

## **Entropy and Community Pattern Analysis<sup>†</sup>**

MICHEL PHIPPS

*Department of Geography, University of Ottawa, Ottawa, Canada*

*(Received 18 May 1981)*

A conceptual framework for the analysis of community types pattern (mosaic) is proposed and discussed in this paper. Within this framework, the pattern is viewed as a system whose freedom to take any particular state from among a set of possible microstates is restrained by ecological factors. This allows for the operationalization of an analytical procedure (PEGASE) based on information theory whereby the entropy of the pattern is reduced by a negentropy provided by various environmental factors.

The introduction of these factors in a stepwise manner gives this method two of its essential characters: (i) its multivariate nature and (ii) the classification of sites in a divisive manner. Given a set of environmental factors, the level of organization of the pattern can be measured by the global environmental negentropy.

Several properties of this model are discussed. Similarly to some early attempts at biophysical classification, the "Pegase" procedure emphasizes the ecological relationships. Indeed, it classifies sites into a set of site type which allows for the best prediction of community types. On the other hand, its basic structure is similar to that of a multiple regression model, without requiring any of the assumptions on data structure and distribution that multiple regression requires.

### **1. Introduction**

The spatial patterns associated with ecological systems can be studied in many different ways. Landscape analysis (Gardiner & Gregory, 1977), biophysical inventories and ecological land classifications (Christian, 1957; Schneider, 1966; Mabbutt, 1968; Hills, 1976) ecological gradient analysis (Whittaker, 1973a; Orloci, 1973; McIntosh, 1967; Strahler, 1978) are examples of such studies. All share the aim of understanding how biological and physical features relate to each other in space and how space can be structured on the basis of these relationships.

These various aspects, however, have not received the same kind of attention. Since the fifties, the need for improved land use planning has

<sup>†</sup> These ideas were developed while the author was engaged in contractual researches for: Atelier de Recherche et d'Application, Toulouse, France; Mont-Ste-Marie Ltd., Ottawa, Canada, and the Forest Administration of Morocco, Rabat, Morocco.

fostered the concern for regionalization. As a result, many ecological land classifications have been produced in various countries (see, for instance, Christian, 1957; Perelman, 1960; Isachenko, 1973; Bertrand, 1969; Sochava, 1970; Hills, 1976; Jurdant, 1976). Most of this research deals with empirical investigations attempting to build classification systems upon biological and physical interdependencies. These relationships are seldom approached as a system, although Ruxton long ago voiced the need for a better understanding of the theoretical aspect of "order and disorder in land" (Ruxton, 1968).

The present paper addresses this aspect of the problem by developing a theoretical and methodological framework for the analysis of plant community patterns. For the purpose of this endeavour, this analysis will be defined as an attempt to determine to what extent the spatial differentiation of plant communities can be considered an ordered pattern. The order sought does not rely on purely spatial configurations but on the correlative variation of vegetation and other distinct features. In this sense one may speak of ecological order.

The nature of this research—a search for order—strongly suggests information theory as the most appropriate conceptual framework. Various concepts such as entropy, negentropy, redundancy and information channel provide useful tools to gain deeper insight into the ecological organization of space.

The first section of this paper focuses on theoretical considerations. It stresses the fact that the conventional entropy maximizing approach does not apply in the present case. It, then, suggests an alternative aimed at acquiring the maximum negentropy from a knowledge of the spatial covariations of the features considered. The second section presents an original procedure based upon this alternative. Finally, an application of this procedure to the analysis of a forest landscape is examined in the third section.

## 2. Theoretical Considerations

Ecologists have long shown interest in information theory. Shortly after its introduction, attempts were made to apply it in various areas such as diversity studies (Margalef, 1956, 1958; Pielou, 1966*a,b,c*) system modeling (Pattern, 1959), classification procedures (Rescigno & Maccaccaro, 1960; Williams, Lambert & Lance, 1966) and the measurement of ecological relationships (Godron, 1968; Guillermin, 1971). More recently, geographers have adopted similar approaches based on entropy—especially on entropy maximizing model—and these have become a well established methodology for the investigation of various geographical problems (see

among others: Leopold & Langbein, 1962; Medvedkov, 1967; Berry & Schwind, 1969; Gurevich, 1969; Wilson, 1970; Semple & Wang, 1970; Marchand, 1972; Batty, 1972; Allaire, Phipps & Stoupy, 1973; Thomas & Reeve, 1976; Webber, 1977).

While ecologists have explored the usefulness of this theory in various directions, it is striking to note that very few applications have been attempted in fields related to physical geography. One reason for this situation might be as follows. The conventional approach to spatial system analysis generally consists of predicting the actual state of a system by maximizing its entropy subject to a set of hypothesized constraints. These are expressed by means of mathematical functions which can be probability functions (Thomas & Reeve, 1976; Webber, 1977) or any general function (Wilson, 1970) such as a global transportation cost, for example (Nijkamp & Paelinck, 1974). In these instances, the observed state of the system provides the means of testing the validity of the hypotheses. Applications to physical systems might differ from this conventional approach in a major way because the constraints acting upon a physical system can often hardly be expressed in mathematical terms. Thus use of the types of functions mentioned above is, for all practical purposes, untenable. This suggests an inverse procedure whereby the constraints acting upon the system are discovered from the analysis of the observed state of the system. Such approaches are common in ecology (Pielou, 1966*b*) and will be followed here.

