

# *Significance testing in ecological null models*

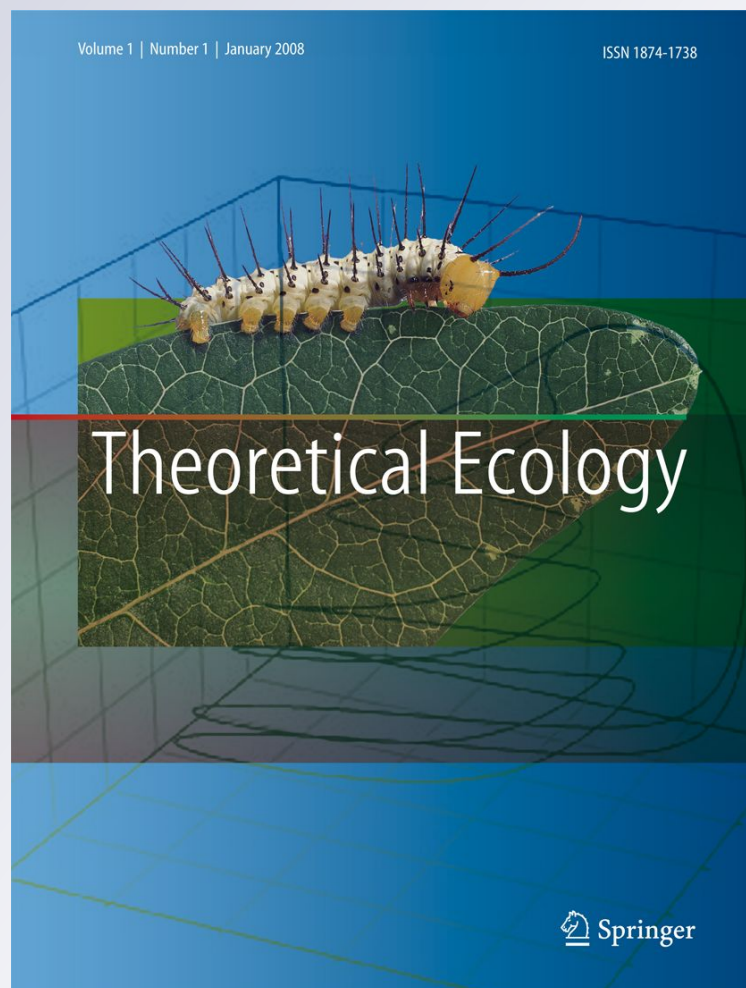
**Joseph A. Veech**

**Theoretical Ecology**

ISSN 1874-1738

Theor Ecol

DOI 10.1007/s12080-012-0159-z



**Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media B.V.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.**

# Significance testing in ecological null models

Joseph A. Veech

Received: 3 August 2011 / Accepted: 10 January 2012

© Springer Science+Business Media B.V. 2012

**Abstract** In the past decade, the use of null models has become widespread in the testing of ecological theory. Along with increasing usage, null models have also become more complex particularly with regard to tests of significance. Despite the complexity, there are essentially only two distinct ways in which tests of significance are conducted. Direct tests derive a  $p$  value directly from the null distribution of a test statistic, as the proportion of the distribution more extreme than the observed value of the test statistic. Indirect tests compare an observed value of a parameter to a null distribution by conducting an additional analysis such as a chi-square test, Kolmogorov–Smirnov test, or regression, although in many cases, this additional step is not necessary. Many kinds of indirect tests require that the null distribution is normal whereas direct tests carry no assumptions about the form of the null distribution. Therefore, when assumptions are violated, indirect tests may have higher type I and II error rates than their counterpart direct tests. A review of 108 null model papers revealed that direct tests were used in 56.5% of studies and indirect tests used in 45.5%. A few studies used both types of test. In general, the randomization algorithms used in most null models should produce normal null distributions, but this could not be confirmed because most studies did not present any description of the null distribution. Researchers should be aware of the differences between direct and indirect tests so as to better use, communicate, and evaluate null models. In many

cases, direct tests should be favored for their simplicity and parsimony.

**Keywords** Critical value · Randomization · Statistical distribution · Test statistic · Type I error

## Introduction

The use of null models has become common in testing ecological theory. Topics and recent examples include trait-mediated coexistence and niche divergence/partitioning (Ackerly et al. 2006; Zimmermann et al. 2009), structure of complex food webs and interaction networks (Bascompte and Melian 2005; Bluthgen et al. 2008; Burns and Zotz 2010), community assembly (Cornwell et al. 2006; Adams 2007), phylogenetic community structure (Helmus et al. 2007; Mouillot et al. 2008; Ingram and Shurin 2009; Kembel 2009), local–regional species diversity relationships (Arita and Rodriguez 2004), and native–exotic diversity relationships (Fridley et al. 2004; Hulme 2008). This list is certainly not exhaustive, but it does indicate the variety of research areas where null models have been applied to test predictions of theory. Along with increased usage, null models have also become increasingly complex, particularly with regard to significance testing.

