Analyzing species distributions among temporary ponds with a permutation test approach to the join-count statistic

Patrick H. Stevens[1],* & David G. Jenkins[2],**

[1]*Department of Mathematics, University of Illinois at Springfield, P.O. Box 19243, Springfield, IL 62794-9243, USA;* [2]*Department of Biology, University of Illinois at Springfield, P.O. Box 19243, Springfield, IL 62794-9243, USA;* *Current address: Department of Statistics, University of Illinois at Urbana-Champaign, 101 Illini Hall, 725 South Wright Street, Champaign, IL 61820, USA;* **Author for correspondence (E-mail: jenkins.david@uis.edu)*

Abstract

The Bluff Springs Sand Ponds (BSSPs) are a set of closely-spaced temporary ponds of varying hydroperiod, depth and surface area. We sampled crustacean communities of 15 ponds throughout hydroperiods in 1996 to examine species distributions among ponds. Although ponds were closely spaced (within ca. 16 ha), most species were present in subsets of the 15 ponds. We then analyzed spatial patterns of 12 crustacean species for complete spatial randomness (CSR) using join-count statistics. However, the join-count was designed for large-samples ($n > 50$), so we further analyzed (by simulation) the join-count and a variation of the join-count (Cliff & Ord, 1981) for small-scale reliability. Simulation results revealed that neither testing distribution was reliable for $n < 30$. We then used a permutation test to analyze species distributions and concluded that some species were distributed non-randomly. Therefore, further investigations of mechanisms causing species distributions (e.g., hydroperiod, physical/chemical conditions, biotic interactions) are clearly prescribed. The permutation test should be useful for studies of species distribution patterns among other temporary waters, and can help focus studies on causal mechanisms of distributions among small numbers of temporary aquatic habitats.

Introduction

The analysis of spatial pattern in community composition is important for the basic and applied ecology of temporary ponds. The detection of patterns can better direct studies on the processes and environmental factors responsible for the patterns (e.g., Schneider & Frost, 1986), and conservation efforts must begin with inventories of species, including their locations (e.g., Simovich, 1998). Distributions of temporary pond species can be presented effectively as tables or maps, but further examination may lead to questions about mechanisms responsible for the distributions (e.g., environmental conditions, biotic interactions, dispersal). A valuable first question may then be 'Is there a pattern to the spatial distributions (i.e., spatial autocorrelation), or are the distributions random?'

Statistical analysis of spatial distributions is best performed with larger numbers of sample sites (e.g., >50; Cliff & Ord, 1981), but studies of multiple temporary ponds can rarely approach these numbers due to practical constraints and the limited numbers of ponds remaining in a human-dominated landscape. Therefore, spatial autocorrelation statistics for the study of temporary ponds would ideally characterize significant spatial pattern for species among a small number of sampling sites. The purpose of this paper is to describe an approach we made to one type of spatial autocorrelation statistic for use with data sets of relatively few sites. We developed this approach during our analyses of species distributions among a set of temporary ponds.

A variety of spatial autocorrelation statistics exist, including join-counts, Moran's I, Geary's C, and K-hat functions. We focus here on one statistic ap-

Table 1. Physical characteristics of the Bluff Springs Sand Ponds in 1996. See Figure 1 for pond locations

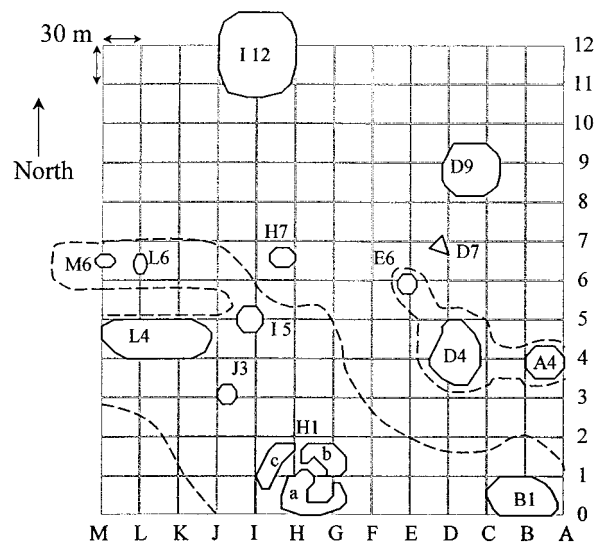| Pond	Initial pooling date	Hydroperiod (days)	Maximum surface area (m^2)	Maximum depth (cm)
A4	April 30	48	350	25
B1	April 30	55	4000	16
D4	April 23	65	3700	41
D7	May 7	41	200	25
D9	April 23	75	1800	46
E6	April 30	48	1200	20
L4	April 30	58	3300	46
H1a	April 30	58	2400	41
H1b	April 30	58	2400	43
H1c	April 30	58	2400	36
I5	April 30	58	800	41
I12	May 7	65	4000	56
J3	April 30	58	900	20
L4	April 30	58	3300	46
L6	May 7	51	2500	48
M6	May 7	51	800	51