## 2.1. ENTROPY OF A CHOROPLETH PATTERN

Let us consider the spatial differentiation of vegetation in a given region. Even if this variation may, in theory, be considered a spatial continuum (McIntosh, 1967; Sobolev & Utekhin, 1973), it is usually described by a set of discrete units (such as community types; Whittaker, 1973*b*) derived from a classification process. This spatial variation can, therefore, be represented by a choropleth pattern to which may be assigned a spatial entropy (Gurevitch, 1969; Pipkin, 1977; Robinson, 1978) given by:

$$H(v) = - \sum_{j=1}^m p_j \text{Log}_e p_j \quad (1)$$

where  $m$  is the number of possible outcomes of the pattern (i.e. the number of community types in the area) and  $p_j$  the probability of the  $j$ th outcome in the spatial unit being observed.

If this unit is considered to be a collection of  $N$  cells, the pattern constitutes a spatial system whose macrostate is defined by  $N$ , the  $m$

outcomes and a frequency array  $f_j (j = 1, m)$  where  $f_j$  is the number of cells associated with the  $j$ th outcome. A microstate is any particular spatial arrangement of the  $N$  cells distributed in the  $m$  outcomes so as to satisfy the values of  $f_j$  (Wilson, 1970; Thomas & Reeve, 1976). In other words, a microstate is a particular spatial configuration which could possibly occur given the macrostate's characteristics. The observed pattern is just one of these numerous arrangements, the number of which is given by:

$$W(v) = \frac{N!}{\prod_{j=1}^m f_j!} . \quad (2)$$

For a system with a finite number of elements, the entropy is given by Boltzmann's formula (Brillouin, 1951, 1962):

$$S(v) = \frac{1}{N} \log_e W(v) . \quad (3)$$

Among the state functions introduced in this section, entropy is, by far, the most important one. As emphasized by most authors, it represents the uncertainty attached to the state of the system, or in other words, the freedom for the system to take various states and its information content as well.

## 2.2. ASSUMPTIONS UNDERLYING THE ENTROPY CONCEPT

At this stage, it is necessary to recall a basic assumption underlying the entropy concept. All the microstates that can be associated with the frequency array  $f_j (j = 1, m)$  are assumed to be equally likely to occur. This implies that anyone of these states, including the observed pattern, has a probability of occurrence given by:

$$p[v] = \frac{1}{W(v)} \quad (4)$$

Furthermore, it implies that the probability of any two cells bearing a particular community type is equal over the whole pattern. This is evident from equation (2) which shows that the number  $W(v)$  can only be attained if all cells of the pattern are interchangeable and indistinguishable from each other with regard to the type of community they bear. This, of course, is tantamount to assuming that the observed area is a uniform space.

This assumption opposes the basic concept of an ecological organization of space according to which cells differ in such characteristics as soil,

topography and mesoclimate, and therefore provide community types with various living conditions. Each community type tends to occupy particular types of sites and this is reflected by the probability of encountering this particular type at various locations. Thus, in contrast to the assumptions implied by the entropy concept, we hypothesize that cells are not comparable with regard to their probability of bearing various vegetation types. They must be distinguished and classified according to homogeneous probability groups. As will be shown later, the search for such groups makes up the bulk of the spatial and ecological analysis to be undertaken.

### 2.3. ECOLOGICAL ORDER OF A CHOROPLETH PATTERN

The inconsistency pinpointed in sections 2.1 and 2.2 between the entropy model as applied so far and the concept of ecological organization allows for formulation of a new set of assumptions upon which further investigation can be based.

If it is assumed that more order exists in the system, then it follows that:

- (i) the actual entropy of the system is less than calculated by equations (1) and (3) and, hence, that the number of possible patterns is less than in equation (2); in other words, the system contains a negentropy which represents a constraint limiting its freedom, thus preventing it from achieving all its potential entropy;
- (ii) since previous results were arrived at by using all the available information contained in the community pattern, only new information sources can provide further negentropy; more precisely this will be provided by other characteristics such as soil, topography, climatic conditions, etc.;
- (iii) since the constraint mentioned above can hardly be expressed in terms of a mathematical function, the hypothesis (i) will be directly tested through a statistical analysis of the system (Wilson, 1970).

Henceforward, the investigation of the pattern will consist of seeking negentropy in other ecological features. For example, the initial set of cells forming the pattern can be split on the basis of the types of soils. If these subsets are characterized by distinct frequency arrays regarding the community types, thus representing a differentiation of the pattern, they reveal an ecological organization of space. The next section will show how this operation reduces the system's entropy. In terms of information theory, we may say that looking at the vegetation through the *information channel* of an ecological characteristic, introduces negentropy into the system, thus revealing some order in the pattern.

## 2.4. INFORMATION CHANNEL AND MUTUAL INFORMATION

Within the framework of this information channel, the relation between the two characteristics ( $v$ : the set of community types and  $s$ : the set of soil types) takes the form of a cross-tabulation where:

$f_{ij}$  denotes the number of cells where the  $j$ th type of community is associated with the  $i$ th type of soil ( $j = 1, m$  and  $i = 1, n$ );

$p_{j/i}$  denotes the conditional probability of the  $j$ th community type given a cell with a type of soil  $i$ ; and

$p_j$  denotes the probability of the same type of community for the whole pattern.

