

A new numerical model for coral distribution

D. J. Abel¹, W. T. Williams², P. W. Sammarco² and J. S. Bunt²

¹ CSIRO Division of Computing Research, Davies Laboratory, Private Bag, Aitkenvale, Queensland 4814, Australia

² Australian Institute of Marine Science, P. M. B. No. 3, Townsville, Queensland 4810, Australia

ABSTRACT: A critical review of previous numerical work on coral distribution has shown that all previous workers have used models which assume the probability of occurrence of a coral to be spatially uniform over the study area. This is not necessarily true. Here we define a new 'constrained occupancy model' which permits spatial variation in probability. A case study involving extensive data (46 transects) on the distribution and abundance of 8 coral species and 2 octocorals throughout the Caribbean was used to test the new model against the commonly used Poisson model. The Poisson model demonstrated that 4 of these species were highly significantly aggregated in distribution and six were random. The new constrained occupancy model, allowing for variation in probability of occurrence, demonstrated that, in fact, 1 tended towards aggregation, most were random, and 2 species were significantly uniform in distribution. We believe that accounting for variation in probability of occurrence in sessile organisms is important when analyzing their distribution, and we consider these findings warrant the attention of field ecologists. While no analytic solution to the new model is known, computational procedures are presented, some of which can be used with a Monte Carlo simulation.

INTRODUCTION

Although there is a formidable corpus of published information concerning coral distribution, the bulk of it is essentially descriptive; few workers have used statistical or other numerical methods to explore the nature of the distributions they describe. Those who have done so have been interested in either or both of two problems. The first is the possibility of establishing significant associations or dissociations between different species (e.g. Bradbury and Young, 1982), a sociological problem with which we are not concerned in this paper. The second seeks to ascertain whether the observed distributions are, or are not, random. Here we first examine the assumptions underlying the methods previously used for this purpose and then define a new model which we believe to be more accurate and appropriate.

The simplest method has been to calculate, in plot samples, the ratio of variance to mean, a quantity variously known as 'relative variance' (Clapham, 1936) or 'coefficient of dispersion' (Blackman, 1942); if this significantly exceeds unity the distribution is accounted 'aggregated', 'contagious', or 'clumped'. It has been used as the sole criterion by Lewis (1970),

Maragos (1974), Dana (1976) and Wallace and Bull (1982). Stimson (1974) and Birkeland et al. (1982) used the plotless nearest-neighbour technique of Clark and Evans (1954) as well as the coefficient of dispersion; Benayahu (1975) also used the 'index of dispersion' of Morisita (1959). Lastly, Grassle (1973), realizing that aggregation is dependent on sample size, used the pattern-analysis technique of Greig-Smith (1952). It is relevant to note that, with the single exception of Stimson, working on *Pocillopora* in Waikiki, Oahu, all have found that the majority of the corals studied showed, at least under some conditions and at some scales of sampling, highly aggregated distributions.

All 4 of the above methods share a common, essentially Poissonian, assumption: that the probability of occurrence of an organism may vary with the size, but not the position, of a sample – i.e. that the probability of occurrence is uniform over the area under study. The history of this assumption has been briefly reviewed by Greig-Smith (1964). 'Student' (W. S. Gossett) (1919) seems to have been the first to point out that if the assumption were not valid, a distribution calculated in the conventional manner would deviate significantly from the Poisson. The only attempt known to us to refine this mathematically was that of Ashby (1935) and Stevens (1937), who were mainly concerned with finding a model which would account for an excessive

Addressee for reprint requests: W. T. Williams

proportion of zero records. Here we define a new model – hereby termed the ‘constrained occupancy model’ – which allows for variation in space of probability of occurrence. Our own interest in the problem arose from the results of a survey, briefly described below, for which we believed the uniformity postulate to be unrealistic.

THE ECOLOGICAL PROBLEM

In 1972, Dr. J. S. Bunt and co-workers carried out a survey of 46 ten-m-line transects on coral reefs at depths of 15 to 42 m (45 to 125 feet) in a number of locations from Florida through the Bahamas and from the Turks and Caicos Islands to Haiti. Details concerning actual site locations, transect orientation, locations of corals and octocorals on individual transects, and rationale for transect length may be found in Bunt and Pepper (1973). Benthic species recorded included 28 species of scleractinian corals, 27 species of gorgonians, 69 sponges (not all unequivocally identified) and a hydrozoan. The data were later subjected to numerical analysis by Bunt *et al.* (1982), and a reduced set was further analysed by Williams *et al.* (1982). A weak geographical pattern was discernible. Long Reef, off the Florida coast (Transects 1 to 7), appeared to be faunistically richer than the general Bahamian region. Those islands which were close to the interface between the Bahamian plateau and the Atlantic Ocean (transects 33 to 38) were more diverse but possessed generally smaller colonies than those further away from this interface (Transects 39 to 46). The 25 transects on Grand Bahama (8 to 32) appeared to encompass most of the variation found elsewhere in the region. A map is given in each of the references cited above.