Figure 1. Bluff Springs Sand Ponds. Ponds are named according to location on the grid, and are drawn as approximate initial size during 1996. Dashed lines indicate approximate locations of highest observed water line: ponds A4, D4, and E6 can merge into one subset in wet years, as can Ponds B1, H1, I5, J3, L4, and L6. Ponds D7, D9, H7, and I12 are each isolated from all other ponds by topography. The site is bounded to the east and south by a berm that separates ponds from roadside ditches adjacent to paved roads (however, a break in the southern ditch near H1 permits merger of pond and ditch water). Pond H7 was not included in the current study.

propriate for the analysis of presence/absence data: join-count statistics. Cliff & Ord (1971, 1973) developed join-count statistics for geographical applications, and Sokal & Oden (1978a, b) demonstrated the utility of these statistics for biological applications. Join-counts have been used in other ecological studies (Gilbert et al., 1994; Real & McElhany, 1976; Sokal & Oden, 1978b) and are well suited for lattice type spatial models, as defined by Cressie (1991). We considered the lattice model (e.g., a chess board) to be most appropriate for analyzing the distributions of aquatic species among temporary ponds: ponds with a particular species present could be considered as black squares in a randomized chess board, separated by white squares (species not present). The K-hat function (Bailey & Gatrell, 1995) has been mainly used in point pattern analysis, which is more appropriate for data positioned among points in a uniform landscape (e.g., trees in a forest).

Materials and methods

The Bluff Springs Sand Ponds (BSSPs) are a set of temporary ponds within a 16 ha forested wetland in Cass County, Illinois. The ponds are surrounded by a ridge of sandy soil that was deposited 10–12 000 years ago (Willman & Frye, 1970). Topography and precipitation combine to yield subsets of ponds that can merge in wet years, and other ponds that are completely isolated (Figure 1). Despite the ponds' proximity, the hydroperiod, initial pooling date, approximate maximum area, and maximum depth varied widely from pond to pond (Table 1).

We sampled 15 ponds weekly, beginning within one week of initial pooling (April 23,1996) to the end of the longest hydroperiod (June 25, 1996). We used a water column sampler (cylinder: 61 cm high × 40 cm diameter), which was forced into the sediment, and all water within the cylinder was sieved on 35 μm. Collected organisms were rinsed off the sieve and preserved with 4% buffered formalin. Samples were examined for species present, and a record of species found in each pond was tabulated.

Results

Seventeen crustacean species were identified, and species presence/absence over the entire hydroperiod (Table 2) was used to calculate join-count statistics.
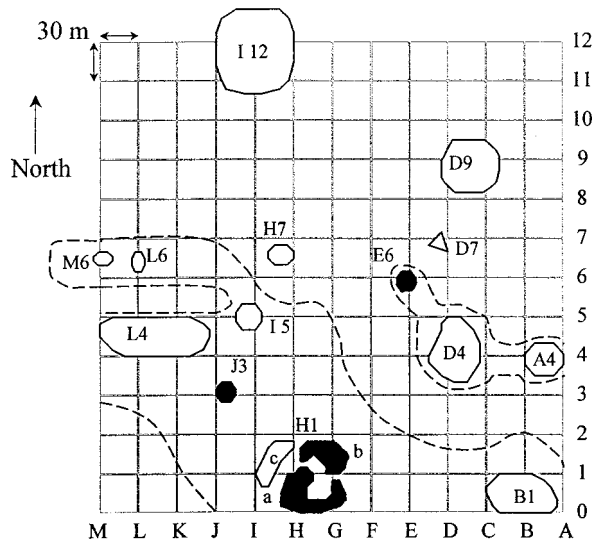
Figure 2. Spatial pattern of *Cryptocyclops bicolor* in the Bluff Springs Sand Ponds during 1996. Black-filled ponds indicate presence, white-filled ponds indicate absence.

