Z-scores unite pairwise indices of ecological similarity and association for binary data

PETR KEIL [iD][1,2,†]

[1]*German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, Leipzig 04103 Germany*
[2]*Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle (Saale) 06120 Germany*

Abstract. Pairwise ecological resemblance, which includes compositional similarity between sites (beta diversity), or associations between species (co-occurrence), can be measured by >70 indices. Classical examples for presence–absence data are Jaccard index or *C*-score. These can be expressed using contingency table matching components *a*, *b*, *c*, and *d*—the joint presences, presences at only one site/species, and joint absences. Using simulations of point patterns for two species with known magnitude of association, I demonstrate that most of the indices converge to a similar value and they describe the simulated association almost identically, as long as they are calculated as a Z-score, that is, as deviation of the index from a null expectation. Further, I show that Z-scores estimate resemblance on average better than raw forms of the indices, particularly in the face of confounding effects of spatial scale and conspecific aggregation. Finally, I show that any single of the matching components, when expressed as Z-score, can be used as an index that performs as good as the classical indices; this also includes joint absences. All this simplifies selection of the right resemblance index, it underscores the advantage of expressing resemblance as deviation from a null expectation, and it revives the potential of joint absences as a meaningful ecological quantity.

† **E-mail:** petr.keil@idiv.de

INTRODUCTION

Ecological resemblance (Legendre and Legendre 2012) is a fundamental ecological quantity that encompasses both similarity of species composition between sites (beta diversity) and co-occurrence patterns among species, and these are the Q- and R-mode analyses, respectively (Legendre and Legendre 2012, Arita 2017). Among the many measures of ecological resemblance, particularly popular, are indices of pairwise similarity or association for binary presence/absence data, with around 80 proposed, as reviewed by Hubálek (1982), Koleff et al. (2003), Rajagopalan and Robb (2005), Legendre and Legendre (2012), Ulrich and Gotelli (2013)

(Appendix S1). These reviews provide guidance for the selection of appropriate indices, based on criteria such as symmetry, additivity, sensitivity to number of sites or species, or interpretability. Even though only a fraction of these indices is widely applied, the task of selecting the right index may still be daunting. Further, different researchers may deploy different indices, hindering synthesis, and comparisons among studies.

All of the indices for binary data can be calculated using *abcd* matching components, which follow contingency tables notation: *a*, the number of joint presences; *b*, the number of presences for only one species (or site); *c*, the number of presences for the other species (or site); and *d*, the number of joint absences. A classical example is

Jaccard index ($a/(a + b + c)$) (Koleff et al. 2003), which has been used mostly in Q-mode analyses. Other examples are C-score (bc) and togetherness (ad) often used in R-mode (Ulrich and Gotelli 2013). However, all of the indices can be used in both Q- and R-mode, that is, for rows or columns of a species-by-site incidence matrix (Hubálek 1982, Legendre and Legendre 2012, Arita 2017).

To standardize the indices, they can be expressed as Z-score (Gotelli and McCabe 2002, Ulrich et al. 2009, Ulrich and Gotelli 2013):

$$Z = \frac{E_{\mathrm{raw}} - E_{\mathrm{exp}}}{\mathrm{SD}_{\mathrm{exp}}} \qquad (1)$$

where E_{raw} is the index calculated on observed data, E_{exp} is the null expectation of the index, and $\mathrm{SD}_{\mathrm{exp}}$ is the standard deviation of the null expectation. This is similar to standardized effect size in meta-analysis (Gurevitch et al. 1992, Ulrich et al. 2009), and it quantifies the deviation of the observations from the null expectation in common units of standard deviation. The null expectation from which E_{exp} and $\mathrm{SD}_{\mathrm{exp}}$ are calculated is usually obtained from null models, that is, by subjecting the data to a repeated randomization that breaks the association among species or sites. Currently, these null models have been well established particularly for binary species associations in R-mode (Gotelli 2000, Ulrich and Gotelli 2013), whereas for abundance data (Ulrich and Gotelli 2010) and for beta diversity assessments in Q-mode, they are still under development and their merits are partly unclear or debated (Chase et al. 2011, Ulrich et al. 2017, Legendre 2019).