We have elected to work with the 10 most abundant species, 8 scleractinian corals and 2 octocorals. These were as follows, presented in order of abundance (total number of colonies in all 46 transects given in brackets):

1. *Montastrea annularis* (Ellis & Solander) (129)
2. *Siderastrea siderea* (Ellis & Solander) (118)
3. *Pseudopterogorgia elisabethae* (Bayer) (Octocoralia) (75)
4. *Agaricia agaricites* (L.) forma *purpurea* (51)
5. *Montastrea cavernosa* (Linnaeus) (46)
6. *Dichocoenia stokesii* (Milne-Edwards & Haime) (45)
7. *Porites astreoides* (Leseur) (39)
8. *Pseudopterogorgia americana* (Gmelin) (Octocorallia) (32)
9. *Agaricia agaricites* (L.) forma *agaricites* (32)
10. *Manicina areolata* (Linnaeus) (31).

The distribution of these 10 species among the 46 transects is summarized in Table 1, which also includes the total number of colonies recorded in each transect. The range of these total numbers is remarkable: from only 4 organisms in Transect 20 to 49 organisms in Transect 34 of the same length. We do not here propose to discuss the reasons for this variation, but we suggest that it is reasonable to postulate that Transect 20 is environmentally less suitable for the settlement of coral larvae and survival of coral spat than is Transect 34, and that the probability of occurrence of *any* coral is lower in Transect 20 than it is in Transect 34. It will now be clear that we are advocating a Bayesian approach. We suggest that, as usual, there is a probability of finding any particular coral species which is dependent on the total frequency of that species; but that there is also, for each transect, a prior probability of finding *any* coral species; and that the set of prior probability is monotonically related to the set of total numbers of organisms in the transects. We now define a model with these properties.

THE NEW ‘CONSTRAINED OCCUPANCY MODEL’

We regard the problem as an example of the ‘occupancy problems’ which are treated in standard texts on probability (e. g. Feller, 1957); these are concerned with the probabilities of assigning r objects to n boxes in a particular way. The solution involves the calculation of N_r , the number of ways in which the given configuration could be obtained, and N_t , the total number of ways in which the r objects can be distributed between the n boxes. The required probability, p , is then given by $p = N_r/N_t$. This type of problem is considered by Bose-Einstein statistics, for which analytic solutions are available. Of the distributions however, no arrangement would be acceptable which required there to be more specimens of a given coral in a given transect than there were organisms recorded for that transect. In other words, our boxes (the transects) are *constrained* in size to the numbers given in the last column in Table 1. We know of only one previous example in the literature of a ‘constrained occupancy’ problem. This was an agricultural problem stated by Williams *et al.* (1973); the objects for arrangement were failures of cows to conceive in a given year, and the boxes were paddocks, each containing 4 cows. No component of N_t was acceptable which would require there to be more than 4 failures in any one paddock; all paddocks were thus constrained to a maximum of 4 failures. They give an analytic solution to which there are, unfortunately, 2 objections. First, their solution assumes that the objects (failures in the above

example or coral colonies in our case) are distinguishable (identifiable as individuals), which would be biologically inappropriate. As a result, their probabilities are underestimated. (We note in passing that we have recalculated the probabilities for the indistinguishable case and that the general conclusions are

not thereby invalidated). Secondly, they give a summation formula to reduce N_i in relation to the constraints. It can be shown that their formula rests on a recurrence relation which will fail if the total number of objects is greater than the sum of any 2 of the constraints. This did not arise in their case, but it would