Five species will not be considered for statistical analyses in this study, because: (a) they appeared in all 15 ponds, so no pattern exists (*Canthocamptus* sp., *Cyclops navus, Daphnia obtusa*); or (b) they appeared in one pond each and also cannot be considered to display a pattern requiring statistical analysis (*Onychodiaptomus sanguineus, Pleuroxus striatus*). The remaining twelve species were tested for complete spatial randomness (CSR). As an example of an interesting pattern that may benefit from statistical analysis, the distribution of the copepod *Cryptocyclops bicolor* is shown in Figure 2.

Join-count analyses

The presence/absence pattern for a species may be considered analogous to that of a mixed-up chess board, with B (black) for species present and W (white) for species absent. Statistics based on binary random variables (where present = 1, absent = 0) can then be calculated to determine whether the pattern of B's and W's are random or show some sort of clustering. Cliff & Ord (1973) developed three join-count statistics, designated BB, BW, and WW. For species presence/absence data, species status x_i is either 1 (corresponding to $B = 1$) or 0 ($W = 0$). BB join-counts represent the statistic for all pairs of sampling sites where both sites have a value of B (species present). BW join-counts represent the statistic for all

pairs of sites where one site is B (species present) and the other is W (species absent). The final join-count, WW, represents the statistic for all pairs of sites where both sites are W (species absent). The three statistics are as follows:

$$BB = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} {}_{i \neq j} w_{ij} x_i x_j,$$

$$BW = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} {}_{i \neq j} w_{ij} (x_i - x_j)^2,$$

and

$$WW = \frac{1}{2} S_0 - (BB + BW),$$

where

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} {}_{i \neq j} \omega_{ij},$$

w_{ij} is the weight (Euclidian distance) between site i and site j, and x_i is the value of the binary random variable at site i ($1 = $ present, $0 = $ absent).

The weight, w_{ij}, between sites i and j is an important aspect of the statistic. For sites separated by a spatial distance, the weight could be the Euclidian distance, squared Euclidian distance, or any other weighting system deemed appropriate by the researchers. In Cliff & Ord's chess board analogy, the weights were binary, depending on the condition of a shared, contiguous border (i.e., nearest neighbors). Weights between sites can be arranged into a weighting matrix for more efficient mathematical computation. It should be noted that any type of weights deemed valid can be used, and should depend on the species and environment under study. When studying presence/absence patterns among habitats, distance-related weights may serve as a representative of dispersal probabilities among habitats, assuming that dispersal probability is inversely related to distance.

We chose Euclidean distances (between all pairs of pond centers) for intersite weights as a simple measure among the closely-spaced ponds. Most join-count analyses have considered only nearest-neighbors as connected. Given the proximity and antiquity of the ponds, we thought it reasonable to expect species to have had dispersal opportunities among all ponds via various vectors. For testing CSR, the null hypothesis assumed a random distribution of the species over the number of ponds where the species was present. For each species, three join-count statistics

Table 2. Species presence/absence by Pond in 1996. Presence is indicated by a value of 1; absence by a value of 0. All species observed are listed here, but species present in all ponds (*Canthocamptus* sp., *Cyclops navus, Daphnia obtusa*), or in one pond only (*Onchydiaptomus sanguineus, Pleuroxus striatus*) were not included in join-count analyses of spatial pattern

Species	A4	B1	D4	D7	D9	E6	H1a	H1b	H1c	I5	I12	J3	L4	L6	M6
Attheyella sp.	1	1	0	0	1	0	1	0	1	1	1	0	0	0	0
Canthocamptus sp.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cryptocyclops bicolor	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0
Cyclops navus	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cyclops nearcticus	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Cyclops haueri	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
Cypridopsis sp.	0	1	1	0	0	0	1	0	0	1	1	0	1	1	0
Cyprois sp.	0	1	1	0	1	0	1	1	1	1	0	1	1	1	0
Daphnia obtusa	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Eubranchipus serratus	0	1	0	0	0	0	1	1	0	1	0	1	0	1	0
Lynceus brachyurus	0	1	0	0	1	0	0	0	0	1	1	1	0	1	0
Onychodiaptomus sanguineus	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Osphranticum labronectum	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0
Pleuroxus striatus	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Scapholebris mucronata	0	0	0	0	1	0	1	1	1	0	0	0	1	1	1
Simocephalus exspinosus	1	1	1	0	1	1	0	1	1	0	1	0	0	0	0
Simocephalus serrulatus	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0

(BB, BW, WW) were calculated and tested as standard normal deviates (SNDs) based on the moments prescribed by Cliff & Ord (1981).

