sciendo

DOI: 10.2478/bile-2024-0010

Biometrical Letters
Vol. 61 (2024), No. 2, 147-159

A review of null models in community ecology: a different robust viewpoint for understanding statistical community ecology

Patricio R. De los Ríos-Escalante[1,2,3], Farhana S. Ghory[4]

[1]Departamento de Ciencias Biológicas y Químicas, Facultad de Recursos Naturales, Universidad Católica de Temuco, Casilla 15-D, Temuco, Chile, e-mail: prios@uct.cl
[2]Núcleo de Estudios Ambientales UC, Temuco, Chile
[3]MSc student in Statistics (Magíster en Estadística), Instituto de Estadística, Facultad de Ciencias, Pontificia Universidad Católica de Valparaíso, Av. Brasil 2830, Valparaíso, Chile
[4]Marine Reference Collection & Collection Center, University of Karachi, Karachi – 75270, Pakistan

SUMMARY

One classic definition of ecology is that it is the science that studies the distribution and abundance of biotic components and their relationship with abiotic components. The use of statistical tools is very important for understanding ecology, especially the distribution and abundance of biotic components. The classic statistical viewpoint was that an ecological community (an interaction of different species in defined time and space) has a determined structure due to biotic and abiotic interactions. Nevertheless, this classic viewpoint has the risk of proneness to type I errors or "false positives". In this situation, null models were proposed that have the premise that community ecology is random, meaning the absence of structure, and the null hypothesis for these models is the absence of regular structured patterns. The present study is a review of null models and their application to aquatic environments. These null models include three main models: for species co-occurrence, asserting that species associations are random; for size overlap, asserting that the size structure of species in the community is random, as a strategy for use of ecological niche; and for niche overlap, asserting that species in a community can share a defined ecological niche with consequent interspecific competition.

Key words: null models, community ecology, error type I, null hypothesis, random.

1. Introduction

One of the main definitions of ecology is that it is the science that studies the distribution and abundance of species and their interaction with abiotic components as determined in time and space (Gotelli, 2008). The use of statistical

tools for understanding ecological processes and properties is very important (Gotelli, 2008; Gotelli & Ellison, 2012).

There are three scales of organization in ecology: population, namely a group of individuals of the same species that interact in defined time and space; community, namely a group of populations that interact in determined time and space; and finally the ecosystem, a group of communities that interact with biotic components in determined time and space (Gotelli, 2008). In the study of community ecology, the classic viewpoint is that the structure of communities is regulated by determined biotic and/or abiotic parameters (Gotelli & Ellison, 2012), and in consequence the null hypothesis for these studies is related to the presence of defined and restricted effects of biotic and/or abiotic parameters on community structures (Strong, 1980; Gotelli & Graves, 1996; Gotelli & Ellison, 2012).

In the last three decades a new viewpoint has been proposed in community statistical ecology, based on null models, with the main assumption as null hypothesis that "community structure has no regulator factors", or that "community structure is random" (Gotelli & Graves, 1996; Gotelli, 2000, 2001). This proposed hypothesis is based on two viewpoints in the philosophy of science: the proposal of Popper (1968) about hypothesis contrasting, and the proposal of Kühn (1962) that described a paradigm which in this situation entails that interspecific competition is an important component of null models, because when a community structure is random there is no interspecific competition; also these models are applied only to field data observations and not to experimental data (Gotelli & Graves, 1996; Gotelli, 2000, 2001; Tondoh, 2006; Tiho & Johens, 2007). Also, it is necessary to remark that null models are different from neutral models in term of semantics, because the proposal of neutral models in community ecology implies that birth, death and migration rates have similar probability (Gotelli & Rhode, 2002).

