

# On hierarchical partitioning of an ecological complexity function

Madhur Anand\*, László Orlóci

*Department of Plant Sciences, The University of Western Ontario, London, Canada, N6A 5B7*

---

## Abstract

We examine a complexity function in terms of its partitionability across hierarchical levels. We use an ecological example and describe two models: partitions by levels in the natural dominance structure of a plant community, and by levels in an analytical hierarchy, constructed from measured species associations. Our results help to answer the frequently asked question of how much the levels of a structured system contribute to total complexity. © 2000 Elsevier Science B.V. All rights reserved.

**Keywords:** Algorithmic complexity; Association; Boreal; Coding; Community; Entropy; Information; Model; Pattern; Scale; Stratification; Vegetation

---

## 1. Introduction

Any observation of a system's complexity is, by its very nature, context and scale dependent. By 'context' we mean the interrelated conditions under which the observations occur, that is, the environmental setting, and by 'scale', the spatio-temporal extent of the observation's resolution. Clearly, context and scale affect perception of complexity in different ways. Ideally, we should like to measure complexity in terms suitable to compare, say, a rainforest, to grassland. But for this, the rules of quantification must be very general. In this paper we use the general complex-

ity function of Anand and Orlóci (1996). We begin by reviewing its definition and relationship to entropy. We proceed to examine complexity and hierarchy by presenting different models of partitioning of the function across levels. We give a worked example from vegetation ecology and close the main text with conclusions.

## 2. Definition of complexity

### 2.1. Complexity and entropy

Complexity has many components and many manifestations (Çambel, 1993). The foremost of these, and often dominant is diversity. This is the number of unique system elements and the manner in which they are proportioned. The simplest and most commonly applied diversity measure is Shannon's (Shannon, 1948) entropy function:

---

\* Corresponding author. Present address: Department of Biology, University of New Mexico, Albuquerque, NM, 87131, USA. Tel.: +1-505-2771727.

E-mail address: manand@sevilleta.unm.edu (M. Anand).

$$H = - \sum_{j=1}^q p_j \log_2 p_j \quad \text{bits} \quad (1)$$

(or  $H/\log_2 e$  nats)

$H$  has served as the basis of a general community theory in ecology, well-explained by Margalef (1968) and Pielou (1975). Properties of Eq. (1) are meticulously dissected by Rényi (1961).

Pursuing the usual ecological interpretation of Eq. (1) we define  $q$  as the number of community elements taken as taxa, character types, or character set types (Orlóci, 1991c; Pillar et al., 1993). Symbol  $p_j$  signifies the proportionate representation of community element  $j$  by units of its kind in the community. Note that the higher the  $H$  value, the greater is the 'evenness' in the community, and thus higher the diversity. For a given value of  $q$ ,  $H$  is maximal when all elements have equal representation, the case of a perfect evenness. When  $p_j$  is taken as one element in a  $q$ -valued distribution then the notion of perfect evenness entails what Kullback (1959) calls an 'equidistribution' and ecologists refer to as the 'most dispersed distribution'.  $H$  is at a minimum when the distribution is most contagious, that is, when one unit represents all the  $q$  community elements, except one. We consider Rényi's (1961) entropy Eq. (2) as a benchmark expression because it generalises the notion of entropy to orders ( $\alpha$ ) from zero up to any value, except exactly Eq. (1):

$$H\alpha = \frac{\log_2 \sum_{j=1}^q p_j^\alpha}{1 - \alpha} \quad \text{bits} \quad (2)$$

In this expression,  $\alpha$  defines the order of entropy. Other symbols have the same meaning as in Eq. (1).  $H\alpha$  is appropriately termed 'generalised diversity', considering that other diversity measures such as richness, Shannon's  $H$  and the log Simpson index (Simpson, 1949) can be derived from it as special cases. In fact, when  $\alpha$  is set equal to 0, 'richness' is defined. The Shannon  $H$  is obtained when  $\alpha$  is allowed to approach 1 and the log Simpson index when  $\alpha$  is set to exactly 2. We should remark that low-order  $H\alpha$ , such as the Shannon and the log Simpson, has low discriminating power between cases (Orlóci, 1991a). In the extreme, at  $\alpha = 0$ ,  $H\alpha$  has absolutely no power of discrimination between cases for which  $q$  is identi-

cal. We note that the choice of  $\alpha$  is arbitrary. For us high-order entropy is preferable, say  $\alpha_{12}$  over  $\alpha_1$  or  $\alpha_2$ . Our reason is simply the fact of dampened rate of change in  $H\alpha$  with respect to  $\alpha$  (Orlóci, 1991a). The consideration of the entire measure profile, rather than single points, is best (Patil and Taille, 1979; Solomon, 1979; Orlóci, 1991a; Tóthmérész, 1995).

