narrow down the range of possibilities.

Being more specific, let

$$\pi_Q = \{\pi_0, \pi_1, \dots, \pi_j, \dots, \pi_w\} \quad (7)$$

where  $\pi_Q$  is the *faunal power set* (with respect to faunal universe,  $Q$ ), the set of all subsets of  $Q$ , where

$\pi_j$  is a *potential fauna*, one distinguishable element of  $\pi_Q$ ,  $j = 0, 1, 2, \dots, w$ ,

$\pi_0$  is an *empty fauna* (where no elements of  $Q$  are present),

$\pi_w$  is a *complete fauna* (where all elements of  $Q$  are present),

and  $w + 1$  is cardinality of  $\pi_Q$ , in the present case

$$w + 1 = 2^{10} = 1024, \text{ because } \sum_{e=0}^{10} \binom{10}{e} = 1024.$$

Let, moreover,  $F_i$  be the frequency distribution of faunal class  $\Phi_i$ ; with respect to  $\pi_Q$  (in brief, a *faunal distribution* of order  $i$ )

$$F_i = \{f_{i0}, f_{i1}, f_{i2}, \dots, f_{ij}, \dots, f_{iw}\} \quad (8)$$

where  $f_{ij}$  is the empirical frequency of  $\pi_j$  in  $\Phi_i$ , and  $\sum_{j=0}^w f_{ij} = m$ . In certain condition (in particular, if  $m \rightarrow \infty$ )  $f_{ij}/m$  may be regarded as a probability estimate of  $\pi_j$  in  $\Phi_i$ ,  $\hat{p}_{ij}$ . Although in our case  $m$  is rather small,  $m = 64$ , it was shown roughly that the results indicated below are essentially the same, if  $m > 64$ ; thus, we can accept  $f_{ij}/m$  as a crude approximation of  $p_{ij}$ . Let  $\hat{P}_i$  the *probability distribution* which corresponds to (8).

Let, moreover,  $\hat{L}_i$  be a *finite scheme* (*sensu* Khinchin), or, much more correctly, an empirical finite scheme with respect to  $\pi_Q$  and  $\hat{P}_i$  such that

$$\hat{L}_i = \left( \begin{array}{ccccccc} \pi_0, & \pi_1, & \dots, & \pi_j, & \dots, & \pi_w \\ \hat{p}_{i0}, & \hat{p}_{i1}, & \dots, & \hat{p}_{ij}, & \dots, & \hat{p}_{iw} \end{array} \right) \quad (9)$$

and,  $\hat{H}_i$  be the *entropy estimate* of  $\hat{L}_i$  (*sensu* Shannon)

$$\hat{H}_i = - \sum_{j=0}^w \hat{p}_{ij} \log \hat{p}_{ij} \quad (10)$$



It is rather easy to see that (10) satisfies all the four postulates mentioned above. Moreover, this way of thinking is very closely related to that of other authors (e.g. Margalef, 1957, etc.), who used, instead, "the statistical mechanics" approach interconnected with the multinomial coefficient. Note that a  $\pi_j$  may be equivalently regarded as a "combination without repetition" (in the sense of combinatorial analysis), a "special event" (in the sense of probability theory), and a "combinational word of alphabet  $Q$ " (in the sense of information theory). Clearly, if  $s_i$  is multinomial (polynomial) coefficient with respect to (8),

$$s_i = \frac{m!}{f_{i0}! f_{i1}! \dots f_{ij}! \dots f_{iw}!} \quad (11)$$

then, with  $\log s_i$ , the postulates (1) - (3) are satisfied, because  $0 \leq \log s_i \leq \log m$ . Further,  $\log s_i \rightarrow \hat{H}_i$ .

