

# The pairwise approach to analysing species co-occurrence

Joseph A. Veech

Department of Biology, Texas State  
University, San Marcos, TX 78666-4684, USA

## ABSTRACT

The analysis of species co-occurrence patterns continues to be a main pursuit of ecologists, primarily because the coexistence of species is fundamentally important in evaluating various theories, principles and concepts. Examples include community assembly, equilibrium versus non-equilibrium organization of communities, resource partitioning and ecological character displacement, the local–regional species diversity relationship, and the metacommunity concept. Traditionally, co-occurrence has been measured and tested at the level of an entire species presence–absence matrix wherein various algorithms are used to randomize matrices and produce statistical null distributions of metrics that quantify structure in the matrix. This approach implicitly recognizes a presence–absence matrix as having some real ecological identity (e.g. a set of species exhibiting nestedness among a set of islands) in addition to being a unit of statistical analysis. An emerging alternative is to test for non-random co-occurrence between paired species. The pairwise approach does not analyse matrix-level structure and thus views a species pair as the fundamental unit of co-occurrence. Inferring process from pattern is very difficult in analyses of co-occurrence; however, the pairwise approach may make this task easier by simplifying the analysis and resulting inferences to associations between paired species.

## Keywords

Assembly rules, competition, C-score, data randomization, ecological community, hypothesis testing, Jared Diamond, pattern analysis, presence–absence matrix, species coexistence.

Correspondence: Joseph A. Veech, Department of Biology, Texas State University, San Marcos, TX 78666-4684, USA.  
E-mail: joseph.veech@txstate.edu

## INTRODUCTION

The analysis of species co-occurrence has a long history played out in studies ranging from community ecology to biogeography. Jared Diamond's classic study of the distributions of terrestrial birds on islands of New Guinea was one of the first analyses of species co-occurrence (Diamond, 1973, 1975). His study emphasized diffuse competition – 'the complex situations resulting from the sum of competitive effects from many other somewhat similar species' (Diamond, 1975, p. 348) – and introduced the idea of assembly rules and forbidden combinations (of species). By definition, testing for diffuse competition requires analysis of co-occurrence among multiple species within an assemblage, not between just two species of a pair. In subsequent years, Diamond's analytical methods were rigorously criticized (Connor & Simberloff, 1979) and confidently defended (Diamond & Gilpin, 1982; Gilpin & Diamond, 1984); see Sanderson

(2000) for a review. As a result of these historical beginnings, studies of species co-occurrence have come to be characterized by the methods utilized (and defended) as well as a primary focus on the entire assemblage (i.e. the presence–absence matrix) as the unit of analysis. The matrix-level approach is well illustrated by the study of nestedness, the way in which smaller sets of species form orderly subsets of increasingly larger sets (Wright & Reeves, 1992; Almeida-Neto *et al.*, 2008; Ulrich *et al.*, 2009).

An emerging alternative to the matrix approach is to examine *pairwise* co-occurrence (Sanderson, 2000; Sfenthourakis *et al.*, 2004, 2006; Veech, 2006; Sanderson *et al.*, 2009; Gotelli & Ulrich, 2010; Collins *et al.*, 2011; Pitta *et al.*, 2012) with methods that can detect patterns of positive, negative or random association between two species (Veech, 2013; see also Ricklefs, 2011, for a pairwise assessment of abundances). Beyond being a more comprehensive statistical assessment, these new methods are also suggesting (sometimes implicitly)

that species may not be organized into neat and tidy communities that exist as discrete spatial entities. At the very least, the local community concept (i.e. Clementsian organization of nature into very similar sets of species repeated at different locales) may be insufficient as a framework for understanding the geographical distribution of species and diversity patterns (Ricklefs, 2008). If so, then analyses of co-occurrence patterns are most appropriate when conducted on species pairs instead of on entire species presence–absence matrices. Of interest, Pielou (1977) described a pairwise approach several decades ago, although it never caught on. In this essay, I describe more recent pairwise approaches and distinguish them from matrix-level approaches and specifically emphasize benefits of the former. I also argue that the study of species co-occurrence now needs to move beyond the view that nature is organized into presence–absence matrices.