Notwithstanding the interpretations of entropy and diversity as if they were dual concepts, Taillie and Patil (1979) have argued the point that entropy and diversity are not completely equivalent concepts, based on their observation that some diversity measures fail to detect the increase in system entropy which they intuitively feel occurs. This suggests an insufficiency in the application of diversity indices. Perhaps even more misleading is the association between entropy-based diversity and complexity, and the implication that the two terms are interchangeable (McIntosh, 1967; Margalef, 1968; Woodwell and Smith, 1969; May, 1979). We refer to discussions by Hinegardner (1983), Koppel (1987) and Atlan (1988) on the clear lack of equivalency between diversity and complexity. We have taken the arguments a step further by providing a mathematical formalization of the intuitive notion that complexity is not entropy, but rather includes entropy plus something else that is not related to disorder, but rather to structure (Anand and Orlóci, 1996). Interestingly, we found that the structure component can be quite meaningful ecologically. While still based on the 'primitives' of communication theory, our definition of complexity goes beyond traditional diversity definitions and is related to the reformulation of Patil and Taille (1979) and Solomon (1979). We give a quick account of our definition in the next section and point out its relation to previous diversity measures later.

## 2.2. Complexity and coding

The central idea in our definition of complexity is that complex things are usually more difficult to describe than simple things. This is in fact an idea implicit in Kolmogorov's (1965) notion of algorithmic complexity. To apply this notion to plant communities, we needed to formalize the

idea of ‘description’. For this a coding theoretical method which translates the community record into a more parsimonious code was an obvious choice. Abramson (1963; p.46) gives formal definition of terms. We refer the reader to Appendix A for the basics of coding theory. For us the ‘message’ is the plant community composition at some defined scale, with  $s_1, s_2, \dots, s_q$  being the community elements, such as the collection of  $q$  species. Phytosociologists refer to the ‘record’ of a community composition by the French word, ‘relevé’. The relevé describes momentary composition, but if repeated at different times in the same place, the description captures the temporal chain of community states, that is, community dynamics. The average code length  $L$  is important for judging parsimony. This is given by the function:

$$L = \sum_{j=1}^q p_j l_j \quad \text{bits} \quad (3)$$

where  $l_j$  is the length of the codeword (number of 0,1 symbols) of the relative quantity  $p_j$ . The  $p_j$  are such that  $p_1 + \dots + p_q = 1$  and all  $p_j$  larger than 0. Obviously, the magnitude of  $L$  depends in some complicated way not only on the number of species ( $q$ ) and their relative abundance (the  $p_j$ ), but also on the codeword length  $l_j$ . We adopt the coding method of Huffman (1952) (please Appendix A for a worked example) which meets our criterion for parsimonious coding, in that low  $p_j$  values are encoded into long codewords and high  $p_j$  values into short codewords, somewhat analogous to Solomon’s (Solomon, 1979) ‘majorization’. What is achieved through our method is a context-independent coding. This overcomes the problem a spurious comparison of communities.

Table 1  
Levels of evenness in a 4-species community<sup>a</sup>

Species proportions	$H$	$L$	$\Delta$
[0.25 0.25 0.25 0.25]	2.00	2.00	0
[0.70 0.10 0.10 0.10]	1.36	1.50	0.14
[0.90 0.03 0.03 0.03]	0.60	1.20	0.6

<sup>a</sup>  $H$  and  $L$  accord with Eq. (2), Eq. (3) in the text.  $\Delta = (L - H)$ .  $H$ ,  $L$  and  $\Delta$  are in bits.

We use the average codeword length  $L$ , the coding effort, as our measure of algorithmic complexity. As coding theory is related to communications theory, we might expect that  $L$  be related to  $H$ . It turns out (Abramson, 1963; see Appendix A for proof) that  $L$  and  $H$  are related in the interesting way of the inequality,

$$L \geq H \quad (4)$$

Clearly, entropy  $H$  of any order is a lower bound for algorithmic complexity  $L$ . Knowing this relationship, we can view  $L$  as a measure of ‘total complexity’ and identify  $H$  as ‘disorder-based complexity’. By this reasoning, complexity includes entropy. The fact that  $L$  possesses also another component,  $\Delta = L - H$  (positive or zero) brings up the pertinent questions: When should ‘total complexity’ exceed ‘disorder-based complexity’? What does  $\Delta$  signify? In answering the first question, using a simple example (see Table 1), we find that  $H$  is equal to  $L$  when the individual (plants) are evenly arranged between the taxa in the community (all  $p_j$ ’s are equal). Under this arrangement,  $H$  reaches maximum value of  $\log_2 q$  bits. By moving away from an even arrangement, the difference becomes increasingly wide. Thus at any time, the community is not at maximum disorder, entropy is insufficient to describe complexity, and the quantity  $\Delta$  takes on special significance.