Table 1. Distribution of species by transects

Region	Transect number	Species										Total no. organisms
		1	2	3	4	5	6	7	8	9	10	
Long Reef	1	1	3	3	–	3	2	–	2	1	–	37
	2	2	2	3	–	2	–	–	–	2	3	30
	3	–	2	–	–	–	3	3	3	2	3	36
	4	5	2	1	–	3	–	1	1	1	1	38
	5	1	2	–	–	1	1	–	1	1	–	29
	6	1	6	1	–	3	1	–	2	2	1	35
	7	2	5	2	–	1	1	2	1	1	1	40
Grand Bahama	8	–	3	3	–	–	–	–	2	1	–	32
	9	1	–	1	–	2	–	1	2	1	–	19
	10	1	2	5	3	2	2	–	1	–	–	36
	11	5	3	–	2	2	1	1	2	1	1	31
	12	2	1	–	1	–	1	–	2	–	–	12
	13	3	3	2	3	–	2	1	1	1	–	33
	14	–	–	3	–	2	–	–	–	–	–	19
	15	–	3	1	2	1	2	–	–	2	–	24
	16	6	4	–	5	1	–	–	5	–	1	36
	17	5	5	–	–	–	–	2	1	–	1	21
	18	6	1	–	5	–	–	3	–	–	–	20
	19	–	2	4	–	–	–	–	–	1	–	8
	20	–	–	1	–	–	–	–	–	1	–	4
	21	–	1	1	2	1	–	–	–	–	–	13
	22	–	1	4	2	–	–	–	–	2	–	19
23	–	2	2	1	2	–	–	–	1	1	25	
24	1	1	2	–	–	1	1	–	–	–	12	
25	1	2	3	3	–	–	1	–	–	–	23	
26	6	2	–	2	1	–	–	–	1	–	15	
27	3	5	–	1	3	1	–	–	–	–	27	
28	1	2	2	3	1	2	1	–	–	–	28	
29	–	1	1	3	–	–	2	2	1	–	22	
30	1	6	2	4	1	1	1	1	–	3	29	
31	–	–	4	–	2	2	–	–	1	–	23	
32	–	2	5	–	2	–	–	–	–	1	26	
Outer islands	33	–	1	1	–	–	–	–	–	1	–	6
	34	9	4	1	2	–	2	2	1	–	3	49
	35	3	7	4	2	1	1	–	–	1	1	30
	36	5	7	3	2	–	2	2	–	2	3	38
	37	4	2	–	1	1	–	1	–	–	–	19
	38	1	3	1	2	–	7	5	1	1	1	45
Inner islands	39	9	2	3	–	2	1	2	–	–	–	38
	40	5	2	–	–	–	–	–	–	1	2	25
	41	16	5	2	–	1	–	1	–	–	–	29
	42	7	1	1	–	–	3	2	–	–	–	22
	43	12	3	–	–	4	2	1	1	–	–	32
	44	–	2	1	–	–	–	–	–	–	1	7
	45	3	3	–	–	1	1	1	–	1	1	26
	46	1	2	2	–	–	3	2	–	1	2	22
Total		129	118	75	51	46	39	40	32	32	31	

arise for every one of our 10 species. If this restriction is not met, the correction term is too large; N_t is then too small, and may even become negative.

We therefore need a solution for r indistinguishable objects partitioned between n boxes, all subject to constraints, all of which may be different. No analytic solution is known, but the senior author has devised an algorithm based on recurrence relations which provides accurate values of N_f , N_t and therefore p . We shall show that the results, though not without intrinsic interest, do not themselves lead to simple ecological interpretation. A variant of the basic algorithm, however, makes possible a Monte Carlo simulation of any number (we have used 5000) of random arrangements of r objects in n boxes subject to the given constraints. From these it is simple to estimate the random frequency expectations, i.e. the expected number of boxes containing 0, 1, 2, etc. objects, based on the 'constrained occupancy model'. This provides the ecological solution required. Details of the algorithms are given in the Appendix.

Observed distributions of corals were tested statistically for departure from randomness utilizing a χ^2 -test. Expected distributions were generated using both the Poisson method and the new model. Where expected values were small, we followed the procedure recommended by Snedecor and Cochran (1967).

RESULTS

The occupancy parameters

The calculated values of N_f , N_t and p for our 10 species are given in Table 2. All the probabilities are extremely low. To understand why this should be, we need to know the maximum values attainable by any of the possible configurations. This can be determined by using the algorithm, but, for large values of r , the computation time required would be financially unac-

Table 2. Values of N_f , N_t and p for the 10 species of corals and octocorals (listed in order of abundance)

