



# A probability-based analysis of temporal and spatial co-occurrence in grassland birds

Joseph A. Veech\*

School of Biological Sciences, University of Northern Colorado, Greeley, CO 80639-0017, USA

## ABSTRACT

**Aim** To test for non-random co-occurrence in 36 species of grassland birds using a new metric and the *C*-score. The analysis used presence–absence data of birds distributed among 305 sites (or landscapes) over a period of 35 years. This analysis departs from traditional analyses of species co-occurrence in its use of temporal data and of individual species' probabilities of occurrence to derive analytically the expected co-occurrence between paired species.

**Location** Great Plains region, USA.

**Methods** Presence–absence data for the bird species were obtained from the North American Breeding Bird Survey. The analysis was restricted to species pairs whose geographic ranges overlapped. Each of 541 species pairs was classified as a positive, negative, or non-significant association depending on the mean difference between the observed and expected frequencies of co-occurrence over the 35-year time-span.

**Results** Of the 541 species pairs that were examined, 202 to 293 (37–54%) were positively associated, depending on which of two null models was used. However, only a few species pairs (<5%) were negatively associated. An additional 89 species pairs did not have overlapping ranges and hence represented *de facto* negative associations. The results from analyses based on *C*-scores generally agreed with the analyses based on the difference between observed and expected co-occurrence, although the latter analyses were slightly more powerful.

**Main conclusions** Grassland birds within the Great Plains region are primarily distributed among landscapes either independently or in conjunction with one another. Only a few species pairs exhibited repulsed or segregated distributions. This indicates that the shared preference for grassland habitat may be more important in producing coexistence than are negative species interactions in preventing it. The large number of non-significant associations may represent random associations and thereby indicate that the presence/absence of other grassland bird species may have little effect on whether a given focal species is also found within the landscape. In a broader context, the probability-based approach used in this study may be useful in future studies of species co-occurrence.

## Keywords

Breeding Bird Survey, dispersal-assembly, Great Plains, habitat selection, landscape, North America, presence–absence matrix, species coexistence, species diversity, species interactions.

\*Correspondence: Joseph A. Veech, School of Biological Sciences, University of Northern Colorado, Greeley, CO 80639-0017, USA.  
E-mail: joseph.veech@unco.edu

## INTRODUCTION

To date, most studies of species co-occurrence have involved the analysis of static species presence–absence matrices. These

analyses are often contentious: researchers sometimes disagree about how to construct the randomized matrices that are required for testing the statistical significance of observed co-occurrence patterns (Gotelli, 2000). There are also a variety

of metrics for measuring species co-occurrence either as a comprehensive property of the entire matrix or as a mean property of species pairs (Gotelli & Graves, 1996; Gotelli, 2000). These metrics and different randomization routines vary in statistical power and rates of Type I and Type II errors (Gotelli, 2000). Ecologists also sometimes disagree as to whether matrices are most appropriately analysed by comparing species (R-mode analysis) or sites (Q-mode analysis) (Gotelli & Graves, 1996). This disagreement on method is miniscule, however, compared with the historical and on-going controversy concerning species co-occurrence patterns and what they tell us about the structure and organization of ecological communities (Diamond, 1975; Connor & Simberloff, 1979; Gilpin & Diamond, 1982; Gotelli & McCabe, 2002; Peres-Neto, 2004; Ulrich, 2004; Bell, 2005; Heino, 2005). Briefly, ecologists differ in their opinions as to whether significantly non-random (or nested) matrices are evidence for deterministic (non-neutral) or stochastic (neutral) processes of community assembly (Gotelli & McCabe, 2002; Ulrich, 2004; Bell, 2005), or whether non-random matrices say anything about process (Connor & Simberloff, 1979; Schluter, 1984).

At the very least, the analysis of species presence–absence data can inform us whether the degree of species co-occurrence in nature is more or less than that expected by chance association among species. That is, are most species randomly associated with one another or are the instances of association either more or less than expected? Moreover, the analysis of temporal presence–absence data might provide additional insight given that, in nature, ecological communities are dynamic. In addition, a probability-based, analytical method of deriving the expected number of instances of co-occurrence between species avoids the controversy over how to randomize matrices. The observed number of instances is then compared with the expected number. In this study, a new metric and method of testing for significant species co-occurrence are presented. The method is applied to an assemblage of 36 grassland-nesting birds distributed among 305 survey sites to determine the number of species pairs representing positive and negative associations. The results of this method are compared with an analysis based on a more commonly used metric of co-occurrence.

## METHODS

### Grassland birds within the Great Plains region

Data on the distribution of grassland birds within the Great Plains region of the United States were obtained from the North American Breeding Bird Survey (BBS). The BBS is an annual survey, conducted in May and June, of birds seen and heard along 39.2-km survey routes. Observers drive the routes (most routes are along secondary rural roads) and stop at 0.8-km intervals to record the presence of bird species during a 3-min time period. The survey has been conducted since 1966 and presently includes more than 3000 routes. Further details of the sampling protocol can be found in Robbins *et al.* (1986),