## PAIRWISE AND MATRIX-LEVEL APPROACHES

Most methods for analysing species co-occurrence can be classified as either pairwise or matrix-level approaches (Table 1). The main difference is that the latter calculate a co-occurrence metric as a property of the entire presence–absence matrix whereas the former examine co-occurrence

‘species-by-species’ (Sfenthourakis *et al.*, 2006; Pitta *et al.*, 2012) to determine whether a particular pair of species is aggregated, segregated or random in occurrence (Gotelli & Ulrich, 2010). A recent approach developed by Arita *et al.* (2012) is best described as a hybrid given that it requires calculations performed on an entire matrix but the metric (proportional species richness) is obtained for each species separately (Table 1). Examples of matrix-level metrics include matrix temperature (Patterson & Atmar, 1986; Atmar & Patterson, 1993), number of unique species combinations (in the matrix), and the variance ratio (Schluter, 1984). Other metrics (e.g. *C*-score, number of checkerboards) can be calculated for species pairs but are typically aggregated (summed or averaged) for entire matrices. The most straightforward way to measure co-occurrence between two species is by the *observed* number of times that the two species co-occur relative to the *expected* number of times (Sanderson, 2000; Sfenthourakis *et al.*, 2004, 2006; Veech, 2006, 2013; Pitta *et al.*, 2012). Some authors have referred to this as the ‘natural metric’ (Sfenthourakis *et al.*, 2004, 2006; Pitta *et al.*, 2012). The expected co-occurrence can be obtained through randomization of the matrix (randomizing species occurrences among sites) (Sanderson, 2000; Sfenthourakis *et al.*, 2006) or through basic probability theory as  $J_{\text{exp}} = (N_1/N) \times (N_2/N) \times N$  (Bowers & Brown, 1982;

**Table 1** Analytical methods previously used to study species co-occurrence (broadly defined). More detailed descriptions of each method can be found in the listed references (although these are not intended to represent a complete list).

Method of analysis	Type of analysis	Description and usage	Reference(s)
Classic null models	Matrix-level	Broad class of null models that simultaneously randomize the occurrences of all species among sites. The models are distinguished by different constraints on the randomization algorithms. Used to find statistically significant co-occurrence patterns for entire presence–absence matrices.	Gotelli (2000), specifically Table 2
Nestedness analysis	Matrix-level	Various algorithms (also sometimes used in classic null models) that randomize species occurrences among sites. Used to detect patterns of nestedness – less species-rich sites as orderly compositional subsets of sites with greater species richness.	Patterson & Atmar (1986), Wright & Reeves (1992), Ulrich <i>et al.</i> (2009)
Network analysis	Matrix-level	Annealing algorithm is applied to a presence–absence matrix (a geographical, taxonomic or guild-based matrix) so as to maximize modularity or the amount of connectedness among nodes (species). Used to find modules (groups of highly connected species) and compartments (spatial clustering of species’ range boundaries).	Carstensen & Olesen (2009), Thébaud (2013)
Range-diversity plots	Hybrid	Calculates and depicts proportional range richness of a species as the mean number of species occurring at the same sites as the focal species. Tests of significance can be conducted through various matrix randomization algorithms. Used to examine co-occurrence in the context of an entire assemblage and its geographical distribution (among other uses).	Arita <i>et al.</i> (2012)
Causality analysis	Pairwise	Compares pairwise co-occurrence from two or more types of presence–absence matrix. The geographical matrix is the traditional form where species are recorded from sampling sites; the ecological matrices consist of species assigned to units representing habitat types or other ecological factors. A fixed–fixed algorithm randomizes species occurrences among units while conserving unit species richness and species incidences. For each matrix, species pairs are classified as positive, negative or no association. Used to examine the ‘causal basis of co-occurrence’ and to identify the ecological factors that most influence co-occurrence.	Sfenthourakis <i>et al.</i> (2006)
Probability model	Pairwise	Applies pure probability-based equations to identify species pairs as having a positive, a negative, a random or an indecisive association. Used to analytically classify (without data randomization) species pairs to the above categories.	Veech (2013)