Most null models are statistical tests of a null hypothesis (occasionally, “null models” are used to simulate an ecological process without any accompanying test of significance). A parameter estimate or test statistic is calculated for a set of observed data. The null model then provides the probability (i.e., a  $p$  value) of obtaining an equal or more extreme value by chance and under the conditions specified by the null model, meant to represent some null hypothesis. Over the years, two distinct approaches for conducting the test of

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s12080-012-0159-z) contains supplementary material, which is available to authorized users.

---

J. A. Veech (✉)

Department of Biology, Texas State University,  
San Marcos, TX 78666, USA  
e-mail: joseph.veech@txstate.edu

significance (i.e., deriving the  $p$  value) have emerged, but heretofore, this critical distinction has not been recognized.

The  $p$  value can be obtained *directly* from a null distribution of a test statistic or *indirectly* from an alternative significance test applied to the null distribution. In both approaches, observed data are randomized or simulated so as to produce the null distribution. In the direct approach, the  $p$  value is taken as the proportion of the null distribution that is more extreme (on either tail) than the observed value of the test statistic. In the indirect approach, an additional step is performed. After producing the null distribution, a test of significance is conducted by comparing the observed value to the mean of the null distribution (or by comparing an entire distribution of observed values to the null distribution) (Fig. 1). In essence, the indirect approach does not recognize the null distribution as a distribution of a test statistic but rather takes the null distribution as a set of simulated parameter estimates that is compared to the observed estimate(s) using chi-square tests,  $t$  tests, regression, or other parametric tests. Essentially, an indirect test takes an inherently non-parametric approach (i.e., data randomization to produce a null distribution) and turns it into a parametric test of significance.

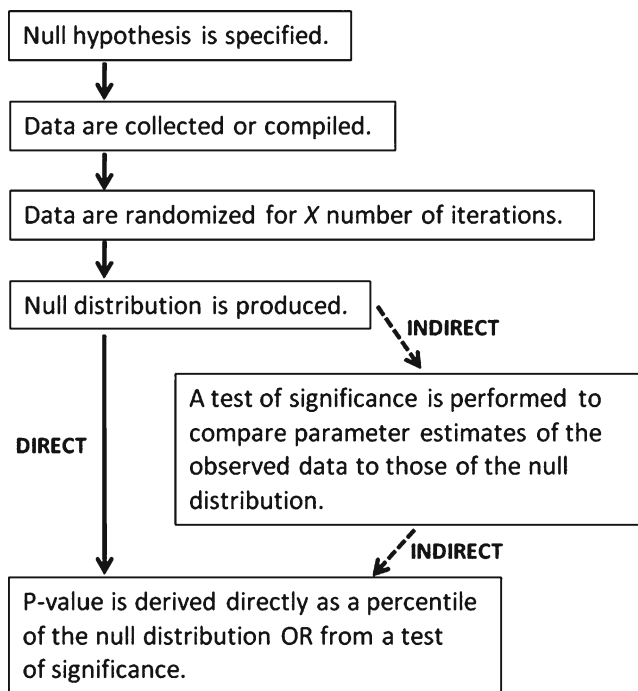
This distinction between direct and indirect tests has important implications for the use and communication of null models. Specifically, indirect tests often require additional assumptions about the form (normality) of the null distributions and all other assumptions that accompany chi-

square tests,  $t$  tests, and regression. When these assumptions are violated, the  $p$  value from a test may not be reliable (Quinn and Keough 2002). Direct tests do not require any assumption about the form of the null distribution; therefore, direct tests are preferable to indirect tests in most cases. Even when assumptions are met, indirect tests may result in unnecessary computational complexity because a direct calculation of the  $p$  value is always possible. The purpose of this review is to elaborate the differences between direct and indirect tests and appeal to researchers to use direct tests for their parsimony and ease of communication.