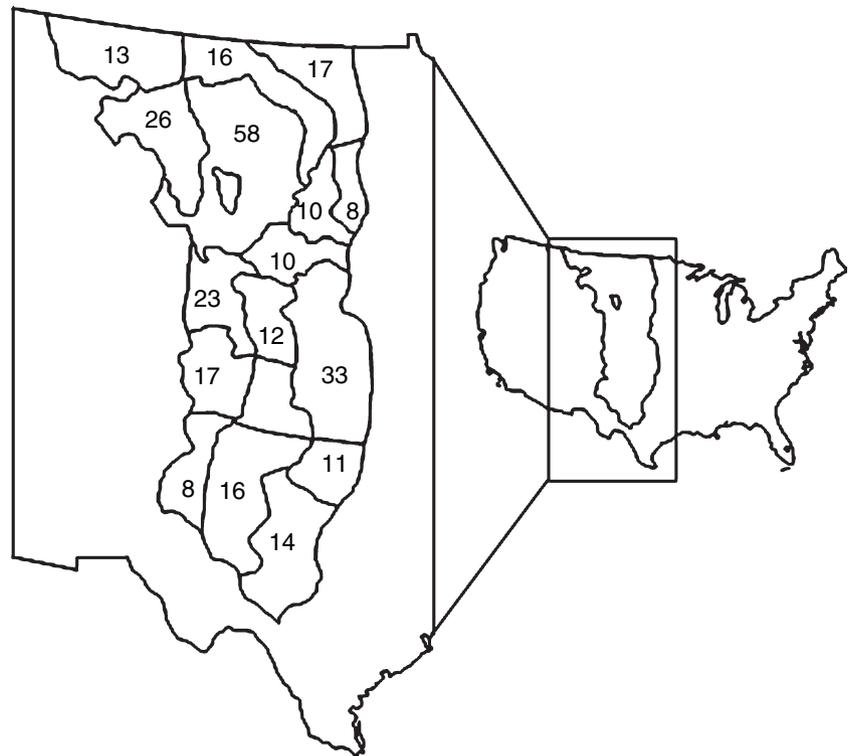
Droege & Sauer (1990) and Sauer *et al.* (2003). This study was limited to grassland birds, as classified by the BBS (Peterjohn & Sauer, 1999; Sauer *et al.*, 2003), because these species could potentially be competing for nest sites and perhaps other resources within their grassland habitat. Of the 118 species that nest in grassland and shrub habitat and occur within the Great Plains region, 47 species had a mean incidence (number of routes on which the species was recorded) per year of > 6.0, which is about 2% of 305. Less common species did not occur frequently enough to enable confident estimation of co-occurrence with other species. Of the 47, only those 36 species having a mean body mass between 10 and 100 g [body mass data obtained from Sibley (2000)] were included (Appendix 1). As is typical in studies of species co-occurrence, all the species belonged to the same broad 'guild' delineated in this study as small- to medium-sized grassland-nesting birds.

Three hundred and five BBS routes were identified within the 'Great Plains', defined as the 17 ecoregional sections in the Great Plains–Palouse Dry Steppe Province, Great Plains Steppe Province, Pecos Valley, Texas High Plains, and Rolling Plains of the Southwest Plateau and Plains Dry Steppe and Shrub Province of the Bailey ecoregional classification system (Bailey, 1995, 1996) (Fig. 1). The Great Plains region is a large relatively intact biome that has a continuous and consistent habitat type (grassland) without major geographical barriers to species dispersal and interaction.

For each route, data were used from surveys conducted between 1967 and 2001, giving a maximum of 35 years of data for any single route; not all routes had all 35 years of data available, although most did. Each of the 36 bird species had a mean incidence per year of > 6.0. The incidence of most species was substantially greater than 6.0 (Appendix 1).

### Analysis of co-occurrence patterns

Species co-occurrence can be analysed using an entire presence–absence matrix or using pairs of species each analysed separately. This latter approach is essentially the analysis of a presence–absence matrix consisting of only two species and is the approach used in this study. The expected frequency of co-occurrence or association between two species is simply their joint probability of occurrence or the product of their individual probabilities. The probability of occurrence of a species is the proportion of routes (sites) that are occupied by the species. Let  $P(1)$  = the probability of occurrence of Species 1, and  $P(2)$  = the probability of occurrence for Species 2. The probability of co-occurrence,  $P(1,2)$ , is then obtained as  $P(1)P(2)$ . Because  $P(\text{only } 1) = P(1) - P(1,2)$ , and  $P(\text{only } 2) = P(2) - P(1,2)$ , while  $P(\text{not } 1) = 1 - P(1)$  and  $P(\text{not } 2) = 1 - P(2)$ , all possible outcomes in a set of  $X$  routes are represented by the equation  $P(\text{only } 1) + P(\text{only } 2) + P(1,2) + [P(\text{not } 1)P(\text{not } 2)] = 1$ . Therefore,  $P(1,2)$  is the expected frequency of co-occurrence if each species in a pair is distributed independently of the other species. That is, the distribution of Species 1 is random with respect to Species 2, and vice versa. Determining expected co-occurrence in this way is equivalent



**Figure 1** Map of the Great Plains region within the United States. The small circular area within the Great Plains is the Black Hills region (not included in the present study). The 17 Bailey ecoregional sections divide the Great Plains region relatively equally. The number within each section is the number of BBS routes within that section. Sections range in area from 30,000 to 197,000 km<sup>2</sup>, with a mean of 80,000 km<sup>2</sup>. The density of BBS routes within the sections ranges from 1.23 to 3.81 per 100 km<sup>2</sup>, with a mean of 2.24 per 100 km<sup>2</sup>.

to the matrix randomization routine that maintains fixed row and column totals; that is, the SIM9 procedure of Gotelli (2000) and the procedure advocated by Connor & Simberloff (1979) and Sanderson *et al.* (1998). Gotelli (2000) examined 10 different randomization routines and found that SIM9 had the lowest rates of Type I and II error, particularly when used with the *C*-score (see below).