We also have:

$$\sum_{j=1}^m f_{ij} = f_{i\cdot} \quad (5)$$

and

$$\sum_{i=1}^n f_{ij} = f_{\cdot j} \quad (6)$$

According to section 2.3, we assume that the characteristic  $s$  introduces some order in the pattern and we define this order as a differentiation of the vegetation probability arrays attached to the various types of soils. We may thus write:

$$p_{j/i} \neq p_j \text{ at least for one } i \quad (7)$$

and test this inequality by using the frequencies of the observed pattern.

Let us first consider  $W_s(v)$ , the number of spatial configurations that can be achieved within the framework of the information channel. Inequality (7) justifies splitting the initial set of cells according to the type of soil. Within each of these subsets, cells combine with each other to give a partial pattern  $i$ . The number  $W_i(v)$  of possible arrangements in forming this partial pattern is given by:

$$W_i(v) = \frac{f_{i\cdot}!}{\prod_{j=1}^m f_{ij}!} \quad (8)$$

The overall pattern will then be the result of the combination of the  $n$  partial patterns and  $W_s(v)$ , the number of arrangements obtainable from this combination is the product of  $W_i(v)$  over  $i = 1, n$ :

$$W_s(v) = \prod_{i=1}^n \left( \frac{f_{i\cdot}!}{\prod_{j=1}^m f_{ij}!} \right) \quad (9)$$

Taking equations (5) and (6) into account,  $W_s(v)$  is equal to or smaller than  $W(v)$  calculated by equation (2). Since entropy is a monotonic function of the number of states, it is clear that the entropy of the system, given by

$$S_s(v) = \frac{1}{N} \log_e W_s(v) \quad (10)$$

has also been reduced.

Furthermore, the relation between  $W(v)$  and  $W_s(v)$  can be expressed by the equation:

$$W(v) = W_s(v) \times r \quad (11)$$

where  $r \geq 1$  is the factor by which the number of possible states has been reduced and which represents a factor of order (Brillouin, 1951). The multiplicative form of equation (11) is consistent with the combinatorial nature of  $W(v)$  and  $W_s(v)$ . The logarithm form of equation (11) gives:

$$\log_e r = \log_e W(v) - \log_e W_s(v). \quad (12)$$

Thus, according to equations (8) and (9):

$$\log_e r = \log_e \left( \frac{N!}{\prod_{j=1}^m f_{.j}!} \right) - \log_e \left[ \prod_{i=1}^n \left( \frac{f_{i.}!}{\prod_{j=1}^m f_{ij}!} \right) \right]. \quad (13)$$

This can also be written:

$$\log_e r = \log_e (N!) - \sum_{j=1}^m \log_e (f_{.j}!) - \sum_{i=1}^n \log_e (f_{i.}!) + \sum_{j=1}^m \sum_{i=1}^n \log_e (f_{ij}!). \quad (14)$$

If numbers  $N$  and  $f$  are large enough, equation (14) can be estimated, using Stirling's formula (Brillouin, 1951), by:

$$\log_e r = N \log_e N + \sum_{j=1}^m \sum_{i=1}^n f_{ij} \log_e f_{ij} - \sum_{j=1}^m f_{.j} \log_e f_{.j} - \sum_{i=1}^n f_{i.} \log_e f_{i.} \quad (15)$$

which calculates the quantity known as the mutual information of the channel (Orloci, 1970; Marchand, 1972).

This quantity measures the amount of information transmitted without noise by the channel (Wilson & Wilson, 1965). That is, the amount of uncertainty regarding the communities which has been removed by a previous knowledge of the ecological characteristic (the type of soil in this

example). As shown above, it also represents the order introduced by this characteristic into the system.

It is of interest to note that  $r$  is related to the  $\chi^2$  statistic by the relation (Kullback, 1959):

$$2 \log_e r \sim \chi^2 \\ (m-1)(n-1)d. \text{ of } f. \quad (16)$$

Equation (16), therefore, allows for the testing of the significance of the mutual information of the channel. By the same token, we are able to test the hypothesis put forward at the beginning of this section. If this hypothesis proves to be true, the ecological characteristic must be regarded as an ecological factor which structures the pattern.

To sum up this section: reviewing some classical results of information theory has shown that:

- (i) the entropy, as calculated in equations (1) and (3) represents a potential freedom for the system which cannot actually be achieved since the system is restrained by an ecological factor;
- (ii) on the basis of the frequencies displayed by the observed pattern, mutual information (equations (12–16)) measures this constraint and allows for the testing of its significance;
- (iii) among other things, mutual information expresses a negentropy, that is a reduction of the number of possible states of the pattern (i.e. of its possible configurations), as well as the amount of order introduced into the spatial system by an ecological factor;
- (iv) the reduction of the system's freedom to adopt various states, entails a higher probability of the observed pattern according to the following set of equations and inequalities:

$$p_s[v] = \frac{1}{W_s(v)} \text{ and } p[v] = \frac{1}{W(v)}$$

and, hence

$$p_s[v] > p[v] \text{ since } W_s(v) < W(v). \quad (17)$$

Finally, the previous theoretical considerations supply the operational framework within which we may attempt to answer the initial question: "To what extent can a pattern of community types be considered an ordered configuration?" From these considerations, it is clear that there will be order insofar as various ecological factors will provide negentropy and, hence, increase the probability of the observed pattern.



