

Species Co-occurrence, an Annotated Bibliography

I have listed below a set of papers describing research in the area of using species lists from a set of sites (primarily islands in an archipelago). The list is primarily constructed of papers that I read while researching this topic. These are listed in bold face, with my comments below each. The list also includes papers which I did not read—or did not read thoroughly enough to comment significantly on them. I deem that these papers are important to the overall discussion and so for completeness have listed them in grey with minimal or no notes as to their content.

Simberloff DS. (1970) Taxonomic diversity of island biotas. *Evolution* 24: 23-47.

Foundation work for discussing island communities.

Diamond JM. (1975) Assembly of species communities. In *Ecology and Evolution of Communities*: 342-444, ML Cody and JM Diamond (eds.), Belknap Press: Cambridge, MA.

102 page monograph on the importance of competition in the structure of avian communities in the Bismark archipelago. The author picks a few examples to support his overall theory, detailing “assembly rules” for these guilds. It is interesting from a naturalist point-of-view, but not from an ecological point-of-view—long on description and conjecture, short on science.

Connor EF and D Simberloff. (1978) Species number and compositional similarity of the Galapagos flora and Avifauna. *Ecological Monographs* 48: 219-248.

The authors discuss alternative thinking on community assembly and site similarity in an island archipelago. Contains an excellent, though nuanced, treatment of Q-mode analysis using incidence matrix data. Has similar findings to Connor & Simberloff 1979, insofar as noting that much of the similarity between islands can be attributed to chance.

Connor EF and D Simberloff. (1979) The assembly of species communities: chance or competition? *Ecology* 60: 1132-1140.

This is the bombshell article that sparked real discussion about the supremacy of competition as an organizing force for species communities. The authors randomly construct sets of island communities from existing species pools according to a set of rules which treat each species independently from every other species (but allow for islands to be able to support a particular number of species and for species to be more or less common). Although the analysis of the constructed communities versus the actual communities is flawed, the authors present ample evidence to support their primary claim. That claim being: it should not be taken for granted that particular species pairs do not exist in nature because of competitive exclusion. They show that the community patterns found in three separate archipelago data sets also occur when species are assembled into communities independently of one another. Their methods include techniques for exploring the ways in which one can sample the space of incidence matrices (species in rows and islands/sites in columns), while preserving row and column sums.

Species Co-occurrence, an Annotated Bibliography

Simberloff D and EF Connor. (1979) Q-mode and R-mode analyses of biogeographic distributions: null hypotheses based on random colonization. In, *Contemporary Quantitative Ecology and Related Ecometrics*: 123-128, GP Patil and ML Rosenzweig (eds), International Cooperative Publishing House: Fairland, MD.

Q-mode assesses similarity of sites, R-mode assesses similarity of species occurrences. Most analyses proceed from the assumption that the R-mode is primarily important.

Brualdi RA. (1980) Matrices of zeros and ones with fixed row and column sum vectors. *Linear Algebra Applications* 33: 159-231.

This is of interest primarily because of the proof that all possible incidence matrices with fixed row and column sums can be arrived at by the process of swapping 2x2 diagonal submatrices.

Simberloff D and EF Connor. (1981) Missing species combinations. *American Naturalist* 118: 215-239.

This paper focuses on species combinations that are not observed in an archipelago (often ascribed to competitive exclusion). The authors suggest that chance can often explain much of these patterns. This paper comprehensively treats historical forays into this realm, with astounding breadth and depth and detail.

Diamond JM and ME Gilpin. (1982) Examination of the “null” model of Connor and Simberloff for species co-occurrence on islands. *Oecologia* 52: 64-74.

Gilpin ME and JM Diamond. (1982) Factors contributing to non-randomness in species co-occurrences on islands. *Oecologia* 52: 75-84.

The authors put up a spirited defense against Connor & Simberloff 1979. They raise several interesting and valid points, but overall fail to overcome the general point of Connor & Simberloff 1979, that species interactive effects (e.g. competition) do not seem to be overly important in structuring island communities. Their valid points are: competitive effects are likely incorporated in the null model, and looking at an entire taxa dilutes the likelihood of finding competitive effects (versus looking at guilds). Some incorporation of species interaction effects is inevitable in forming any null model, since it is impossible to determine at this time what forces structured the communities that now appear on the islands—interaction effects likely played *some* role. The authors’ attempt to create an alternative to Connor & Simberloff’s community construction technique is severely flawed and its analysis is not well thought out—there is little of value in the second half of this paper.