The null models are based on Monte Carlo randomizations or combinations of species reported in a community with a minimum of 1000 combinations (Gotelli, 2000, 2001), based on species presence–absence matrices, size average

of each species in a community, or species abundances in a community (Gotelli & Graves, 1996; Gotelli & Entsminser, 2015; Gotelli et al., 2015). Such models are not prone to type I errors, meaning "false positives", and are more robust than those based on a traditional statistical ecology viewpoint (Gotelli & Graves, 1996; Gotelli, 2000, 2001; Tondoh, 2006; Tiho & Johens, 2007). The first software used for null models was Ecosim, which was originally created as freely available, but in recent years has changed to private license (Gotelli & Entsminser, 2015); nevertheless, for R programming (R Development Core Team, 2023) the library EcosimR has been developed, which uses the R language for calculating null models (Gotelli et al., 2015). The aim of the present study is to review null models and their descriptions, based on aquatic community data and using the R package EcosimR (Gotelli et al., 2015).

2. Species co-occurrence null models

2.1. Theoretical basis

This kind of null model is based on a species presence–absence matrix, with sites in columns and species in rows; when a species is present the corresponding field is filled with one, and when the species is absent it is filled with zero (Gotelli, 2000, 2001). The bases of these models are the proposals of checkerboard, checkerboard C-score, variance ratio, and species combination number (Gotelli, 2000). The literature on null models reveals that checkerboard C-score is the main algorithm for species co-occurrence null models (Gotelli, 2000; Tondoh, 2006; Tiho & Johens, 2007). The C-score has this equation (Gotelli, 2000):

$$CU = (R_i - S)(R_j - S)$$

where R_i and R_j are the row totals for species i and j, and S is the number of sites for both species (Gotelli, 2000; Tiho & Johens, 2007).

In accordance with Gotelli (2000) it is possible to construct nine algorithm models based on Monte Carlo randomization, the possibilities being that rows (species) and/or columns (sites) can be fixed, meaning that data are preserved, or

that rows and/or columns can be equiprobable, so that all data are suitable, or else that the data are proportional to the sum for each row or column. Given these options, the most robust results are provided by the following combinations (Gotelli, 2000; Tiho & Johens, 2007):

Fixed-fixed: this null model implies that the row and the column sums of the original matrix are preserved, meaning that each random community contains the same number of species as the original community (fixed column), and each species occurs with the same frequency as in the original community (fixed row). According to Gotelli (2000) this model is not prone to type I errors (false rejection of the null hypothesis) and it also has good power for detecting non-randomness.

Fixed-equiprobable: this algorithm means that only the row sums are fixed, whereas the columns are treated as equiprobable. This null model treats all the samples (columns) as equally suitable for all species.

Fixed-proportional: this algorithm means that the species occurrence totals are preserved as in the original community, and the probability that a species occurs in a sample (= column) is proportional to the column total for that sample.

An example of the application of null models to real data (cf. De los Ríos-Escalante et al., 2023a) concerns the El Lenguado beach site (23°46'09.5"S, 70°28'22.2"W) visited on December 28, 2019. Random quadrants (10 × 10 cm) were established in the studied sites (N = 90 for each site) during low tide. Five species were reported: *Allopetrolisthes punctatus* (3.82 ± 3.67), *Petrolisthes granulosus* (0.61 ± 1.22 ind), *P. tuberculosus* (0.28 ± 0.64 ind), *Betaeus truncatus* (0.07 ± 0.31 ind) and *Homolaspis plana* (0.03 ± 0.18); the original data are given in supplementary material S1 and the codes in supplementary material S2. Results are presented in Table 1 and Fig. 1.

2.2. Hypothesis proposal

Null hypothesis: the decapod species associations are not structured (or are random).

Alternate hypothesis: the decapod species associations are structured (or are not random).

Table 1: Results of null model analysis of species co-occurrence for decapods collected at the studied site (P-values lower than 0.05 denote structured patterns in species associations)

Model	Observed index	Mean index	Standard effect size	Variance	P
Fixed-fixed	263.5	234.95	6.338	20.287	0.001
Fixed-proportional	263.5	143.23	4.962	587.610	0.001
Fixed-equiprobable	263.5	125.83	8.410	268.020	0.001

Note: the observed index is the result of all simulations according to the model; the standard effect size is the correction of calculations based on sample size.

