Oecologia (Berl.) 36, 317–326 (1978)

Oecologia

The Detection of Non-Random Patterns of Distribution of Species along a Gradient

A.J. Underwood

Zoology Building, School of Biological Sciences, University of Sydney,
Sydney, N.S.W. 2006, Australia

Summary. A simple method for detecting non-random patterns of distribution of the boundaries of species is described. The method uses transects running across a community, where the number of upper and lower boundaries of species in each quadrat is recorded. The expected number of quadrats containing one or more boundaries can be calculated from the binomial distribution. The mean deviation of observed from expected number of such quadrats, for a set of transects, can be tested for departures from zero. Significant departures greater than zero indicate regular dispersion of boundaries. A mean deviation significantly less than zero indicates clustering of the boundaries. The method is unbiased and thus corrects previously published methods.

Introduction

Pielou (1975a, b) has devised a test to detect non-random patterns of distribution of species along an environmental gradient. The technique has been used to investigate the structure of salt-marsh communities (Pielou and Routledge, 1976) and is theoretically of great value for other communities. There has been much discussion of the arrangement of species' boundaries, particularly with respect to the grouping of plant species into natural communities (for example, the discussion in Whittaker, 1967). Pielou's (1975a, b) test was designed to distinguish between random, regular and clumped patterns of distribution of the boundaries of coexisting and overlapping species. If the species were grouped into definable communities, their boundaries should coincide, giving a clumped pattern. If assemblages of species were haphazard, however, their boundaries should be spatially distributed at random. The alternative arrangements were illustrated by Pielou and Routledge (1976, Fig. 2).

0029-8549/78/0036/0317/$02.00

Here, I show that the statistical test developed by Pielou (1975a, b) is biased, and further, that the probability distribution used was not correct. These two considerations make the technique unsuitable for the detection of non-random patterns of distribution of boundaries of a set of species.

Accordingly, I have developed a new method of detecting such patterns, which is based on similar premises to the one developed by Pielou (1975a, b) and which uses similar sample data. Details of the use of this method in an analysis of intertidal community structure will be presented elsewhere (Underwood, 1978).

The Pielou Method

Pielou and Routledge (1976) sampled salt-marsh plants in belt transects of 50 cm width, divided into contiguous quadrats of length 1 m. The transects ran from the landward to the seaward ends of a marsh and were placed at least 10 m apart, to ensure the independence of samples. In each transect, the species present in each quadrat were recorded. The upper and lower boundaries of each species were thus determined as the upper and lower quadrats in which the species occurred.

Consider only the upper boundaries of the species. In any one transect, there are Q quadrats and k species whose upper boundaries are determined. This excludes any species which extends above the uppermost quadrat. Let the number of quadrats containing at least one upslope boundary be U. Pielou (1975a, b) determined that, under the null hypothesis that the upper boundaries are distributed at random to the quadrats, the probability of any value of U is given by:

$$Pr(U=u)=p_u=\frac{\binom{Q}{u}\binom{k-1}{u-1}}{\binom{Q+k-1}{k}}, \quad u=1,2,\dots \min(Q,k) \tag{1}$$

Using this distribution, a sign test was constructed to determine departures from random allocation of boundaries to quadrats. For any transect, the median value of U, say U_{med}, given Q and k, was calculated as follows:

$$U_1 < U_{\mathrm{med}} < U_2, \quad \text{and} \quad U_2 - U_1 = 1 \tag{2}$$

where

$$\sum_{u=1}^{U1} p_u < 0.50 \quad \text{and} \quad \sum_{u=1}^{U2} p_u > 0.50.$$

For a collection of transects, $\mathbf{U}$ is the number of times U exceeded the expected median U_{med}. Then, given random dispersal of upper boundaries in each transect, the distribution of $\mathbf{U}$ was considered to be binomial with $P=0.5$. A significantly low value of $\mathbf{U}$ indicates that the upper boundaries were clustered, whereas a significantly high value of $\mathbf{U}$ indicates regularity of spacing of upper boundaries. The same procedure can be applied to the lower boundaries of the species in each transect (for the full details and examples, see Pielou and Routledge, 1976).