### Examples of direct and indirect tests

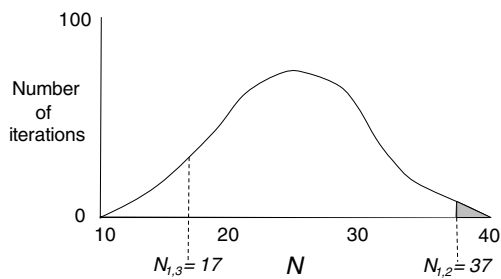
The primary purpose of most null models is to produce data in a form that would be expected in the absence of the particular process or force that presumably (according to the alternative hypothesis) causes structure in the observed data. Inasmuch as they exist and affect the observed data, other processes are allowed to contribute structure to the null pattern as well. Indeed the null model and pattern should include these other processes so that the process of interest can be isolated and tested (Tokeshi 1986; Wilson 1995; Gotelli and Graves 1996; Gotelli and McGill 2006; Moore and Swihart 2007). Typically, the null distribution is created via a repetitive process (1,000–10,000 iterations) of data randomization. A given parameter of interest (e.g., mean) is calculated for each random dataset and compiled into the null distribution. The null distribution represents the values of this parameter that could be obtained if the process of interest (specified in the null hypothesis) had no structure-producing effect on the data.

In a direct test, the  $p$  value is the proportion of random datasets that produce a value of the test statistic that is as extreme or more extreme (on either end of the null distribution, corresponding to one-tailed tests) than the observed value (Fig. 2). In an indirect test, the  $p$  value is not derived directly from the null distribution. Instead, a test of significance is applied to the null distribution. For example, some researchers (Thornber and Gaines 2004; Frank van Veen and Murrell 2005; Farias and Jaksic 2009; Manzaneda and Rey 2009; Okuda et al. 2009) have fit confidence intervals directly on the expected value or mean of the null distribution. In this approach, the observed value (i.e., the parameter estimated on the observed data) is then compared to this confidence interval. If the observed value is outside the confidence interval (greater than the upper confidence limit or less than the lower limit), then the  $p$  value is reported as the significance level (or twice the significance level if the test is two-tailed). This method of conducting tests of significance is only valid if the null distribution is normal (or approximately normal) because parametric confidence intervals assume a normal distribution.



**Fig. 1** Steps in using a null model to test a null hypothesis. An indirect test differs from a direct test in having an additional step, specifically using a statistical test of inference to obtain a  $p$  value instead of obtaining it directly from the null distribution





**Fig. 2** In this hypothetical null model, a direct test is used to compare the observed co-occurrence of two species ( $N_{1,2}=37$ ) among  $N$  total number of sampling sites to the co-occurrence that would be expected if the two species were randomly dispersed among the sites (i.e., the null distribution). The proportion of the null distribution to the left of an observed value provides the probability that the two species would co-occur at fewer sites based solely on random dispersion whereas the proportion to the right of an observed value (gray shading) represents the probability that the two species would occur at more sites based solely on random dispersion. In this example, species 1 and 2 are significantly and positively associated ( $P_{1,2}=0.035$ ) whereas species 1 and 3 ( $N_{1,3}=17$ ) are neither positively nor negatively associated ( $P_{1,3}=0.215$ ). See Veech (2006) for more details

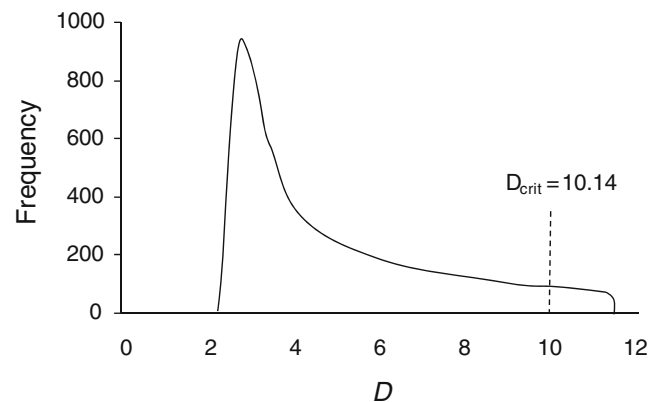
Indirect tests have also been conducted by calculating standardized effect sizes (SES) from null distributions and using the SES to infer significance (Ulrich 2004; Arrington et al. 2005; Ulrich and Gotelli 2007; Azeria et al. 2009; Englund et al. 2009), although this also requires normality of the null distribution. In this approach, the observed value of the parameter estimate (e.g.,  $C$  score in analyses of nestedness of species presence/absence matrices) is standardized on the null distribution by subtracting the expected value (mean of the null distribution) from the observed value and then dividing by the standard deviation of the null distribution. The observed value is significant (at an alpha level of 0.05) if its standardized value is greater than 1.96 (or less than  $-1.96$ ) as these correspond to the 95% confidence limits for a standard normal distribution.