Species No.	N_f	N_t	p
1	0.3628×10^{34}	0.5840×10^{42}	0.6211×10^{-8}
2	0.2771×10^{32}	0.2163×10^{41}	0.1281×10^{-8}
3	0.1809×10^{30}	0.1907×10^{34}	0.9489×10^{-4}
4	0.1623×10^{24}	0.5050×10^{28}	0.3125×10^{-4}
5	0.5422×10^{24}	0.1971×10^{27}	0.2751×10^{-2}
6	0.5388×10^{23}	0.9990×10^{26}	0.5395×10^{-3}
7	0.7188×10^{22}	0.2811×10^{25}	0.2557×10^{-3}
8	0.2703×10^{20}	0.4556×10^{22}	0.5933×10^{-2}
9	0.1291×10^{19}	0.4556×10^{22}	0.2834×10^{-3}
10	0.4396×10^{19}	0.1896×10^{22}	0.2319×10^{-2}

N_f Number of ways in which the observed configuration can be obtained.
 N_t Total number of ways in which r coral colonies can be distributed among 46 transects
 $p = N_f/N_t$

ceptable. We have therefore chosen two values of r for which to calculate exemplary maximum probabilities: $r = 32$ and $r = 45$. The values are, respectively, 0.2605×10^{-1} and 0.6044×10^{-2} . It is evident that, with numbers on these scales, the probability of any specified configuration is low; such probabilities have little ecological meaning. It is of interest though, to determine what configuration corresponds to these maximum probabilities. In Table 3, we list the number of cells with 0, 1, 2, etc. records corresponding to the maximum probability for $r = 32$ (Column 1); any permutation of these, subject to the constraints, will produce the same probability. The corresponding number of $r = 45$ are given in Column 3 of the table. In Columns 2 and 4, the random expectations are given which have been obtained by generating 5000 random allocations for $r = 32$ and $r = 45$, respectively. The closeness of Columns 1 with 2 and Columns 3 with 4 can hardly be fortuitous. χ^2 values for the 2 tables have extraordinarily low values, corresponding to prob-

Table 3. Frequency distributions calculated using maximum probability and random expectation approaches for 2 examples of r

Records per cell	$r = 32$		$r = 45$	
	Maximum probability frequency 1	Random expectation 2	Maximum probability frequency 3	Random expectation 4
0	26	26.86	23	22.97
1	12	11.30	11	11.60
2	5	4.74	6	5.86
3	2	1.88	3	2.92
≥ 4	1	1.22	3	2.65
		$\chi^2 = 0.132$ $p > 0.99$		$\chi^2 = 0.083$ $p > 0.99$

Table 4. χ^2 - tests for all species assuming Poisson vs. constrained occupancy models

Species	Poisson			Constrained occupancy		
	d. f.	χ^2	P	d. f.	χ^2	P
<i>Montastrea annularis</i>	7	70.34	P < .001	9	8.82	0.30 < P < 0.50
<i>Siderastrea siderea</i>	6	6.50	0.25 < P < 0.50	8	25.71	0.001 < P < 0.01
<i>Pseudopterogorgia elisabethae</i>	5	4.49	0.25 < P < 0.50	7	6.24	0.50 < P < 0.70
<i>Agaricia agaricites</i> f. <i>purpurea</i>	4	18.43	P < .005	5	8.03	0.10 < P < 0.20
<i>Montastrea cavernosa</i>	3	2.36	0.5 < P < 0.75	5	3.81	0.50 < P < 0.70
<i>Dichocoenia stokesii</i>	3	3.45	0.25 < P < 0.50	5	3.20	0.50 < P < 0.70
<i>Porites astreoides</i>	3	2.06	0.50 < P < 0.75	4	2.86	0.50 < P < 0.70
<i>Pseudopterogorgia americana</i>	3	3.43	0.25 < P < 0.50	4	1.68	0.70 < P < 0.80
<i>Agaricia agaricites</i> f. <i>agaricites</i>	3	3.02	0.25 < P < 0.50	4	11.88	0.01 < P < 0.02
<i>Manicina areolata</i>	3	12.13	0.005 < P < 0.01	4	8.06	0.05 < P < 0.10

abilities of greater than 0.99. It is normally assumed that probabilities greater than about 0.90 represent resemblance too close to be attributable to chance. It therefore appears that the most probable configuration is the random one. We believe that these 2 examples are suggestive of a general pattern which might also hold for other values of r .

