




Species richness correlates of raw and standardized co-occurrence metrics

Werner Ulrich¹  | Yasuhiro Kubota^{2,3}  | Buntarou Kusumoto^{2,4} |
Andres Baselga⁵  | Hanna Tuomisto⁶ | Nicholas J. Gotelli⁷

¹Chair of Ecology and Biogeography, Nicolaus Copernicus University in Toruń, Toruń, Poland

²Faculty of Science, University of the Ryukyus, Nishihara, Okinawa, Japan

³Marine and Terrestrial Field Ecology, Tropical Biosphere Research Center University of the Ryukyus, Nishihara, Okinawa, Japan

⁴Center for Strategic Research Project, University of the Ryukyus, Nishihara, Okinawa, Japan

⁵Facultad de Biología, Universidad de Santiago de Compostela, Santiago de Compostela, La Coruña, Spain

⁶Department of Biology, University of Turku, Turku, Finland

⁷Department of Biology, University of Vermont, Burlington, Vermont

Correspondence

Werner Ulrich, Chair of Ecology and Biogeography, Nicolaus Copernicus University in Toruń, Lwowska 1, 87-100 Toruń, Poland.
Email: ulrichw@umk.pl

Funding information

Polish National Science Centre, Grant No.: 2014/13/B/NZ8/04681; European Regional Development Fund (ERDF), Grant No.: CGL2016-76637-P

Editor: Pedro Peres-Neto

Abstract

Measuring β -diversity and changes in species composition across multiple sites and environments is a major research focus in macroecology, and a variety of metrics have been proposed to quantify species co-occurrence patterns in a species \times site occurrence matrix. However, indices of β -diversity and species co-occurrence are often statistically dependent on the number of species in an assemblage. We compared the results of several common co-occurrence metrics with patterns generated by a spatially explicit neutral model simulation. We found that all measures of co-occurrence and β -diversity, whether raw, rescaled or standardized by a null model expectation, were highly correlated with the total species richness of the landscape. The one important exception were the effect sizes of the fixed–fixed null model algorithm, which preserves row and column sums of the original matrix during matrix randomization. Our results call for a careful interpretation of meta-analyses of assemblages that differ widely in species richness. At a minimum, observed species richness should be used as a statistical covariate in regression analyses, and results of the fixed–fixed algorithm should be compared carefully with the results of other randomization tests.

KEYWORDS

community composition, co-occurrences, diversity, metacommunity, null models, statistical inference

Many authors have tried to infer associations between variation in species composition and habitat variability (Guo et al., 2016), latitude (Qian, Chen, Mao, & Ouyang, 2013; Xu, Chen, Liu, & Ma, 2015), climate (Keil et al., 2012) and other abiotic factors (Alahuhta et al., 2017; Tuomisto, Ruokolainen, & Yli-Halla, 2003). The standard tool in such inference is a regression analysis, in which environmental variables serve as predictors to explain the variability in the pattern of species co-occurrence. Three common indices are proportional species turnover ($\beta_p = 1 - \frac{\alpha}{\gamma}$), where α is the average local richness and γ is the richness pooled over the same local sites (Tuomisto, 2010), the C-score, a

normalized count of all 2×2 checkerboard submatrices within a species \times sites presence–absence matrix (Stone & Roberts, 1990), and the β_{sim} measure, which represents the degree of species replacement among sites in a way that is independent of the richness difference between the sites (Simpson, 1943, 1960). The popularity of the last of these measures has increased since Baselga (2010) partitioned the classical Sørensen index of compositional similarity such that β_{sim} was one of the two components. Finally, we included the NODF (nestedness by overlap and decreasing fill) metric of nestedness (Almeida-Neto, Guimarães, Guimarães, Loyola, & Ulrich, 2008), the ordered loss of

species among sites, a pattern widely believed to be opposed to species turnover (but see Ulrich, Kryszeński et al., 2017).