What is it about the community that could be causing the extra coding effort? What kind of complexity are we detecting by the quantity  $\Delta$ ? We show in our earlier paper that  $\Delta$  is statistically linked to properties such as vertical layering and nonlinear species responses, typically complex and universal in vegetation. Thus we concluded that  $L$  reflects not only entropy but, in addition, properties emergent under the effect of synergistic interactions and exchanges, beyond disorder. This implies the involvement of ‘design’ (Dawkins, 1986) in general and ‘self-organization’ (Nicholis and Prigogine, 1989) in particular. We termed the quantity  $\Delta$  as ‘structural complexity’ to distinguish it from ‘disorder-based complexity’. This is common sense, since at maximum disorder, there is no room for structure and hence  $\Delta = 0$ .



Fig. 1. Vertical stratification in a mixed conifer-deciduous forest stand in boreal Ontario. Photo taken near Elk Lake, Ontario, July 1995.

### 2.3. Complexity and hierarchy

So far our discussion has centred on the problem of finding a context-free measure, and not on considerations of community organisation. We give an example for the latter in Figs. 1 and 2. It is clear that complexity is related to the interaction of different levels or strata. Our model of complexity incorporates the assumption that organisation is hierarchical (Pattee, 1973; Allen and Starr, 1982; Goodall, 1986). To reflect on this further, we quote Simon (1962): ‘... If there are important systems in the world that are complex without being hierarchic, they may to a considerable extent escape our observation and our understanding’. Studies have indeed shown the importance of the effect of levels of organisation on perception of ecological complexity (Ziegler, 1979; Allen and Wyleto, 1983; Goodall, 1986; Müller, 1992; Pillar et al., 1993; Li and Müller, 1995). But these have their own definitions of complexity and hierarchy (levels). What is clear, however, is that the model to be effective cannot be a simple additive model, if for no other reason than the unalterable fact that less general levels are constrained by more general levels in complex systems.

The interdependence of levels is what characterises a nested hierarchy (Allen and Starr, 1982; Orlóci, 1991c). But even in this case, closer and closer inspection of a community (at lower and lower levels) may not lead simply to clearer and clearer resolution of its complexity. That is, complexity may not be completely decomposable among levels. This requires a relaxation of the definition of ‘hierarchical’ which we find in Li and Müller’s (1995) discussion of partial decomposition.

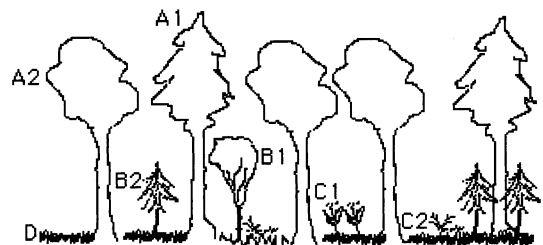


Fig. 2. Vertical stratification in an idealised community (modelled after Dansereau, 1957). Seven strata (A1, A2, B1, B2, C1, C2, D) are recognised based upon height classes of vegetation.

### 3. Modes of partitioning

As we have already pointed out, the method of hierarchical analysis of complexity will depend on the complexity function and on how the hierarchical levels are defined. Having discussed the former in brief here and in detail elsewhere (Anand and Orlóci, 1996), we now discuss the latter. We have to accept the fact that hierarchical levels emerge from observations by abstraction following decisions about how observations are made and how systems are described (Allen and Wyleto, 1983). But this does not clarify things, because the number of choices is limitless. What would be more useful would be some reasoning about why one model might be more appropriate than others in a specific case. We now discuss the different hierarchical models of partitioning.

Considering Figs. 1 and 2, the simplest partitions would directly follow the material stratification of the community by layers with each stratum representing a distinct analytical level. Partitions on this basis would reveal the complexity contribution of individual strata as if the strata were unconnected. We know that this is not a realistic condition upon which to build an analytical partition of complexity. In nature interactions between community strata are very strong which should not be excluded as a source for complexity.

In another model we would block strata in spirit not unlike the blocking of contiguous quadrats in Greig-Smith's (Greig-Smith, 1952; Greig-Smith, 1961) method of pattern analysis. If we proceeded along this alternative we would be looking for peaks in contributions by strata. The exercise would however be trivial, since in most cases there are so few strata to be blocked.

A more realistic possibility would be to follow a 'directional decomposition'. This would be done to reveal contributions of the community strata to total complexity under constraint of the direction in which the partition is approached. The directional constraint would appear at first sight too confining, but in actual fact it can be very natural in the sense that each higher stratum has an overbearing effect upon all the lower strata. If we

think of this in terms of the structure pictured in Fig. 2, the direction to follow from stratum D to A1 is logical by reasoning that it is stratum D in which the total effect of stratification in the stand is most confounded. This is because of light interception, gradation in atmospheric effects, and other interactions.