### 3. Methodological Considerations

#### 3.1. GENERAL PROPERTIES OF THE MODEL

With these main objectives in mind, a procedure for a spatial and ecological analysis of the pattern has been developed from the basic approach discussed above. Assuming that the ecological organization of the pattern  $v$  rests upon several characteristics ( $s_1, s_2 \dots s_l$ ) one may write:

$$v = \psi(s_1, s_2, \dots s_k, \dots s_l). \quad (18)$$

In addition to its multivariate nature, the function  $\psi$  must satisfy several other conditions which will now be reviewed.

The first condition deals with the variable scale problem. All components of equation (18) are treated as nominal scale variables. On one hand, this enables us to handle characteristics corresponding to various scale types together in the same manner. On the other hand, some information is lost when quantitative variables are treated as nominal ones. However, this loss seems more acceptable than creating artifacts by treating all characteristics as quantitative and ascribing to some of them non-existent quantitative properties.

The second property deals with the need to consider the many interactions that occur between  $v$  and the set of environmental characteristics. The effect of a particular state of a characteristic  $s_k$  on  $v$  can change according to its combination with the state of other characteristics. This point is crucial, since most primary ecological factors responsible for the spatial variation of community types (such as the temperature) remain largely unknown at the scale of the community pattern. Because of this, the analysis rests upon surrogates. The function must therefore take interactions into consideration so as to enable us to interpret, with particular combinations of surrogates, those effects actually due to the primary factors.

The third and last condition refers to the stochastic nature of the function. Undoubtedly, the model sought should include a degree of uncertainty although the goal of the method is to reduce it. There are two reasons for this. On one hand, the set of ecological characteristics might not contain all the relevant information with regard to the spatial distribution of the communities. On the other hand, the very nature of the relationship may well be seen as a non-deterministic one. As Gould (1961) has stated, stochastic models account for both the true uncertainty of a system and the lack of information about it.

These three conditions are satisfied by an information channel whereby a set of states of environmental characteristics relates to the set of community types through a stochastic matrix. The difference between the

function  $\psi$  and the channel previously examined in this paper lies in the structure of the set of environmental states. In the first case, we had a set of states regarding a single characteristic whereas, in the second, the model is based upon a set of complex states resulting from the combination of several characteristics. The main problem with this procedure is, therefore, in providing the means of setting out these combinations.

### 3.2. NEGENTROPY MAXIMIZING PROCEDURE: PEGASE<sup>†</sup>

This procedure was first applied in an analysis of the ecological organization of rural landscapes (Allaire *et al.*, 1973). It is designed to reach the maximum negentropy on  $v$  which the combination of the characteristics  $s_k$  ( $k = 1, l$ ) can provide. More precisely, the procedure will construct these combinations by bringing in the characteristics in a stepwise fashion. This raises two questions, namely, the question of path and that of the number of possible combinations. Depending on the order in which the characteristics are brought into the combinations, there are numerous ways of achieving the above mentioned goal. From among these numerous paths the procedure will choose the most efficient one, that is the path which reaches the objective most rapidly. This means that, at each step, the characteristic to be brought in will correspond to the one which provides the largest amount of negentropy.

The second deals with the vast number of possible combinations that can be formed from a relatively small number of environmental characteristics. Although the optimized path mentioned above somewhat reduces this number, the latter must be limited to those combinations which are really significant in terms of negentropy created.

The application of both principles—optimization of the path and economy in setting up the combinations—defines more precisely the stepwise process to be developed.

This process starts with a sample of  $N$  cells, each of which is defined with regard to:

- (i) the community type;
- (ii) the state of each characteristic  $s_k$ .

Let  $S_0(v)$  be the initial entropy associated with the pattern  $v$ , and calculated according to equation (3).

<sup>†</sup> The procedure has been integrated in a computer routine called "Pegase" which stands for:

"Partition d'un Ensemble Géographique  
pour l'Analyse Spatiale Ecologique".

### 3.2.1. *First step*

All  $l$  characteristics  $s_k$  are successively tested against the community types distribution. In all cases the mutual information is calculated. Let  $s_h$  be the characteristic that shows the largest significant mutual information. The initial set of cells is then split into subsets corresponding to the states of  $s_h$ . Nevertheless, two or more subsets are kept together when they have frequency vectors that do not differ significantly from each other according to the Kullback criterion ( $m-1$  d. of  $f$ ) previously mentioned. This results in reducing the number of groups formed. At the same time, this differential information regarding the state of  $s_h$  remains available for use in subsequent steps.

At the first step, these operations secure the most efficient progression toward maximum negentropy, while keeping the number of subsets at its lowest level.

### 3.2.2. *Subsequent steps*

Exactly the same procedure is applied independently to each of the subsets produced in the preceding step. The process continues, a subset produced at any step becoming an initial subset at the following step, until a stopping rule intervenes.

### 3.2.3. *Stopping rules*

For the process to stop, certain conditions are required, The stopping rules correspond to three types of situations.