Given that these indices (and many others) are normalized for differences in local species richness, many authors have used them directly in macroecological analyses. However, constraints on the covariation of α and γ possibly make these indices nonlinearly related to overall species richness (Kraft et al., 2011) and sample grain (Tuomisto & Ruokolainen, 2012). As a result of these correlations, algebraic normalization of these indices may not remove the effects of species richness (Ulrich, Baselga et al., 2017), making the interpretation of observed associations between patterns of co-occurrence and environmental variables challenging.

Here, we systematically examine the relationship between overall species richness and the C-score, β_{sim} and NODF, three widely used indices of community structure. To understand this relationship in a simple unstructured community, we applied a spatially explicit ecological drift model (Ulrich, Jabot, & Gotelli, 2017) and took samples of 10 patches each from two sets of metacommunities (100 with fixed and 100 with varying matrix fill; see Supporting Information supplement A for modelling details). The number of species in the metacommunity varied randomly between 5 and 100. Initial total numbers of individuals per species ranged between 292 and 15,972, and patches had between 226 and 998 individuals. The 10 local patches were drawn randomly from the landscape of 100 contiguous local patches in the neutral model landscape. We compared the species co-occurrence indices of the 10-patch samples and of the whole metacommunities with those obtained by the following three different ways of randomizing species occurrences: (a) occurrence probabilities were equal among patches and species (equiprobable–equiprobable null model); (b) occurrence probabilities were proportional to occurrence totals of rows (= species) and columns (= sites; proportional–proportional null model; Ulrich & Gotelli, 2012); and (c) occurrences were placed randomly, but marginal row and column totals matched those in the original matrix (fixed–fixed independent swap null model). For each of the three null models and three species co-occurrence indices, we examined the relationship between the index and either the total sample richness (γ_S) or the metacommunity richness (γ_M) of the matrix as predictor variables. The metacommunity richness can be considered the real causal variable, with total sample richness being dependent on it. Results on sample richness are shown in the Supporting Information (supplement B) for comparison, as the relevant metacommunity species richness is usually unknown for real datasets.

In ecological drift models, γ_M and α are approximately linked by a logarithmic function [$\alpha = a \times \ln(\gamma_M) + b$] (W.U., unpublished), with parameters a and b depending on model settings. Given that $\alpha = \text{fill} \times \gamma_M$, the arguments below refer to average species richness (α) and to matrix fill as well. Results for samples and whole model communities were qualitatively identical, and we show the results for the whole model communities below and those for the samples in the Supporting Information (supplement B).

Many authors have used raw scores of β_P and β_{sim} for pattern detection because these two indices are invariant in combination with null models that retain total numbers of occurrences (matrix fill) in the

randomization. Here, we first show that the raw scores of four common metrics still depend heavily and in a nonlinear manner on total richness, γ_M (Figure 1; Supporting Information supplement B: Figures B7–B12); hence, they are also correlated with sample species richness, γ_S (Supporting Information Figures B1–B16). Importantly, any model (regression analysis and ANOVA) aiming at inferring the dependence of co-occurrence patterns on environmental variables, in which species richness (γ_M or γ_S) serves as covariate, will not fully remove richness dependences. In the model:

$$Y = \omega_1 \gamma + \varepsilon_\gamma \quad (1)$$

where Y is the co-occurrence metric, ω is the slope parameter, and ε has a normal error structure; the quotient $\Pi Y = Y/\gamma$ should be independent of γ . In our model of ecological drift, this linear relationship is not found, irrespective of model settings (Supporting Information Figures B1–B12). The nonlinearity may differ across metacommunities and cannot be assumed a priori.

Null model analysis (Gotelli & Graves, 1996, Gotelli & Ulrich 2012) is a common strategy to control for sampling and richness effects in co-occurrence indices. Originally, these analyses were used to generate a simple p -value, which is the tail probability of obtaining the observed pattern or one that is more extreme, given the null hypothesis that is embodied in the particular randomization algorithm [$p(\text{data}|\text{H}_0)$, where H_0 is the null expectation]. For meta-analysis, the p -value can be converted to either a net effect size, NES (= observed metric minus average of simulated metrics), or to a standardized effect size ($\text{SES} = \text{NES}/\sigma_N$), where σ_N is the standard deviation of the simulated null values. The SES measures the relative position of the observation in the tail of the distribution, with the SES interval (−2.0, 2.0) corresponding to roughly the 95% confidence interval of the normalized data.