Bias and the Probability Distribution of U

The binomial test devised by Pielou (1975a, b) is biased. The bias stems from the discreteness of the distribution for values of U [Eq. (1), above]. For example, in a transect with $Q=23$ quadrats and $k=14$ upper boundaries, it is found from Equation (1) that:

$$\sum_{u=1}^{8} p_u = 0.3737 \quad \text{and} \quad \sum_{u=1}^{9} p_u = 0.6508.$$

Thus, the expected median of U is between 8 and 9, say 8.5. This example was taken from Pielou and Routledge (1976, p. 315). The binomial probability for U observed to exceed or be less than U_{med} is not $P=0.5$. It is obvious that:

$$Pr(U \leqq 8) = 0.3737 \quad \text{and hence} \quad Pr(U \geqq 9) = 0.6263$$

and, thus, there is a bias in the binomial test. This bias always favours significance in the direction of regularity, from the definition of U_{med}. The cumulative probability of any value of U less than or equal to U_1 [in Eq. (2), above] is always less than 0.5. Thus, the probability of exceeding the expected median, U_{med}, is always greater than 0.5. This bias was recognized by Pielou (pers. comm.) and was assumed to be small.

The distribution of U in Equation (1) is, however, incorrect and attempts to devise alternative, unbiased tests using U_{med} are therefore pointless. I have tested the probability distribution of U by computer simulations, generating k variates from a uniform distribution between 0 and Q. The number of such variates in each unit interval 0–1, 1–2, ... $(Q-1)-Q$, thus corresponds to the random allocation of k boundaries to Q quadrats. The number of quadrats containing at least one of the k variates is U. Numerous combinations of Q (over the range 10–80) and of k (over the range 3–40) were simulated 100 times, which represents 100 transects of each condition, and gives an observed frequency distribution of U for each combination of Q and k. Results of two such simulations are given, as examples, in Table 1, with tests of the agreement with expected frequencies from Pielou's (1975a, b) derivation [as in Eq. (1), above]. In all combinations of Q and k tested, there were significant departures of the frequency distribution of U from that expected.

Pielou (pers. comm.) considered that the distances between boundaries are distributed as random variates from a negative exponential distribution, and

Table 1. Frequency distributions of numbers of boundaries in quadrats, in 100 simulations (U is the number of quadrats containing at least one boundary in each simulation)

$Q=20$, $k=10$

$U=$	$\leqq 5$	6	7	8	9	10
Observed frequency in 100 simulations	2	7	21	35	29	6
Expected frequency [from Eq. (1)]	12.0	24.4	32.5	22.6	7.5	0.9

$\chi^2(4df) = 114.6$, $P < 0.001$

$Q=20$, $k=5$

$U=$	$\leqq 3$	4	5
Observed frequency in 100 simulations	9	32	59
Expected frequency [from Eq. (1)]	17.9	45.6	36.5

$\chi^2(2df) = 22.4$, $P < 0.01$

that the number of species in a transect (k) is a random variate from a Poisson distribution. This definition of random dispersion of boundaries along a gradient can be shown to be mathematically identical to the simulations using variates from a uniform distribution, as described above. The simulations using variates from a uniform distribution should therefore have given the probability distribution of U derived by Pielou (1975a, b), if this were correct.

The probability distribution in Equation (1) is incorrect because it is based on the probabilities of random distribution of indistinguishable balls into boxes derived from Bose-Einsein combinatorial formulae (see Feller, 1968, p. 20). These are inappropriate, and the alternative combinatorials for distinguishable events (the Maxwell-Bolzmann formulae in Feller (1968)) should be used. The reasons for use of the Maxwell-Bolzmann formulations are illustrated in Figure 1, where the possible results of independent distribution of 3 balls into 3 boxes are illustrated. When the balls are indistinguishable, there are 10 possible results (Fig. 1A). The probability of arriving at any value of U is, however, identical to that when the balls are distinguishable (Fig. 1B), except in the very rare set of natural phenomena to which the Bose-Einstein formulae apply. In the case of distinguishable balls, there are more possible results ($Q^k = 27$, in the example illustrated). Even when the balls are indistinguishable, the probabilities of arriving at any value of U should be based on the number of results possible for distinguishable balls. Equation (1) assumes that each possible result has the same probability, according to Bose-Einstein theory. Even in this small example, there are large discrepancies between the two theoretical distributions of U.