Results of the join-count tests are summarized in Table 3. Because we considered this an exploratory analysis (of spatial pattern, and then of join-count statistics), we were wary of a Type II error, and so we chose significance levels of $p \leq 0.10$ to indicate spatial autocorrelation. Significance with a positive SND for the BB statistic and/or negative SND for the BW statistic indicates that a distribution was more widely dispersed than expected at random (i.e., negative spatial autocorrelation). *Atheyella* sp., *Cyclops nearcticus*, *Simocephalus exspinosus*, *Simocephalus serrulatus*, and *Osphranticum labronectum* had negative spatial autocorrelation of presence/absence distributions (Table 3). Significant positive autocorrelation (i.e., a species is more spatially clustered than expected at random) is indicated by negative SNDs for the BB statistic and/or positive SNDs for the WW statistic. *Cryptocyclops bicolor, Cyclops haueri, Cyprois*, and *Scapholeberis mucronata* had distributions displaying positive spatial autocorrelation (Table 3). It should be noted that other combinations of significance among the three statistics may not clearly indicate the type of autocorrelation present.

The join-counts (above) were tested as standard normal deviates. However, Cliff & Ord (1973) showed that join-counts are asymptotically-normal distributed. Because of this, join-counts of spatial arrays with few sites (e.g., many temporary pond studies, including ours) may not be valid. Therefore, Cliff & Ord (1981) performed small-sample-size simulations, and found that the normal assumption was inappropriate for hypothesis testing at smaller sample sizes ($10 < n < 50$). Based on these results, they concluded that certain small-sample corrections would make a better approximation than the asymptotic normal approximation. Cliff & Ord (1973) also recommended other guidelines for using these small-sample corrections: (1) the corrections are not recommended when one locality figures in more than about 30% of the joins; (2) the corrections are only valid for $\alpha > 0.10$ in each tail; (3) they recommended using a value (e.g., the proportion of sites labeled B) > 0.5. For example, in a system with 4 of 10 sampling sites having a value of B (proportion $= 0.4$), it would be better to test the W value since the proportion of W scores would be 0.6; and (4) they recommended evaluating join-count statistics only when the proportion value is between 0.2 and 0.8 (Cliff & Ord, 1981).

Although Cliff & Ord's small sample corrections worked reasonably well for the data sets they ana-

Table 3. Summary of join-count analysis results, using standard normal deviates. See text for explanation of positive (+) or negative (−) spatial autocorrelation designations. $P =$ number of ponds in which a species was present in 1996. Values shown for the BB, BW, and WW statistics are the significance levels (p-values) of the statistics; asterisks denote p-values considered significant (≤ 0.100) for this study

Species	Autocorrelation	P	BB	BW	WW
Attheyella sp.	−	7	0.100*	0.330	0.079*
Cryptocyclops bicolor	+	4	0.032*	0.122	0.057*
Cyclops nearcticus	−	2	0.063*	0.023*	0.019*
Cyclops haueri	+	2	0.259	0.066*	0.067*
Cypridopsis sp.		7	0.144	0.350	0.251
Cyprois sp.	+	10	0.068*	0.041*	0.318
Eubranchipus serratus		6	0.129	0.342	0.317
Lynceus brachyurus	−	9	0.233	0.147	0.015*
Osphranticum labronectum	−	3	0.329	0.008*	0.014*
Scapholeberis mucronata	+	7	0.200	0.100*	0.452
Simocephalus exspinosus	−	8	0.422	0.013*	0.027*
Simocephalus serrulatus	−	3	0.034	0.170	0.104