Indirect tests are sometimes carried out as Kolmogorov–Smirnov tests or regressions of observed vs. expected values. K–S tests are based on a test statistic (often denoted as  $D$  and asymptotically distributed as a chi-square) that represents the difference between an observed cumulative distribution and any hypothetical cumulative distribution (e.g., normal, exponential). Practitioners that use this approach compare their observed data to the distribution produced by the null model and infer statistical significance (i.e., non-random structure) in their observed data if the K–S test statistic is significant (e.g., Schoener 1984; Buston and Cant 2006; Kohda et al. 2008). Regression has also been used to assess the relationship between observed data and expected data (produced by the null model). These researchers are typically interested in whether the null model (and the processes that it simulates) can produce the pattern in the observed data as evidenced by a large or significant coefficient of determination or positive regression coefficient (e.g., McCain 2004, 2005; Burns

2006; Jankowski and Weyhenmeyer 2006). These approaches are indirect in that the  $p$  value comes from the K–S test (or regression) instead of directly as a percentile of the null distribution of a test statistic. For example, the indirect regression-based approach of McCain (2004, 2005) for testing species diversity gradients has an alternative in the direct test of Veech (2000) and Connolly et al. (2003).

### Type I and II error

In any statistical test, type I (rejecting a true null hypothesis) and II (accepting a false null hypothesis) errors can arise when assumptions of the test are violated particularly if the violations alter the form of the test statistic's distribution. Indeed, a major assumption is typically that the test statistic ( $t$ ,  $F$ ,  $\chi^2$  among others) follows a specific distribution. Therefore, indirect tests from null models must adhere to the assumptions just as they would if the test of significance were being applied outside the context of a null model (Fig. 3). However, direct tests do not rely on any assumptions about the form of the null distribution. In direct tests, the  $p$  value is simply some percentile of the null distribution regardless of its shape. To be sure, direct tests have some rate of type I and II error. The type I error rate is inherent in the alpha or significance level chosen by the researcher, and the type II error rate is also determined by the structure of the data and the power of the test. In general, when applied to particular datasets, all significance tests inherently have type I and II error rates at some nominal



**Fig. 3** The null distribution for a hypothetical test statistic,  $D$ , simulated as a Poisson distribution,  $P(x) = [(\lambda^x/x!)e^{-\lambda} \times 10] + 2$  where  $\lambda=0.05$  and  $x=\text{rnd}[0, 1,000]$ , with 10,000 iterations. Multiplication by 10 rescales the variable, and the addition of 2 places a minimum bound. At the significance level,  $\alpha=0.05$ , the critical value of the test statistic from a direct test, is 10.14. Any observed value  $\geq 10.14$  would be significant ( $p<0.05$ ). The null distribution is clearly non-normal, so an indirect test should not be used. For example, an observed value of 10.14 represents an SES of 1.65 with  $p=0.0985$  (not  $p=0.05$ ); an indirect test based on SES would conclude no significant effect ( $p>0.05$ ) and thereby represents a type II error. Type I errors are also possible when the SES is significant while the  $p$  value from a direct test is not

level. However, use of indirect tests on non-normal null distributions can increase these rates of error compared to their rates when using direct tests (Fig. 3).

When a null distribution is normal, the  $p$  value obtained from a direct test is approximately equivalent to the  $p$  value obtained from a simple indirect test such as converting an SES value to its corresponding  $p$  value from a standard normal distribution. In a greater context, previous authors (Fisher 1935; Hoeffding 1952; Kempthorne and Doerfler 1969; Romano 1989; Efron and Tibshirani 1998; Manly 1998) have demonstrated that randomization (permutation, resampling, Monte Carlo) tests and their parametric counterparts have equivalent error rates when the randomized distribution of the test statistic is normal. Therefore, the direct and indirect tests of null models have the same rates of type I and II error when the null distribution is normal. Although many parametric tests have greater power than non-parametric *rank-based* tests (e.g., Mann–Whitney test, Kruskal–Wallis test), the direct test described here is not rank based and so does not have de facto less power than the indirect (parametric) test.