The Probability Distribution of U

Any attempt to formulate the correct probability distribution of U based on Maxwell-Bolzmann combinatorial formulae involves a large repetitive set of calculations. In general, the probability of getting a particular set of results in a transect is much easier to calculate, but is considerably less useful. Consider k boundaries in Q quadrats, where U is the number of quadrats containing at least one boundary, and u_i ($i = 1, 2, \ldots Q$) is the number of boundaries in the i-th quadrat.

The number of ways of arranging k boundaries in Q quadrats is Q^k. The number of ways of arranging k boundaries in Q quadrats with occupancy numbers $u_1, \ldots u_i \ldots u_Q$ is given by:

$$\frac{k!}{u_1!\, u_2! \ldots u_i! \ldots u_Q!}.$$

Hence, the probability of getting exactly that set of values of u_i, in the observed order, is given by:

$$\frac{k!}{(u_1! \ldots u_i! \ldots u_Q!)\, Q^k}.$$

Let the number of different values of u_i such that $u_i \geqq 1$ be G. The frequency of quadrats containing each of the G different values is g_j, $j=1, 2, \ldots G$. The number of empty quadrats is g_0 (i.e. $g_0 = Q - U$). Thus,

$$Q = \sum_{j=0}^{G} g_j.$$

The number of different ways of ordering the observed set of u_i values is then given by:

$$\frac{Q!}{g_0!\, g_1! \ldots g_j! \ldots g_G!}.$$

The probability of attaining the observed set of u_i values is thus given by:

$$Pr(R) = \frac{k!\,Q!}{(u_1! \ldots u_i! \ldots u_Q!)(g_0!\, g_1! \ldots g_j! \ldots g_G!)\, Q^k}. \tag{3}$$

For a set of T transects, the probability of getting a particular set of u_i values in each transect can be calculated, using Equation (3). These probabilities could be combined, using Fisher's (1958) summation, as follows:

$$P = -2 \sum_{t}^{T} \log_e (Pr(R_t)) \tag{4}$$

where $Pr(R_t)$ is that for the t-th transect, calculated from Equation (3), and $t = 1, 2, \ldots T$.

This combined probability, P, over all T transects, is distributed as χ^2 with $2T$ degrees of freedom. It would be necessary to use Equation (3) for all possible sets of values of u_i which deviate, from random allocation of boundaries to the quadrats, more than the observed set, in order to construct a significance test for each transect. The sum of such a set of probabilities would then serve as R_t to calculate P in Equation (4). This would provide a test for random distribution of boundaries. A significantly high value of P [from Eq. (4)] compared with the tabled value of χ^2 would demonstrate that the species' boundaries were not distributed at random to the quadrats over a set of sampled transects. It is not, however, clear whether the departure from a random distribution is in the direction of clustering or in the direction of regularity. The use of such a statistic is obviously severely limited.

Alternatively, the formulations leading to Equation (3) could be used to generate the probability distribution of U in any transect. This is, however, difficult to do in practice. For each value of U ($U=1, 2, \ldots \min(Q, k)$), there are several possible values of G, several different sets of u_i and several possible sets of G_j. These would have to be calculated for any set of U, Q, and k. The general formula for the probability distribution of U, given Q and k, was derived by Feller (1968, p. 64) as follows:

$$Pr(U=u)=\frac{Q!}{u!(Q-u)!\,Q^k}\sum_{v=0}^{u}(-1)^v\frac{u!}{v!(u-v)!}(u-v)^k$$

where $v=0, 1 \ldots u$; $u=1, 2 \ldots \min(Q, k)$.