IV. 3  $\hat{H}_i$ , of course, is a measure of uncertainty of faunal distribution of order  $i$ . The value of this measure obviously depends on the order, i.e. on the size of sampling units. If, for instance, the sampling volume, say  $V_u$ , is large enough to contain all the taxa considered to be the elements of a faunal (or biotal) universe, and this is true for all (or, at least, most of) sampling units of that particular size, then,  $f_{uw} = m$ ,  $\hat{p}_{uw} = 1$ , and because

$$\hat{L}_u = \begin{pmatrix} \pi_0, \pi_1, \dots, \pi_j, \dots, \pi_w \\ 0, 0, \dots, 0, \dots, 1 \end{pmatrix} \quad (12)$$

then  $H_u = 0$  (or, at least  $H_u \rightarrow 0$ ). If we know that  $V_u$  size (more correctly the smallest size of that type), then we know that the whole range between  $V_u$  and  $V_z$ ,  $|V_u, V_z|$  is absolutely homogenous (from a faunal point of view), where  $V_z$  is the volume of maximum size (e.g. the whole volume of a pond). We know, further, that if  $V_i < V_u$ , then  $H_i > 0$ , since there is no other potential fauna than  $\pi_w$  (except, in an extreme sense,  $\pi_0$ ) which can be such hegemonic as  $\pi_w$  in (12). This shows clearly that faunal diversity (or faunal uncertainty) can be studied properly by a set of sampling volumes. If the volume of sampling is increasing, one might learn the state of faunal diversity in space.

Let us call such an increasing process a *spatial process*, and in particular, our empirical process

$$S_x = \begin{bmatrix} V_1 \rightarrow V_2 \rightarrow \dots \rightarrow V_i \rightarrow \dots \\ U_1 \rightarrow U_2 \rightarrow \dots \rightarrow U_i \rightarrow \dots \\ \hat{L}_1 \rightarrow \hat{L}_2 \rightarrow \dots \rightarrow \hat{L}_i \rightarrow \dots \\ \hat{H}_1 \rightarrow \hat{H}_2 \rightarrow \dots \rightarrow \hat{H}_i \rightarrow \dots \end{bmatrix} \quad (13)$$

a *spatial diversity process*, an obviously crude approximation of a continuous stochastic process. (This process,  $S_x$ , may be considered to be a sub-case of the class of "two-valued stochastic processes".)

Of course, we are interested in  $\widehat{H}_x$ ,

$$\widehat{H}_x = \{\widehat{H}_1, \widehat{H}_2, \dots, \widehat{H}_i, \dots, \widehat{H}_8\} \quad (14)$$

the *entropy distribution* with respect to  $S_x$ , characterizing the distribution of faunal diversity in space.

IV.4 Consider then table 2 and figure 1. It is shown that by the successive increasing of sample volume entropy increases to a certain volume ("maximum volume"), then decreases and tends to zero at a certain volume ("minimum volume"), where, like in (12), every taxon can in great probability be found in every sampling unit and thus the order (in this respect) is complete.

IV.5 Interpreting this trend of  $\widehat{H}_x$ , especially the concepts maximum volume,  $V_{\max}$ , and minimum volume,  $V_{\min}$ , should carefully be considered. It is easy to see that  $V_{\max}$  literally means "a size of maximum entropy", and  $V_{\min}$  does "a size of minimum entropy".

The problem of  $V_{\min}$  is clearly in close relation to the famous minimum area ("minimiareal") problem of plant sociology. In fact, this is a three-dimensional generalization of that concept. (Similarly, we can think of a one-dimensional generalization of minimum area, namely "minimum linea", in terms of terrestrial communities.) It is well-known that how many efforts have been made (in particular, by means of the rather sterile and non-probabilistic approach of "species-area functions") to determine minimum area (cf. Greig Smith, 1964). We believe that by a further development of the present approach (by means of "gain of information" methods; cf. Kullback, 1959) this is quite possible.

Table 2.

Sampling volume	$m\widehat{H}_i$
$V_1$	137,96
$V_2$	173,02
$V_3$	202,23
$V_4$	177,40
$V_5$	194,07
$V_6$	156,49
$V_7$	145,94
$V_8$	0,00

Table 2 The  $m\widehat{H}_i$ -values [of  $\widehat{H}_x(14)$ ] characterizing faunal diversity vs. sampling volumes (cf. II. 3)