It is still unclear whether the transformation of a co-occurrence index with null model analysis to an NES or SES will necessarily remove the dependence on overall species richness or other sampling properties of the matrix. Legendre and Legendre (2012) and Ulrich and Gotelli (2012) cautioned that the standard deviation of randomized metric values might decrease with the square root of species richness owing to effects of statistical averaging. Ulrich, Baselga et al. (2017) identified a Narcissus effect in species co-occurrence analysis, namely that the correlation between observed and randomized values is mediated by overall richness. Our results support these concerns (Figure 2; Supporting Information Figures B1, B2, B5, B6, B7, B8, B11, B12); the NES and SES of null models that do not retain marginal totals during randomization co-vary with overall species richness. The precise pattern of this relationship depends on the particular index and null model algorithm, but is usually nonlinear. In spite of the identical neutral model settings generating the diversity landscape, the proportional–proportional null model generated a tendency towards increased turnover (decreased nestedness) at higher species richness (Figure 2; Supporting Information Figures B1, B2, B7, B8), whereas the equiprobable null model generated the opposite pattern (Supporting Information Figures B5, B6, B11, B12). Such contrasting outcomes make an unequivocal interpretation of NES and SES scores challenging. Previous meta-analyses based

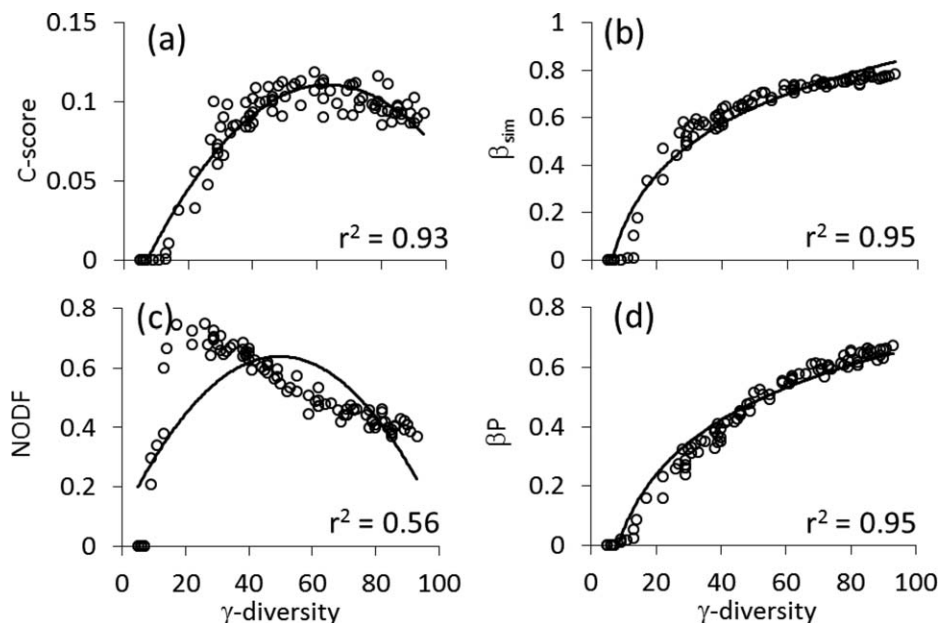


FIGURE 1 Raw scores of four important species co-occurrence indices depended in a nonlinear (a, c, quadratic; b, d, logarithmic) manner on total species richness in 100 fully sampled neutral communities. r^2 values refer to ordinary least squares fits

on these two algorithms may need to be re-evaluated. We note that these two null models encompass other algorithms enumerated by Gotelli (2000), and our concerns refer to these models as well.