Wright SJ and CC Biehl. (1982) Island biogeographic distributions: testing for random, regular, and aggregated patterns of species occurrence. *American Naturalist* 119: 345-357.

A “shared island” test and criticism of Connor & Simberloff 1978, rebuked by Simberloff and Connor 1984.

Species Co-occurrence, an Annotated Bibliography

Simberloff D and EF Connor. (1984) Inferring competition from biogeographic data: a reply to Wright and Biehl. *American Naturalist* 124: 429-436.

This seems to be of interest only if you read Wright & Biehl and became confused about Simberloff & Connor 1979.

Gilpin ME and JM Diamond. (1984) Are species co-occurrences on islands non-random, and are null hypotheses useful in community ecology? pp297-315.

Connor EF and D Simberloff. (1984) Neutral models of species' co-occurrence patterns. pp316-331.

and Rejoinders: pp332-343.

In *Ecological Communities Conceptual Issues and the Evidence*, DR Strong jr, D Simberloff, LG Abele, AB Thistle (eds), Princeton University Press, Princeton, NJ.

Not much new thinking in these articles, just restatements of old arguments and poking little holes in little details of the opposing argument. These articles seem to exist only because there was blood in the water and people love spectacle—fascinating reading from a history of science point-of-view, but very little new thought. Connor & Simberloff make the point that if row and column sums are not fixed in the null model (as they are not in Gilpin & Diamond's model), it is then possible to randomly create degenerate matrices—i.e. islands with no species and species that do not occur on any island. Gilpin & Diamond respond as if this is an attack on their character & moral fiber. (No, I'm not kidding.)

Schluter D. (1984) A variance test for detecting species associations, with some example applications. *Ecology* 65:: 998-1005.

Develops the variance ratio statistic, later used in Gotelli 2000.

Connor E and D Simberloff. (1986) Competition, scientific method, and null hypotheses in ecology. *American Scientist* 74: 155-162.

Connor's final words on the subject of science versus description in the use of biogeographic incidence data in ecology. Emphasis on the need for rigorous examination of non-experimental data. Also an excellent article for understanding the authors' philosophy of science as it applies to ecology.

Ryti RT and ME Gilpin. (1987) The comparative analysis of species occurrence patterns on archipelagos. *Oecologia* 73: 282-287.

This article describes an analysis technique which is tangential to the co-occurrence analyses—creating three broadly-interpretable statistics from an incidence matrix. It doesn't specifically address co-occurrence patterns in the context of competition. It doesn't seem as though the proposed technique ever caught on either.

Wilson JB. (1987) Methods for detecting non-randomness in species co-occurrences: a contribution. *Oecologia* 73: 579-582.

Tests Gilpin & Diamond's 1982 methodology on randomly constructed matrices—finds significant structure. Concludes that Gilpin & Diamond's technique is fatally flawed and develops an alternative. The alternative is a

construction technique whereby a 1 is assigned to the row or column which has the most outstanding 1's to be placed—this technique suffered from no “hang-ups” in any of Wilson's trials. It is subsequently shown (Gotelli 2000) that this technique is strongly biased toward certain matrix arrangements.

Gilpin ME and JM Diamond. (1987) Comments on Wilson's Null Model. *Oecologia* 74: 159-160.

Gilpin & Diamond claim (rather unconvincingly) that Wilson has misunderstood their model & that his tests are invalid. They further claim that Wilson's technique is flawed because it can't detect perfect checkerboards. This is Gilpin & Diamond blowing smoke and hoping that no one figures it out.

Roberts A and L Stone. (1990) Island-sharing by archipelago species. *Oecologia* 83: 560-567.

Some deep-thinking, independent voices arrive on the scene. The authors explain at some length why the chi-squared analysis performed in Connor & Simberloff 1979 (and also by D&G 1982) was not appropriate—in short, because the observations compared cannot be considered independent. The authors also develop a “sharing index” and apply it to the Vanuatu dataset—the sharing index is simple and meaningful; it is never used again because it cannot suggest the cause or direction of departure from independent assembly. This work is the first to cite a reference to a proof (Brualdi 1980) that the swap technique can create every possible incidence matrix with fixed row & column sums.

Stone L and A Roberts. (1990) The checkerboard score and species distributions. *Oecologia* 85: 74-79.