In every null model, the structure of the null distribution (i. e., its shape) is due to the observed data, randomization process, and the parameter of interest. Most null models likely produce null distributions that are normal (Fig. 2) given that the normal distribution is the outcome of most unconstrained randomization routines (Manly 1998). This is also the reason that bootstrapping (resampling with replacement) generally works for estimating non-parametric measures of spread (e. g., variance, standard error, confidence intervals) for a dataset (Efron and Tibshirani 1998). However, when the randomization process of a null model is restricted, contingent, and heavy with stipulations about how the data are randomized (e.g., Bascompte and Melian 2005; Christian et al. 2006; Wonham and Pachepsky 2006; Blanchet et al. 2010), then the resulting null distributions may be other than normal. In addition, some metrics may give non-normal distributions (Efron and Tibshirani 1998). For example, the net relatedness index and nearest taxon index of Kembel and Hubbell (2006) often produce distributions that are right skewed (Swenson et al. 2006). The convex hull volume is a multivariate metric used to examine the effects of habitat filtering on the dispersion of species traits in plant communities (Cornwell et al. 2006). For these types of metrics, it may be more appropriate to use direct tests, as in Prado and Lewinsohn (2004), Cornwell et al. (2006), and Swenson et al. (2006), given that direct tests do not require that the null distribution has any particular form.

Often times, indirect tests are not needed; the researcher can appropriately derive a  $p$  value directly from the null distribution (i.e., the direct test). However, a particularly useful application of indirect tests is in calculating standardized effect sizes. These allow comparisons among multiple tests based on different null distributions (Gotelli and McCabe 2002). This comparison is possible because the

null distributions all get converted to a standard normal distribution removing any effect of scale or unit of measurement in producing differences among the null distributions and possible differences among test outcomes. Standardized effect sizes from different null distributions are directly comparable. Multiple SES values can be averaged to test for patterns over multiple null model tests in a framework similar to meta-analysis (Gotelli and McCabe 2002; Gotelli and Rohde 2002; Nipperess and Beattie 2004; Horner-Devine et al. 2007). However, normality of the null distribution is required when calculating SES (Fig. 3). Standardization alone does not normalize a distribution; it only rescales the distribution to have mean=0 and SD=1. Therefore, although most null models likely produce null distributions that are normal, verifying the normality of a null distribution is a good idea prior to calculating standardized effect sizes.

## Literature review

I conducted a literature review to assess the frequency of use of direct and indirect tests in studies using null models. I examined 108 papers from seven ecological journals (Appendix S1). I found papers by doing a keyword search in ISI Web of Science on “null distribution” or “null model” appearing in the title or abstract of papers published in the seven journals between 2000 and 2010. I read the Methods and Results sections of each paper to identify the type of test that was used. Papers that were clearly review articles or that used null models as process-simulation models (and hence did not do significance testing) were excluded. The 108 papers were compiled from *American Naturalist*, *Ecology*, *Ecology Letters*, *Global Ecology and Biogeography*, *Journal of Animal Ecology*, *Oecologia*, and *Oikos*. This list does not include all the journals that publish null model papers; however, it does include many of the major publishers and is a representative set of ecological journals.

Direct tests were used in 56.5% of the studies, and indirect tests were used in 45.4% (4.6% of the studies used both types of test). Because most authors did not provide any description of the form of their null distributions, it was not possible to assess how often indirect tests may have been misapplied. Inappropriate use of indirect tests may not be a widespread problem if most null models tend to use unrestricted randomization algorithms and test statistics that are free to vary over a wide range of possible values. Nonetheless, researchers that use indirect tests in their null models need to be aware of their limitations.

## Conclusion

Null models have a long history of testing the predictions of ecological theory. In fact, the null model approach was born

out of the 1970s debates concerning the role of interspecific competition in determining the coexistence of species in the same ecological communities (Diamond 1975; Connor and Simberloff 1979, 1983; Diamond and Gilpin 1982; Gilpin and Diamond 1982). As recognized by these authors, the possibility that competition might prevent coexistence stems directly from one of the very first theories of ecology: the principle of competitive exclusion and limiting similarity (Gause 1934; Hutchinson 1959). Since then, null models have proven to be valuable tools.

The earliest null models all used indirect tests. For example, Simberloff (1970) was perhaps the first to generate a null model distribution by randomization; he used a *t* test to compare whether observed species/genus ratios were significantly different from the set of ratios represented by the null distribution. Many other studies (e.g., Simberloff 1978; Connor and Simberloff 1979; Strong et al. 1979; Simberloff and Boecklen 1981; all of the null models included in Strong et al. 1984) also used randomization or analytical probability-based solutions to find expected values (of a test statistic) that the observed value could be compared to through an indirect test. Tokeshi (1986) appears to be the first researcher to have used a direct test (sensu Fig. 2) and thus to have also recognized that the distributions produced by null models are essentially distributions of test statistics.