Veech, 2006, 2013; Araújo *et al.*, 2011), where  $J_{\text{exp}}$  represents the expected number of sites (or samples) that have species 1 and 2,  $N_1$  is the number of sites that have species 1,  $N_2$  is the number of sites that have species 2, and  $N$  is the total number of sites. Note that  $N_1/N$  and  $N_2/N$  are the marginal probabilities of occurrence of species 1 and 2, respectively, also known as their incidence rates.

When one wishes to test for statistical significance of an observed co-occurrence metric (pairwise or matrix-level) then the presence-absence matrix is often randomized to produce a null distribution of the metric as a basis for comparison (i.e. deriving a  $P$ -value). There are many algorithms for conducting the randomization and these vary in their Type I and II error rates (Gotelli, 2000). Perhaps, the most common and oft-recommended algorithm is the fixed-fixed (F-F) randomization; the column and row sums of the matrix remain fixed at their original values during the randomization (Gotelli, 2000). That is, the randomization does not alter the species incidence rates (row sums) or richness of the sampling sites (column sums). Incidence rates and richness are implicitly assumed to be real ecological properties of the species assemblage that should not be altered when testing for non-random patterns of co-occurrence in the matrix or between species in a pair. The opposite of the F-F algorithm is the equiprobable-equiprobable (E-E) randomization in which column and row sums are free to vary. All species are assumed to be equally probable in occurring at a given sampling site (differences in observed incidence rates are ignored during the randomization) and all sampling sites are assumed to be equally probable in having a given species (differences in observed richness among sites are ignored). The logic behind these different randomization algorithms has been thoroughly discussed elsewhere (Connor & Simberloff, 1979; Gotelli & Graves, 1996) and will not be repeated in this essay. However, my main point, relevant to this essay, is that no one randomization algorithm is more logical or ecologically realistic than any other; the algorithms primarily differ in their statistical properties (Gotelli, 2000; Ulrich & Gotelli, 2013). See Ulrich & Gotelli (2013) for a complete review and evaluation of 15 metrics used to measure species co-occurrence in conjunction with matrix randomization.

When testing matrix-level co-occurrence patterns, it is obviously necessary to randomize the entire presence-absence matrix in order to determine if the co-occurrence metric is significantly different from random. However, this need not be the case when using a pairwise approach. In its most basic form, the pairwise approach is solely concerned with testing pairs of species one-by-one as though there are  $S_p$  number of  $2 \times N$  matrices (species by sampling site) each examined separately, where  $S_p$  could represent all possible species pairs,  $S_p = N!/[2 \times (N-2)!]$ , or some pre-selected group of pairs. Any of the randomization algorithms and matrix-level metrics could be applied to these  $2 \times N$  matrices. However, a recently proposed pairwise method circumvents the need for randomization.