The authors invent a new metric for measuring the impact of negative species interactions: the C-score, which measures the average degree of species segregation and gives more weight to segregated widespread species than to uncommon ones. The authors make use of an undocumented fill technique (by B. Milne) in addition to the usual swap technique and re-analyze the Vanuatu and Antilles datasets.

Bolger DT, AC Alberts, and ME Soulé. (1991) Occurrence patterns of bird species in habitat fragments: sampling, extinction, and nested species subsets. *American Naturalist* 137: 155-166.

I never got hold of a copy of this.

Stone L and A Roberts. (1992) Competitive exclusion or species aggregation? *Oecologia* 91: 419-424.

Stone & Roberts squeeze out one more paper, using two new metrics and their old C-score (Stone & Roberts 1990). Not much added value—just the inclusion of metrics specifically aimed at capturing species aggregations instead of species segregations.

Manly BFJ. (1995) A note on the analysis of species co-occurrences. *Ecology* 76: 1109-1115.

The author bemoans the inefficiency of techniques requiring thousands of swaps and proposes using an unbiased sample drawn from the set of all sequential swaps of length n which contain the original incidence matrix. However, this is not an unbiased sample of the set of *all possible* incidence matrices with fixed row and column sums.

Gotelli NJ and GR Graves. (1996) Chapter 7. Co-occurrence. pp 153-205 in *Null Models in Ecology*. Smithsonian Institution Press, Washington, DC 368p.

Worth reading simply for the introduction, but also contains much well-reasoned thought on the topic. The authors put their fingers on the heart of the controversy: can assembly rules be inferred from presence/absence data, without experimental data? After reviewing null models for R-mode analysis, conclude that either fixed row & proportional columns or proportional rows & columns are best—although see also Gotelli 2000.

Sanderson JG, MP Moulton, and RG Selfridge. (1998) Null matrices and the analysis of species co-occurrences. *Oecologia* 116: 275-283.

The authors propose a recursive construction technique hereafter referred to as the knights tour. A promising technique, but I would like to have seen their algorithm sketched out in a bit of detail, instead of having to glean hints about specifics from the text. Their analysis makes some of the same independence assumptions as were debunked by Roberts and Stone 1990. See also Gotelli & Entsminger's comments in 2001.

Gotelli NJ. (2000) Null model analysis of species co-occurrence patterns. *Ecology* 81(9): 2606-2621.

The author provides an excellent overview of incidence matrix randomization techniques and co-occurrence indices. He then systematically tests nine incidence matrix construction techniques (null models) and four co-occurrence indices for their performance, insofar as type I and type II error. Uses sequential swaps, like Manly 1995, but the sample need not include the original matrix (again, this is not an unbiased sample of the set of all possible matrices with fixed row & column sums). Concludes that row sums (species) should always be fixed, and that the C-score yields the best type-II error rate with a tolerable type-I error rate.

The author also suggests that both “island lists” and “sample lists” are appropriate for these kinds of analyses. This is the first time I have seen anyone suggest that lists of species taken from samples (in relatively homogeneous habitat) are testable using these techniques, which have historically only been applied to species lists believed to be exhaustively complete.

Gotelli NJ and GL Entsminger. (2001) Swap and fill algorithms in null model analysis: rethinking the knight's tour. *Oecologia* 129: 281-291.

The authors think a little deeper about Sanderson, et al's algorithm & find a way to remove a perceived bias (whether the bias was real or not is difficult to

Species Co-occurrence, an Annotated Bibliography

determine from the sketchy way in which the algorithm was outlined in Sanderson, et al 1998). Certainly a construction technique is preferable to a swapping technique from the perspective of eliminating bias in the selection of test matrices—the downside is that this construction technique is vastly more expensive in compute cycles than any of the swap techniques. Gotelli’s “random knight’s tour” is certainly a large step in the right direction, although he, too, does not supply enough details on his algorithm to make plain whether there are any lingering sources of bias. Still, I find this algorithm more attractive than any swap algorithm I’ve seen, with the possible exception of Zaman and Simberloff 2002, which has its own problems.

Zaman A and D Simberloff. (2002) Random binary matrices in biogeographical ecology – instituting a good neighbor policy. *Environmental and Ecological Statistics* 9: 405-421.

The authors show excellent thinking about how to adjust swap techniques to reduce their bias. Unfortunately, their brilliance is marred by having only sample analysis which uses a metric which doesn’t measure what they claim it does (using rather trivial counter-examples). This is largely due to the inclusion of a term which counts the number of sites/islands on which a species does *not* occur.