The probability of occurrence of a species is calculated as its incidence (number of routes on which it occurs) divided by the number of routes on which it could occur ( $N_1$  for Species 1 and  $N_2$  for Species 2). The values  $N_1$  and  $N_2$  affect  $P(1)$  and  $P(2)$ , and hence directly determine the expected frequency of co-occurrence. Therefore, two different null models were used to obtain  $N_1$  and  $N_2$ . In the first model, the Bailey ecoregional sections (from among the 17 composing the Great Plains) were identified for which the geographic range of a species overlapped at least 50% of the ecoregional section (Appendix 1). The total number of routes within this set of ecoregions was then taken as the potential number of routes on which the species could occur ( $N_1$  and  $N_2$  for Species 1 and 2). Bailey ecoregions are defined by climate, vegetation, topography, and geology such that routes within an ecoregion are relatively similar in these characteristics; thus, unoccupied routes can be assumed to be potentially suitable for a species even if it has not been recorded. Then,  $N_1$  and  $N_2$  were used to calculate  $P(1)$  and  $P(2)$  respectively, which, multiplied together, give  $P(1,2)$ , the expected frequency of co-occurrence. In the second null model, the total number of routes in the entire Great Plains region was used to define the number of routes on which

each species could potentially occur ( $N_1 = N_2 = 305$ ), and hence values of  $P(1)$ ,  $P(2)$ , and  $P(1,2)$  were obtained that were different from those of null model 1.

For each year (1967–2001), the observed frequency of co-occurrence of the species in a pair was determined as the number of routes on which both species were recorded ( $Q_{\text{obs}}$ ) divided by the number of routes on which they could have co-occurred,  $N_{1,2}$ . Null model 2 assumes that  $N_{1,2} = 305$  for all species pairs. However, for null model 1,  $N_{1,2}$  was defined as the number of routes in the set of Bailey ecoregions shared by both species, again using the 50% overlap criterion. Note that, for null model 1,  $N_{1,2}$  must be  $\leq$  both  $N_1$  and  $N_2$ .

Because observed and expected frequencies of co-occurrence were obtained for multiple years (typically 35), it was possible to calculate the standard deviation and 95% confidence intervals for the mean difference between the observed and expected co-occurrence ( $O - E$ ) for each species pair. For most species pairs,  $O - E$  values were normally distributed (64% of pairs had skewness values from  $-0.5$  to  $0.5$ , and 78% had kurtosis values from  $-1$  to  $1$ ), which thus permitted use of the parametric SD in subsequent tests of significance (see below). However, a bootstrap procedure was also used to obtain the non-parametric SD: the  $O - E$  values for a species pair were re-sampled with replacement 1000 times, each time the SD was calculated. The non-parametric SD was determined as the mean of the 1000 SD values (Manly, 1997). Thus, for each species pair and null model, a mean  $O - E$  value, parametric SD, and the non-parametric SD were available.

The test for a significantly positive or negative mean  $O - E$  value was conducted by comparing the value with a normal distribution with a mean of zero and standard deviation equal to the actual standard deviation estimated either parametrically or by bootstrapping. In this test, the  $P$  value is the proportion of the distribution greater than a positive  $O - E$  value or less than a negative  $O - E$  value. The association between the species in a pair was then categorized as negative, positive, or non-significant based on the sign of the mean  $O - E$  value.

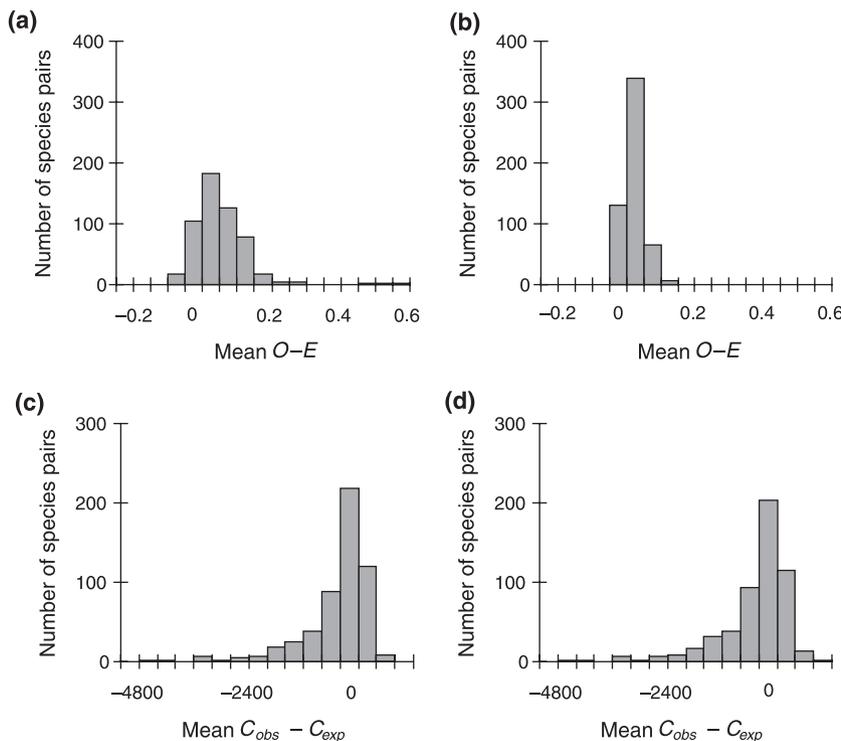
Species co-occurrence was also analysed using the  $C$ -score (Stone & Roberts, 1990). Gotelli (2000) found the  $C$ -score to have good statistical properties, including power. For a pair of species,  $i$  and  $j$ , the  $C$ -score is calculated as  $(N_i - Q_{\text{obs}})(N_j - Q_{\text{obs}})$ , where  $N_i$  and  $N_j$  represent the observed number of sites (BBS routes) occupied by species  $i$  and  $j$  respectively (Gotelli & McCabe, 2002). The two null models each provided an expected co-occurrence,  $Q_{\text{exp}}$  [that is equal to  $P(1,2) N_{1,2}$ ], and hence an expected  $C$ -score. For each species pair in each year, the observed  $C$ -score and the two expected  $C$ -scores were calculated, and  $C_{\text{obs}} - C_{\text{exp}}$  obtained for each of the two null models. I obtained the mean and bootstrap SD for each species pair (over all 35 years) and then categorized the species association as done for the mean  $O - E$  values. The  $C$ -score measures disassociation between two species, so a mean  $C_{\text{obs}} - C_{\text{exp}}$  value that is significantly positive actually represents a negative association (i.e., the observed  $C$ -score is greater than the expected), and a value that is significantly negative represents a positive association. Results from the analysis based on  $C$ -scores were compared with results from using  $O - E$  as a metric for measuring co-occurrence.