lyzed, their analyses of binary join-count statistics only included sample sizes of 25 or larger (Cliff & Ord, 1973). Their recommendations for sample sizes as small as 10 were based on simulation analyses they performed on Moran's I, which was developed for interval class variables; not for binary variables (Cliff & Ord, 1971). In addition, their weighting matrix for the binary join-count simulations was also binary (e.g., weight for adjacent black squares on a chess board = 1). Cliff & Ord justified this because 'non-binary weights make the distribution less lumpy and somewhat easier to approximate' (Cliff & Ord, 1973). In other words, they used a worst-case scenario, assuming other conditions would approach normality faster. However, two important questions arise. First, how fast do other, more ecologically-realistic types of weighting matrices approach normality? Our system used a weighting matrix of non-binary values and the statistics should therefore approach normality faster than when a binary weighting matrix is used. Secondly, are Cliff & Ord's small-scale corrections valid for those types of weighting matrices? The small-scale corrections may only be valid for join-counts that use a binary weighting matrix. To answer these questions, we analyzed the join-count statistics for simulated conditions of small sample size (i.e., few sites), and using a non-binary, completely-connected weighting matrix (Euclidean distances), as we expected these conditions to be of greatest value in temporary pond ecology.

Simulations

Simulations were run using various sample sizes to investigate the validity of the normal and small-scale distribution with a non-binary, completely-connected matrix. For the following simulations, P represents the number of sites with a species present. For each sample size $n = 10–20, 22, 24, 26, 28$, and 30, the BB and BW join-count statistics were simulated for each value of P, except for the cases of $P = 1, n−1$, and n: a species being present ($P = 1$) or absent ($P = n−1$) at only one site; or present at all 15 sites ($P = n$) cannot be considered to display a complex pattern requiring statistical analysis. SAS/Base and SAS/IML was used as a programming language, and simulations proceeded as follows:

1. Generate a random non-binary, completely-connected weighting matrix. The random number generator used was the *rannor* or *ranuni* function provided in the SAS/Base software (1994).
2. Generate random placement of the P presence locations, using the same random number generators as above.
3. Calculate the appropriate join-count.
4. Classify the join-count into one of eleven mutually exclusive classes based on the appropriate theorized distribution. The eleven classes of join-count statistics were: $y < L(0.001)$; $L(0.001) \leq y < L(0.005)$; $L(0.005) \leq y < L(0.025)$; $L(0.025) \leq y < L(0.05)$; $L(0.05) \leq y < L(0.1)$; $L(0.1)$

$\le y <$ L(0.9); L(0 .9) $\le y <$ L(0.95); L(0.95) $\le y <$ L(0.975); L(0.975) $\le y <$ L(0.995); L(0.995) $\le y <$ L(0.999); and L(0.999) $\le y$, where L(α) is the lower 100α percentile, and y is the number of the simulations which fall into each class.

5. After generating and classifying each of 1567 simulations, test the empirical distribution against the theoretical distribution (either normal or small-scale) using a chi-square goodness-of-fit statistic with 10 deg of freedom (11 classes–1). Simulations were replicated to a sample size of 1567 based on a maximum error of estimation of 0.05(σ^2).

Complete details of the simulation results can be found in Stevens (1998). The results indicated that when using non-binary, completely-connected weighting matrices (e.g., Euclidean distances among all ponds), the BB join-count is more reliably tested as a standard normal deviate for $n > 23$ sites, and the BW join-count cannot be tested as a standard normal deviate for $n < 30$. In addition, the small-scale corrections recommended by Cliff & Ord were not reliable when the weighting matrix was completely-connected and non-binary.

An empirically-based permutation test approach

Based on the above analyses, we concluded that analyses of spatial patterns among few sites and using the usual join-counts (as standard normal deviates or the recommended small-scale corrections) are not reliable when using a completely-connected non-binary weighting matrix (e.g., Euclidean distances). This presents a dilemma for ecological investigations of temporary aquatic habitats, and other ecosystems, that may benefit from the use of join-count statistics to examine species' spatial patterns. We think it is unrealistic to weight join-count statistics by a species' presence or absence in nearest-neighbor sites only. This approach essentially assumes that organisms disperse only among nearest neighbors. We thought it was more realistic to assume that all sites may be connected by dispersal (non-binary, completely-connected weighting matrix), especially when sites are closely spaced, as at the Bluff Springs Sand Ponds. Of course, the value of inter-site connection weights implies knowledge of dispersal rates between sites. In the absence of more information, we considered it reasonable to use inter-site distance in the weighting matrix, on the assumption that distance is related to dispersal proba-

bilities among sites. Because we favored a non-binary, completely-connected weighting matrix for ecological systems, and because the small-scale corrections of Cliff & Ord (1973) do not apply for that case, we used a permutation test to examine the join-counts and then compared them to the standard normal deviate results. The permutation test was based on empirical distributions in the BSSP presence/absence data, and could be applied to other empirical data sets as well.