The distinction between direct and indirect tests presented in this paper will enable clearer communication and understanding of null models. In the past decade, use of null models has become much more common. Null models have become more intricate with details on the tests of significance not always explained well, particularly for indirect tests. Simply stating whether the test is direct or indirect will greatly assist others in understanding and evaluating the null model. Most importantly, researchers should consider carefully whether an indirect test is needed. In many cases, direct tests will suffice and be easier to communicate.

**Acknowledgments** I thank Ben Bolker for his very insightful comments on the topic of this manuscript. Sean Connolly, Nick Gotelli, Spyros Sfenthourakis, Werner Ulrich, Diego Vázquez, and two anonymous reviewers also provided helpful comments and suggestions on one or more previous versions of the manuscript.

## References

- Ackerly DD, Schilck DW, Webb CO (2006) Niche evolution and adaptive radiation: testing the order of trait divergence. *Ecology* 87:S50–S61. doi:10.1890/0012-9658(2006)87[50:NEAART]2.0.CO;2
- Adams DC (2007) Organization of *Plethodon* salamander communities: guild-based community assembly. *Ecology* 88:1292–1299. doi:10.1890/06-0697
- Arita HT, Rodriguez P (2004) Local-regional relationships and the geographical distribution of species. *Global Ecol Biogeogr* 13:15–21. doi:10.1111/j.1466-882X.2004.00067.x
- Arrington DA, Winemiller KO, Layman CA (2005) Community assembly at the patch scale in a species rich tropical river. *Oecologia* 144:157–167. doi:10.1007/s00442-005-0014-7
- Azeria ET, Fortin D, Lemaître J, Janssen P, Hébert C, Darveau M, Cumming SG (2009) Fine-scale structure and cross-taxon congruence of bird and beetle assemblages in an old-growth boreal forest mosaic. *Global Ecol Biogeogr* 18:333–345. doi:10.1111/j.1466-8238.2009.00454.x
- Bascompte J, Melian CJ (2005) Simple trophic modules for complex food webs. *Ecology* 86:2868–2873. doi:10.1890/05-0101
- Blanchet S, Grenouillet G, Beauchard O, Tedesco PA, Leprieux F, Durr HH, Busson F, Oberdorff T, Brosse S (2010) Non-native species disrupt the worldwide patterns of freshwater fish body size: implications for Bergmann's rule. *Ecol Lett* 13:421–431. doi:10.1111/j.1461-0248.2009.01432.x
- Bluthgen N, Frund J, Vazquez DP, Menzel F (2008) What do interaction network metrics tell us about specialization and biological traits? *Ecology* 89:3387–3399. doi:10.1890/07-2121.1
- Burns KC (2006) A simple null model predicts fruit-frugivore interactions in a temperate rainforest. *Oikos* 115:427–432. doi:10.1111/j.2006.0030-1299.15068.x
- Burns KC, Zotz G (2010) A hierarchical framework for investigating epiphyte assemblages: networks, meta-communities, and scale. *Ecology* 91:377–385. doi:10.1890/08-2004.1
- Buston PM, Cant MA (2006) A new perspective on size hierarchies in nature: patterns, causes, and consequences. *Oecologia* 149:362–372. doi:10.1007/s00442-006-0442-z
- Christian KA, Tracy CR, Tracy CR (2006) Evaluating thermoregulation in reptiles: an appropriate null model. *Am Nat* 168:421–430
- Connolly SR, Bellwood DR, Hughes TP (2003) Indo-Pacific biodiversity of coral reefs: deviations from a mid-domain model. *Ecology* 84:2178–2190. doi:10.1890/02-0254
- Connor EF, Simberloff D (1979) The assembly of species communities: chance or competition? *Ecology* 60:1132–1140. doi:10.2307/1936961
- Connor EF, Simberloff D (1983) Interspecific competition and species co-occurrence patterns on islands: null models and the evaluation of evidence. *Oikos* 41:455–465. doi:10.2307/3544105
- Cornwell WK, Schilck DW, Ackerly DD (2006) A trait-based test for habitat filtering: convex hull volume. *Ecology* 87:1465–1471. doi:10.1890/0012-9658(2006)87[1465:ATTFHF]2.0.CO;2
- Diamond JM (1975) Assembly of species communities. In: Cody ML, Diamond JM (eds) *Ecology and evolution of communities*. Harvard University Press, Cambridge
- Diamond JM, Gilpin ME (1982) Examination of the “null” model of Connor and Simberloff for species co-occurrences on islands. *Oecologia* 52:64–74. doi:10.1007/BF00349013
- Efron B, Tibshirani R (1998) *An introduction to the bootstrap*. CRC Press, New York
- Englund G, Johansson F, Olofsson P, Salonsaari J, Ohman J (2009) Predation leads to assembly rules in fragmented fish communities. *Ecol Lett* 12:663–671. doi:10.1111/j.1461-0248.2009.01322.x
- Farias AA, Jaksic FM (2009) Hierarchical determinants of the functional richness, evenness and divergence of a vertebrate predator assemblage. *Oikos* 118:591–603. doi:10.1111/j.1600-0706.2008.16859.x
- Fisher RA (1935) *The design of experiments*. Oliver and Boyd, Edinburgh
- Frank van Veen FJ, Murrell DJ (2005) A simple explanation for universal scaling relations in food webs. *Ecology* 86:3258–3263. doi:10.1890/05-0943
- Fridley JD, Brown RL, Bruno JF (2004) Null models of exotic invasion and scale-dependent patterns of native and exotic species richness. *Ecology* 85:3215–3222. doi:10.1890/03-0676