From a group of 36 species, there are 630 pairs of species possible. However, 89 pairs were excluded from the analysis because  $N_{1,2}$  for these pairs was zero; that is, the geographic ranges of the two species did not overlap. These species pairs represent negative associations that might represent the legacy of past interactions over evolutionary time. The focus of this study is the analysis of species co-occurrence over much shorter ecological time-scales in relatively small landscapes.

In this study, the landscape surrounding a BBS route is the spatial scale that corresponds to 'sites' of a species presence-absence matrix. Data were not available to assess the co-occurrence of bird species at a spatial level less than that of the landscape surrounding a survey route. Because of this, negative associations may have been slightly underestimated and positive associations overestimated. That is, two species occurring on opposite ends of a BBS survey route would be considered to co-occur within the landscape although not to co-occur at a smaller scale. Thus, observed co-occurrence is inflated in that case. However, with highly vagile organisms such as birds, assessment of co-occurrence at small spatial scales (e.g. ha instead of  $\text{km}^2$ ) can lead to false absences, which leads to the opposite problem of overestimated negative associations. To be cautious, the results of this study should be considered within the context of the spatial scale that was used.

**RESULTS**

For null model 1, the mean  $O - E$  values for the 541 species pairs ranged from  $-0.16$  to  $0.58$  (Fig. 2a). For null model 2, the values ranged from  $-0.02$  to  $0.16$  (Fig. 2b). Of the 541 species



**Figure 2** The mean values of pairwise co-occurrence as measured by (a)  $O - E$  for null model 1, (b)  $O - E$  for null model 2, (c)  $C_{\text{obs}} - C_{\text{exp}}$  for null model 1, and (d)  $C_{\text{obs}} - C_{\text{exp}}$  for null model 2. Negative values of  $C_{\text{obs}} - C_{\text{exp}}$  represent positive associations, and positive values represent negative associations. A total of 541 species pairs were analysed. See text for abbreviations.

pairs that were tested, 202–293 pairs (37–54%) represented species that were positively associated, depending on which null model, standard deviation, and significance level were used (Table 1). Negative associations were never represented by more than 6% of the total number of pairs. A large percentage of species pairs could not be classified as either positive or negative associations, in that the mean  $O - E$  value was non-significant (Table 1).

Overall, the two null models did not differ by much. However, the numbers of non-significant associations were slightly lower for null model 2 (Table 1). In addition, null model 2 revealed almost twice as many negative associations (although these were uncommon) as model 1, regardless of the type of SD used and the significance level. Within a given null model, use of the bootstrap SD resulted in slightly greater numbers of significant associations (positive and negative) than were found by using the parametric SD (Table 1). As expected, significance level had the strongest effect on the classification of species pairs. The higher significance level of 0.05 provided a test that had more power for detecting positive and negative associations than did the 0.01 significance level (Table 1).

The mean  $C$ -scores ( $C_{\text{obs}} - C_{\text{exp}}$ ) also revealed many positive associations and very few negative associations, with most species associations being non-significant (Table 2). Mean  $C_{\text{obs}} - C_{\text{exp}}$  values ranged from  $-4691$  to  $720$  for null model 1 (Fig. 2c) and from  $-4691$  to  $892$  for null model 2 (Fig. 2d). Overall, the co-occurrence analysis based on  $C_{\text{obs}} - C_{\text{exp}}$  agreed with the analysis based on  $O - E$ , although the latter was

slightly more powerful in finding greater numbers of negative and positive associations.