Under these conditions we can formulate a directional decomposition algorithm based upon methods analogous to Orlóci's (Orlóci, 1978) ranking techniques. In the simplest case, we have the total complexity specific to stratum A1 as the complexity in the sample {A1, A2, C1, C2, D} minus the complexity in the reduced sample without A1, {A2, C1, C2, D}. Similar reasoning gives us the complexity value specific to A2 after A1 is removed from the pool, to B1 after A1 and A2 are removed, and so on. In the end, a perfectly additive partitioning of complexity is achieved.

We recognise all of the above schemes as being potentially useful, but they rely heavily on a rigid physical structure in the community. To brake away from this constraint, we consider a nested hierarchical scheme in which the structure we actually see is steps removed from any concrete, material structure in the community. The hierarchy is analytically nested, built upon information that an association matrix conveys (2nd order structure) based on the observed stratal affiliation of species (1st order structure). Our *modus operandi* entails description of the community in terms of occupancy scores of species and stratal affiliation, computation of an association matrix (from Euclidean distances), a nested hierarchical cluster analysis (Orlóci, 1991b), and lastly, partition of complexity according to the hierarchic structure so created. The analytical nested hierarchy is the model in terms of which ecologists think of a community when they focus attention upon the invisible interactions that give coherence to the community strata.

### 4. A worked example

In the following we examine properties that exist only at the community level, not traceable directly from information about individual species

Table 2

A typical phytosociological relevé of a 10 m<sup>2</sup> plot showing stratal affiliation of the species and their abundance estimates<sup>a</sup>

Strata	I.D. no.	Species	Cover abundance
A1	1	<i>Pinus bankstana</i>	3
	2	<i>Populus tremuloides</i>	2
A2	3	<i>Abies balsamea</i>	1
	4	<i>Picea mariana</i>	1
	1	<i>Pinus banksiana</i>	1
	2	<i>Populus tremuloides</i>	1
B1	3	<i>Abies balsamea</i>	2
	5	<i>Amelanchier</i> sp.	1
	6	<i>Corylus cornuta</i>	0.1
	4	<i>Picea mariana</i>	0.1
	1	<i>Pinus banksiana</i>	0.1
	2	<i>Populus tremuloides</i>	2
	7	<i>Prunus pensylvanica</i>	1
	8	<i>Salix discolor</i>	1
	9	<i>Sorbus decora</i>	0.01
B2	3	<i>Abies balsam</i>	0.01
	5	<i>Amelanchier</i> sp.	0.1
	10	<i>Betula papyrifera</i>	0.01
	11	<i>Comptonia peregrina</i>	1
	6	<i>Corylus comuta</i>	0.1
	12	<i>Diervillea lonicera</i>	1
	13	<i>Kalmia angustifolia</i>	3
	4	<i>Picea mariana</i>	0.01
	1	<i>Pinus banksiana</i>	0.01
	2	<i>Populus tremuloides</i>	1
	7	<i>Primus pensylvanica</i>	1
	8	<i>Salix discolor</i>	1
	14	<i>Vaccinium</i> spp.	4
C1	15	<i>Aster</i> sp.	1
	10	<i>Betula papyrifera</i>	0.01
	16	<i>Cornus canadensis</i>	1
	17	<i>Cypripedium acaule</i>	0.01
	18	<i>Glyceria striata</i>	0.01
	19	<i>Mazanthemum canadense</i>	2
	20	<i>Oryzopsis pungens</i>	0.1
	21	<i>Polygala paucifolia</i>	0.1
	22	<i>Trientalis borealis</i>	0.1
C2	16	<i>Cornus canadensis</i>	1
	23	<i>Epigaea repens</i>	0.01
	24	<i>Gaultheria hispidula</i>	1
	25	<i>Unnaea borealis</i>	1
	26	<i>Pyrola rotundifolia</i>	0.01
D	27	Bareground	1
	28	<i>Cladonia rangiferina</i>	1
	29	<i>Dicranum polysetum</i>	1
	30	<i>Pleurozium schreben</i>	3
	31	<i>Sphagnum</i> sp.	0.01

<sup>a</sup> Data were collected in July 1995 near Elk Lake, Ontario. Cover-abundances are measured on the Braun–Blanquet scale.

Table 3

Directional decomposition of complexity.  $H$  is Rényi's entropy of order 1 Eq. (2),  $SI$  is Brillouin's information Eq. (6),  $L$  is average codeword length Eq. (3) and  $SL$  is a weighted  $L$  Eq. (5)<sup>a</sup>

Level	$H$	$SI$	dev (%)	$L$	$SL$	dev (%)
D	2.05242	32.8387	44.0964 (8.67)	2.2125	34	66 (6.36)
C2, D	3.07741	76.9351	74.4159 (14.52)	4	100	469 (13.4)
C1, C2, D	3.78378	151.351	182.955 (35.71)	5.975	239	469 (45.2)
B2, C1, C2, D	4.51765	334.306	90.43 (17.65)	9.56757	708	(22.1)
B1, B2, C1, C2, D	4.61669	424.736	46.502 (9.03)	10.1957	938	77 (7.43)
B1, B2, C1, C2, D	4.57513	471.238	41.153 (8.03)	9.85437	1015	22 (2.12)
A1, A2, B1, B2, C1, C2, D	14.45557	1512.391		19.01739	110371	

<sup>a</sup> The deviations columns give strata-specific contributions of  $SI$  and  $SL$  in absolute (dev) and percentage of total (%) terms, e.g.: for Strata D,  $\text{dev}(SI) = SI_{C2,D} - SI_D$ .

behaviour. We consider directional decomposition based on physical strata and hierarchical decomposition based on levels that we construct from a species association matrix as already explained above.