## PROBABILISTIC MODEL OF CO-OCCURRENCE

The probabilistic model of species co-occurrence (Veech, 2013) is based on calculating the number of ways that two species can co-occur at exactly  $j$  number of sampling sites given that each species occurs at  $N_1$  and  $N_2$  number of sites out of a total of  $N$ . This quantity is the product of the following combinations:  $a = C(N, j)$ ,  $b = C(N - j, N_2 - j)$ ,  $c = C(N - N_2, N_1 - j)$ , where  $a$  represents the number of unique arrangements of species 1 and 2 co-occurring at  $j$  number of sites among all sites,  $b$  is the number of ways that species 1 can be placed among the remaining sites that do not already have species 2, and  $c$  is the number of ways that species 2 can be placed among the remaining sites that do not already have species 1. The product,  $abc$ , is then divided by the number of unique ways that species 1 and 2 can be arranged among all sites without regard for  $j$ . This quantity is  $d \times e$ , where  $d = C(N, N_2)$  and  $e = C(N, N_1)$ . The probability,  $p_j = abc/de$ , represents the probability that species 1 and 2 co-occur at exactly  $j$  number of sites. The probability,  $p_j$ , is then calculated for all  $j$  satisfying the inequality:  $\max\{0, N_1 + N_2 - N\} \leq j \leq \min\{N_1, N_2\}$ . Finally, the quantity  $\sum p_j$  for all  $j < J_{\text{obs}}$  represents the probability that the two species would co-occur at fewer than  $J_{\text{obs}}$  sites (the number of sites where the two species actually co-occur). Likewise,  $\sum p_j$  for all  $j > J_{\text{obs}}$  is the probability that the two species would co-occur at greater than  $J_{\text{obs}}$  sites. Without requiring any randomization of the data, the probabilistic model analytically determines the probability that two species of a pair co-occur at an observed frequency ( $J_{\text{obs}}$ ) greater than ( $P_{\text{gt}}$ ) or less than ( $P_{\text{lt}}$ ) the frequency expected if the two species were distributed randomly of each other ( $J_{\text{exp}}$ ). Species pairs can be classified as positive, negative or random associations based upon the values of  $P_{\text{gt}}$  (or  $P_{\text{lt}}$ ) relative to a pre-defined significance level. The model has several desirable statistical properties including rigorous control of Type I and II error probabilities (Veech, 2013).

The probabilistic model is essentially an analytical analogue of an F-E randomization algorithm in which species incidences are fixed but sampling sites are equiprobable in their chances of having a given species. (Granted, the researcher has control over which sites are included in the set of  $N$  and it is generally best practice not to include sites outside the species' geographical range.) Therefore, the probabilistic model examines co-occurrence without regard for any constraints imposed by variation in species richness among sampling sites or variation in any other site characteristic (e.g. area). In this way, the probabilistic model implicitly recognizes the biological validity of incidence rates (which are fixed in the model) but does not subsume co-occurrence patterns under patterns of species richness. In fact, the contrary may be more realistic: co-occurrence patterns determine differences among sampling sites (ecological communities) in richness and composition and thus there is no reason that sites must be equiprobable in the probabilistic model. That is, site richness can be an unconstrained variable in the model if one views richness as the consequence of co-occurrence, instead of vice versa. Nonetheless,

the F-E randomization algorithm (in null models) is most appropriately applied to data from equal-sized sampling units (sites) of similar ecological conditions. In situations where sites vary in a meaningful environmental property, the equiprobable assumption might not be justified. I am working on a modification of the probabilistic model that will remove the assumption of equiprobable sites (see below).

## INDEPENDENCE OF SPECIES IN CO-OCCURRENCE ANALYSES

Previous authors have correctly pointed out that the co-occurrence of two species in a pair may not be completely independent of a third species (Zaman & Simberloff, 2002; Sfenthourakis *et al.*, 2006; Gotelli & Ulrich, 2010; Pitta *et al.*, 2012). That is, if species  $x$  and  $y$  tend to positively co-occur while  $x$  and  $z$  also positively co-occur then species  $y$  and  $z$  might also be expected to co-occur positively (the same scenario can also be presented with negative co-occurrence). This non-independence could be problematic for interpreting pairwise co-occurrence patterns involving multiple species pairs (Zaman & Simberloff, 2002); for example,  $x$  and  $y$  may be positively and significantly associated with one another simply because each is significantly associated with  $z$ . However, this need not always be the case. Species  $x$  and  $y$  might co-occur at different sites than do  $x$  and  $z$  such that  $y$  and  $z$  never or rarely co-occur (Sfenthourakis *et al.*, 2006). From the perspective of set theory, co-occurrence of  $x$ ,  $y$  and  $z$  with each other (in pairs) is non-independent only when the intersection (number of co-occurrence sites) of  $x$  and  $y$  is large and intersection of  $x$  and  $z$  is large compared with each set  $x$ ,  $y$  and  $z$ . In that case, intersection of  $y$  and  $z$  would also be large and the three species would co-occur at a large proportion of the total number of sites and the species pairs ( $xy$ ,  $xz$  and  $yz$ ) would not be independent. Researchers using a pairwise approach should be cognizant of such non-independence when tallying the number of positive and negative associations in a dataset.