## DISCUSSION

At a landscape scale, species of grassland birds appear to be distributed in conjunction with one another much more often than they are segregated from one another. Depending on the null model and significance level that were applied, more than 50% of the species pairs represented positive associations whereas less than 5% represented negative associations. The tendency for species to co-occur more often than expected could arise from a preference for the same habitat or from local adaptation to similar environmental conditions (Gotelli *et al.*, 1997; Peres-Neto *et al.*, 2001; Peres-Neto, 2004; Bell, 2005). The 36 bird species included in this study all prefer open habitat consisting of substantial grass cover and low-lying shrubs that can be used for nesting. The large number of positive associations may also indicate that the landscapes surrounding the survey routes are not saturated. That is, competitively inferior species may be coexisting with stronger competitors because nesting sites and food supplies are not exhausted by the stronger competitors. The competitive abilities of the grassland bird species were not analysed in this study, however, so this is only speculation. The low number of negative associations also supports the contention that inter-specific competition (if it is occurring) is generally not preventing species coexistence in the landscapes.

**Table 1** Tests for pairwise species associations using the mean observed – expected frequency of co-occurrence ( $O - E$ )

Type of association	Model 1 ( $\alpha = 0.05$ ) (%)	Model 2 ( $\alpha = 0.05$ ) (%)	Model 1 ( $\alpha = 0.01$ ) (%)	Model 2 ( $\alpha = 0.01$ ) (%)
Standard deviations estimated parametrically				
Positive	276 (51.0)	278 (51.4)	202 (37.3)	202 (37.3)
Negative	18 (3.3)	29 (5.3)	1 (0.2)	3 (0.6)
Non-significant	247 (45.7)	234 (43.3)	338 (62.5)	336 (62.1)
Standard deviations estimated by bootstrapping				
Positive	288 (53.2)	293 (54.2)	210 (38.8)	212 (39.2)
Negative	18 (3.3)	32 (5.9)	3 (0.6)	6 (1.1)
Non-significant	235 (43.4)	216 (39.9)	328 (60.6)	323 (59.7)

The numbers of species pairs representing positive, negative, and non-significant associations are shown for both null models evaluated using alpha levels of 0.05 and 0.01. Species pairs were assigned to the three types of association based on the outcome of the tests using either a parametric standard deviation or a bootstrap standard deviation.

**Table 2** Tests for pairwise species associations using the mean observed – expected  $C$ -scores ( $C_{\text{obs}} - C_{\text{exp}}$ )

Type of association	Model 1 ( $\alpha = 0.05$ ) (%)	Model 2 ( $\alpha = 0.05$ ) (%)	Model 1 ( $\alpha = 0.01$ ) (%)	Model 2 ( $\alpha = 0.01$ ) (%)
Positive	230 (42.5)	225 (41.6)	138 (25.5)	132 (24.4)
Negative	1 (0.2)	6 (1.1)	0	0
Non-significant	310 (57.3)	310 (57.3)	403 (74.5)	409 (75.6)

The numbers of species pairs representing positive, negative, and non-significant associations are shown for both null models evaluated using alpha levels of 0.05 and 0.01. The tests used standard deviations that were estimated by bootstrapping.

Given the paucity of negative associations, the same analyses of co-occurrence were conducted on a subset of the species (sparrows; family Emberizidae) that might be more likely to exhibit negative associations as a result of competition based on similar body sizes and diet. Among the 36 grassland bird species, there were 11 species of sparrows. Their body masses ranged from 10.5 to 38 g with a mean of 19.5 g (Appendix 1). Sparrows are primarily insectivorous during summer and granivorous during winter. Thus, they may form a more exclusive trophic assemblage than do all 36 grassland birds, and hence may be more likely to be associated negatively. Their similar body masses and diets predict negative associations. Of the 53 pairs of sparrow species that were analysed with the  $O - E$  metric, most represented either non-significant (40–52%) or positive (48–54%) associations, depending on the significance level (0.05 or 0.01) that was applied. Thus, the frequency of non-significant associations is about the same among sparrows as it is among all 36 grassland birds. Furthermore, non-significant associations are much more common than negative associations. At most, only 6% of the sparrow pairs represented negative associations. This indicates that interspecific competition for food and nesting resources may not have much influence on the coexistence of sparrows at a landscape scale. A similar lack of predicted negative associations was found in a recent study. Sfenthourakis *et al.* (2006) analysed presence–absence matrices for a wide variety of taxa and found that the frequency of negative associations among congeners was no greater than that among non-congeners.

Schluter (1984) cautioned against the inference of particular processes from patterns of species co-occurrence. Similarly, there are limits to inferring the *absence* of particular processes from co-occurrence patterns. However, at the very least, analysis of co-occurrence patterns should be able to tell us whether species co-occur at a greater or lower rate than expected as a result of chance association (Gotelli & McCabe, 2002). If two species co-occur at a greater rate, then any number of possible factors (e.g. shared habitat preference, avoidance of a shared predator, shared and relatively unlimited food resource) may be involved. Conversely, if two species co-occur at a lower rate than expected by chance, very different factors (e.g. different habitat preferences, competition for resources) must be involved. Finally, if two species co-occur at a rate about equal to that expected based on their individual rates of incidence (i.e.  $O - E = 0$ ) then the entire suite of processes (habitat selection, predator avoidance, interspecific competition) may be occurring without any one process having an overriding influence on coexistence.

Regardless of the co-occurrence metric, null model, or significance level, a large number of the 541 species pairs were not significantly associated in either a positive or negative way (Tables 1 & 2). Non-significant associations between species may have occurred in either of two ways. First, the mean  $O - E$  value (over all 35 years) for a species pair could be near zero with a relatively small standard deviation (or coefficient of variation). In such cases, the species co-occur each year at a

frequency that is similar to the expected frequency if each species is distributed independently of the other species. Second, the mean  $O - E$  value for a species pair might have a relatively large standard deviation (or CV), which would indicate substantial year-to-year variation in observed co-occurrence relative to expected co-occurrence. That is, some years the species pair might have relatively large positive  $O - E$  values and other years the pair might have relatively large negative  $O - E$  values.