Comparison with the Poisson distribution

The complete set of χ^2 values calculated following both the Poisson and constrained occupancy models is given in Table 4. For 5 of the 10 species both models suggest that the distribution is random. For 3 (*Montastrea annularis*, *Agaricia agaricites* f. *purpurea*, and *Manicina areolata*), the Poisson model shows highly significant deviations from randomness, whereas the occupancy model suggests that the distributions are random or exhibit only marginally significant deviation (in the case of *Manicina*). As examples, we com-

pare the expectations for *Montastrea* and *Manicina* in Table 5. We note that the Poisson model would expect *Montastrea* to be sharply peaked in the region of 2 or 3 specimens per transect, with few zeros or high values. Compared with this, the observed distribution has far too many zeros and too many high values; it would be highly significantly aggregated. In contrast, the observed distribution does not differ significantly from the random expectations under the constrained occupancy model. We feel justified in questioning whether some of the aggregated populations on record may be artefacts of an inappropriate use of the Poisson model. There is much less difference between the two models for the uncommon *Manicina*, though even here a highly significant deviation from expectation has been converted into one which is at best marginally significant.

In 2 cases however, the distribution appears to be random on the Poisson model but non-random on the occupancy model (Table 4); these cases, together with *Manicina*, which is non-random on both, require closer examination.

Table 5. Comparison of expected distributions and χ^2 -tests using Poisson and Constrained Occupancy models for 2 corals

No. per transect	<i>Montastrea annularis</i> (r = 129)			<i>Manicina areolata</i> (r = 31)		
	Observed	Poisson expectation	Constrained occupancy model	Observed	Poisson expectation	Constrained occupancy model
0	14	2.80	11.17	27	23.46	27.26
1	11	7.84	8.77	12	15.81	11.21
2	3	11.00	6.64	2	5.26	4.57
3	4	10.28	4.49	5	1.18	1.84
4 (or 4+)	1	7.21	3.71	0	0.49	1.12
5	5	4.04	2.73			
6	5	1.89	2.04			
7	1	0.76	1.19			
8 (or 8+)	4	0.26	2.36			
		$\chi^2 = 70.34$ (P < 0.001)	8.82 (0.3 < P < 0.5)		$\chi^2 = 12.13$ (P < 0.01)	8.06 (.05 < P < .10)

Species showing deviations from expected distributions under the new model

The 3 species exhibiting significant or marginally significant deviation from expectation based upon the constrained occupancy model were *Siderastrea siderea*, *Agaricia agaricites* f. *agaricites*, and *Manicina areolata*. Detailed results for these corals are given in Table 6. Since the distinction between the 2 forms of

Table 6. Species with deviations from random expectation under the constrained occupancy model

Records per cell	Actual frequency	Random expectation	Contribution to χ^2
<i>Siderastrea siderea</i>			
0	4	12.53	5.80
1	8	9.26	0.17
2	16	6.75	12.68
3	8	4.90	1.96
4	2	3.64	0.74
5	4	2.56	0.81
6	2	1.86	0.01
7	2	1.30	0.38
≥8	0	3.16	3.16
$\chi^2_{(8)} = 25.71$ (0.001 < P < 0.01)			
<i>Agaricia agaricites</i> f. <i>agaricites</i>			
0	20	26.86	1.75
1	20	11.30	6.70
2	6	4.74	0.33
3	0	1.88	1.88
≥4	0	1.22	1.22
$\chi^2_{(4)} = 11.88$ (0.01 < P < 0.02)			
<i>Manicina areolata</i>			
0	27	27.26	0.00
1	12	11.21	0.06
2	2	4.57	1.45
3	5	1.84	5.43
≥4	0	1.12	1.12
$\chi^2_{(4)} = 8.06$ (0.05 < P < 0.10)			

Agaricia agaricites might be regarded as taxonomically controversial, we recalculated expected and χ^2 values using combined records, but the result did not substantially change the values shown in Table 6.

The unexpected feature to emerge here is that, of the 3 species showing deviations from random expectation, only *Manicina* shows any tendency to aggregation, i.e. an excess of observations in the tail of the distribution. The other 2 deviate in the direction of an abnormally uniform or 'regular' distribution, i.e. there is an excess of observations towards the centre of the distribution. These results are strikingly similar to those of Stimson (1974), whose colonies showed random or regular distributions according to the analytical technique used. It may be that the assumption of

uniform spatial probability distribution was met in his study area, implying that the Poissonian model is more appropriate in some areas than in others.