While working on a separate project, I noticed that several of the classical resemblance indices for binary data gave surprisingly similar answers when expressed as Z-scores, although in their raw form, each of them might have a unique way of capturing resemblance (Koleff et al. 2003). This led me to a suspicion that that maybe the actual mathematical formula of resemblance indices does not matter that much, if at least some of the important matching components are in the formula, and as long as the index is expressed as a Z-score. To test this, and to figure out which of the components actually really need to be in a resemblance index, I devised a simulation exercise, as described below and in Appendix S2. My first aim was to examine how

similar to each other are the existing resemblance indices when applied to simulations with varying and known resemblance, and how similar they are when converted to their Z-score variants. My second aim was to find out how well the raw indices reflect the parameter in the simulation that regulated the magnitude of resemblance. Finally, I also aimed to identify which of the four matching components are most important for capturing the simulated resemblance, both in their raw form and as Z-scores.

Methods

I set this exercise to be in R-mode; that is, it examined resemblance indices as measures of between-species associations (Gotelli 2000). I simulated spatially explicit distributions of pairs of species as two point patterns (Wiegand and Moloney 2014), while varying the magnitude of association between the two species, and while varying conspecific aggregation and number of individuals per species. I modeled the between-species association as being dependent on spatial distance, and it was controlled by a single parameter alpha, with negative values for segregation, zero for independence, and positive for attraction (Appendix S1: Fig. S1, S2). All of this resulted in 612 unique combinations of simulation parameters. I then aggregated the point patterns to grids of varying grain (resolution), and I used these to create binary vectors with 0 for absence and 1 for presence of each species in each grid cell. Details on the simulation procedure are in Appendix S2.

For each simulated pair of species, I quantified the observed resemblance E_{raw} between the two binary vectors using 74 published indices (Appendix S1) and 4 matching components $abcd$. For all of these indices and matching components, I also calculated their Z-scores, which I did by running 400 realizations of null model that randomly reshuffled presences within each binary vector (sim2 algorithm of Gotelli and Ellison 2013), irrespectively of the values in the other vector. Importantly, the locations over which the presences were reshuffled also included double zeroes, that is, locations where none of the species was originally present, but which were still part of the spatial domain of the simulations. These 400 realizations were then used to

calculate the SD_{exp} and E_{exp} and the Z-score using Eq. 1. Further details are in Appendix S2.

I then calculated correlation matrix of Spearman rank correlations between each pair of indices, with one correlation matrix for raw indices and one for Z-scores; I summarized these correlation matrices as boxplots and also in a network graph (R package qgraph). I also measured Spearman rank correlation between each index (or their Z-scores) and the simulated association alpha. For all evaluations in this study, I used absolute values of the rank correlations, since I was interested in the magnitude, not the direction of the correlation, and since some indices represent similarity (attraction), while others represent dissimilarity (segregation), yet others reflect both. All code and data used are available at https://github.com/petrkeil/Z-scores.

RESULTS

Across all of the 74 indices and 4 matching components, Z-scores were considerably more similar to each other than were the raw metrics, as expressed by the between-index pairwise Spearman correlations (Fig. 1). This is further obvious when the correlations are visualized in a network graph (Fig. 2) where almost all Z-scores aggregate in the middle of the main cluster. Exceptions to this overall trend are indices 49, 50, 53, 78, and 79 (see Appendix S1 for their exact definition), all complex indices with little or no use in ecology (Hubálek 1982, Rajagopalan and Robb 2005). Note that from these analyses I omitted indices no. 61, 76, and 77 (Appendix S1) since they produced too many NA values.