The community in the example is a mixed conifer-deciduous stand located in boreal Ontario (Fig. 1). The specific stratal structure is given in Fig. 2 and Table 2. We applied directional decomposition and nested decomposition as described above. The complexity to be partitioned is the weighted total complexity,

$$SL_k = f_{\cdot k} L_k, \quad k = 1, 2, \dots, c \quad (5)$$

In this  $f_{\cdot k}$  indicates total species abundance and  $L_k$  stands for total complexity (3.3) specific to hierarchical level  $k$ , with  $c$  being the number of hierarchical levels. For partitioning the entropy-based component of complexity we use Brillouin's (1962) information ( $SI$ ),

$$SI_k = f_{\cdot k} \quad (6)$$

Symbols  $f_{\cdot k}$  and  $c$  are defined the same as in (3.6).  $H_k$  represents Rényi's entropy of order one specific to hierarchical level  $k$ .  $SL$  and  $SI$  are the partitionable equivalents of average complexity ( $L$ ), and

entropy ( $H$ ). In the relationship  $SL = SI + S\Delta$ ,  $S\Delta$  is a positive quantity or zero.

The results from directional decomposition based on the data in Table 1 are presented in Table 3. The analysis identifies stratum C1, containing herb and some chamaephyte shade tolerant species, as the most significant contributor to total com-

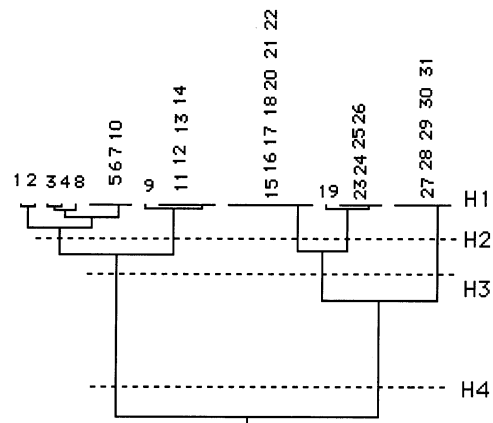


Fig. 3. The nested hierarchy of species emerging from cluster analysis of an association matrix. Four hierarchical levels (H1–H4) are shown with 31, 5, 3, 2 species groups, respectively. Please refer to Table 2 for species names corresponding to numbers 1–31.

Table 4

Partitioning of quantities through analytical nested decomposition.  $H$  is Shannon's entropy function Eq. (2),  $SI$  is Brillouin information Eq. (6),  $L$  is average codeword length Eq. (3) and  $SL$  is a weighted  $L$  Eq. (5)<sup>a</sup>

Level	$H$	$SI$	dev %	$L$	$SL$	dev %
H4	0.932112	107.193	38.838 (7.58)	1	115	40 (3.86)
H3	1.26983	146.031	89.353 (17.44)	1.34783	155	86 (8.29)
H2	0.89266	235.384	277.007 (54.06)	2.09565	241 (76.76)	
H1	14.0949	512.391		9.01739	1037	

<sup>a</sup> The deviations columns give strata-specific contributions of  $SI$  and  $SL$  similar to Table 3.

plexity ( $SL$ ). Interestingly the tree layers, A1 and A2, are the least significant contributors. The latter suggests that those elements in the community, which are most dominant, need not be those which are most directly responsible for its complexity, albeit they may as a canopy sustain complexity.

Cluster analysis of species associations yielded the analytical structure in Fig. 3. This third-order structure reflects the effect of dominance sorting on complexity in species-groups which expand their membership as the hierarchic level increases. The results are presented in Table 4. On this basis we can evaluate contributions of the species groups to complexity. The species grouping contributing most to total complexity occurs at level H2 where, interestingly, woody, herbaceous and bryophyte species dominate the groups. That a high order analytical technique finds that the community is best described through a sharp recognition of these three life-form based groups is reassuring to the ecologist observer.

## 5. Conclusion

While a context-free measurement of community complexity is achievable using  $L$ , information can be gained by examining the behaviour of the complexity function at various levels of the system. Our results show that there is varia-

tion in complexity over different levels. Based on these results direct comparisons can be made between levels to detect which level contributes most to total complexity. We described in detail two different partitioning schemes, directional through physical strata and hierarchical nested based on an association matrix. We find it interesting that in directional decomposition the most suppressed strata, the herbaceous understory, contributes most significantly to community complexity. We argue that attributing complexity to specific community properties should not necessarily rely on direct observations of structure, like stratification in a forest stand. Higher-order analytical constructs which capture the effect of stratal interactions upon complexity, not obvious from direct inspection, are more revealing of how a complex system exhibits properties not detectable from an inventory of its parts.