Of course, non-independence of species pairs might represent ecological reality. For example, positive co-occurrence should exist among two or more competing prey species and a predator species when there is predator-mediated coexistence. This describes the situation where competitively inferior prey species are able to coexist with a competitively dominant prey species whose abundance is limited by the predator (Paine, 1974; Caswell, 1978; Shurin & Allen, 2001). Non-random co-occurrence of three or more species (i.e. non-independence of species pairs) is a characteristic of a dataset more so than the method of analysis. Indeed, some methods (e.g. network analysis) could be useful in identifying sets of closely linked species (see below).

## TESTABLE PREDICTIONS

Several authors have recently suggested that the study of species co-occurrence could be improved conceptually and heuristically if researchers would specify a priori the co-occur-

rence patterns expected (predicted) in the data (Sfenthourakis *et al.*, 2006; Gotelli & Ulrich, 2010; Veech, 2013). Historically, analysis of co-occurrence has consisted primarily of post hoc interpretation; that is, testing for non-random pattern and then proposing an explanation instead of boldly and precisely predicting patterns based upon knowledge of ecological processes and principles. At the very least, researchers should be able to develop predictions based on species traits (e.g. habitat preferences, competitive ability, climatic tolerances) and site properties. The importance of prediction (i.e. formal hypothesis testing) is that it provides a more rigorous test of proposed explanation.

A previous study by Sfenthourakis *et al.* (2006) is a step in the right direction. Although they did not make precise predictions for how specified species should be associated, they did propose and use an insightful framework to distinguish between ecological and geographical factors (processes). This framework involves compiling an 'ecological matrix' in addition to the typical 'geographical matrix'. The geographical matrix is a species  $\times$  sampling site matrix. The ecological matrix replaces actual physical sampling locations with particular environmental variables or characteristics (e.g. low elevation, coniferous vegetation, calcareous substrate) that have been recorded for each site (see Table 4 in Sfenthourakis *et al.*, 2006). For example, if species 1 was found at sampling sites that have coniferous vegetation then the column representing 'coniferous vegetation' and the row representing 'species 1' is given a value of '1' to denote the species presence for that habitat variable. Applying a pairwise approach on both the geographical and ecological matrices allows a researcher to predict and interpret patterns of co-occurrence (see Table 1 in Sfenthourakis *et al.*, 2006). For example, competitive exclusion should lead to two potentially competing species having a positive pairwise association in the ecological matrix (because they have similar niches) but a negative pairwise association in the geographical matrix (i.e. no syntopy anywhere) (Sfenthourakis *et al.*, 2006).

Much of the insight and benefit of the Sfenthourakis *et al.* (2006) framework is a result of the comprehensive characterization of environmental conditions at each site. They used an ecological matrix that had 41 environmental variables representing all types of factors that could be important to the distribution of the study organisms, terrestrial isopods. They also explicitly used a pairwise approach for analysing co-occurrence instead of a matrix-level approach. This further facilitated their ability to classify each species pair as a positive, negative or neutral association. In a somewhat analogous way, the probabilistic model can also be used to examine the relative roles of different environmental variables on species co-occurrence. The set of sampling sites ( $N_X$ ) used in the analysis can be specified so as to only include sites that are similar for a given variable,  $X$ . The association between two species is then examined for  $N_X$  and classified as positive, negative or random along with calculating the difference between observed and expected co-occurrence,  $J_{\text{obs}} - J_{\text{exp}}$ . The same steps are repeated for the set of remaining sites not in  $N_X$ . Subsequent comparison of

the results of the two analyses (and multiple analyses of this type) then may indicate which environmental factors are most responsible for the observed co-occurrence (or lack thereof). The environmental variables could even be species richness and the presence of a particular third species.