Gotelli NJ and DJ McCabe. (2002) Species co-occurrence: A meta-analysis of J. M. Diamond's assembly rules model. *Ecology* 83 (8): 2091-2096.

The authors selected 96 datasets from studies with a goal of constructing complete species lists. They analyzed these using a null model of fixed row & column sums and three co-occurrence indices (C-score, CHECKER, and COMBO). Standardizing these scores showed that, in general, there are fewer species pairs co-occurring than would be expected by chance.

Manly, B., and J. G. Sanderson. (2002) A note on null models: justifying the methodology. *Ecology* 83:580–582.

I did not read this paper, but included the reference because it is the target of a paper I did read, below: Gotelli and Entsminger 2003.

Gotelli NJ and K Rohde. (2002) Co-occurrence of ectoparasites of marine fishes: a null model analysis. *Ecology Letters* 5: 86-94.

Contains of lovely summary of the state of species co-occurrence literature and applies it in a meaningful way to a new kind of islands inhabited by a new kind of fauna.

Gotelli NJ and GL Entsminger. (2003) Swap algorithms in null model analysis. *Ecology* 84 (2): 532-535.

The authors effectively refute criticisms from Manly and Sanderson 2002 (not one that I read, but listed above for completeness). G&E attempted to reproduce M&S’s results using prior and current releases of their EcoSim software, but were unable to generate even a single type-I error, whereas M&S had generated 20 (of 20) in their implementation.

Osnas EE and CD Ankney. (2003) Null models of North American prairie duck communities: local habitat conditions and temporal scale influence community patterns. *Evolutionary Ecology Research* 5: 913-32.

The authors use abundance data, randomly redistributed to form a null model of community assembly. The overall technique is quite clever, and likely useful. However, some of the specific decisions on which birds count and which birds don't, as well as how to treat observations of the different sexes, don't seem to make much sense.

Miklos, I and J Podani. (2004) Randomization of presence-absence matrices: comments and new algorithms. *Ecology* 85: 86-92.

The authors propose a new technique for examining the space of incidence matrices with fixed row & column sums. Their summary of the state of the problem is good, but their description of their algorithm and why it is better than existing algorithms needs work. They may have a solution, but their description is opaque enough that I don't think anyone knows for sure. They certainly don't demonstrate that their algorithm samples the space uniformly or completely—not that any other algorithm has demonstrated both of these either.

Ulrich, W. (2004) Species co-occurrences and neutral models: reassessing J.M. Diamond's assembly rules. *Oikos* 107: 603-9.

This is very similar to the Bell 2005 paper (below), but not as good.

Bell, G. (2005) The co-distribution of species in relation to the neutral theory of community ecology. *Ecology* 86(7): 1757-70.

As a proponent of Hubbell's neutral community model, Bell tests neutral communities for "independent assembly" using incidence matrix techniques. He finds that there are both more co-occurrences and more mutually exclusive species than the null model predicts. This lends an extra notion of veracity to neutral theory—although I personally find it more likely to apply to long-lived plants (or other sedentary organisms in which precedence plays a large role).

Ribichich, AM. (2005) From null community to non-randomly structured actual plant assemblages: parsimony analysis of species co-occurrences. *Ecography* 28: 88-98.

The abominable English aside, this paper presents some interesting ideas that aren't all that convincing in the final wash. In short, starting with a number of nearby communities, a single pre-historical source was hypothesized and parsimony used to guess how the communities evolved over time into the present communities. Flaws are too numerous to list, but the big ones include inadequacy of sampling, ophiophobia, and the improbability of the root assumption: that all of the communities came from the same source community.

Gotelli NJ and GL Entsminger. (2005) EcoSim: Null models software for ecology. Version 7. Acquired Intelligence Inc. & Kesey-Bear. Jericho, VT 05465. <http://garyentsminger.com/ecosim.htm>.

This free software package allows one to perform many kinds of tests on incidence & abundance matrices. The co-occurrence analysis includes all nine kinds of construction constraints and three sampling algorithms (sequential swap, independent swap, and random knight's tour). Also has options for dealing with degenerate matrices when they occur. (Also see Gotelli 2000.) Also contains analysis tools for "macroecology", "niche overlap", "size overlap", "species diversity", "guild structure", and a set of "standard tests".