- Gause GF (1934) The struggle for existence. Williams and Wilkins, Baltimore
- Gilpin ME, Diamond JM (1982) Factors contributing to non-randomness in species co-occurrences on islands. *Oecologia* 52:75–84. doi:10.1007/BF00349014
- Gotelli NJ, Graves GR (1996) Null models in ecology. Smithsonian Institution Press, Washington, D. C
- Gotelli NJ, McCabe DJ (2002) Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. *Ecology* 83:2091–2096. doi:10.1890/0012-9658(2002)083[2091:SCOAMAJ]2.0.CO;2
- Gotelli NJ, McGill BJ (2006) Null versus neutral models: what's the difference? *Ecography* 29:793–800. doi:10.1111/j.2006.0906-7590.04714.x
- Gotelli NJ, Rohde K (2002) Co-occurrence of ectoparasites of marine fishes: a null model analysis. *Ecol Lett* 5:86–94. doi:10.1046/j.1461-0248.2002.00288.x
- Helmus MR, Savage K, Diebel MW, Maxted JT, Ives AR (2007) Separating the determinants of phylogenetic community structure. *Ecol Lett* 10:917–925. doi:10.1111/j.1461-0248.2007.01083.x
- Hoeffding W (1952) The large-sample power of tests based on permutations of observations. *Ann Math Stat* 23:169–192. doi:10.1214/aoms/117729436
- Horner-Devine MC, Silver JM, Leibold MA, Bohannon BJM, Colwell RK, Fuhrman JA, Green JL, Kuske CR, Martiny JBH, Muyzer G, Ovres L, Reysenbach A, Smith VH (2007) A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* 88:1345–1353. doi:10.1890/06-0286
- Hulme PE (2008) Contrasting alien and native plant species-area relationships: the importance of spatial grain and extent. *Global Ecol Biogeogr* 17:641–647. doi:10.1111/j.1466-8238.2008.00404.x
- Hutchinson GE (1959) Homage to Santa Rosalia, or why are there so many kinds of animals? *Am Nat* 93:145–159
- Ingram T, Shurin JB (2009) Trait-based assembly and phylogenetic structure in northeast Pacific rockfish assemblages. *Ecology* 90:2444–2453. doi:10.1890/08-1841.1
- Jankowski T, Weyhenmeyer GA (2006) The role of spatial scale and area in determining richness-altitude gradients in Swedish lake phytoplankton communities. *Oikos* 115:433–442. doi:10.1111/j.2006.0030-1299.15295.x
- Kembel SW (2009) Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecol Lett* 12:949–960. doi:10.1111/j.1461-0248.2009.01354.x
- Kembel SW, Hubbell SP (2006) The phylogenetic structure of a neotropical forest tree community. *Ecology* 87:S86–S99. doi:10.1890/0012-9658(2006)87[86:TPSOAN]2.0.CO;2
- Kempthorne O, Doerfler TE (1969) The behavior of some significance tests under experimental randomization. *Biometrika* 56:231–248. doi:10.1093/biomet/56.2.231
- Kohda M, Shibata J, Awata S, Gomagano D, Takeyama T, Hori M, Heg D (2008) Niche differentiation depends on body size in a cichlid fish: a model system of a community structured according to size regularities. *J Anim Ecol* 77:859–868. doi:10.1111/j.1365-2656.2008.01414.x
- Manly BFJ (1998) Randomization, bootstrap and Monte Carlo methods in biology. Chapman & Hall, New York
- Manzaneda AJ, Rey PJ (2009) Assessing ecological specialization of an ant-seed dispersal mutualism through a wide geographic range. *Ecology* 90:3009–3022. doi:10.1890/08-2274.1
- McCain CM (2004) The mid-domain effect applied to elevational gradients: species richness of small mammals in Costa Rica. *J Biogeogr* 3:19–31. doi:10.1046/j.0305-0270.2003.00992.x
- McCain CM (2005) Elevational gradients in diversity of small mammals. *Ecology* 86:366–372. doi:10.1890/03-3147
- Moore JE, Swihart RK (2007) Toward ecologically explicit null models of nestedness. *Oecologia* 152:763–777. doi:10.1007/s00442-007-0696-0