Predictions about species co-occurrence can also be based on species traits, regardless of whether the analysis uses information about site characteristics. It is relatively common to test for non-random co-occurrence patterns in subsets of species within the same taxonomic units (e.g. genera or families) to determine if evolutionary relatedness (and hence trait similarity) affects co-occurrence. Such tests rely on the notion that evolutionarily closely related species are more likely to be potential competitors (i.e. have relatively greater niche overlap) than are distantly related species (Elton, 1946; Denno *et al.*, 1995; Losos, 2008; Valiente-Banuet & Verdú, 2008; Wiens *et al.*, 2010; Ricklefs, 2011; Violle *et al.*, 2011). However, researchers could go even further by identifying a priori particular traits that might mediate (or prevent) coexistence. Exact predictions could then be made for specific pairs based on their traits. Further, species traits (or at least trait–environment interactions) can also be formally incorporated into the co-occurrence analysis (e.g. Sfenthourakis *et al.*, 2006). I am currently developing an extension of the probabilistic model that will allow users to specify a probability that a particular species could occur at a given site. In this way,  $N$  (total number of occupiable sites) is adjusted based on either species traits or environmental properties of the sites.

The pairwise approach does not rely on the overly simplistic assumption that species affect one another only as pairs isolated from all other species. Processes such as predator-mediated coexistence involve more than two species and hence the resultant co-occurrence pattern will reflect that. Pairwise analyses might still be used in those scenarios to test hypotheses. However, new methods such as network analysis (Carstensen & Olesen, 2009; Araújo *et al.*, 2011; Fontaine *et al.*, 2011; Carstensen *et al.*, 2012) that test for non-random clusters of species among sites (i.e. modules and compartments; Thébault, 2013) may be ideal for uncovering the co-occurrence patterns among multiple (more than two) species (Table 1). These methods primarily attempt to identify the clusters (modules) and compartments within a matrix in a way roughly analogous to how pairwise approaches attempt to identify the positive, negative and random species associations. These are essentially tests for structure within a matrix but not tests for any emergent property of a matrix (see next section). As such, network analysis does not carry the hidden assumption that a presence–absence matrix is a real ecological entity, unless of course an entire matrix is found to be one large module.

### PRESENCE–ABSENCE MATRICES: NECESSITY OR CONVENIENCE?

Is the species presence–absence matrix a necessity for analysing co-occurrence and thinking about how species assemblages are

geographically structured or is it simply a convenient way to organize and store data? The pairwise approach suggests the latter. If causative mechanisms can be tested and found by analysing species associations one-by-one then there is no need to think that the presence–absence matrix has any meaningful ecological property (in and of itself). This argument also covers multispecies associations – finding non-random clusters or modules of species within a matrix probably indicates that some ecological process is at work producing the association although it does not mean that the entire matrix itself has any emergent properties. Matrix-level metrics simply summarize the co-occurrence patterns (or lack thereof) that exist among the species; the metrics themselves are not measuring any kind of co-occurrence pattern that exists only at the matrix level. One way to see this is to consider that the value of a matrix-level metric typically will not change much when a given species or site is removed. If some emergent property of the matrix was being measured, then even a very slight or small change in the matrix should lead to a dramatic change in the metric – in which case, a presence–absence matrix truly would be more than the ‘sum of its parts’. Although the study of co-occurrence and community assembly is historically rooted in the analysis of presence–absence matrices, there is no compelling reason to keep searching for nonrandom structure at the level of an entire matrix.

### CONCLUSIONS

Using a pairwise approach, species interactions can be classified as positive, negative or random (neutral). Although, as with any statistical test, some species pairs might not be classifiable simply as a result of the test having low power (Sfenthourakis *et al.*, 2006; Veech, 2013). Together these interactions represent all the possible ways that two species could associate in nature, and presumably each type of association represents the outcome of real ecological and evolutionary processes. Matrix-level metrics do not allow for classifying pairs as positive, negative or random associations. Certainly, in any species assemblage or set of communities, there may be meaningful associations (i.e. modules and compartments) that simultaneously involve multiple species. Nonetheless, this does not detract from using a pairwise approach. Indeed, a recent study using network analysis showed that many species are only weakly linked to most other species in a network (Araújo *et al.*, 2011). This suggests that multispecies interactions (e.g. diffuse competition) may not be relevant (see also Pitta *et al.*, 2012).