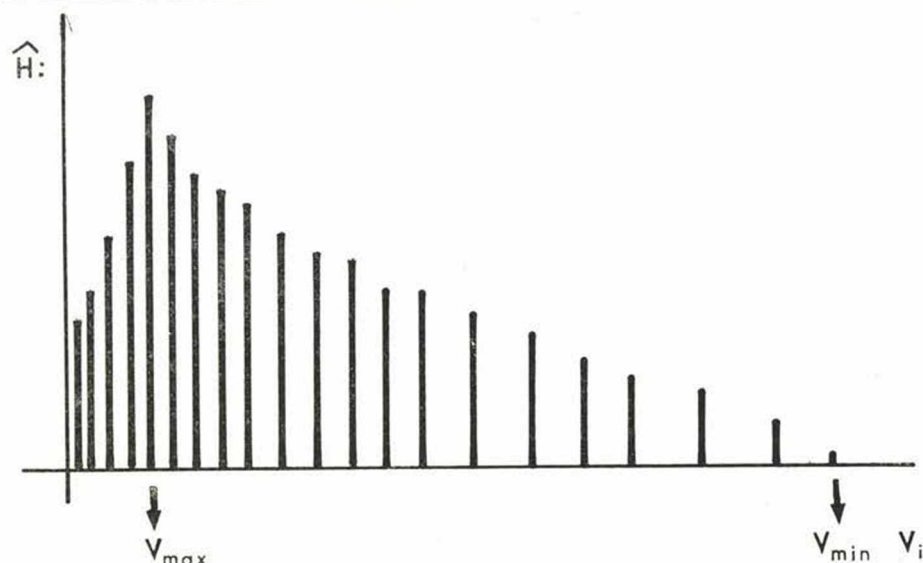


Figure 1 The general ("idealized") trend of function  $\widehat{H}_x$  [cf. (14);  $H_i$  values (10) vs.  $V_i$  values]; the empirical values of our data in table 2 represent one of its special cases.

Naturally, if our general model is not faunistic (i.e. not of the Bernoulli type) but "quantitative" (i.e. the number of individuals are also considered) we can think of a special non-faunistic "characteristic volume" namely, a "representative minimum volume". Whatever a "representative minimum volume",  $V_{\min}^{(r)}$ , is, the inequality relation  $V_{\min}^{(r)} > V_{\min}$  must be true.

It is, moreover, possible, that a  $V_{\min}$ -size (or state) is permitted only is a maximum size of the sampling volumes (i.e. in the size of the particular stand), if some populations are located only "in small subsets" in that particular stand. In this case we can speak of a "trivial  $V_{\min}$ ". The supposed commonness of trivial  $V_{\min}$  in nature shows clearly, again, the necessity of a probabilistic approach. Instead of "minimum topographical unit" as such ( $V_{\min}$ , minimum area, etc.), it seems to be much better to think of a set of such units at different probability levels. In the near future our intention is to study such a set of "characteristic volumes" carefully, and we hope that many "structural" or "pattern" problems of biocenoses (e.g. homogeneity heterogeneity, etc.) can be usefully approached in such a way.

IV.6 Beside the problem of  $V_{\min}$ , the question remains how to interpret the ideal  $\Gamma$ -like distribution of  $\widehat{H}_x$  showing in fig. 1, and the somewhat surprising existence of  $V_{\max}$ .

Concerning  $V_{\max}$ , suppose that  $m \rightarrow \infty$ . If all events (potential faunas) are equiprobable, then, according to the Hartley formula, at a particular sampling size,  $H = \log m$ . In our data, of course, we can think only of a "relative maximum" at  $V_{\max}$ , or, better, at an interval,  $V_{\max}$ -range,  $\overline{V_{\max}}$ .

The simplest way of explanation of  $\overline{V_{\max}}$  is related to the values of faunal valences (cf. III.2), or more correctly, to the frequency distributions of them with respect to  $S_x$ . (a) It is clear that if the sampling volume is "small", then a few "small numbers" of the possible values (0–10) are represented. For instance, in our  $U_1$  (where  $V_1$  is 2 ml.), there are only three sorts of  $n_{ik}$  values (namely, 0, 1, 2) whose frequency distribution is:

$$f(n_{.10}) = 17; \quad f(n_{.11}) = 36; \quad f(n_{.12}) = 11$$

(b) Similarly if the sampling volume is "large", then, again, a few "larger numbers" of the possible values are permitted. We can guess, for instance, that approaching  $V_8$  more closely, the frequency of 10's grows exponentially. It is rather easy to see (although our present evidence is indirect) that in both cases  $\widehat{H}_i$ -value is small, because our uncertainty with respect to  $F$ 's is rather small, too. Now, contrary to (a) and (b), (c) there is always an interval, where the opposite is true, since in that interval the majority of faunal valence values may occur (see table 1), and, consequently, the corresponding value of  $\widehat{H}_x$  is rather high.