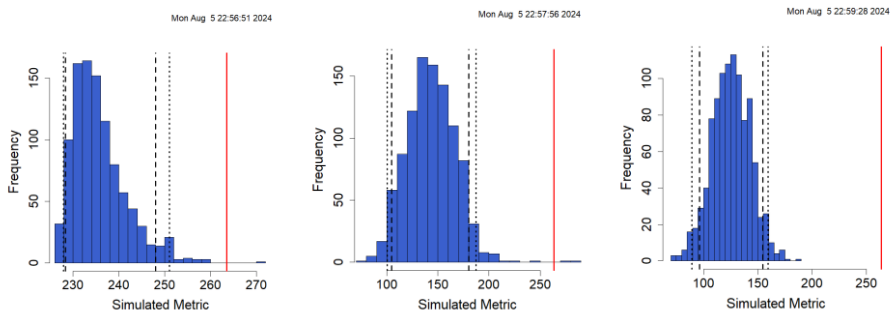

Figure 1: Graphs of species co-occurrence null models (fixed-fixed: left; fixed-proportional: center; fixed-equiprobable: right), for decapod data (cf. De los Ríos-Escalante et al., 2023); red line corresponds to observed mean

The results presented show that species associations are structured (or not random) for all models; then the null hypothesis is accepted.

3. Niche overlap null models

3.1. Theoretical basis

For niche overlap analysis, an individual matrix is built in which rows and columns represent species and sites, respectively, and this is used to test whether the niche overlap significantly differs from the corresponding value under the null hypothesis (random assemblage). These analyses were applied to data from the

second field collection, and were based on the Pianka and Czekanowski indices. The equation for the Pianka index is

$$O_{kl} = \frac{\sum_i^n p_{il} p_{ik}}{\sqrt{\sum_i^n p_{il}^2 \sum_i^n p_{ij}^2}}$$

where p_{il} and p_{ij} are the sites occupied by species l and j.

The Czekanowski index has the following equation:

$$O_{il} = O_{ij} = 1.0 - 0.5 \sum_{1=1}^{n} |p_{il} - p_{ij}|$$

where p_{il} and p_{ij} are the sites occupied by species l and j.

Both equations are used for the estimation of niche sharing based on specific algorithms, and both models reveal the probability of niche overlap compared with the niche overlap of the theoretically simulated community (Gotelli et al., 2015). The niche amplitude can be retained or reshuffled. The specialization of each species is preserved when the niche amplitude is retained. By contrast, the niche amplitude algorithm uses a wide utilization gradient of specialization when it is reshuffled. Furthermore, zero participation in the observed matrix can be maintained or omitted. We used the RA3 algorithm, which retains the amplitude and reshuffles the zero conditions (Gotelli et al., 2015; Carvajal-Quintero et al., 2015).

An example is presented using the same data as the previous example (the intertidal decapod community at a rocky shore in the south of Antofagasta, Chile; cf. De los Ríos-Escalante et al., 2023a). The R code is provided in supplementary material S2. Results are presented in Table 2 and Fig. 2.

3.2. Hypothesis proposal

Null hypothesis: the decapod species overlap in ecological niche (or it is random).

Alternate hypothesis: the decapod species do not overlap in ecological niche (or it is not random).

Table 2: Results of null model analysis of species niche overlap for decapods collected at the studied site (P-values lower than 0.05 denote niche overlap)

Model	Observed index	Mean index	Standard effect size	Variance	P
Pianka index	0.009	0.194	-3.795	0.002	0.995
Czekanowski index	0.130	0.142	-0.434	0.001	0.656