- (i) The first type is a matter of technical convenience. The process has to be stopped when the number of cells in a subset is less than the minimum needed to implement the statistical operations required by the procedure.
- (ii) The second occurs when the entropy of a subset is low enough to consider, for all practical purposes, that the corresponding group of cells is almost determined regarding the community type. A minimum value can be set up under which a subset is considered a terminal one.
- (iii) The third type occurs when the entropy of the subset is above the minimum value but when none of the available characteristics (those which have not been used yet) adds significant negentropy. This means, of course, that a certain amount of uncertainty which does not depend on the known information remains. It can be true uncertainty or lack of information, as stated before.

The end product of the process is a number of subsets which could not be further subdivided. These subsets are related to each other within a



dendrogram (see Fig. 1). This graph shows that the strategy is that of a divisive classification (Williams *et al.*, 1966; Spence & Taylor, 1970). Each terminal subset represents a combination of environmental features, as well as the class of cells defined by this combination. Such a combination is made up of the different states of characteristics which have been used along the branch of the dendrogram ending in that particular subset. It also defines the complex state of environmental conditions prevailing in the class of cells.

Moreover, the method's theoretical basis suggests that the fundamental property of these classes of sites lies in their homogeneity regarding the community types probabilities. In other words, cells within a particular class are indistinguishable from each other whereas they are not comparable from one class to another one.

#### 4. Application of the Method

The "Pegase" procedure has been applied to a number of cases representing different types of problems. The study of the organization of a forest landscape located in Mont-Ste-Marie (Québec) provides an example<sup>†</sup>. Prior to the application of the analysis scheme, the vegetation was mapped over an area of approximately 14 km<sup>2</sup> in the hilly region of the Laurentides (Canadian shield). Toward this aim, a set of 10 community types was first identified through procedures usually followed for this purpose. These included the sampling of 63 sites, and then a numerical classification based on an information criterion (Dale & Anderson, 1972). The latter yielded 14 community types of which 4 were discarded because of their very low occurrence. An ordination by correspondence analysis (Escofier-Cordier, 1969; Hill, 1974; Legendre & Legendre, 1979, p. 130)—also known as reciprocal ordering (Orloci, 1978)—revealed several cenoclines which, after being interpreted as ecoclines, suggested some of the main abiotic features controlling the spatial distribution of vegetation. The community types were then mapped at a scale of 1/8400 following several reconnaissance transects and interpretation of various aerial photographs. The 10 community types displayed on the map are: *beech-red oak*; *sugar maple-beech*; *s. maple-white birch*; *r. oak-red and white pines*; *r. oak-large tooth poplar*; *s. maple-basswood*; *r. oak-ironwood*; *s. maple-ironwood*; *balsam fir-white spruce*; *w. pine-b. fir*.

<sup>†</sup> This example is taken from C. Morin's work: "Modèle d'organisation spatiale et écologique des types de communautés végétales de la région du Mont-Ste-Marie", M.A. Thesis, University of Ottawa, 1979.

For clarity, the procedure has been presented in terms of cells and Boltzmann's entropy formula for a finite set of elements. In practice, it is often difficult to determine homogeneous cells. Instead, we may use a sample of sites drawn from an infinite population. Entropy and the related concepts will then apply according to Shannon's formula. For calculation purposes, probabilities are estimated by relative frequencies.

The sample of sites used in the "Pegase" procedure consisted of 1475 sites taken from the map by applying a systematic square sampling grid with a distance of approximately 106 m between neighbouring sites. In addition to the community type, the information for each site included 9 environmental features: *elevation, direction of slope, slope gradient, relative elevation, site profile, site drainage, soil parent material, soil drainage and depth of overburden material*. From the initial 1475 sites, 1215 were defined for all information items and were finally used.

Figure 1 shows the upper part of the corresponding dendrogram, revealing the landscape model of organization. A thorough understanding of this ecological organization can be achieved by interpreting this figure and by pointing out the shift of the frequency values which occurs along the progression toward greater negentropy. However, for the purpose of this paper, the emphasis is placed on that progression itself.

At any point in the process, the system's entropy can be calculated according to:

$$H_t(v/c) = \frac{1}{N} \sum_{u=1}^t [H_u(v/c) \times f_u] \quad (19)$$

where  $t$  is the number of subsets provided by the combination  $C$  of all the environmental features;  $H_u(v/c_u)$  and  $f_u$  are the entropy and the size of the subset corresponding to the particular combination  $c_u$ .

In this example, the initial entropy

$$H_0(v) = 2.17 \quad (20)$$

decreases along with the procedure to a final value

$$H_T(v/c) = 1.01 \quad (21)$$

following the elaboration of 124 subsets from an initial set of 1215 cells.

Between the initial and the final entropy value, intermediate values can show the progress of the negentropy as a function of the number of subsets ( $t$ ). The negentropy gain can also be expressed by the redundancy:

$$R_t = 1 - \frac{H_t(v/c)}{H_0(v)} \quad (22)$$

and represented as a negative exponential function of  $t$ .

At this stage, the following problem must be raised: the divisive strategy used in "Pegase" requires starting with a large sample, thus allowing the process to run as long as possible without restraint from technical stops due to the limited number of cells in some subsets. Although the initial set of cells is generally large, such stops are frequent toward the end of the procedure. This makes it difficult to estimate the level of order which actually exists directly from the experimental data. However, it is possible to derive a theoretical function within the domain of validity of the experimental function, that is to say, in the domain where the technical stopping rule does not operate. This theoretical function is then extended by continuity providing an estimate of the final redundancy of the system.