- Mouillot D, Krasnov BR, Poulin R (2008) High intervality explained by phylogenetic constraints in host-parasite webs. *Ecology* 89:2043–2051. doi:10.1890/07-1241.1
- Nipperess DA, Beattie AJ (2004) Morphological dispersion of Rhytidoponera assemblages: the importance of spatial scale and null model. *Ecology* 85:2728–2736. doi:10.1890/03-0741
- Okuda T, Noda T, Yamamoto T, Hori M, Nakaoka M (2009) Latitudinal gradients in species richness in assemblages of sessile animals in rocky intertidal zone: mechanisms determining scale-dependent variability. *J Anim Ecol* 78:328–337. doi:10.1111/j.1365-2656.2008.01495.x
- Prado PI, Lewinsohn TM (2004) Compartments in insect-plant associations and their consequences for community structure. *J Anim Ecol* 73:1168–1178. doi:10.1111/j.0021-8790.2004.00891.x
- Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. Cambridge University Press, NY
- Romano JP (1989) Bootstrap and randomization tests of some non-parametric hypotheses. *Ann Stat* 17:141–159. doi:10.1214/aos/1176347007
- Schoener TW (1984) Size differences among sympatric bird-eating hawks: a worldwide survey. In: Strong DR, Simberloff D, Abele LG, Thistle AB (eds) Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, pp 254–281
- Simberloff D (1978) Using island biogeographic distributions to determine if colonization is stochastic. *Am Nat* 112:713–726.
- Simberloff DS (1970) Taxonomic diversity of island biotas. *Evolution* 24:23–47. doi:10.2307/2406712
- Simberloff D, Boecklen W (1981) Santa Rosalia reconsidered: size ratios and competition. *Evolution* 35:1206–1228. doi:10.2307/2408133
- Strong DR, Szyska LA, Simberloff DS (1979) Test of community-wide character displacement against null hypotheses. *Evolution* 33:897–913. doi:10.2307/2407653
- Strong DR, Simberloff D, Abele LG, Thistle AB (1984) Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton
- Swenson NG, Enquist BJ, Pither J, Thompson J, Zimmerman JK (2006) The problem and promise of scale dependency in community phylogenetics. *Ecology* 87:2418–2424. doi:10.1890/0012-9658(2006)87[2418:TPAPOS]2.0.CO;2
- Thomber CS, Gaines SD (2004) Population demographics in species with biphasic life cycles. *Ecology* 85:1661–1674. doi:10.1890/02-4101
- Tokeshi M (1986) Resource utilization, overlap, and temporal community dynamics: a null model analysis of an epiphytic chironomid community. *J Anim Ecol* 55:491–506. doi:10.2307/4733
- Ulrich W (2004) Species co-occurrences and neutral models: reassessing J. M. Diamond's assembly rules. *Oikos* 107:603–609. doi:10.1111/j.0030-1299.2004.12981.x
- Ulrich W, Gotelli NJ (2007) Null model analysis of species nestedness patterns. *Ecology* 88:1824–1831. doi:10.1890/06-1208.1
- Veech JA (2000) A null model for detecting nonrandom patterns of species richness along spatial gradients. *Ecology* 81:1143–1149. doi:10.1890/0012-9658(2000)081[1143:ANMFDN]2.0.CO;2
- Veech JA (2006) A probability-based analysis of temporal and spatial co-occurrence in grassland birds. *J Biogeogr* 33:2145–2153. doi:10.1111/j.1365-2699.2006.01571.x
- Wilson JB (1995) Null models for assembly rules: the Jack Horner effect is more insidious than the Narcissus effect. *Oikos* 72:139–144. doi:10.2307/3546047
- Wonham MJ, Pachepsky E (2006) A null model of temporal trends in biological invasion records. *Ecol Lett* 9:663–672. doi:10.1111/j.1461-0248.2006.00913.x
- Zimmermann Y, Ramirez SR, Eltz T (2009) Chemical niche differentiation among sympatric species of orchid bees. *Ecology* 90:2994–3008. doi:10.1890/08-1858.1