## Acknowledgements

The authors thank the Chinese Academy of Sciences and Professor Zu Yuangang for support at the ISEM'95 conference. The Natural Sciences and Engineering Research Council of Canada provided operating grant to L.O. and graduate scholarship to M.A. We are grateful for the helpful comments of the editor and two reviewers.



## Appendix A

### A.1. Basic coding-theoretical definitions (after Abramson, 1963)

Let the set of symbols comprising a given alphabet be called  $S = \{s_1, s_2, \dots, s_q\}$ .

Then a *code* is an invertible mapping of all possible sequences of the symbols of  $S$  into sequences of symbols of some other alphabet  $X = \{x_1, x_2, \dots, x_r\}$ .

$S$ , symbolizes the *source alphabet* and  $X$  the *code alphabet*.

A block code is a code which maps each of the symbols of the source alphabet  $S$  into a fixed sequence of symbols of the code alphabet  $X$ . These fixed sequences of the code alphabet (sequences of  $x_j$ ) are called *code words* and are denoted by  $X_i$ .

The *n*th extension of a block code which maps the symbols  $s_i$  into the code words  $X_i$  is the block code which maps the sequences of source symbols  $\{s_{i1}, s_{i2}, \dots, s_{in}\}$  into the sequences of code words  $\{X_{i1}, X_{i2}, \dots, X_{in}\}$ .

A block code is said to be *nonsingular* if all the words of the code are distinct. A block code is *uniquely decodable* if, and only if the *n*th extension is nonsingular for every finite  $n$ .

A uniquely decodable code is said to be *instantaneous* if it is possible to decode each word in a sequence without reference to succeeding code symbols.

### A.2. Average length $L$

For a given source alphabet and a given code alphabet, we can construct many instantaneous and uniquely decodable codes. The natural criterion for selection among these from the standpoint of economy of expression is length. We are interested in the uniquely decodable code with the smallest possible *average length*.

Let a block code transform the source symbols  $\{S_1, S_2, \dots, S_q\}$  into the code words  $\{X_1, X_2, \dots, X_q\}$ . Let the probabilities of the source symbols be  $\{p_1, p_2, \dots, p_q\}$  and let the lengths

of the code words be  $\{l_1, l_2, l_q\}$ . Then  $L$  is the *average length* of the code and is given by

$$L = \sum_{j=1}^q p_j l_j \quad \text{bits} \quad (\text{C.1})$$

Consider a uniquely decodable code which maps the symbols from a source  $S$  into code words composed of symbols from a binary code alphabet. This code is *compact* if its average length  $L$  is less than or equal to the average length  $L$  of all other uniquely decodable for the same source and the same code alphabet.

### A.3. Relationship between $H$ and $L$

We make use of the Kraft Inequality (Kraft, 1949),

$$\sum_{i=1}^q r^{-l_i} \leq 1 \quad (\text{C.2})$$

where  $r$  is the number of different symbols in the code and  $l_1, l_2, \dots, l_q$  the code word lengths of an instantaneous code. For simplicity we will consider codes with  $r = 2$  [binary].

Recall that

$$H = - \sum_{i=1}^q p_i \log_2 p_i \quad (\text{C.3})$$

Let  $Q_1, Q_2, \dots, Q_q$  be any  $q$  numbers such that  $Q_i \geq 0$  for all  $i$  and  $\sum_{i=1}^q Q_i = 1$

It is not difficult to prove (Abramson, 1963) that

$$\sum_{i=1}^q p_i \log_2 \frac{1}{p_i} \leq \sum_{i=1}^q p_i \log_2 \frac{1}{Q_i} \quad (\text{C.4})$$

with equality if and only if  $p_i = Q_i$  for all  $i$ .

Relation C.4 is valid for any set of nonnegative numbers  $Q_i$  which sum to 1. We may choose therefore,

$$Q_i = \frac{2^{-l}}{\sum_{i=1}^q 2^{-l}} \quad (\text{C.5})$$

and obtain

$$\begin{aligned}
H &\leq - \sum_{i=1}^q p_i (\log_2 2^{-l_i}) + \sum_{i=1}^q p_i \left( \log_2 \sum_{j=1}^q 2^{-l_j} \right) \\
&\leq \log_2 2 \times \sum_{i=1}^q p_i l_i + \log_2 \left( \sum_{j=1}^q 2^{-l_j} \right) \\
&\leq L + \log_2 \left( \sum_{j=1}^q 2^{-l_j} \right) \quad (\text{C.6})
\end{aligned}$$

The Kraft inequality tells us that argument of the second logarithm on the right of C.6 must be less than or equal to 1. The logarithm is therefore less than or equal to 0, and

$$H \leq L \quad (\text{C.7})$$

#### A.4. Huffman coding

We require a description of the plant community at a desired spatial scale commonly known in ecology as a ‘coenostate’ or ‘releve’. This includes each species ( $s_j$ ) and its relative abundance ( $p_j$ ). For the sake of facile illustration, suppose the relevé is as below. The example and its explanation is adapted from Abramson (1963) and Hamming (1980).