An empirically-based permutation test is randomized sampling of all possible arrangements, within the limits of the number of ponds occupied by a species. For instance, if a species was present in 7 of the 15 ponds, then the permutation test randomly assigns the species to seven locations among the 15 possibilities. This is performed for many replications, and a join-count statistic is calculated for each replication. A frequency distribution of the random join-count values can then examined for the rare large or small values (tails of the distribution) that correspond to 1%, 5% or 10% probabilities. The join-counts for the actual study can then be compared to the simulated percentage points. Full details of the permutation tests can be found in Stevens (1998).

The results of both the normal (SND) and empirical (permutation) join-count significance tests for the BSSP data were very similar (Table 4). The normal statistics indicate significance at almost the same levels as the empirical distribution. The significance values of the BB statistic based on normal and empirical distributions are identical. The WW statistics indicate significance for the same species, although the levels differ in a few cases, and the BW statistics are also similar in most cases (though not identical). Therefore, it would appear that the normal and empirical distributions give very similar results, despite the simulation results.

One should ask then: why bother with a permutation test, if it provides the same answer as the normal assumption? An answer may be that both distributions happened to provide similar results for the data we collected, and that other data sets may not lead to the same coincidence. To address this possibility, we made a final simulation. We generated a random set of BB join-count statistics for the range of P values obtainable with the BSSP data ($2 \le P \le 13$), and using either the empirical or normal distributions. We compared the distributions with Chi-square goodness-of-fit tests (Table 5). All chi-square scores were significant ($\alpha = 0.05$, $\chi^2 = 18.307$), indicating that the distributions are different. We concluded that the

Table 4. Comparison of join-count results using standard normal deviate (SND) and empirically-based permutation (EMP) tests. ns = not significant, and values shown are significance levels of tests, where values ≤ 0.100 were considered significant for this study

Species	*B B*		*B W*		*W W*	
	SND	EMP	SND	EMP	SND	EMP
Attheyella sp.	ns	ns	ns	ns	0.10	0.10
Cryptocyclops bicolor	0.05	0.05	ns	ns	0.10	0.05
Cyclops nearcticus	0.10	0.10	0.025	0.10	0.025	0.05
Cyclops haueri	ns	ns	0.10	0.05	0.10	0.025
Cypridopsis sp.	ns	ns	ns	ns	ns	ns
Cyprois sp.	0.10	0.10	0.05	0.10	ns	ns
Eubranchipus serratus	ns	ns	ns	ns	ns	ns
Lynceus brachyurus	ns	ns	ns	ns	0.025	0.025
Osphranticum labronectum	ns	ns	0.01	0.025	0.025	0.05
Scapholeberis mucronata	ns	ns	0.10	ns	ns	ns
Simocephalus exspinosus	ns	ns	0.025	0.05	0.05	0.05
Simocephalus serrulatus	0.05	0.05	ns	ns	ns	ns

empirically-based permutation tests would be more valid than comparisons to a normal distribution in this case, and that the two distributions coincidentally provided similar significance values for our data set. It is likely that other empirical distributions (i.e., other weighting matrices) would yield significance values more different from analyses based on normal distributions, which would only strengthen the argument for using a permutation test for join-count statistics under small-sample situations.

Using permutation tests on join-counts can be very useful for temporary ponds research, and should foster experimental approaches on causative mechanisms of non-random distributions uncovered by this technique. Our approach to join-count statistics can be used when relatively few ponds (e.g., 15) have been censused; an important concern for studies that require large efforts among rare habitats. A generalized version of the SAS program used to perform the permutation test is available at http://www.uis.edu/ jenkins/joinct. The program requires SAS/Base and SAS/IML, and is designed for binary random variables (such as species presence/absence) and any type of weighting matrix the researcher defines. Users of the program should cite this paper. Interested readers should also note that commercially-available statistics programs also include permutation tests of spatial autocorrelation statistics (e.g., the S Plus Spatial Stats module includes permutation tests of Moran's I).