In this example, the theoretical function is expressed by:

$$R_t = 0.6 - 0.615 e^{-0.0257t} \quad (23)$$

and is represented in Fig. 2.

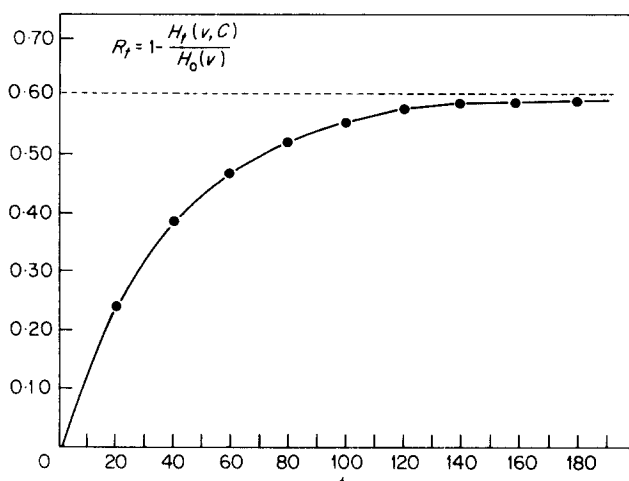


FIG. 2. Function linking the progress toward greater negentropy and the number of subsets.

Equation (23) shows that:

$$R_t \rightarrow 0.60 \text{ when } t \rightarrow \infty$$

This asymptotical value constitutes the best estimate of the landscape's degree of ecological organization.

Let us consider the set of  $t = 124$  terminal groups. The corresponding redundancy is given by equation (23):

$$R_T = 0.60. \quad (24)$$

From this value, the negentropy can be calculated by

$$I_T(v/c) = H_0(v) \times R_T \quad (25)$$

and, hence:

$$\log_e r = N \times H_0(v) \times R_T \quad (26)$$

which gives:

$$\log_e r = 1582 \quad (27)$$

with

$$\begin{aligned} H_0(v) &= 2.17 \\ R_T &= 0.60 \end{aligned} \quad (28)$$

and

$$N = 1215.$$

As noted, previously,  $r$  represents the factor by which the *a priori* number of possible patterns is reduced by the information channel. We can therefore write:

$$W(v) = W_C(v) \times e^{1582} \quad (29)$$

and this equation is a measure of the ecological order introduced into the spatial system, given the information which has been used in this analysis.

This overwhelming number has to be compared to  $W(v) = e^{2637}$  and  $W_C(v) = e^{1055}$  which are respectively the initial and final number of possible arrangements of the pattern. In fact, using the procedure does not require handling such numbers since it works on

$$H(v) = \frac{1}{N} \log_e W(v)$$

a number of a magnitude order of 1 and whose maximum is  $\log_e m$  (2.30 in the case of this example).

## 5. Discussion

This analysis scheme has been referred to as a spatial analysis. However, space itself (e.g. the  $x$  and  $y$  co-ordinates of the plane) is by no means considered a structuring feature of the landscape. As pointed out in the introduction, "the order sought does not rely on purely spatial configurations but on the correlative variation of vegetation and other distinct



features". This remark suggests that the analysis relates less to spatial relationships, as discussed for instance by Pielou (1975) or Matern (1979), than to ecological relationships. One must therefore determine to what extent this scheme deserves to be termed spatial.

We may first notice that the systematic sampling technique maximizes the spatial information by giving each geometric subdivision of the space an even representation in the sample. Because of this, the sample provides a most accurate image of the actual pattern. Ultimately, this approach attempts to establish the probability with which the known ecological constraints should have given rise to the observed pattern. Moreover, since the procedure operates as a site classification, it entails a great number of chorological considerations. Of particular interest is the fact that these choric classes contain sites which are comparable with regard to their probability of bearing the various community types. This property suggests the use of these classes as operational spatial units for management purposes. All these aspects, in our opinion, strongly emphasize the spatial slant of this approach.

Any attempt to inquire into the ecological organization of land requires a preliminary choice of characteristics which, by hypothesis, structure the landscape. This is evident from the land classification literature where, in most examples discussed, features chosen *a priori* are assumed to be the most effective ones in structuring the landscape (Gardiner & Gregory, 1977; Schneider, 1966; Mabbutt, 1968). How well these characteristics account for the real organization of space is a question neither novel nor, to any extent, particular to any one approach. The preceding section of this paper makes it clear that the ecological order shown in the pattern rests on the set of features introduced in the procedure. However, information theory implies a measure of each characteristic's structuring effectiveness, thus giving this method an heuristic value in contrast to many ecological land classifications.