Source  $S$

$S_j$	$P_j$
$S_1$	0.4
$S_2$	0.2
$S_3$	0.2
$S_4$	0.1
$S_5$	0.1

Huffman (1952) developed the method which we use for constructing compact codes for the case of a binary code alphabet. We use his method. How do we translate this ‘source’ into a parsimonious code by Huffman coding?

Consider the source  $S$  with  $q$  symbols and symbol probabilities,  $p_1, p_2, \dots, p_q$ , as in the example,  $q = 5$ . Let the symbols be ordered so that  $p_1 \geq p_2 \geq \dots, p_q$  as in the example. By regard-

ing the last two symbols of  $S$  as combined into one symbol [by adding the correspond  $p_i$ ’s], we may obtain a new source from  $S$  containing only  $q - 1$  symbols. This new source is referred to as a reduction of  $S$ . The symbols of this reduction may be reordered, and again we may combine the two least probable symbols to form a reduction of this reduction of  $S$ . By proceeding in this manner, we construct a sequence of sources, each containing one fewer symbols than the previous one, until we arrive at a source with only two symbols. For the example above, the sequence of reduced sources is as follows:

	$S_1$	$S_2$	$S_3$	$S_4$
$S_1$	0.4	0.4	0.4	0.6
$S_2$	0.2	0.2	0.4	0.4
$S_3$	0.2	0.2	0.2	
$S_4$	0.1	0.2		
$S_5$	0.1			

Construction of this sequence of reduced sources is the first step in Huffman coding. The second step entails the recognition that a binary compact instantaneous code for the last reduced source [only two symbols] is trivial: it consists of two words, 0 and 1. Working backwards, in general, the compact instantaneous code for  $S_{j-1}$  is formed from  $S_j$  by the following rule:

‘One of the symbols of  $S_j$  say  $s_x$ , is formed from two symbols of the preceding source  $S_{j-1}$ . We call these two symbols  $s_{x0}$  and  $s_{x1}$ . Each of the other symbols of  $S_j$  corresponds to one of the remaining symbols of  $S_{j-1}$ . Then the compact instantaneous code for  $S_{j-1}$  from the code for  $S_j$  as follows: We assign to each symbol of  $S_{j-1}$  (except  $s_{x0}$  and  $s_{x1}$ ) the code word used by the corresponding symbol of  $S_j$ . The code words used by  $s_{x0}$  and  $s_{x1}$  are formed by adding a 0 and 1, respectively, to the code word used for  $s_x$ .’ (Abramson, 1963)

The compact Huffman code for the example is illustrated below.

	$S_1$			$S_2$			$S_3$			$S_4$		
	$p_i$	$[X_i]$	$l_i$	$p_i$	$[X_i]$	$l_i$	$p_i$	$[X_i]$	$l_i$	$p_i$	$[X_i]$	$l_i$
$S_1$	0.4	[1]	1	0.4	[1]	1	0.4	[1]	1	0.6	[0]	1
$S_2$	0.2	[01]	2	0.2	[01]	2	0.4	[00]	2	0.4	[1]	1
$S_3$	0.2	[000]	3	0.2	[000]		3	0.2	[01]	2		
$S_4$	0.1	[0010]	4	0.2	[001]		3					
$S_5$	0.1	[0011]	4									

The general case should be obvious from the example. When all  $p_i$  are of the form  $(1/2)^{\alpha_i}$ , where  $\alpha_i$  is an integer, the coding algorithm is simplified, the code word lengths  $l_i$  taken *a priori* as  $\alpha_i$  and the codes constructed by simply creating unique combinations of the code alphabet [0,1].