DISCUSSION

Results from the 2 models yield 2 different interpretations concerning the distribution of the corals and octocorals examined in this case study. Analysis by the newly derived 'constrained occupancy model' indicates that these corals are generally randomly distributed with two species uniformly distributed. In only 1 case was a tendency towards an aggregated distribution found. An analysis utilizing the Poisson model would have yielded the conclusion that 3 of the species were aggregated in distribution, 7 were random, and none were uniform. We believe that those derived from the Poisson model may not be correct due to violation of the assumption that the probability of occurrence is uniform over the area of the study. The new constrained occupancy model accounts for variation from this assumption. The Poisson solution is analytic and easily amenable to hand-calculation; the occupancy model, which requires a Monte Carlo simulation, unfortunately, is not. Large-scale computing facilities are, however, becoming increasingly available to ecologists.

It should be recalled that, although the data are extensive in geographic coverage for the Caribbean, they are restricted to the 15 to 42 m (45 to 125 feet) depth range. Therefore, the reader is advised to use care in drawing any general conclusions concerning coral and octocoral distribution in the Caribbean, particularly in shallower water, on the basis of this case study alone. Our main concern here is that other reports of aggregated coral distributions based on Poisson-type models should be accepted with caution.

We believe that the 'constrained occupancy model' is more appropriate than the Poisson model for analysis of distribution of corals and octocorals as well as perhaps other sessile benthic organisms and that our findings merit the attention of field ecologists.

APPENDIX

Algorithms for the computation of 'constrained occupancy' parameters and for the generation of constrained random samples (D. J. Abel).

Computation of N_i

N_i can be determined efficiently by use of a recurrence relation giving the number of ways in which a

total of h objects can be allocated to cells 1 through $(k+1)$. This is calculated from the number of ways g objects can be allocated to cells 1 through k . Let the number of cells be n and the number of objects r . Denote the maximum number of objects able to be allocated to cells by $c_1, c_2, \dots, c_i, \dots, c_n$.

Consider a partial allocation such that a total of g objects has been allocated to the first k cells and $(h-g)$ objects to the $(k+1)$ th cell. A total of $(r-h)$ objects remains to be allocated to the cells $(k+2)$ through n . For such an allocation to be feasible, the following conditions must be met:

$$0 \leq g \leq \sum_{i=1}^k c_i \quad \forall k \in (1, 2, \dots, n)$$

$$0 \leq (h-g) \leq c_{k+1} \quad \forall k \in (1, 2, \dots, n-2)$$

$$0 \leq (r-h) \leq \sum_{i=k+2}^n c_i \quad \forall k \in (1, 2, \dots, n-1)$$

We define δ_{kgh} as taking the value 1 if these conditions hold for a given k, g , and h , and 0 otherwise.

If there are d_{kg} ways of allocating objects to the first k cells, then for a given k and g , each feasible allocating of objects to the $(k+1)$ the cell gives rise to an allocation of h objects (for some h) to the first $(k+1)$ cells, with d_{kg} ways of making the allocation to the $(k+1)$ cells. Considering all possible ways of arriving at a total allocation of h objects (for a given h) to the first $(k+1)$ cells, we have

$$d_{(k+1)h} = \sum_{g=0}^r d_{kg} \delta_{kgh} \quad \forall k \in (1, 2, \dots, n-1)$$

As a total of r objects must always be allocated to the n cells

$$d_{1r} = 1$$

$$d_{1g} = 0 \quad \forall g \neq r$$

N_t is defined as the total number of ways of allocating a total of r objects to the n cells. Thus,

$$N_t = \sum_{i=g'}^{g''} d_{ni}$$

where $g' = \min(0, r - \sum_{i=1}^{n-1} c_i)$

$$g'' = \max(r, c_n)$$

Computation of N_t

N_t can be calculated by a form of exhaustive enumeration, using a tree search to form all different permutations of an observed set of occupancies for each cell.

Denote the observed set of occupancies by b . To generate only different permutations of the elements of b , the tree search considers allocations of the different members of b . Let d be the set of different elements of b , and e be the count of times the corresponding element of d appears in b .

To reduce the solution times of a full tree search explicitly generating all permutations, two mechanisms are applied. First, as each allocation is made to a cell within a permutation, a test is made for the remaining occupancies to determine whether each can be allocated to all remaining cells. If in any starting sequence this condition is met, the number of permutations with the allocations made at that time can be trivially solved as a classical occupancy problem. Secondly, to force this condition to apply early in the tree, d is sorted into descending order and c in ascending order. However, computational difficulties may remain, as clearly there can arise cases where classical occupancy problems are not encountered until late in the tree. Our experience on a CDC Cyber 76 computer suggests that the algorithm becomes prohibitively expensive if the largest observed occupancy cannot be allocated to 8 or more cells with 7 or more different occupancies.