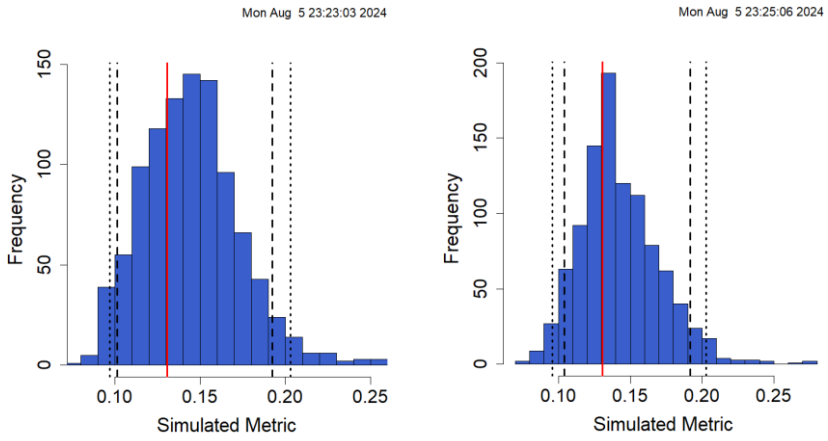

Figure 2: Results of Pianka (left) and Czekanowski index (right) for decapod data (cf. De los Ríos-Escalante et al., 2023); red line corresponds to observed mean

The results reveal that the species do not overlap in ecological niche, and in consequence there is no interspecific competition, and the community is not random. Thus there is a structured pattern, because the analyzed species have specific ecological niches.

4. Size ratio null model

4.1. Theoretical basis

The null model analysis was applied to investigate “size structure” for the total of specimens collected. The specific aim was to determine any non-random patterns in the size overlap of the various species, and the analysis is based on the

assumption that species with similar sizes do not coexist because they share many resources (Gotelli & Entsminger, 2015). The data for such an analysis consist of a matrix in which each species is a row, and each site is a column (Gotelli & Graves, 1996; Gotelli & Entsminger, 2015). Entries in the matrix represent the mean size, i.e. the mean length, of each species. The original matrix is then reshuffled to produce random patterns that would be expected in the absence of competitive interactions. The following algorithms were used for that purpose: size uniform, size uniform user, and size uniform pool, combined with metrics of minimum difference and variance ratio (Gotelli & Entsminger, 2015; Gotelli et al., 2015), thus giving a total of six different simulations.

The site was El Quisco beach, 32°24’00’’ S 71°42’00’’ W (cf. Andrade et al., 2023) at low tide. Five species were reported: *Homolaspis plana* (24.0 ± 5.3 mm), *Cyclograpsus cinereus* (8.8 ± 1.8 mm), *Gaudichaudia gaudichaudi* (13.6 ± 5.4 mm), *Petrolisthes granulosus* (10.2 ± 2.0 mm) and *Betaeus truncatus* (29.9 ± 4.6 mm). The original data are given in supplementary material S2, and the codes in supplementary material S3. The results are presented in Table 3 and Fig. 3.

Table 3: Results of null model analysis of size overlap for decapods collected at the studied site (P-values greater than 0.05 denote size overlap)

Algorithm	Metric	Mean index	Observed index	Variance	Standard effect size	P
Size uniform	Min. difference	1.264	1.400	1.055	0.131	0.380
Size uniform user	Min. difference	1.063	1.400	0.737	0.391	0.300
Size source pool	Min. difference	1.677	1.400	1.852	-0.203	0.480
Size uniform	Variance ratio	0.193	0.072	0.036	-0.636	0.755
Size uniform user	Variance ratio	0.093	0.072	0.043	0.015	0.389
Size source pool	Variance ratio	0.152	0.072	0.046	-0.371	0.510

4.2. Hypothesis proposal

Null hypothesis: the decapod species have similar size ratio (or it is random).

Alternate hypothesis: the decapod species do not have similar size ratio (or it is not random).

The results revealed the absence of size overlap; thus the null hypothesis is rejected, and the size ratio of species is not random.