A closer examination of the results revealed that most non-significant associations were of the latter type. Of the 338 species pairs with non-significant mean  $O - E$  values (null model 1,  $\alpha = 0.01$ ), 302 pairs had a CV > 50%, which thus indicated a relatively large SD compared with the mean. Similarly, with null model 2 and  $\alpha = 0.01$ , 293 of 336 species pairs had a CV > 50%. Therefore, the non-significant associations mostly represent species that were inconsistently associated over the 35 years from 1967 to 2001. Some years the two species of a pair might have co-occurred more than expected, whereas other years the two species might have co-occurred less than expected. Only a few species pairs tended to co-occur each year at the expected frequency. However, the temporal inconsistency in co-occurrence is evidence that, over long time-spans (e.g. three decades or more), grassland bird species are often distributed independently of one another.

The independent co-occurrence of a species pair is equivalent to saying that the two species have spatial distributions that are random relative to each other. These temporally stochastic associations are consistent with the dispersal-assembly perspective of how species are distributed among communities and landscapes (Hubbell, 2001). The overall lack of negative associations is at odds with the niche-assembly perspective of species coexistence and distribution. The dispersal-assembly perspective asserts that most ecological communities are open and in a non-equilibrium state in which species come and go. Dispersal and stochastic local extinction govern species coexistence more than do species interactions (Hubbell, 2001). The niche-assembly perspective asserts that species interactions and niche partitioning determine species coexistence in an interactive equilibrium in which community membership is relatively steady. Some species pairs (or combinations) are able to coexist and others are not, such that the latter exhibit segregation or negative associations (i.e. co-occurrence that is lower than expected). However, because there is a limit to inferring process (or lack of) from particular patterns of co-occurrence, a thorough comparison of dispersal assembly and niche assembly must come from other types of analyses.

In this study, the use of temporal data emphasizes that species co-occurrence is a dynamic pattern. Even when a species pair can be categorized as a positive or negative association, there is annual variation in the extent to which the species co-occur. Some of this variation might represent sampling error, but much of it probably reflects the dynamic nature of food resources, predator populations, landscape disturbance, weather, and other factors that may regulate bird

populations. For instance, if food resources are extra-abundant in a particular year then two competing species may be able to coexist in some landscapes. The analysis of species presence–absence data may often miss important population dynamics that more accurately reflect the actual ecological processes involved in species coexistence. For example, two species may exhibit greater than expected co-occurrence and yet have opposite trends in abundance over time. Likewise, some negative and non-significant associations might actually represent species with similar population trajectories in the landscapes or communities in which they co-occur.

The  $O - E$  metric was more proficient (more powerful) than the  $C_{\text{obs}} - C_{\text{exp}}$  metric at revealing positive and negative associations among the bird species. The corollary of this is that the  $O - E$  metric might be more prone to Type I error (finding a significant positive or negative association that does not truly exist) than is the  $C_{\text{obs}} - C_{\text{exp}}$  metric. However, a comparison of  $O - E$  with  $C_{\text{obs}} - C_{\text{exp}}$  was not the main purpose of the study, and therefore it is not appropriate to conclude that  $O - E$  is better than  $C_{\text{obs}} - C_{\text{exp}}$  at finding non-random co-occurrence patterns.

In this study, species co-occurrence was analysed in a way that is substantially different from the traditional approach of calculating a co-occurrence metric for an entire presence–absence matrix. There are three main differences. (1) The  $O - E$  metric was calculated for species pairs, and subsequent significance testing was conducted on species pairs, not on entire matrices. Many other co-occurrence metrics are also pairwise, such as the  $C$ -score (Stone & Roberts, 1990), the statistic used by Pleasants (1990), the binary covariance (Bell, 2005), and the number of co-occurrences (Sfenthourakis *et al.*, 2006). Typically, researchers have averaged the pairwise metric over all species in the matrix. In the present study, the goal was to determine the percentage of species pairs that were either positively or negatively associated – hence the focus on pairwise co-occurrence instead of co-occurrence as a community property. Sfenthourakis *et al.* (2006) used a species-by-species approach in their recent study, although it was not based on analytically derived co-occurrence probabilities as in the present study. (2) Temporal presence–absence data were used in the present study. The temporal data allowed for estimation of confidence intervals around the mean observed co-occurrence of species within a pair (over all 35 years), and thus direct significance testing of the difference between observed and expected co-occurrence. The use of temporal data provided a test for non-random co-occurrence that is potentially more powerful than tests based on static data. Using a similar co-occurrence metric applied to static data, Sfenthourakis *et al.* (2006) typically found significant associations (positive and negative) in only about 3–6% of the species pairs that they analysed. (3) The analysis of species co-occurrence in the present study avoided the potential complications of creating randomized presence–absence matrices. Expected co-occurrence was obtained analytically from probabilities of occurrence of each species in a pair. Because expected co-occurrence is derived from the same

matrix as is observed co-occurrence, this probability-based approach conserves species incidence and richness of sites (i.e. row and column totals of the presence–absence matrix), an important statistical feature of co-occurrence analyses (Gotelli, 2000). The direct use of probabilities to measure and test for non-random species co-occurrence fits within the larger realm of interpreting biological processes and patterns in terms of probability (Denny & Gaines, 2000).