Generation of a random sample

A random sample of allocations to the n cells can be generated by a Monte Carlo process which proceeds cell by cell. We denote a set of ordered allocations by a .

Consider allocation to the first cell within a single sample. For that allocation and the full set of allocations to be feasible, we require that a_1 be bounded such that

$$g'' \leq a_1 \leq g'$$

where $g'' = \max(0, r - \sum_{i=2}^n c_i)$

$$g' = \min(c_1, r)$$

For each a_1 within this range, the algorithm for N_t can be applied to determine N_{1a} , the number of ways of allocating a_1 objects to the first cell and a total of $(r-a_1)$ objects to the cells 2 through n . We observe that in a randomly-generated set of allocations with all sets equiprobable, the probability of having a given a_1 objects in the first cell is given by

$$N_{1a} / \sum_{i=g'}^{g''} N_{1i}$$

A value for a_1 can thus be sampled by generating a random number s under a rectangular distribution in the range $(0,1)$ and choosing the smallest a_1 for which

the cumulative probability exceeds s . More formally, the number of objects allocated to the first cell is given by

$$a_1 = k$$

where k is the smallest integer such that

$$\sum_{j=g''}^k N_{1j} \geq s \sum_{j=g'}^{g''} N_{1j}$$

Allocation of objects to the cells 2 through $(n-1)$ must consider the objects already allocated before the cell is reached. The bounds on the number of objects able to be allocated to the i th cell are given by

$$g_i'' \leq a_i \leq g_i'$$

where $g_i'' = \max(0, r - \sum_{j=i+1}^n c_j - \sum_{j=1}^{i-1} a_j)$

$$g_i' = \max(c_i, r - \sum_{j=1}^{i-1} a_j)$$

The number of objects allocated to the cells 2 through $(n-1)$ are then given by

$$a_i = k$$

where k is the smallest integer such that

$$\sum_{j=g_i''}^k N_{ij} \geq s \sum_{j=g_i'}^{g_i''} N_{ij}$$

For the n th cell

$$a_n = r - \sum_{j=1}^{n-1} a_j$$

Where several samples are to be generated, it is obviously desirable to generate a matrix of all possible N_i values that could be referenced for a given r objects and a given n cells.

Schematic sequence for N_i algorithm

- (i) From b , form d and c , with d in descending sequence
- (ii) Sort c into ascending sequence
- (iii) Cell under consideration indexed by i ; occupancy for current cell indexed by j_i
 $N_i \leftarrow 0$; $i \leftarrow 1$; $j_i \leftarrow 0$
- (iv) Determine next occupancy for cell i
Find smallest k such that
 $e_k > 0$; $k > j_i$; $d_k \leq c_i$
(If none remains, all possible occupancies for i within permutations within current allocations to cells 1 through $(i-1)$ have been evaluated. Move to previous cell. If none, go to (vii).)
- (v) Effect allocation to cell
 $j_i \leftarrow k$
 $e_k \leftarrow e_k - 1$

- (vi) Has a complete permutation been considered?
If $i \neq n$, go to (viii)
 $N_i \leftarrow N_i + 1$
- (vii) Deallocate current cell and move to previous cell
 $k \leftarrow j_i$; $e_k \leftarrow e_k + 1$; $j_i \leftarrow 0$
Beware of case where current cell is first. When this is reached all permutations have been considered
If $i = 1$, exit
 $i \leftarrow i - 1$
Go to (iv)
- (viii) Can all remaining occupancies be allocated to cells $(i+1)$ through n ?
If so, solve a classical occupancy problem.
Find k as $\max(d_i)$ with $e_i \neq 0$
If $k > c_{i+1}$ go to (ix)
Solve for N_i' , the result obtained from the classical occupancy problem for unallocated d to $(n-i)$ cells.
 $N_i \leftarrow N_i + N_i'$
Go to (vii)
- (ix) Move to next cell
 $i \leftarrow i + 1$
 $j_i \leftarrow 0$
Go to (iv)

Availability of computer programs

FORTRAN programs implementing the algorithms described above and designed for use on a CDC Cyber 76 computer are available from the senior author.