None of the indices gave a perfect correlation with the simulated underlaying association represented by parameter alpha (Fig. 3), likely because of the confounding effect of conspecific aggregation and varying spatial grain. However, expressing the raw indices as Z-score increased their median correlation with alpha from 0.69 to 0.79 (Fig. 3), and it dramatically reduced variation in the performance of the indices, from the interquartile range of 0.18 to interquantile range of 0.009 (Fig. 3). Thus, even though different indices gave different answers concerning the magnitude of association between two species, they gave very similar answers when used as Z-score.
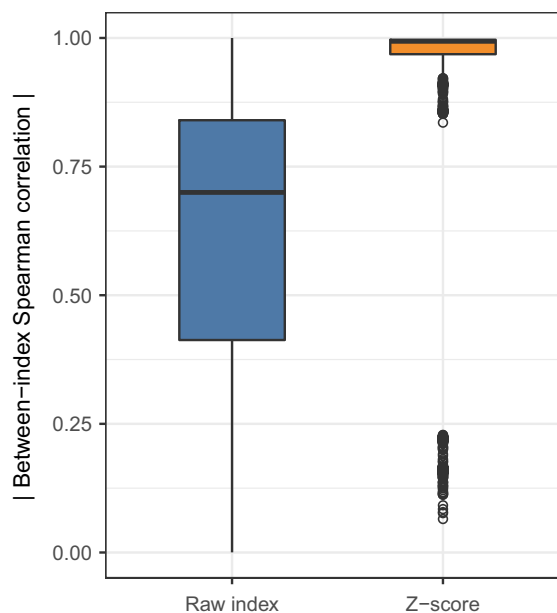

Fig. 1. Summary of Spearman correlations (their absolute value) among raw metrics and among Z-scores, calculated on simulated species associations. Boxplots summarize the lower triangular part of an index-by-index correlation matrix. Boxes show 25%, 50%, and 75% quantiles, whiskers show span of the data but do not go further than 1.5 of the interquartile range. Omitted are indices no. 61, 76, and 77 (Appendix S1) since they produced too many NA values.

Another striking result was the performance of single matching components alone—not only using them as Z-scores substantially improved their correlation with alpha (from median of 0.41 to 0.79), but they also gave almost identical answers (reduced interquartile range of the correlation from 0.35 to 0.001). Thus, each of the four *abcd* components sufficiently captures the underlaying simulated association (Fig. 3). In Appendix S3, I show why. Surprisingly, this also includes the number of joint absences *d* (double zeroes), a quantity that has been deemed uninformative for ecological purposes (Legendre and Legendre 2012).

DISCUSSION

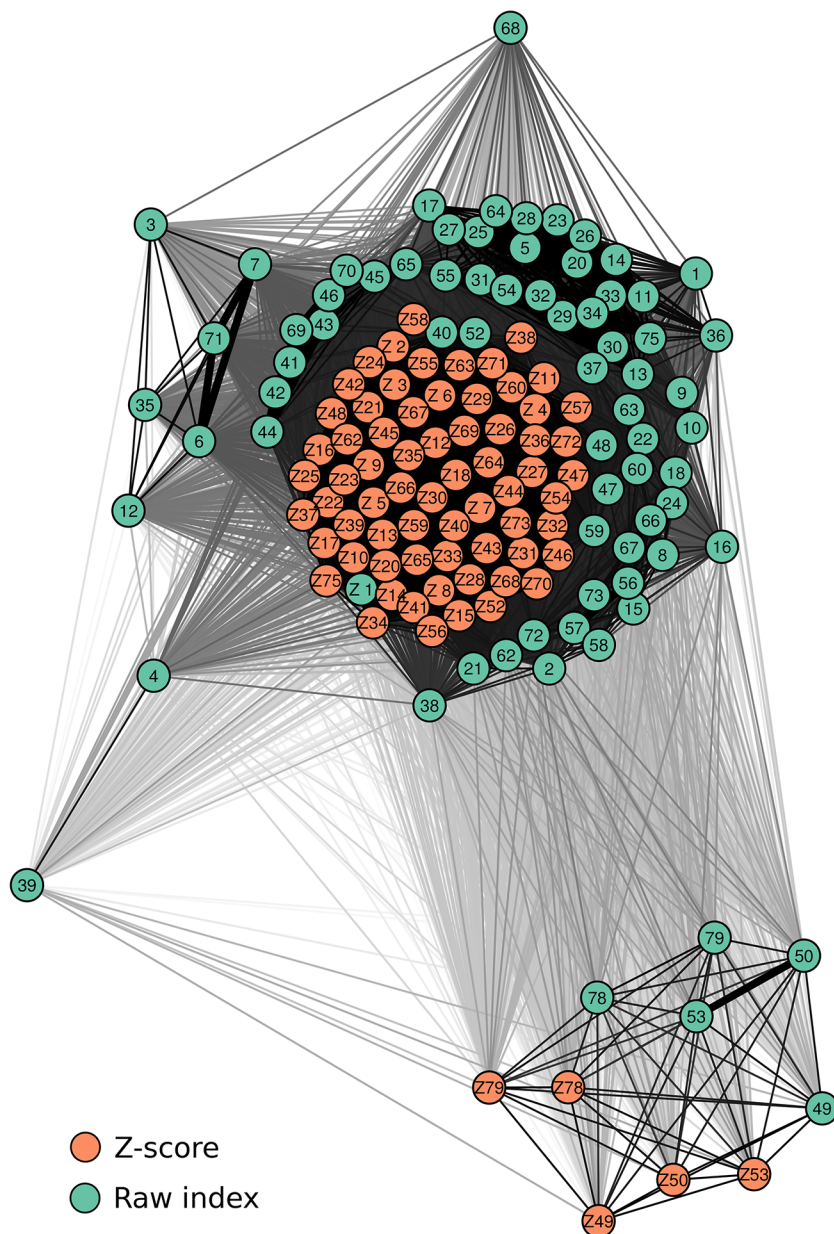Majority of resemblance indices gave a range of values when expressed in their raw