## References

- Abramson, N., 1963. Information theory and coding. McGraw-Hill, Inc, New York.
- Allen, T.F.H., Starr, T.B., 1982. Hierarchy: Perspectives for Ecological Complexity. The University of Chicago Press, Chicago, IL.
- Allen, T.F.H., Wyleto, E.P., 1983. A hierarchical model for the complexity of plant communities. *J. Theor. Biol.* 101, 529–540.
- Anand, M., Orlóci, L., 1996. Complexity in plant communities: notion and quantification. *J. Theor. Biol.* 179, 179–186.
- Atlan, H., 1988. Measures of biologically meaningful complexity. In: Peliti, L. and, Vulpiani, A. (Eds.), Measures of Complexity. Springer-Verlag, Berlin, pp. 112–127.
- Brillouin, L., 1962. Science and Information Theory. Academic Press, New York.
- Çambel, A.B., 1993. Applied Chaos Theory: A Paradigm for Complexity. Academic Press, New York.
- Dansereau, P., 1957. Biogeography: An Ecological Perspective. The Ronald Press Company, New York.
- Dawkins, R., 1986. The Blind Watchmaker. Penguin Books, Suffolk.
- Goodall, D.W., 1986. Classification and ordination: their nature and role in taxonomy and community studies. *Coenoses* 1, 3–9.
- Greig-Smith, P., 1961. Data on pattern within plant communities. I. The analysis of pattern. *J. Ecol.* 49, 695–702.
- Hamming, R.W., 1980. Coding and Information Theory. Prentice-Hall, Inc., New Jersey.
- Hinegardner, R., 1983. Biological complexity. *J. Theor. Biol.* 104, 7–20.
- Huffman, D.A., 1952. A method for the construction of minimum redundancy codes. *Proc. Inst. Radio Eng.* 40, 1098–1101.
- Koppel, M., 1987. Complexity, depth and sophistication. *Complex Syst.* 1, 1087–1091.
- Kraft, L.G., 1949. A Device for Quantizing, Grouping, and Coding Amplitude Modulated Pulses. MSc thesis, Electrical Engineering Department, Massachusetts Institute of Technology March.
- Kullback, S., 1959. Information theory and Statistics. Wiley, New York.
- Li, B.L., Müller, F., 1995. Application of the pansystem entropy concept to hierarchical ecosystem analysis. *Adv.Syst. Sci. Appl.* 1, 13–18.
- Margalef, R., 1968. Perspectives in Ecological Theory. The University of Chicago Press, Chicago, IL.
- May, R.M., 1979. The structure and dynamics of ecological communities. In: Anderson, R.M., et al. (Eds.), Population Dynamics. Blackwell, Oxford.
- McIntosh, R.P., 1967. An index of diversity and the relation of certain concepts to diversity. *Ecology* 48, 392–404.
- Müller, F., 1992. Hierarchical approaches to ecosystem theory. *Ecol. Model.* 63, 215–242.
- Nicholis, G., Prigogine, I., 1989. Exploring Complexity. Freeman and Co, New York.
- Orlóci, L., 1978. Multivariate Analysis in Vegetation Research. Dr W. Junk, bv, the Hague.
- Orlóci, L., 1991. Entropy and Information. In: Ecological Computations Series (ECS), vol. 3. SPB Academic Publishing bv, The Hague.
- Orlóci, L., 1991. Ecological programs for Instructional Computing on the Macintosh. In: Ecological Computations Series (ECS), vol. 2. SPB Academic Publishing bv, The Hague, p. 131.
- Orlóci, L., 1991. On character-based plant community analysis: choice, arrangement, comparison. *Coenoses* 6, 103–107.
- Patil, G.P., Taille, C., 1979. An overview of diversity. In: Grassle, J.F., Patil, G.P., Smith, W., Taille, C. (Eds.), Ecological Diversity in Theory. International Co-operative Publishing House, Fairland, MD.

- Pattee, H.H., 1973. *Hierarchy Theory: The Challenge of Complex Systems*, G. Braziller, New York.
- Pielou, E.C., 1975. *Ecological Diversity*. Wiley, New York.
- Pillar, de Patta, V., Orlóci, L., 1993. Taxonomy and perception in vegetation analysis, *Coenoses* 8, 53–66.
- Rényi, A., 1961. On measures on entropy and information. In: Neyman, J. (Ed.), *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst. Technol.* 27, 379–423.
- Simon, H.A., 1962. The architecture of complexity. *Proceedings of the American Philosophical Society* 106, 467–482.
- Tóthmérész, B., 1995. Comparison of different methods of diversity ordering. *J. Veg. Sci.* 6, 283–290.
- Simpson, E.H., 1949. Measurement of diversity. *Nature* 163, 688.
- Solomon, D.L., 1979. A comparative approach to species diversity. In: Grassle, J.F., Patil, G.P., Smith, W., Taillie, C. (Eds.), *Ecological Diversity in Theory and Practice*. International Co-operative Publishing House, Fairland, MD.
- Taillie, C., Patil, G.P., 1979. Diversity, entropy and the Hardy–Weinberg equilibrium. In: Grassle, J.F., Patil, G.P., Smith, W., Taillie, C. (Eds.), *Ecological Diversity in Theory and Practice*. International Co-operative Publishing House, Fairland, MD.
- Woodwell, G.M., Smith, H.H., (eds.) 1969. Diversity and stability in ecological systems. *Brookhaven Symposium, Biology*. 22.
- Ziegler, B.P., 1979. Multilevel multiformalism modeling: an ecosystem example. In: (Halfon, E (Ed.), *Theoretical Systems Ecology*. Academic Press, New York.