The probability of occurrence for a species depends on the number of sites in which the species could potentially exist. This is the main difference between the two null models used in this study. Even though the probability of occurrence of a species differed between the two models for all species (Appendix 1), there were only slight differences in the results obtained using the two null models (Table 1). Recall that in null model 1,  $N$  (number of routes occupied) is determined by the ecoregions within the species geographic range, but in null model 2,  $N = 305$  (entire Great Plains region) for all species. For the 36 species, the probability of occurrence ranged from 0.051 to 0.506 (mean 0.231) in null model 1, and from 0.021 to 0.506 (mean 0.131) in null model 2. Using more than one null model can often provide additional insight even when the results from the models do not completely agree (Gotelli, 2000). In the present study, the results did agree. This indicates that the results are fairly robust to the sets of BBS routes used to calculate the probability of occurrence for each species.

The probability-based metric of species co-occurrence used in this study was the observed minus expected frequency of co-occurrence ( $O - E$ ). Unlike some other co-occurrence metrics (e.g.  $C$ -score, number of checkerboards), it has a straightforward interpretation. It is the proportion (or percentage) of sites co-occupied by two species either above or below that expected as a result of chance. For example, an  $O - E$  value of 0.1 means that the two species occur together in 10% more sites than expected. This is informative apart from any test of significance, because it directly indicates the extent to which two species depart from having a random relationship (Veech, 2005). When that departure is large (either negative or positive) then ecological and evolutionary reasons can be sought to explain it. In this way, the  $O - E$  metric may be useful in future studies of species co-occurrence.

## ACKNOWLEDGEMENTS

I thank J. Chave, N. Gotelli, and several anonymous referees for their comments on an earlier version of this manuscript. In addition, I thank the many past and present volunteers of the North American Breeding Bird Survey for their continuing efforts in monitoring bird diversity and abundance.

## REFERENCES

- Bailey, R.G. (1995) *Description of the ecoregions of the United States*. Miscellaneous Publication No. 1391, United States Department of Agriculture Forest Service, Washington, DC.

- Bailey, R.G. (1996) *Ecosystem geography*. Springer-Verlag, New York, NY.
- Bell, G. (2005) The co-distribution of species in relation to the neutral theory of community ecology. *Ecology*, **86**, 1757–1770.
- Connor, E.F. & Simberloff, D. (1979) The assembly of species communities: chance or competition? *Ecology*, **60**, 1132–1140.
- Denny, M. & Gaines, S. (2000) *Chance in biology, using probability to explore nature*. Princeton University Press, Princeton, NJ.
- Diamond, J.M. (1975) Assembly of species communities. *Ecology and evolution of communities* (ed. by M.L. Cody and J.M. Diamond), pp. 342–444. Harvard University Press, Cambridge, MA, USA.
- Droege, S. & Sauer, J.R. (1990) *North American Breeding Bird Survey annual summary 1989*. Biological Report 90, US Fish and Wildlife Service, Washington, DC.
- Gilpin, M.E. & Diamond, J.M. (1982) Factors contributing to non-randomness in species co-occurrences on islands. *Oecologia*, **52**, 75–84.
- Gotelli, N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606–2621.
- Gotelli, N.J. & Graves, G.R. (1996) *Null models in ecology*. Smithsonian Institution Press, Washington, DC.
- Gotelli, N.J. & McCabe, D.J. (2002) Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rule model. *Ecology*, **83**, 2091–2096.
- Gotelli, N.J., Buckley, N.J. & Wiens, J.A. (1997) Co-occurrence of Australian land birds: Diamond's assembly rules revisited. *Oikos*, **80**, 311–324.
- Heino, J. (2005) Metacommunity patterns of highly diverse stream midges: gradients, checkerboards, and nestedness, or is there only randomness? *Ecological Entomology*, **30**, 590–599.
- Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ.
- Manly, B.F.J. (1997) *Randomization, bootstrap and Monte Carlo methods in biology*. Chapman and Hall, New York, NY.
- Peres-Neto, P.R. (2004) Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. *Oecologia*, **140**, 352–360.
- Peres-Neto, P.R., Olden, J.D. & Jackson, D.A. (2001) Environmentally constrained null models: site suitability as occupancy criterion. *Oikos*, **93**, 110–120.
- Peterjohn, B.G. & Sauer, J.R. (1999) Population status of North American grassland birds from the North American Breeding Bird Survey, 1966–1996. *Ecology and conservation of grassland birds of the western hemisphere* (ed. by P.D. Vickery and J.R. Herkert), pp. 27–44. Cooper Ornithological Society, Allen Press, Lawrence, KA.
- Pleasants, J.M. (1990) Null model tests for competitive displacement: the fallacy of not focusing on the whole community. *Ecology*, **71**, 1078–1084.
- Robbins, C.S., Bystrak, D. & Geissler, P.H. (1986) *The Breeding Bird Survey: its first fifteen years, 1965–1979*. Resource Publication 157, US Department of the Interior, Fish and Wildlife Service, Washington, DC.
- Sanderson, J.G., Moulton, M.P. & Selfridge, R.G. (1998) Null matrices and the analysis of species co-occurrences. *Oecologia*, **116**, 275–283.
- Sauer, J.R., Hines, J.E. & Fallon, J. (2003) *The North American Breeding Bird Survey, results and analysis 1966–2002*. Version 2003.1, USGS Patuxent Wildlife Research Center, Laurel, MD.
- Schluter, D. (1984) A variance test for detecting species associations, with some example applications. *Ecology*, **65**, 998–1005.
- Sfenthourakis, S., Tzanatos, E. & Giokas, S. (2006) Species co-occurrence: the case of congeneric species and a causal approach to patterns of species association. *Global Ecology and Biogeography*, **15**, 39–49.
- Sibley, D.A. (2000) *The Sibley guide to birds*. Knopf Publishing, New York, NY.
- Stone, L. & Roberts, A. (1990) The checkerboard score and species distributions. *Oecologia*, **85**, 74–79.
- Ulrich, W. (2004) Species co-occurrences and neutral models: reassessing J. M. Diamond's assembly rules. *Oikos*, **107**, 603–609.
- Veech, J.A. (2005) Analyzing patterns of species diversity as departures from random expectations. *Oikos*, **108**, 149–155.