In contrast, NES and SES scores of the fixed-fixed null model were only weakly correlated with total species richness (Figure 3; Supporting Information Figures B3, B4, B9, B10, B13), making this null model the only one that allows for a comparison of species co-occurrences across metacommunities of different total species richness. Additionally, and in accordance with our intuition, this algorithm correctly identified the neutral test communities as being random (Figure 3; Supporting Information Figures B3, B4, B9, B10), whereas the other null models pointed to either segregation or nestedness, respectively (Supporting Information Figures B1, B2, B5, B6, B7, B8, B11, B12). These results suggest that previous meta-analyses that have used the fixed-fixed null model may have given the most reliable results in assessing the proportion of non-random species associations among metacommunities of different species richness.

Our results demand caution, not only when interpreting the results of pattern analysis, but also when using index values in subsequent

statistical analyses. Given that raw metric scores and standardized effect sizes of most null model analyses might be related to species richness, it may be hard to compare differences among metacommunities that differ greatly in species richness. We found that most richness effects were nonlinear and thus cannot be removed by simple linear normalization (Figure 1; Supporting Information Figures B1–B12). The most reliable method seems to be the fixed-fixed null model, which has previously been advocated based on its performance in benchmark tests with both structured and random artificial matrices (Gotelli, 2000; Ulrich & Gotelli, 2013).

Importantly, except for β_{sim} , the indices used here are unaffected by the transposition of rows (= species) and columns (= sites). Thus, the same dependence on species richness that we described also holds for a dependence on sample size, which will arise when comparing matrices of different dimension. At a minimum, meta-analyses that use matrices of different sizes should use species richness and site number as statistical covariates, and these meta-analyses should include a further comparison with the fixed-fixed algorithm, which is relatively invariant to these sampling effects.

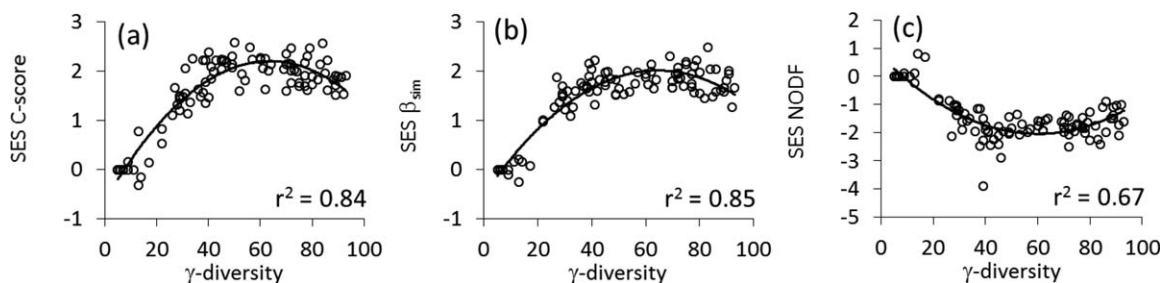


FIGURE 2 Standardized effect sizes (SES; proportional null model) of (a) the C-score, (b) β_{sim} and (c) NODF depend in a nonlinear manner on total species richness in 100 fully sampled neutral metacommunities. r^2 values refer to ordinary least squares fits of quadratic functions. Note that β_P is invariant in combination with a null model that fixes total numbers of occurrences

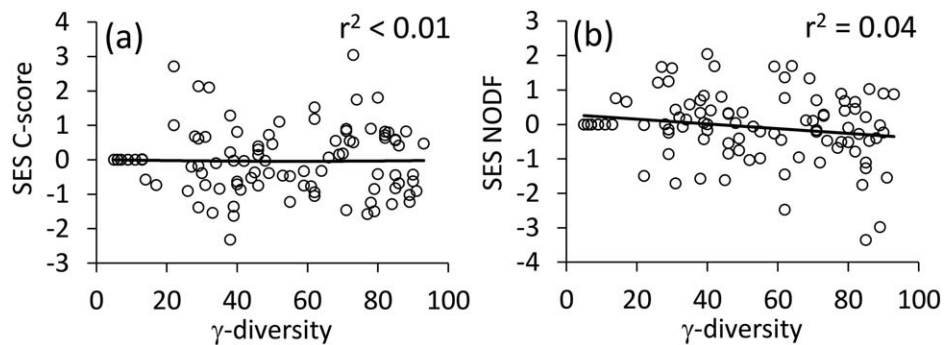


FIGURE 3 Standardized effect sizes (SES; fixed–fixed null model) of (a) the C-score and (b) NODF do not depend on total species richness in 100 fully sampled neutral metacommunities with fixed abundance. r^2 values refer to ordinary least squares fits of quadratic functions. Note that β_p and β_{sim} are invariant in combination with the fixed–fixed null model