Ecological significance

As indicated by both normal (SND) and permutation test results, some species were distributed non-randomly among the Bluff Springs Sand Ponds, despite the relatively small spatial scale and high density of ponds at the site (15 ponds in 16 ha). Further analyses of additional years may reveal different patterns, especially when other years have different hydrological conditions (e.g., earlier filling of ponds, longer hydroperiod). Therefore, it would be hasty to conclude that species distributions revealed in this single-year study are definitive.

Also, interesting patterns can sometimes be revealed more simply than by this statistical approach. For example, *Eubranchipus serratus* did not have significant spatial autocorrelation in any of the above analyses, but comparison of Table 2 and Figure 1 shows that it was present during 1996 only in ponds that can connect in years of high precipitation. We used simple Euclidean distances as weights in our analyses, based on the assumption that dispersal vectors among the closely-spaced ponds could have historically transported species among all ponds. This assumption may not be appropriate for *Eubranchipus serratus*, or conditions in 1996 may have varied among pond subsets in ways that were important for this species.

However, visual inspection of maps could not clearly identify non-random distributions for most species. The scientific method requires that null hy-

Table 5. Empirical (Emp) and normal distributions of simulated BB join-count values, expressed as percentiles in simulations of $N = 1000$. Distributions were simulated for each of the presence (P) values analyzed for Bluff Springs Sand Ponds data. The simulated empirical and normal distributions were compared by χ^2 goodness-of-fit tests: all χ^2 scores were significant ($p = 0.05$, $\chi^2 = 18.3$)

Percentiles	$P = 2$		$P = 3$		$P = 4$		$P = 5$		$P = 6$		$P = 7$	
	Emp.	Normal	Emp.	Normal	Emp.	Normal	Emp.	Normal	Emp.	Normal	Emp.	Normal
< 0.005	0	0	5	7	2	4	7	13	3	4	3	9
0.005–< 0.010	13	0	10	6	7	5	8	11	7	2	2	4
0.010–< 0.025	15	11	11	13	18	11	16	8	8	23	19	24
0.025–< 0.050	22	20	23	34	25	34	18	26	26	25	24	23
0.050–< 0.100	42	62	67	37	49	58	61	58	55	41	38	51
0.100–< 0.900	820	800	782	812	784	788	795	784	807	786	820	793
0.900–< 0.950	49	55	50	36	53	35	60	47	47	59	50	65
0.950–< 0.975	25	20	25	22	36	39	12	26	28	32	24	16
0.975–< 0.990	0	22	15	13	15	14	11	21	6	23	13	13
0.990–< 0.995	14	0	9	9	4	12	6	6	6	5	1	1
> 0.995	0	10	3	11	7	0	6	0	7	0	6	1
χ^2 values:	42.51		44.52		24.95		30.25		43.62		42.78	

Percentiles	$P = 8$		$P = 9$		$P = 10$		$P = 11$		$P = 12$		$P = 13$	
	Emp.	Normal	Emp.	Normal	Emp.	Normal	Emp.	Normal	Emp.	Normal	Emp.	Normal
< 0.005	4	5	3	10	6	9	4	11	4	6	0	8
0.005–< 0.010	5	8	5	11	2	3	3	10	3	15	19	5
0.010–< 0.025	17	8	14	12	17	21	16	23	14	26	12	16
0.025–< 0.050	29	31	28	25	29	24	23	22	20	27	19	49
0.050–< 0.100	47	41	43	53	54	62	65	58	51	41	63	8
0.100–< 0.900	807	800	798	791	802	779	795	775	804	823	809	855
0.900–< 0.950	49	58	49	62	46	65	44	80	62	43	29	59
0.950–< 0.975	20	32	38	26	23	30	20	18	19	19	36	0
0.975–< 0.990	10	13	12	9	13	5	16	3	13	0	6	0
0.990–< 0.995	8	4	3	1	4	2	8	0	0	0	0	0
> 0.995	4	0	7	0	4	0	6	0	10	0	7	0
χ^2 values:	23.11		24.08		26.84		85.65		28.89		462.42	

potheses (e.g., random distributions) are disproved before developing further hypotheses to explain observations. We now possess a more substantial indication that interesting ecological phenomena exist for multiple species in the Bluff Springs Sand Ponds than was possible by mapping alone. The analyses described above permit more specific questions to be addressed regarding causes of the 1996 patterns, for more species, than was possible in the absence of our analyses. For example, why was *Cryptocyclops bicolor* spatially clustered in 1996? What conditions existed in those ponds that permitted those populations to hatch and grow, while other populations did not? We have a more clear agenda for future studies (e.g., additional presence/absence surveys, pond con-

ditions analyses) and experiments to test hypotheses (e.g., transplant and culture experiments).