This is, of course, a very crude explanation of such a concept as  $V_{\max}$ , for obvious reason. Faunal valence is related only indirectly to combinational relations (although, in the case of  $U_1$ , for instance, it can be definitely stated that at most 56 events are permitted from 1024), and, there is special sort of diversity (valential diversity of both types) not to be considered here. More precisely (but still in a rather indirect way) we can study table 3, where the number of possible and empirical ("realized") combinational events are compared. Table 3 shows essentially the same relation-

Table 3

Number of possible events	$\binom{10}{0}$	$\binom{10}{1}$	$\binom{10}{2}$	$\binom{10}{3}$	$\binom{10}{4}$	$\binom{10}{5}$	$\binom{10}{6}$	$\binom{10}{7}$	$\binom{10}{8}$	$\binom{10}{9}$	$\binom{10}{10}$
Sampling volume	Number of realized events										
$V_1$	1	6	8	0	0	0	0	0	0	0	0
$V_2$	1	6	10	7	4	0	0	0	0	0	0
$V_3$	0	2	7	9	8	3	1	0	0	0	0
$V_4$	0	0	1	4	7	5	4	2	0	0	0
$V_5$	0	1	1	2	5	7	7	6	1	0	0
$V_6$	0	0	0	0	1	4	4	4	3	3	1
$V_7$	0	0	0	0	0	0	1	3	5	4	1
$V_8$	0	0	0	0	0	0	0	0	0	0	1

Table 3 Comparison between the numbers of possible and realized events (combinations). Columns represent binomial coefficients,  $\binom{K}{e}$ , where  $K$  is number of taxa in faunal universe  $Q$  [cf. II. 2, (3)],  $e$  is a possible value of a faunal valence,  $n_{ik}$  (cf. III. 2). An entry of the table shows the number of empirically possible (permitted) events at a given sampling volume (cf. II. 3) and at a given value of  $e$ .

ship as before concerning three ("small", "mixed", and "large") range of sampling sizes influencing uncertainty distribution. It is quite understandable, for instance, why  $\hat{H}_x$  in size  $V_3$  is (at least, empirically) at a maximum, since three sorts of combinational classes (where  $n_{.3k} = 2, 3, 4$ ) are approximately equiprobable.

Of course, the only direct way of the interpretation of  $\hat{H}_x$  is to study the compound table of  $L_i$ 's. Unfortunately, because such a table consists of  $8 \times 1024$  cells, this is too lengthy for the present publication. (It is very easy to imagine that if a biotal universe has a cardinality more than 20, which is surely "not too rich" a biota from a biological point of view, all similar tables [from order  $10^6$  - ] are too voluminous for any publication.) This is the reason why  $\hat{H}_x$  is considered to be practically important.

IV.7 But its importance has a theoretical aspect, too. We believe that the general trend of  $\hat{H}_x$  discussed above is general enough (at least in qualitative terms) of being characteristic for all possible objects (e.g. all biocenoses, or even any "ornamentation" of a finite set of symbols). One of us studied many types of terrestrial plant communities and obtained results of the same type as it is shown by fig. 1, with "quantitative modifications" of the particular object (Juhász-Nagy, 1967).

IV.8 In this paper we have attempted to outline the simplest problems which represented the starting questions of our investigations. We had to treat only a very restricted range of these problems, however, and by far without a claim of entirety. The cause of restrictions mentioned in II.1 was our endeavour to attain the comparatively simplest approach owing not only to temporal, counting and capacity delimitations but also for the sake of a clearer view for future development.

IV.9 It is our pleasant duty to express our gratitude to Dr. J. Szabó (Zoology Dept., Debrecen Univ.) for study facilities, and to Mr. Gy. Dezső agronomist, and Mr. F. Péter, fishery master, for their cordial help.