Our results do also call for caution when using indices of β -diversity and nestedness in environmental analyses, even after accounting for possible sample size effects. If environmental or spatial variables are correlated with species richness (either γ or α), a particular linear model might imply environmental effects simply because of the collinear variable structure. Null model analyses do not secure for such misinterpretations. We advise careful checking for richness effects on metrics of community structure and the focal environmental variables.

Our results are consistent with the analyses of Ulrich, Baselga et al. (2017), who argued that diversity and species turnover are so tightly linked that it may be impossible to tease these effects apart completely. In this respect, we highlight that our simulations are based on a set of specific assumptions. We compare patterns among neutral metacommunities of different richness. Comparisons of samples within single metacommunities might exhibit contrasting relationships between co-occurrence and sample richness (cf. Henriques-Silva, Lindo, & Peres-Neto, 2013; Jacobson & Peres-Neto, 2010; Ulrich & Gotelli 2012). Furthermore, ecological drift, commonly assumed to be the most basic stochastic generator of ecological communities (Alonso, Etienne, & McKane, 2006), intimately links species richness (γ) and matrix fill (equivalent to α) by the specific set of parameters, particularly the degree of dispersal limitation. Our aim was to show that neutral metacommunities generated by identical parameter settings inevitably exhibit a relationship between the pattern of species co-occurrence and γ (and α). Other types of community assemblies (e.g., Ulrich & Gotelli, 2013) that break off the constraints on richness and fill might display different relationships. As we do not know how observed metacommunities have assembled, our results leave the possibility that the observed variation in the pattern of species co-occurrence among metacommunities is largely attributable to variation in richness.

This does not mean that turnover and nestedness do not exist as independent patterns. Metacommunities of identical richness, numbers of study sites, and matrix fill can exhibit very different internal structures. The four metrics tested here (either as raw scores or in combination with a null model) would be perfectly suited to identify these differences. However, any (environmental) comparison of metacommunities of different species richness, numbers of sites, or total

numbers of occurrences might require additional information on the process of community assembly, possible environmental constraints and sophisticated statistical analyses. The present contribution is a call for the development of such techniques and for more rigorous co-occurrence analyses that always include possible richness effects.

ACKNOWLEDGMENTS

W.U. was supported by the Polish National Science Centre (2014/13/B/NZ8/04681). A.B. is supported by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (ERDF) through grant CGL2016–76637-P.

ORCID

Werner Ulrich  <http://orcid.org/0000-0002-8715-6619>

Yasuhiro Kubota  <http://orcid.org/0000-0002-5723-4962>

Andres Baselga  <http://orcid.org/0000-0001-7914-7109>

REFERENCES

- Alahuhta, J., Kosten, S., Akasaka, M., Auderset, D., Azzella, M., Bolpagni, R., ... Heino, J. (2017). Global variation in the beta diversity of lake macrophytes is driven by environmental heterogeneity rather than latitude. *Journal of Biogeography*, 44, 1758–1769.
- Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Jr., Loyola, R. D., & Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: Reconciling concept and measurement. *Oikos*, 117, 1227–1239.
- Alonso, D., Etienne, R. S., & McKane, A. J. (2006). The merits of neutral theory. *Trends in Ecology and Evolution*, 21, 451–457.
- Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, 19, 134–143.
- 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*. Washington, D.C.: Smithsonian Institution Press.
- Gotelli, N. J., & Ulrich, W. (2012). Statistical challenges in null model analysis. *Oikos*, 121, 171–180.
- Guo, H., Chamberlain, S. A., Elhaik, E., Jalli, I., Lynes, A.-R., Marczak, L., ... Pennings, S. C. (2016). Geographic variation in plant community structures of salt marsh: Species, functional, and phylogenetic perspectives. *PLoS One*, 10, e0127781.