The choice of weights in the analyses described above is very important, and should be carefully considered (and explicitly described). Alternative weighting systems (e.g., precipitation-driven connectance among ponds, or coefficients based on hydroperiod) may be viewed as tests of underlying assumptions regarding causes of distributions, and could be used to indicate the most fruitful approaches to subsequent research.

Beyond the bounds of the Bluff Springs Sand Ponds, the permutation approach to the join-count statistics described here may prove useful as a tool in biotic inventories and surveys of rare, threatened, and endangered species. Often, these species inhabit rela-

tively few, isolated sites. If habitat requirements need to be assessed, or if habitats need to be managed, the most scientifically-defensible approach would be to first demonstrate that the species is occupying habitats non-randomly.

In summary, join-count statistics are not restricted to the common nearest-neighbor setting for large numbers of sites. Join-count statistics, and especially the permutation approach we have used, can be based on a non-binary, completely-connected weighting matrix (e.g., Euclidean distances) for few sample sites. The choice of weights represents assumptions about habitats, and so can be used as a statistical approach to testing the importance of those assumptions (e.g., intersite dispersal) to species' presence/absence patterns. Comparisons of those weighting approaches among multiple sites should be useful in statistical evaluations of dispersal and its role in regulating community structure (Ricklefs, 1987; Jenkins & Buikema, 1998), as well as local processes regulating community structure (e.g., Schneider & Frost, 1996).

Acknowledgements

References

Bailey TC and Gatrell AC (1995) Interactive Spatial Data Analysis. John Wiley & Sons, New York

Cliff AD and Ord JK (1971) Evaluating the percentage points of a spatial autocorrelation coefficient. Geograph Anal 3: 51–62

Cliff AD and Ord JK (1973) Spatial Processes: Models and Applications. Pion, London

Cliff AD and Ord JK (1981) Spatial Accesses: Models and Applications. Pion, London

Cressie NAC (1991) Statistics for Spatial Data. John Wiley & Sons, New York

Gilbert GS, Hubbell SP and Foster RB (1994) Density and distance-to-adult effects of a canker diseases of trees in a moist tropical forest. Oecologia 98:100–108

Jenkins DG and Buikema AL Jr. (1998) Do similar communities develop in similar sites? A test with zooplankton structure and function. Ecol Monogr 68: 421–443

Real LA and McElhany P (1996) Spatial pattern and process in plant-pathogen interactions. Ecology 77: 1011–1025

Ricklefs RE (1987) Community diversity: relative roles of local and regional processes. Science 235: 167–171

SAS Institute Inc. 1994. SAS/Base User's Guide, version 6 edition. SAS Institute Inc., Cary, NC

Schneider DW and Frost TM (1996) Habitat duration and community structure in temporary ponds. J North Am Benthol Soc 15: 64–86

Simovich M (1998) Crustacean biodiversity and endemism in California's ephemereal wetlands. In: Witham C, Bauder E, Belk D, Ferre WR and Ornduff (eds.), Ecology, Conservation, and Management of Vernal Pool Ecosystems. Proceedings from a 1996 Conference, California Native Plant Society, pp. 107–118

Sokal RR and Oden NL (1978a) Spatial autocorrelation in biology. 1. Methodology. Biol J Linnean Soc 10: 199–228

Sokal RR and Oden NL (1978b) Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. Biol J Linnean Soc 10: 229–249

Stevens PH (1998) The effects of completely-connected lattices on small-sample spatial autocorrelation hypothesis testing. Masters Thesis, University of Illinois at Springfield.

Willman HB and Frye JC (1970) Pleistocene stratigraphy of Illinois. Illinois State Geologicla Survey Bulletin 94. Urbana, IL, 204 pp