## V. Summary

In simple situation (II) and by a simple information theory model system (III, IV.) we studied a new type of diversity (called faunal diversity) and a spatial diversity process of this type whose "qualitative trend" is shown by table 2 and fig 1. We believe that this trend is a very general one (characteristic for all possible communities), with "quantitative modifications" of the particular objects, moreover, that these features can be analyzed in several ways and they are interconnected with many other properties of synbiological phenomena.

## REFERENCES

- Andrewartha, H. — Birch, L. 1954. The distribution and abundance of animals. Univ. Chicago Press, Chicago, Ill.  
 Black, G. A. — Dobzhansky, Th. — Pavan, C. 1950. Some attempts to estimate species diversity and populations diversity of trees in Amazonian forests. Bot. Gaz. 111: 413 — 425.

- Connell, J. H. — Orias, E. 1964. The ecological regulation of species diversity. *Amer. Naturalist* **98**: 399–414.
- Elton, C. S. 1949. Population interspersions; an essay on animal community studies. *J. Ecol.* **37**: 1–23.
- Elton, C. S. 1958. The ecology of invasions by animals and plants. Methuen, London, England.
- Goodall, D. W. 1952. Quantitative aspects of plant distribution. *Biol. Rev.* **27**: 194–245.
- Greig-Smith, P. 1964: Quantitative plant ecology. Butterworths, London.
- Grice, G. D. — Hart, A. D. 1962: The abundance, seasonal occurrence and distribution of the epizooplankton between New York and Bermuda. *Ecol. Monogr.* **32**: 287–309.
- Hairston, N. G. 1959: Species abundance and community organization. *Ecology* **40**: 404–416.
- Hairston, N. G. 1964. Studies on the organization of animal communities. Jubilee Symposium Supplement. *J. Ecol.* **52**: 227–239.
- Hutchinson, G. E. 1953. The concept of pattern in ecology. *Proc. Acad. Nat. Sci. Philadelphia*, **105**: 1–12.
- Hutchinson, G. E. 1958. Concluding remarks. Symposium on Quantitative Biology **22**: 415–427.
- Juhász-Nagy, P. 1967. On some „characteristic areas” of plant community stands. Proceedings of the Colloquium on Information Theory org. by the Bolyai Matematikai Soc.: 269–282. Akadémiai Kiadó, Budapest.
- Klopper, P. H. 1959. Environmental determinants of faunal diversity. *Amer. Naturalist* **93**: 337–342.
- Klopper, P. H. — MacArthur R. H. 1960. Niche size and faunal diversity. *Amer. Naturalist* **95**: 293–300.
- Kullback, S. 1959. Information Theory and Statistics. J. Wiley, N. Y.
- Lloyd, M. — Ghelardy, R. J. 1964. A table for calculating the equitability component of species diversity. *J. Anim. Ecol.* **33**: 217–225.
- MacArthur, R. H. — MacArthur, J. W. 1961. On bird species diversity. *Ecology* **42**: 594–598.
- Margalef, R. 1957. Information theory in ecology. (reprinted in 1958). *Gen. Syst.* **3**: 37–71.
- Margalef, R. 1962. “Diversità” dello zooplankton del Lago Maggiore. *Mem. Ist. Ital. Idrobiol.* **15**: 137–151.
- McIntosh, R. P. 1967. An index of diversity and the relation of certain concepts to diversity. *Ecology*. **48**: 392–404.
- Pianka, E. R. 1967. On lizard species diversity North American Flatland Deserts. *Ecology* **48**: 333–351.
- Preston, F. 1948. The commonness and rarity of species. *Ecology* **29**: 254–283.
- Preston, F. 1962. The canonical distribution of commonness and rarity. I. *Ecology* **43**: 185–216 II. *Ecology* **43**: 410–432.
- Simpson, E. H. 1949. Measurement of diversity. *Nature*, **1963**: 688.
- Sokal, R. R. 1961: Distance as a measure of taxonomic similarity. *Syst. Zool.* **10**: 71–79.
- Williams, C. B. 1964. Patterns in the balance of nature. Academic Press, New York — London.