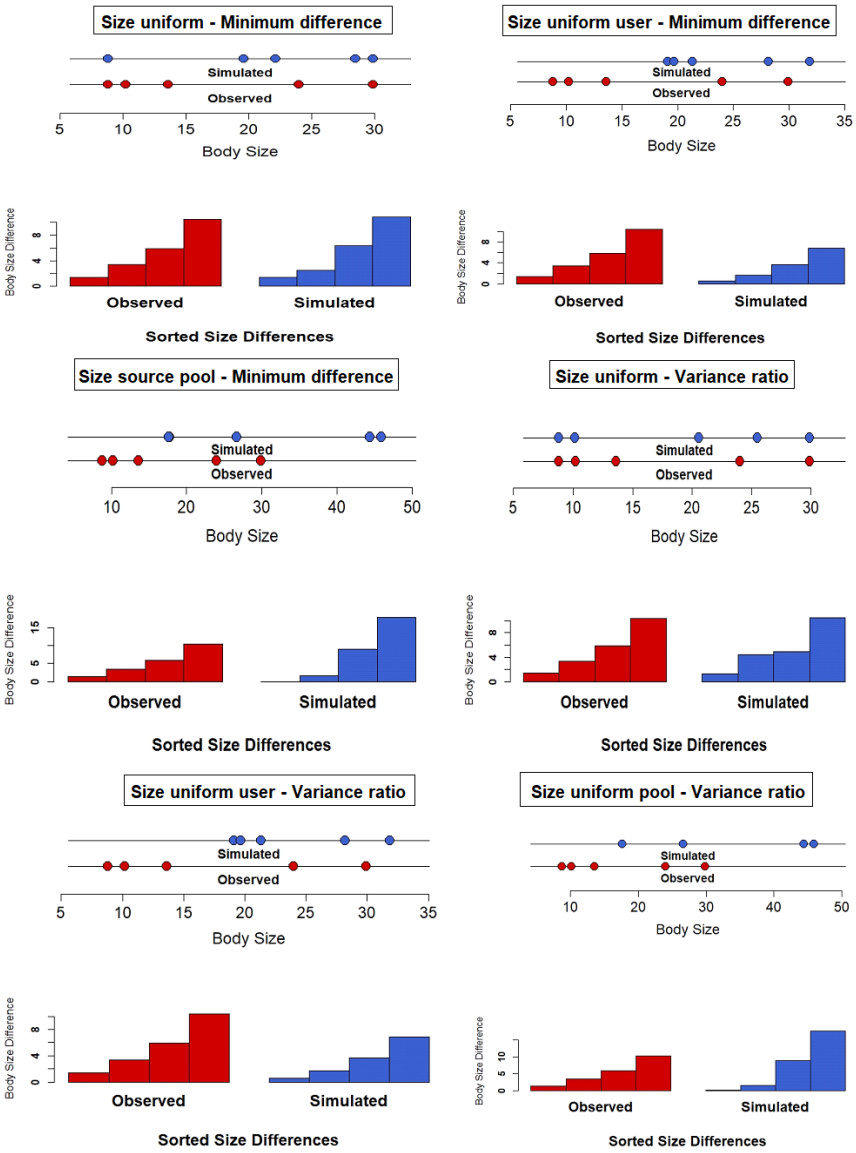

Figure 3: Graphs of results of size overlap null model analysis for decapods collected at the studied site

5. Discussion

Prior work has revealed that null models may be applicable to any kind of aquatic environment, such as inland water communities in southern Chilean rivers in rainy forests (Solís-Lufí et al., 2022; Figueroa & De los Ríos-Escalante, 2023) or semi-arid rivers in Algeria (Gharbi et al., 2023). The results for intertidal invertebrates in the north of Chile are different in the case of species co-occurrence, because this is random due to the presence of many repeated species in many sites, whereas in the case of niche overlap similar results are obtained (De los Ríos-Escalante et al., 2023b), although if we consider only decapods based on observations for the north of Chile, the species associations are random, and there is no overlap of ecological niche, then there is random absence (De los Ríos-Escalante et al., 2023a). Similar results for size ratio were obtained for decapods along the intertidal coast in the north of Chile (De los Ríos-Escalante et al., 2020). Nevertheless, the available literature on null models in Chilean inland waters reveals that species associations are random, and niche overlap is variable: it can be present or absent (De los Ríos-Escalante et al., 2020; Figueroa & De los Ríos-Escalante, 2023), probably because the species diversity in Chilean inland waters is low and there is much repetition between sites in comparison to marine environments. For Algerian rivers, similar results were reported concerning random presence in the case of species co-occurrence and absence of ecological niche overlap, with consequent absence of interspecific competition and random presence (Gharbi et al., 2024).