An example of the calculations necessary for the simple case of $Q=10$, $k=5$ is given in Table 2. As k increases, the calculations become more laborious, but they do fit the observed distribution of U from computer simulations. This is demonstrated, in Table 3, for two examples. The goodness-of-fit in each example should be compared with those in Table 1, where the Pielou (1975a, b) formulation was used.

An alternative approach to the whole problem is used below, to avoid the necessity of repetitive and tedious calculations based on Maxwell-Bolzmann probabilities.

The Expected Value of *U*

The expected value of U, given Q and k, can be calculated, using the binomial or Poisson distributions. For each quadrat, the probability of getting each species' boundary, by chance, is Q^{-1}. From the binomial distribution, the probability of getting none of the k boundaries in a single quadrat is:

Table 2. Example calculation of the probability distribution of U (the number of quadrats containing at least one species' boundary) when Q (the number of quadrats) = 5, and k (the number of species) = 5

U	No. of boundaries in each quadrat (u_i)					No. of different values, $u_i \geq 1$	Frequency of each of the G values			$Pr(R)$ [from Eq. (3)]	$Pr(U)$
	u_1	u_2	u_3	u_4	u_5	G	g_0	g_1	g_2		
1	5	0	0	0	0	1	4	1	–	0.0016	0.0016
2	4	1	0	0	0	2	3	1	1	0.0320	0.0960
	3	2	0	0	0	2	3	1	1	0.0640	
3	3	1	1	0	0	2	2	2	1	0.1920	0.4800
	2	2	1	0	0	2	2	1	2	0.2880	
4	2	1	1	1	0	2	1	3	1	0.3840	0.3840
5	1	1	1	1	1	1	0	5	–	0.0384	0.0384

Table 3. Frequency distributions of numbers of boundaries in quadrats, in 100 simulations (U is the number of quadrats containing at least one boundary in each simulation)

$Q=20$, $k=10$					
$U=$	$\leqq 6$	7	8	9	10
Observed frequency in 100 simulations	9	21	35	29	6
Expected frequency [using Eq. (3)]	7.5	22.3	37.1	26.7	6.4
	$\chi^2(4df)=0.72$, $P>0.90$				
$Q=20$, $k=5$					
$U=$		$\leqq 3$	4	5	
Observed frequency in 100 simulations		9	32	59	
Expected frequency [using Eq. (3)]		5.5	36.3	58.1	
	$\chi^2(2df)=2.75$, $P>0.25$				

$$Pr(0)=(1-1/Q)^k.$$

Thus, the probability of getting at least one boundary is:

$$Pr(\geqq 1)=1-(1-1/Q)^k.$$

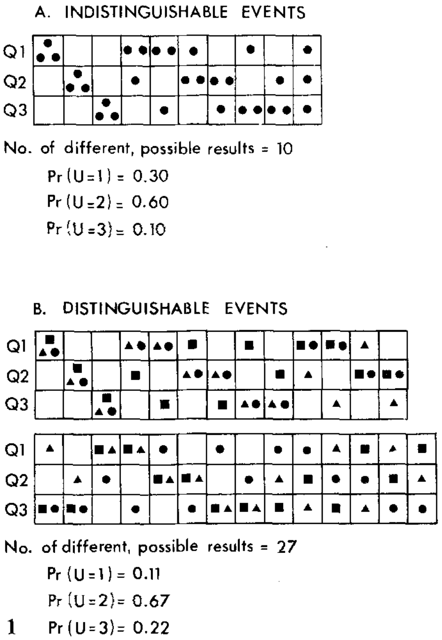
Thus, the expected number of quadrats containing at least one boundary, $E(U)$, is:

$$E(U)=Q(1-(1-1/Q)^k). \tag{5}$$

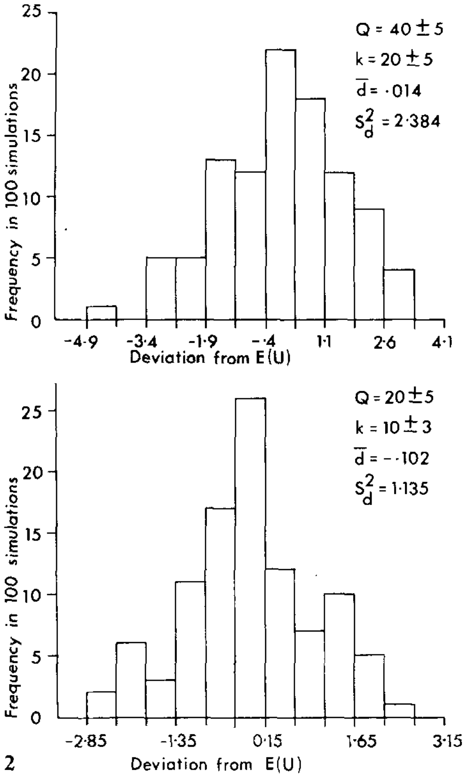
This is in agreement with the expected value from the probability distribution for U, given Q and k, derived by Feller (1968) which was presented earlier.

In any transect, the deviation between the observed number of quadrats with at least one boundary (U obs) and the expected number ($E(U)$) can be calculated. For a set of transects, there is a sample of deviations from expected values. If the species boundaries are distributed at random, the mean deviation of the sample should be zero.

I have examined the distribution of the deviations from expected values, in computer simulations. Two different conditions were simulated, firstly where the number of quadrats (Q) and species (k) were kept constant over 100 simulations, thus representing 100 transects. The second set of conditions was the more realistic field situation, where the number of quadrats will vary from transect to transect, because of topographical variations in the substratum. The number of species may also vary, particularly if the species have a patchy horizontal distribution. Simulations were done with combinations of Q in the range 10–80, and k over the range 3–40. In most cases, except where Q and k were both small, deviations from the binomial expected values [i.e. Eq. (5)] were normally distributed. Two examples are illustrated in Figure 2. Thus, in a set of T



1



2

Fig. 1A and B. The random allocation of $k=3$ balls to $Q=3$ boxes. **A** Indistinguishable balls. Bose-Einstein probability formulae give each possible result the same probability of occurrence. This is incorrect for boundaries of species in quadrats. **B** The balls are distinguishable. Maxwell-Bolzmann probability formulae give each possible result the same probability of occurrence. These probabilities also apply to the results with indistinguishable balls (as in **A**)

Fig. 2. Frequency distributions of deviations from expected values in computer simulations. $E(U)$ was calculated from the binomial distribution [as in Eq. (5)], for each of 100 simulations. The number of quadrats (Q) and the number of species (k) varied, at random, over the ranges given for each set of simulations. The distributions are approximately normal

transects, the expected number of quadrats containing at least one boundary is:

$$E(U_i)=Q_i(1-(1-1/Q_i)^{k_i})$$

where Q_i is the number of quadrats and k_i the number of species in the i-th transect ($i=1, 2, \dots T$). The observed number of quadrats containing at least one boundary is U obs i for the i-th transect. Hence, the deviation from expected is: $d_i=U$ obs $i-E(U_i)$. The mean deviation is:

$$\bar{d}=\frac{\sum_{i=1}^{T} d_i}{T}$$

with standard error:

$$SE_d = \sqrt{\frac{\sum_{i=1}^{T} d_i 2 - \frac{\left(\sum_{i=1}^{T} d_i\right)^2}{T}}{(T-1).}}$$

The null hypothesis that the boundaries of the species are distributed at random can be tested by a t-test, where:

$$t = \frac{\bar{d}}{SE_d} \quad \text{with } (T-1) \text{ degree of freedom.}$$

If $\bar{d}$ is significantly greater than zero, the species' boundaries are distributed more regularly than would be expected by chance. In contrast, if $\bar{d}$ is less than zero, the species' boundaries are clumped.