Fig. 2. Representation of Spearman correlations among raw indices (green) and their Z-scores (orange) in a network graph, produced by R package qgraph. Each point represents an index, and lines between points are correlations between indices, with thicker and darker lines indicating stronger correlations. The main result here is that most of the Z-scores unite at the center of the graph, reflecting their stronger pairwise correlations. Index numbers correspond to indices in Appendix S1. Note that the graphing algorithm does not allow points to overlap, sometimes pushing even strongly correlated indices apart.

form, and they varied in their correlation with each other and with the simulated association. This is not surprising, since the indices have been designed to reflect various facets of resemblance.

For instance, C-score ($b \times c$) was designed to specifically capture the magnitude of segregation between species (negative association), while togetherness ($a \times d$) captures positive
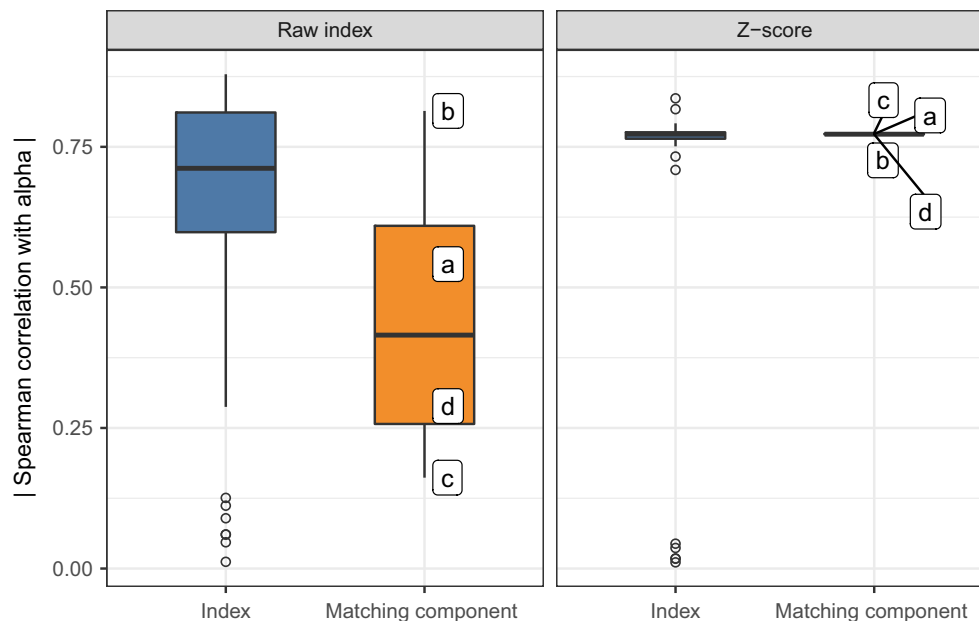
Fig. 3. Correlation of resemblance indices with the parameter alpha that determines the magnitude of attraction (or segregation) of two simulated point patterns. The correlation is measured by absolute value of the Spearman rank-rank correlation. Shown are existing published indices for binary data ($N = 74$; Appendix S1) and the four matching components a, b, c, and d. Left panel shows raw values of the indices, and right panel shows their Z-scores.