As mentioned above, "Pegase" is a divisive classification. More precisely, it operates according to a process which can be depicted as *divisive*, *hierarchic*, *multivariate* and *monothetic* (Williams *et al.*, 1966; Spence & Taylor, 1970). There are large numbers of these classifications and it is of interest to see how this particular one compares to others with respect to their basic structure. Many early attempts to derive regions by divisive classifications were "genetic and deductive in nature" (Mabbutt, 1968). They were based on procedures involving a great deal of subjectivity (Spence & Taylor, 1970). However, it is interesting to point out, (following Grigg, 1961), that "the criteria on which the regional delimitation was made were frequently a property not of the thing regionalized but some

element which was presumed to be largely responsible for the distribution studied." In this respect, the method proposed here bears strong resemblance to these works and to some predictive classification schemes designed to predict the state of an environmental factor from the presence of particular species (MacNaughton, 1963; Orloci, 1978, p. 237). In combining several environmental features so as to lessen the entropy of the community types pattern, it establishes multi-featured environmental regions designed to provide the best prediction of areas based on a different feature: the vegetation. This predictability is measured by the final entropy of the system. It constitutes the fundamental property of the function  $\psi$ . Since the combination building process implies the selection of a few combinations—the most effective ones—from among an overwhelming number of possibilities, it contributes to a large extent to the heuristic value of this approach. Moreover, the mathematical formalization of the method and the mathematical tools used in the procedure make it possible to somehow overcome the criticism of subjectivity previously mentioned (Spence & Taylor, 1970).

If compared to other numerical divisive procedures, such as Association Analysis, Edwards and Cavalli-Sforza's Classification, McNaughton's Classification and Information Analysis (Williams *et al.*, 1966; Whittaker, 1973*b*; Goodall, 1973), "Pegase" draws its originality from the characteristic mentioned above. All other procedures are meant to delineate, on the basis of floristic criteria, groups of vegetation samples with a relative homogeneity with respect to their floristic composition. In contrast to "Pegase", these methods therefore make use of classification criteria which are of the same nature as the groups to be derived. The procedure proposed here places emphasis on the relationships (i.e. the information channel) between the two sets of elements: the environmental characteristics on one hand and the community types pattern on the other. To what extent these environmental characteristics allow an accurate prediction of vegetation is a fundamental concern here, since the answer to this question represents the understanding of the ecological and spatial relationships between both sets of elements.

Because of its predictive nature, "Pegase" recalls to mind the multiple regression model. However, in contrast to this latter, "Pegase" operates without any restriction on data structure and distribution. This is of interest in such application as spatial and ecological analysis where, as stated previously, most variables lack of the continuity and linearity required by multiple regression model.

Finally, "Pegase" also provides a tool which allows for the reduction of a system's entropy without requiring a mathematical expression of the

system's constraints. This, in our view, is crucial to consider if information theory is to become of common use in physical geography.

## 6. Conclusion

This paper proposes a theoretical framework within which to investigate vegetation choropleth pattern. Although this framework provides a tool that measures the degree of ecological organization of such a pattern, it should not be reduced to this. Quoting Gurevitch (1969) one could say that it was "designed to construct a conceptual system and a structural-mathematical edifice that will make possible deeper, more comprehensive and subtler investigations of geographical phenomena and processes".

Indeed, this information approach to spatial and ecological analysis provides the foundations for possible further developments. By allowing the measurement of ecological order in an area, it could provide the means of assessing organizational changes over time, due to natural or man-made causes (Pielou, 1966*b*). From another standpoint it affords the possibility of making an in depth analysis of the ecological relationships underlying order previously pointed out. This could lead to a modelization of the landscape in the form of a spatial stochastic matrix, thus allowing further comparisons between models of organization and finally, between different landscapes.

Because its strategy is based on a divisive classification process, the procedure also embodies chorologic developments. As was shown above terminal subsets are groups of sites that form noncontiguous sections of the space. Altogether these sections constitute a new choropleth pattern whose areal classes contain sites which are comparable with regard to their probability *vis-à-vis* the different community types. The hypothesis that each particular section of the pattern could be considered a homogeneous subdivision of space for the purpose of planning can be advanced on this basis.

In conclusion, this approach contributes to the development of the theoretical background of spatial and ecological analysis. At the same time, by helping us to understand the fundamental question of "order and disorder" in land, it should improve the predictive value of concrete research operations such as landscape analysis and biophysical inventory along the line defined by Ruxton (1968).

## REFERENCES

- ALLAIRE, G., PHIPPS, M. & STOUPEY, M. (1973). *Espace Géogr.* 3, 185.  
BATTY, M. (1972). *Area* 4, 231.