## BIOSKETCH

**Joseph A. Veech** is primarily interested in the ecological processes that influence species diversity within communities, landscapes, and ecoregions. He uses a variety of research approaches, including spatial modelling, null models and randomization tests, meta-analysis, and experimentation.

---

Editor: Walter Jetz

## APPENDIX

**Appendix 1** List of the 36 grassland bird species used in the study. Body mass, mean incidence (number of routes occupied per year over 35 years), number of Bailey ecoregions in which the species geographic range overlapped >50% of the ecoregion, and the probabilities of occurrence for each species are provided

Common name	Scientific name	Body mass (g)	Mean incidence	Ecoregions	<i>P</i> (occurrence), null model 1	<i>P</i> (occurrence), null model 2
Cassin's sparrow	<i>Aimophila cassinii</i>	19	39.7	9	0.351	0.131
Baird's sparrow	<i>Ammodramus bairdii</i>	17.5	17.4	4	0.242	0.057
Grasshopper sparrow	<i>Ammodramus savannarum</i>	17	115.4	17	0.380	0.380
Sprague's pipit	<i>Anthus spragueii</i>	25	8.2	4	0.114	0.027
Cedar waxwing	<i>Bombycilla cedrorum</i>	32	9.8	9	0.051	0.032
Lark bunting	<i>Calamospiza melanocorys</i>	38	96.3	13	0.379	0.317
McCown's longspur	<i>Calcarius mccownii</i>	23	7.9	3	0.266	0.138
Chestnut-collared longspur	<i>Calcarius ornatus</i>	19	41.5	7	0.286	0.137
Northern cardinal	<i>Cardinalis cardinalis</i>	45	37.5	7	0.407	0.123
House finch	<i>Carpodacus mexicanus</i>	21	20.5	15	0.081	0.068
Yellow-billed cuckoo	<i>Coccyzus americanus</i>	65	41.5	11	0.310	0.137
Black-billed cuckoo	<i>Coccyzus erythrophthalmus</i>	52	21.0	7	0.138	0.069
Yellow warbler	<i>Dendroica petechia</i>	9.5	59.3	13	0.234	0.195
Bobolink	<i>Dolichonyx oryzivorus</i>	43	42.0	8	0.266	0.138
Gray catbird	<i>Dumetella carolinensis</i>	37	30.5	12	0.128	0.100
Common yellowthroat	<i>Geothlypis trichas</i>	10	65.6	14	0.247	0.216
Yellow-breasted chat	<i>Icteria virens</i>	25	14.1	9	0.081	0.046
Orchard oriole	<i>Icterus spurius</i>	19	64.0	16	0.226	0.211
Sage thrasher	<i>Oreoscoptes montanus</i>	43	6.5	1	0.249	0.021
Savannah sparrow	<i>Passerculus sandwichensis</i>	20	35.5	9	0.189	0.117
Lazuli bunting	<i>Passerina amoena</i>	16	10.3	6	0.067	0.034
Blue grosbeak	<i>Passerina caerulea</i>	28	43.9	14	0.189	0.145
Painted bunting	<i>Passerina ciris</i>	16	18.8	4	0.459	0.062
Indigo bunting	<i>Passerina cyanea</i>	15	15.8	8	0.154	0.052
Vesper sparrow	<i>Poocetes gramineus</i>	26	59.6	13	0.259	0.196
Dickcissel	<i>Spiza americana</i>	27	68.1	15	0.265	0.224
Brewer's sparrow	<i>Spizella breweri</i>	10.5	21.1	5	0.234	0.069
Clay-coloured sparrow	<i>Spizella pallida</i>	12	29.1	6	0.211	0.096
Field sparrow	<i>Spizella pusilla</i>	12.5	32.3	12	0.154	0.106
Eastern meadowlark	<i>Sturnella magna</i>	90	33.7	9	0.291	0.111
Western meadowlark	<i>Sturnella neglecta</i>	97	153.8	17	0.506	0.506
Bewick's wren	<i>Thryomanes bewickii</i>	10	19.0	8	0.169	0.062
Curve-billed thrasher	<i>Toxostoma curvirostre</i>	79	6.3	4	0.167	0.021
Brown thrasher	<i>Toxostoma rufum</i>	69	87.9	14	0.331	0.289
Bell's vireo	<i>Vireo belli</i>	8.5	15.2	7	0.173	0.050
Red-eyed vireo	<i>Vireo olivaceus</i>	17	10.9	7	0.075	0.036