Null models only provide results as to whether the studied data are random or structured, without more specification of potential causes if there is structured presence (Gotelli & Graves, 1996; Gotelli, 2000, 2001; Gotelli et al., 2015; Vaughan et al., 2017; Kohli et al., 2018; Supriya et al., 2020; Semler et al., 2022); this can be a disadvantage if the potential causes of the results of null models can be explained in more detail (Blanchet et al., 2020). In spite of the disadvantages described by Blanchet et al. (2020), concerning the fact that null models do not consider the effects of abiotic factors on species regulation, the recent literature on null models confirms that these models are not prone to type I errors, meaning

that null models are more robust (Vaughan et al., 2017; Barger 2023). On this basis, it is suggested that if null models are used, it will be necessary to manage specifically biotic and abiotic ecological parameters for the studied sites if the results of null models imply random presence, and to complement them with exploratory multivariate data analysis that can provide a potential explanation of random presence according to the null model, or explain the causes of random absence in a zooplankton community such as was reported for lakes with different trophic status and water quality (De los Ríos-Escalante et al., 2020). Also, the results of null models can be influenced by sample size, in term of the species reported that can be associated with a surface sample or a number of samples (Wang et al., 2020); in this context, it would probably be necessary to apply null models with different sizes of ecological samples at the spatial scale to determine the accuracy of the models.

In conclusion, the present study suggests that null models are robust as a tool for the statistical study of community ecology, due to the fact that they are not prone to type I errors (false rejection of the null hypothesis), but in situations when null models reveal random presence due to a low number of species and many repeated species, for example, it is suggested to complement them with exploratory multivariate statistical analysis.

Acknowledgments

The study was financed by the project MECESUP UCT 0804. The authors express their gratitude to M.I. and S.M.A., for improving the manuscript. We acknowledge the valuable contribution and inspiration of Prof. Marco Antonio Retamal (Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Chile) (1941–2019).

REFERENCES

Andrade D., Ventura M.J., Stella C., De los Ríos-Escalante P. (2022): Size overlap in intertidal decapod communities on a central Chilean rocky beach (El Quisco, 32°24’S, Valparaíso region, Chile). Crustaceana, 95(4): 489–495. https://doi.org/10.1163/15685403-bja10203

Barger M.A. (2023): Comparison of methods to detect interspecific competition among parasites in depauperate communities. Journal of Helminthology, 97: e105. doi:10.1017/S0022149X23000901

Blanchet F.G., Cazelles K., Gravel D., (2020): Co-occurrence is not evidence of ecological interactions. Ecology Letters, 23(7): 1050–063. https://doi.org/10.1111/ele.13525

Carvajal-Quintero J.D., Escobar F., Alvarado F., Villa-Navarro F.A., Jaramillo-Villa U., Maldonado-Ocampo J.A. (2015): Variation in freshwater fish assemblages along a regional elevation gradient in the northern Andes, Colombia. Ecology and Evolution, 2(13): 2608-2620. https://doi.org/10.1002/ece3.1539

De los Ríos-Escalante P., Figueroa-Muñoz G., Retamal M.A., Vega-Aguayo R., Esse C. (2020): Size overlap in intertidal decapod communities along the Chilean coast. Scientia Marina, 84: 151–154. DOI: https://doi.org/10.3989/scimar.04973.14A

De los Ríos-Escalante P., Ibañez-Arancibia E., Esse C., Zúñiga O., Fajardo M., Ghory F.S. (2023a): Null models to explain highly diverse community of intertidal decapods on a temperate rocky coast in Antofagasta, northern Chile (23º27′S). Journal of Crustacean Biology, 43(1), 1–6. https://doi.org/10.1093/jcbiol/ruad048

De los Ríos-Escalante P., Esse C., Stella C., Adikesavan P., Zúñiga O. (2023b): Null and neutral models for explaining spatial variations in intertidal invertebrate species diversity on rocky shores in Antofagasta town (23°S, Chile). Animal Biology 73(1): 113–125. https://doi.org/10.1163/15707563-bja10099

Figueroa D., De los Ríos-Escalante P. (2023): Null models for explaining macroinvertebrate communities in northern Patagonian Cautín river (39°S, Araucania region, Chile). Spixiana, 45: 169-176.