- BERRY, B. & SCHWIND, P. (1969). *Geogr. Analys.* **1**, 5.
- BERTRAND, G. (1969). *Rev. Géogr. Pyr. S.O.* **29**, 249.
- BRILLOUIN, L. (1951). *J. appl. Phys.* **22**, 338.
- BRILLOUIN, L. (1962). *Science and Information Theory*. 2nd edition, New York: Academic Press.
- CHRISTIAN, C. S. (1957). *Proc. 9th Pacific Sc. Congress* **20**, 74.
- DALE, M. B. & ANDERSON, D. J. (1972). *J. Ecol.* **60**, 639.
- ESCOFIER-CORDIER, B. (1969). *Cah. Bur. Univ. Rech. Oper. Univ. Paris.* **13**, 25.
- GARDINER, V. & GREGORY, K. J. (1977). *Progr. Phys. Geogr.* **1**, 1.
- GODRON, M. (1968). *Oecologia Plantarum* **3**, 185.
- GOODALL, D. W. (1973). In: *Handbook of Vegetation Science, V: Ordination and Classification of Communities* (R. Whittaker, ed.), p. 575. The Hague: Dr. W. Junk B. V.
- GOULD, P. (1961). In: *Spatial Organization: The Geographer's View of the World* (Abler, Adams and Gould, eds), p. 93, New York: Prentice Hall.
- GRIGG, D. (1961). In: *Integrated Models in Geography* (R. S. Chorley and P. Haggett, eds) p. 461. London: Methuen.
- GUILLERM, J. L. (1971). *Oecologia Plantarum* **6**, 209.
- GUREVICH, B. L. (1969). *Soviet Geography* **10**, 387.
- HILL, M. O. (1974). *Appl. Statist.* **23**, 340.
- HILLS, G. A. (1976). In: *Ecological Land Classification*. p. 73. Ottawa: Environment Canada.
- ISACHENKO, A. G. (1973). *Soviet Geography* **14**, 229.
- JURDANT, M. (1976). In: *Ecological Land Classification*. p. 173, Ottawa: Environment Canada.
- KULLBACK, S. (1959). *Information Theory and Statistics*. New York: John Wiley.
- LEGENDRE, L. & LEGENDRE, P. (1979). *Ecologie numérique*. Vol. 2, Paris: Masson.
- LEOPOLD, L. B. & LANGBEIN, W. B. (1962). *U.S. Geol. Surv. Prof. Papers* **500 A**.
- MABBUTT, J. A. (1968). In: *Land Evaluation* (G. A. Stewart, ed.), p. 11. Canberra: McMillan.
- MCINTOSH, R. P. (1967). *Bot. Rev.* **33**, 130.
- MACNAUGHTON, P. (1963). *Biometrics* **19**, 364.
- MARCHAND, B. (1972). *Geogr. Analys.* **4**, 234.
- MARGALEF, R. (1956). *Invest. Pesquera.* **3**, 99.
- MARGALEF, R. (1958). *Gener. Syst.* **3**, 36.
- MATERN, B. (1979). In: *Spatial and Temporal Analysis in Ecology* (R. M. Cormack and J. K. Ord, eds), p. 271. Fairland, Maryland: International Co-operative Publishing House.
- MEDVEDKOV, Y. (1967). *Papers Reg. Sc. Assoc.* **18**, 165.
- NIJKAMP, P. & PAELINCK, J. H. P. (1974). *Papers 13th Eur. Congr. Reg. Sc. Assoc.* **33**, 13.
- ORLOCI, L. (1970). *J. theor. Biol.* **29**, 173.
- ORLOCI, L. (1973). In: *Handbook of Vegetation Science, V: Ordination and Classification of Communities* (R. Whittaker, ed.), p. 249. The Hague: Dr. W. Junk B. V.
- ORLOCI, L. (1978). *Multivariate Analysis in Vegetation Research*. The Hague: Dr. W. Junk B. V.
- PATTEN, B. C. (1959). *Ecology* **40**, 221.
- PERELMAN, A. I. (1960). *Soviet Geography* **2**, 63.
- PIELOU, E. C. (1966a). *J. theor. Biol.* **13**, 131.
- PIELOU, E. C. (1966b). *J. theor. Biol.* **10**, 370.
- PIELOU, E. C. (1966c). *Am. Nat.* **100**, 463.
- PIELOU, E. C. (1975). *Ecological Diversity*. New York: John Wiley.
- PIPKIN, J. S. (1977). *Geogr. Analys.* **9**, 187.
- RESCIGNO, A. & MACCACCARO, W. B. (1960). In: *Sympos. on Information Theory*. London.
- ROBINSON, V. B. (1978). *Prof. Geogr.* **30**, 74.
- RUXTON, B. P. (1968). In: *Land Evaluation* (G. A. Stewart, ed.), p. 29. Canberra: McMillan.
- SCHIEDER, S. J. (1966). *Proc. 2nd Intern. Sympos. Photo-Interpr.* VI, 28. Paris: Technip.
- SEMPLE, R. K. & WANG, L. H. (1970). *Depart. of Geogr. Discus. Papers.* **5**, Univ. of Toronto.

- SOBOLEV, L. N. & UTEKIN, V. D. (1973). In: *Handbook of Vegetation Science, V: Ordination and Classification of Communities* (R. Whittaker, ed), p. 75. The Hague: Dr. W. Junk B. V.
- SOCHAVA, V. B. (1970). *Papers 5th Congr. Geogr. Soc. USSR*, Leningrad.
- SPENCE, N. A. & TAYLOR, P. J. (1970). *Prog. Geography* **2**, 3.
- STRAHLER, A. H. (1978). *J. Biogeogr.* **5**, 403.
- THOMAS, R. W. & REEVE, D. E. (1976). *Geogr. Analys.* **9**, 113.
- WEBBER, M. J. (1977). *Ann. Assoc. Amer. Geogr.* **67**, 254.
- WHITTAKER, R. H. (1973a). In: *Handbook of Vegetation Science, V: Ordination and Classification of Communities* (R. Whittaker, ed.), p. 7. The Hague: Dr. W. Junk B. V.
- WHITTAKER, R. H. (1973b). In: *Handbook of Vegetation Science, V: Ordination and Classification of Communities* (R. Whittaker, ed.), p. 323. The Hague: Dr. W. Junk B. V.
- WILSON, A. G. (1970). *Operational Res. Quarterly* **21**, 247.
- WILSON, I. G. & WILSON, M. E. (1965). In: *Information, Computers and System Design*. p. 251., New York: John Wiley.
- WILLIAMS, W. T., LAMBERT, J. M. & LANCE, G. N. (1966). *J. Ecol.* **54**, 427.