This test is unbiased and has the advantage that information on the size of the deviation in each transect is used to estimate the mean deviation. Any test using only the sign of the deviation in each transect ignores the magnitude of departure from the expected, random distribution of species to quadrats. Alternative, nonparametric tests could be used to determine if the median deviation in a set of transects differed significantly from zero. Because my simulations show that the deviations from expected are distributed nearly normally, I favour the use of the simple t-test.

Use of the Method on Data from Transects

An example of the use of this method is given in Table 4. The data come from a survey of intertidal rock platforms in Britain and will be discussed in full

Table 4. Data from transects run across an intertidal rock-platform

Transect (i)	No. of quadrats (Q_i)	No. of upper boundaries (k_i)	No. of quadrats containing at least one upper boundary		Deviation from expected $d_i = U_{\text{obs}\,i} - E(U_i)$
			Expected[a] $E(U_i)$	Observed $U_{\text{obs}\,i}$	
1	31	20	14.91	15	0.09
2	34	21	15.83	17	1.17
3	35	21	15.96	14	−1.96
4	40	21	16.49	16	−0.49
5	20	19	12.45	15	2.55
6	19	19	12.20	14	1.80
7	22	21	13.72	14	0.28
8	24	20	13.75	15	1.25
9	24	21	14.18	13	−1.18
10	30	20	14.77	16	1.23

[a] $E(U_i)$ calculated using Equation (5).

Mean deviation ($\bar{d}$) = 0.47; standard error (SE_d) = 0.44; $t = 1.07$; $P > 0.20$

elsewhere (Underwood, 1978). Ten transects of contiguous quadrats, each 0.5 m square, were sampled at Dunollie, Scotland, in July 1977. The transects were approximately 10 m apart horizontally, to ensure their independence, and ran from the highest levels reached by intertidal organisms, to low water level (below the upper limit of the kelp *Laminaria*). The presence in each quadrat of all the species of algae, barnacles, and sea-anemones was recorded. The shore rose monotonically from low water level, so that every quadrat was vertically higher than those quadrats nearer to low water level. The number of quadrats varied from 19–40, and the number of species from 19–21. The mean deviation of observed from expected numbers of quadrats containing at least one upper boundary was not significantly different from zero (Table 4 shows the data for the upper boundaries of species in the ten transects). Thus, there is no evidence that the upper boundaries of species on that shore were distributed other than randomly.

The technique described here provides a relatively simple, and unbiased method of detecting non-random spatial patterns of dispersion of boundaries of species in any community which can be sampled by belt transects of contiguous quadrats.

Acknowledgements. This work was supported by a University of Sydney Research Grant and a grant from the Australian Research Grants Committee. I gratefully acknowledge the useful discussions with Drs. H. Barnes, G.J. Caughley, G. Cooney, J. Robinson, P.F. Sale, the assistance with computing and analysis of P.A. Cameron and the advice, by correspondence, from Professor E.C. Pielou. Much of the mystery surrounding probability theory was patiently explained to me by Dr. P.A. Underwood.

References

Feller, W.: An introduction to probability theory and its applications, Vol. 1, 3rd ed. New York: Wiley 1968

Fisher, R.A.: Statistical methods for research workers, 13th ed. New York: Hafner 1958

Pielou, E.C.: Ecological models on an environmental gradient. In: Proc. International Conf. Applied Statistics (R.P. Gupta, ed.) (1975a)

Pielou, E.C.: Ecological diversity. New York: Wiley 1975b

Pielou, E.C., Routledge, R.D.: Salt marsh vegetation: Latitudinal gradients in the zonation patterns. Oecologia (Berl.) **24**, 311–321 (1976)

Underwood, A.J.: A refutation of critical tidal levels as determinants of the structure of intertidal communities on British shores. J. Exptl. Mar. Biol. Ecol., **34** (1978, in press)

Whittaker, R.H.: Gradient analysis of vegetation. Biol. Rev. **42**, 207–264 (1967)

Received January 10, 1978