- Henriques-Silva, R., Lindo, Z., & Peres-Neto, P. R. (2013). A community of meta-communities: Exploring patterns in species distributions across large geographical areas. *Ecology*, *94*, 627–639.
- Jacobson, B., & Peres-Neto, P. R. (2010). Quantifying and disentangling dispersal in metacommunities: How close have we come? How far is there to go? *Landscape Ecology*, *25*, 495–507.
- Keil, P., Schweiger, O., Kühn, I., Kunin, W. E., Kuussaari, M., Settele, J., ... Storch, D. (2012). Patterns of beta diversity in Europe: The role of climate, land cover and distance across scales. *Journal of Biogeography*, *39*, 1473–1486.
- Kraft, N. J. B., Comita, L. S., Chase, J. M., Sanders, N. J., Swenson, N. G., Crist, T. O., ... Myers, J. A. (2011). Disentangling the drivers of β diversity along latitudinal and elevational gradients. *Science*, *333*, 1755–1758.
- Legendre, P., & Legendre, L. (2012). *Numerical ecology* (3rd ed.). Amsterdam, The Netherlands: Elsevier.
- Qian, H., Chen, S., Mao, L., & Ouyang, Z. (2013). Drivers of β diversity along latitudinal gradients revisited. *Global Ecology and Biogeography*, *22*, 659–670.
- Simpson, G. G. (1943). Mammals and the nature of continents. *American Journal of Science*, *241*, 1–31.
- Simpson, G. G. (1960). Notes on the measurement of faunal resemblance. *American Journal of Science*, *258*, 300–311.
- Stone, L., & Roberts, A. (1990). The checkerboard score and species distributions. *Oecologia*, *85*, 74–79.
- Tuomisto, H. (2010). A diversity of beta diversities: Straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, *33*, 2–22.
- Tuomisto, H., & Ruokolainen, K. (2012). Comment on 'Disentangling the drivers of β -diversity along latitudinal and elevational gradients'. *Science*, *335*, 1573.
- Tuomisto, H., Ruokolainen, K., & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of Western Amazonian forests. *Science*, *299*, 241–244.
- Ulrich, W., & Gotelli, N. J. (2012). A null model algorithm for presence-absence matrices based on proportional resampling. *Ecological Modelling*, *244*, 20–27.
- Ulrich, W., & Gotelli, N. J. (2013). Pattern detection in null model analysis. *Oikos*, *122*, 2–18.
- Ulrich, W., Baselga, A., Kusumoto, B., Shiono, T., Tuomisto, H., & Kubota, Y. (2017). The tangled link between β - and γ -diversity: A Narcissus effect weakens statistical inferences in null model analyses of diversity patterns. *Global Ecology and Biogeography*, *26*, 1–5.
- Ulrich, W., Jabot, F., & Gotelli, N. (2017). Competitive interactions change the pattern of species co-occurrences under neutral dispersal. *Oikos*, *126*, 91–100.
- Ulrich, W., Kryszeński, W., Sewerniak, P., Puchałka, R., Strona, G., & Gotelli, N. G. (2017). A comprehensive framework for the study of species co-occurrences, nestedness, and turnover. *Oikos*, *126*, 1607–1616.
- Xu, W., Chen, G., Liu, C., & Ma, K. (2015). Latitudinal differences in species abundance distributions, rather than spatial aggregation, explain beta diversity along latitudinal gradients. *Global Ecology and Biogeography*, *24*, 1170–1180.

BIOSKETCH

The authors are particularly interested in the assembly of ecological communities at different spatial and temporal scales and in the variability of species richness at a global scale. They try to disentangle the interplay between patterns of species associations in plant and animal communities, environmental (and human) impact, and evolutionary constraints.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Ulrich W, Kubota Y, Kusumoto B, Baselga A, Tuomisto H, Gotelli NJ. Species richness correlates of raw and standardized co-occurrence metrics. *Global Ecol Biogeogr*. 2018;27:395–399. <https://doi.org/10.1111/geb.12711>