LITERATURE CITED

- Ashby, E. (1935). The quantitative analysis of vegetation. *Ann. Bot.* 49: 779-802
- Benayahu, Y. (1975). Quantitative characteristics of community structure of stony corals, soft corals and algae in the northern Gulf of Eilat (Red Sea). M. Sc. thesis, Tel Aviv University (in Hebrew; English abstract sighted)
- Birkeland, C., Rowley, D., Randall, R. H. (1982). Coral recruitment patterns at Guam. In: Gomez, E. D., Birkeland, C. E., Buddemeier, R. W., Johannes, R. E., Marsh, J. A., Jr., Tsuda, R. T. (ed.) *The reef and man. Proceedings of the 4th international Coral Reef Symposium, Manila, 1981*
- Blackman, G. E. (1942). Statistical and ecological studies in the distribution of species in plant communities. I. Dispersion as a factor in the study of changes in plant communities. *Ann. Bot. N.S.* 6: 351-370
- Bradbury, R. H., Young, P. C. (1982). The race and the swift revisited, or is aggression between corals important? In: Gomez, E., Birkeland, C. E., Buddemeier, R. W., Johannes, R. E., Marsh, J. A., Jr., Tsuda, R. T. (ed.) *The reef and man. Proceedings of the 4th International Coral Reef Symposium, Manila, 1981*

- Bunt, J. S., Pepper, S. (1973). *In situ* studies on the tropical benthos. III. Transect data and species lists. Technical Rept. US-RSMAS No. 73509
- Bunt, J. S., Williams, W. T., Chalker, B. E. (1982). Coral associations at depths of 45 to 125 feet in the Bahamian region. In: Gomez, E., Birkeland, C. E., Buddemeier, R. W., Johannes, R. E., Marsh, J. A., Jr., Tsuda, R. T. (ed.) The reef and man. Proceedings of the 4th International Coral Reef Symposium, Manila 1981
- Clapham, A. R. (1936). Over-dispersion in grassland communities and the use of statistical methods in plant ecology. *J. Ecol.* 24: 232–251
- Clark, P. J., Evans, F. C. (1954). Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology* 35: 445–453
- Dana, T. F. (1976). Reef-coral dispersion patterns and environmental variables on a Caribbean coral reef. *Bull. mar. Sci.* 26: 1–13
- Feller, W. (1957). Introduction to probability theory, Vol. I. John Wiley & Sons, New York.
- Grassle, J. F. (1973). Variety in coral reef communities. In: Endean, R., Jones, O. A. (ed.) Biology and geology of coral reefs. Academic Press, New York, p. 247–270
- Greig-Smith, P. (1952). The use of random and contiguous quadrats in the study of the structure of plant communities. *Ann. Bot. N.S.* 16: 293–316
- Greig-Smith, P. (1964). Quantitative plant ecology, 2nd ed. Butterworths, London
- Lewis, J. B. (1970). Spatial distribution and pattern of some Atlantic reef corals. *Nature, Lond.* 227: 1158–1159
- Maragos, J. E. (1974). Coral communities on a seaward reef slope, Fanning Island. *Pacif. Sci.* 28: 257–278
- Morisita, M. (1959). Measuring of the dispersion of individuals and analysis of the distributional patterns. *Mem. Fac. Sci. Kyushu Univ. Ser. E.* 2: 215–235
- Snedecor, G. W., Cochran, W. G. (1967). Statistical methods, 6th ed. Iowa State University Press, Ames, Iowa
- Stevens, W. L. (1937). Significance of grouping. *Ann. Eugen.* 8: 57–69
- Stimson, J. (1974). An analysis of the pattern of dispersion of the hermatypic coral *Pocillopora meandrina* var. *nobilis* Verrill. *Ecology* 55: 445–449
- 'Student' (1919). An explanation of deviations from Poisson's law in practice. *Biometrika* 12: 211–215
- Wallace, C. C., Bull, G. D. (1982). Patterns of juvenile coral recruitment on a reef front during a spring-summer spawning period. In: Gomez, E., Birkeland, C. E., Buddemeier, R. W., Johannes, R. E., Marsh, J. A., Jr., Tsuda, R. T. (ed.) The reef and man. Proceedings of the 4th International Coral Reef Symposium, Manila, 1981
- Williams, W. T., Haydock, K. P., Edey, L. A., Ritson, J. B. (1973). Analysis of a fertility trial with Droughtmaster cows. *Aust. J. agric. Res.* 22: 979–991
- Williams, W. T., Clay, H. J., Bunt, J. S. (1982). The analysis, in marine ecology, of three-dimensional data matrices with one dimension of variable length. *J. exp. mar. Biol. Ecol.* 60: 189–196

This paper was presented by Dr. R. P. M. Bak; it was accepted for printing on April 8, 1983