Gharbi M., Ziane N., Baaloudj A., De los Ríos-Escalante P., Esse C., Chaib S., Houhamdi M. (2024): Distribution of benthic macroinvertebrate communities in different kind of inland water bodies in northeastern Algeria. Brazilian Journal of Biology, 84: e273662. https://doi.org/10.1590/1519-6984.273662

Gotelli N.J., Graves G.R. (1996): Null models in ecology. Smithsonian Institution Press, Washington, D.C. 1-388 pp.

Gotelli N.J. (2000): Null model analysis of species co-occurrence patterns. Ecology. 81(9): 2606-2621. https://doi.org/10.1890/0012-9658(2000)081[2606:NMAOSC]2.0.CO;2

Gotelli N.J. (2001): Research frontiers in null model analysis. Global Ecology and Biogeography, 10(9): 337–343. https://doi.org/10.1046/j.1466-822X.2001.00249.x

Gotelli, N.J. (2008): A primer of ecology. 4th Edition. Oxford University Press, 290 p.

Gotelli, N.J., Ellison A.M. (2012): A Primer of Ecological Statistics. 2nd Edition. Oxford University Press, 640 p.

Gotelli, Entsminser (2015): EcoSim: Null models software for ecology. DOI:10.5281/zenodo.16504

Gotelli N.J., Hart E., Ellison A. (2015): EcoSimR: null model analysis for ecological data. https://github.com/GotelliLab/EcoSimR/issues

Kohli B.A., Terry R.C., Rowe R.J. (2018): A trait-based framework for discerning drivers of species co-occurrence across heterogeneous landscapes. Ecography, 41(129): 1921–1933. DOI:10.1111/ecog.03747

Kuhn T.S. (1962): The Structure of Scientific Revolutions, 2nd ed., Univ. Chicago Press. 222 p.

Popper K.R. (1968): The Logic of Scientific Discovery, Hutchinson, London. 545 p.

R Development Core Team, (2023): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org

Semler R.F., Sanders N.J., CaraDana P.J., Baird A.H., Jing X., Robinson J.P.W., Graham N.A.J., Keith S.A., (2022): Reef fishes weaken dietary preferences after coral mortality, altering resource overlap. Journal of Animal Ecology, 91(10): 2125–2134. https://doi.org/10.1111/1365-2656.13796

Solis-Lufi K., Suazo M.J., Avila-Salem M.E., Maldonado-Murúa C., Aponte H., Farias J., De los Ríos-Escalante P. (2022): Community structure of benthic invertebrates in the Allipen river basin, North Patagonia, Araucania region (39°S, Chile). Brazilian Journal of Biology, 82: e232805. https://doi.org/10.1590/1519-6984.232805

Strong D.R. (1980): Null hypothesis in ecology. Synthese, 43: 271-285. https://doi.org/10.1007/BF00413928

Supriya K., Price T.D., Moreau C.S. (2020): Competition with insectivorous ants as a contributor to low songbird diversity at low elevations in the eastern Himalaya. Ecology and Evolution, 10(10): 4280–4290. DOI: 10.1002/ece3.6196

Tiho S., Johens G. (2007): Co-occurrence of earthworms in urban surroundings: a null model analysis of community structure. European Journal of Soil Biology, 43: 84-90. https://doi.org/10.1016/j.ejsobi.2006.10.004

Tondoh J.E. (2006): Seasonal changes in earthworm diversity and community structure in Central Côte d'Ivoire. European Journal of Soil Biology, 42: supplement, 1, pp. s334-s340.979. https://doi.org/10.1016/j.ejsobi.2006.09.003

Vaughan I.P., Gotelli N.J., Memmott J., Pearson C.E., Woodward G., Symodson W.O.C. (2018): econullnetr: An r package using null models to analyse the structure of ecological networks and identify resource selection. Methods in Ecology and Evolution, 9(3): 728–733. DOI: 10.1111/2041-210X.12907

Wang J., Zhang C., Xue Y., Xu B., Chen Y., Ren Y. (2020): Effect of sampling intensity on understanding species co-occurrence pattern of fish community using null model analysis. Ecological Indicators, 119: 106814. https://doi.org/10.1016/j.ecolind. 2020.106814.