The main conceptual limitation on the analysis of co-occurrence patterns is that we cannot directly infer the exact process(es) responsible for any particular pattern (Schluter, 1984). This is probably a more severe limitation when analyses are matrix-based than when they are based on species pairs. The pairwise approach is very amenable to hypothesis-testing in that predictions about co-occurrence can be made a priori for paired species based on their traits and/or the properties of the sampling sites.

## ACKNOWLEDGEMENTS

I thank Yoni Gavish, Ivan Castro-Arellano, Dan McGarvey, Michael Huston, Dominique Gravel, Spyros Sfenthourakis, Werner Ulrich and an anonymous referee for providing insightful comments that helped me improve the manuscript.

## REFERENCES

- Almeido-Neto, M., Guimarães, P., Guimarães, P.R., Loyola, R.D. & Ulrich, W. (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, **117**, 1227–1239.
- Araújo, M.B., Rozenfeld, A., Rahbek, C. & Marquet, P.A. (2011) Using species co-occurrence networks to assess the impacts of climate change. *Ecography*, **34**, 897–908.
- Arita, H.T., Christen, A., Rodríguez, P. & Soberón, J. (2012) The presence–absence matrix reloaded: the use and interpretation of range–diversity plots. *Global Ecology and Biogeography*, **21**, 282–292.
- Atmar, W. & Patterson, B.D. (1993) The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, **96**, 373–382.
- Bowers, M.A. & Brown, J.H. (1982) Body size and coexistence in desert rodents: chance or community structure? *Ecology*, **63**, 391–400.
- Carstensen, D.W. & Olesen, J.M. (2009) Wallacea and its nectarivorous birds: nestedness and modules. *Journal of Biogeography*, **36**, 1540–1550.
- Carstensen, D.W., Dalsgaard, B., Svenning, J.-C., Rahbek, C., Fjeldså, J., Sutherland, W.J. & Olesen, J.M. (2012) Biogeographical modules and island roles: a comparison of Wallacea and the West Indies. *Journal of Biogeography*, **39**, 739–749.
- Caswell, H. (1978) Predator-mediated coexistence: non-equilibrium model. *The American Naturalist*, **112**, 127–154.
- Collins, M.D., Simberloff, D. & Connor, E.F. (2011) Binary matrices and checkerboard distributions of birds in the Bismarck Archipelago. *Journal of Biogeography*, **38**, 2373–2383.
- Connor, E.F. & Simberloff, D. (1979) The assembly of species communities: chance or competition? *Ecology*, **60**, 1132–1140.
- Denno, R.F., McClure, M.S. & Ott, J.R. (1995) Interspecific interactions in phytophagous insects: competition re-examined and resurrected. *Annual Review of Entomology*, **40**, 297–331.
- Diamond, J.M. (1973) Distributional ecology of New Guinea birds. *Science*, **179**, 759–769.
- 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.
- Diamond, J.M. & Gilpin, M.E. (1982) Examination of the “null” model of Connor and Simberloff for species co-occurrences on islands. *Oecologia*, **52**, 64–74.
- Elton, C. (1946) Competition and the structure of ecological communities. *Journal of Animal Ecology*, **15**, 54–68.
- Fontaine, C., Guimarães, P.R., Kéfi, S., Loeuille, N., Memmott, J., van der Putten, W.H., van Veen, F.J.F. & Thébaud, E. (2011) The ecological and evolutionary implications of merging different types of networks. *Ecology Letters*, **11**, 1170–1181.
- Gilpin, M.E. & Diamond, J.M. (1984) Are species co-occurrences on islands non-random and are null hypotheses useful in community ecology? *Ecological communities: conceptual issues and the evidence* (ed. by D.R. Strong, D. Simberloff, L.G. Abele and A.B. Thistle), pp. 297–315. Princeton University Press, Princeton, NJ.
- 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. & Ulrich, W. (2010) The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia*, **162**, 463–477.
- Losos, J.B. (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, **11**, 995–1007.
- Paine, R.T. (1974) Intertidal community structure: experimental studies on relationship between a dominant competitor and its principal predator. *Oecologia*, **15**, 93–120.
- Patterson, B.D. & Atmar, W. (1986) Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society*, **28**, 65–82.
- Pielou, E.C. (1977) *Mathematical ecology*. Wiley Interscience, New York.
- Pitta, E., Giokas, S. & Sfenthourakis, S. (2012) Significant pairwise co-occurrence patterns are not the rule in the majority of biotic communities. *Diversity*, **4**, 179–193.
- Ricklefs, R.E. (2008) Disintegration of the ecological community. *The American Naturalist*, **172**, 741–750.
- Ricklefs, R.E. (2011) Applying a regional community concept to forest birds of eastern North America. *Proceedings of the National Academy of Sciences USA*, **108**, 2300–2305.
- Sanderson, J.G. (2000) Testing ecological patterns: a well-known algorithm from computer science aids the evaluation of species distributions. *American Scientist*, **88**, 332–339.
- Sanderson, J.G., Diamond, J.M. & Pimm, S.L. (2009) Pairwise co-existence of Bismarck and Solomon landbird species. *Evolutionary Ecology Research*, **11**, 771–786.
- Schluter, D. (1984) A variance test for detecting species associations with some example applications. *Ecology*, **65**, 998–1005.
- Sfenthourakis, S., Giokas, S. & Tzanos, E. (2004) From sampling stations to archipelagos: investigating aspects of the assemblage of insular biota. *Global Ecology and Biogeography*, **13**, 23–35.
- Sfenthourakis, S., Tzanos, 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.
- Shurin, J.B. & Allen, E.G. (2001) Effects of competition, predation, and dispersal on species richness at local and regional scales. *The American Naturalist*, **158**, 624–637.
- Thébault, E. (2013) Identifying compartments in presence–absence matrices and bipartite networks: insights into modularity measures. *Journal of Biogeography*, **40**, 759–768.
- Ulrich, W. & Gotelli, N.J. (2013) Pattern detection in null model analysis. *Oikos*, **122**, 2–18.
- Ulrich, W., Almeida-Neto, M. & Gotelli, N.J. (2009) A consumer's guide to nestedness analysis. *Oikos*, **118**, 3–17.
- Valiente-Banuet, A. & Verdú, M. (2008) Temporal shifts from facilitation to competition occur between closely related taxa. *Journal of Ecology*, **96**, 489–494.
- Veech, J.A. (2006) A probability-based analysis of temporal and spatial co-occurrence in grassland birds. *Journal of Biogeography*, **33**, 2145–2153.
- Veech, J.A. (2013) A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, **22**, 252–260.
- Violle, C., Nemergut, D.R., Pu, Z. & Jiang, L. (2011) Phylogenetic limiting similarity and competitive exclusion. *Ecology Letters*, **14**, 782–787.
- Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B., Cornell, H.V., Damschen, E.I., Davies, T.J., Grytnes, J.A., Harrison, S.P., Hawkins, B.A., Holt, R.D., McCain, C.M. & Stephens, P.R. (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, **13**, 1310–1324.
- Wright, D.H. & Reeves, J.H. (1992) On the meaning and measurement of nestedness of species assemblages. *Oecologia*, **92**, 416–428.
- Zaman, A. & Simberloff, D. (2002) Random binary matrices in biogeographical ecology – instituting a good neighbor policy. *Environmental and Ecological Statistics*, **9**, 405–421.

## BIOSKETCH

**Joseph Veech** is broadly interested in the ecological and evolutionary factors that affect species distribution and abundance at a wide range of spatial scales. He also works to develop new statistical methods for the testing of pattern and evaluates previously proposed methods. Most recently, he has taken an interest in promoting the use of probability theory in ecological research.

---

Editor: Miguel Araújo