association (Ulrich and Gotelli 2013), and yet Pearson tetrachoric correlation $((a \times d - b \times c) / (((a + b) \times (c + d) \times (a + c) \times (b + d))^{0.5}))$ captures both positive and negative associations in a single number (Hubálek 1982). However, most of the indices converge when expressed as Z-scores —the Z-scores become strongly correlated with each other (hence the title) and they also, on average, better reflect the simulated association than the raw metrics. In a way, when a null model is deployed in order to move from the raw similarity index to the Z-score, one actually relaxes the need for the index to have a clear interpretation, since that shifts from the index to the null model. For example, Jaccard index is interpreted as proportional (or percentage) overlap of occurrences of two species, but it becomes a mere summary statistic (sensu Wiegand and Moloney 2014) when contrasted with expectations from a null model, where the null model now bears all the information/meaning.

The uncovered convergence and good performance of Z-scores is reassuring to the field currently flooded with beta diversity and association indices. As long as an index is presented as a Z-score, it matters little which one is used—within the classical and most often used indices (Jaccard, Sorenson, Simpson, C-score), one can't really go wrong, and one can even compare Z-scores from different studies, irrespectively of the actual index used, as long as the null model is the same. Further, there may be instances when we know the precise ecological interpretation of what a null model does; in such case, it might be simpler to rely on the meaning of the null model, rather than on the meaning of the metric, since meanings of some metrics might be complex, or opaque. My results also imply that indices which include double zeroes (e.g., the simple matching coefficient), are a good option, even though in their raw form they are heavily biased by any arbitrary addition of joint absences, for example, due to arbitrary spatial delineation of the study area. I propose that this united behavior occurs because each of the matching components *abcd* carries equal information about the deviation from the null expectation (Stevens 1938; Appendix S3). This leads me

to a proposition that any of the matching components can be used as a valid index of ecological resemblance, if Z-scored, which has an alluring touch of minimalism.

Here, I should add that one reason for the varying, and slightly inferior, performance (measured by their correlation with parameter alpha) of the raw metric could be that the association could have been obscured by varying numbers of individuals (Wiegand and Moloney 2014), varying spatial grain (Araújo and Rozenfeld 2014), and perhaps also by varying conspecific aggregation. Further, the limited ability of the raw indices to recover the undelaying association is also expected since the simulated association is realistically (Wagner 2003, Wiegand and Moloney 2014) distance-dependent, while the evaluated indices are naively spatially implicit. My results corroborate that Z-scores and null models can, at least partially, account for these confounding effects, which is something that has been known in point pattern analysis (Wiegand and Moloney 2014) and which has also been appreciated in co-occurrence analysis, for example, in accounting for the confounding effect of species richness (Ulrich et al. 2018).

I stress that the generality of my finding critically depends on the particular null model. For a two-species situation with binary data, my results are general, since there really are only two ways to randomize the presences: Either they are randomized over all n sites or species ($n = a + b + c + d$), or the sites with double zeroes (d) are excluded. The latter variant will likely reduce performance of indices that do include double zeroes in their formula, but I also struggle to see a clear biological interpretation of such null model. However, the problem of null model selection will become increasingly important, and complex, once more than two species are added to the community matrix (Gotelli 2000), once abundances are used instead of binary occurrences (Ulrich and Gotelli 2010), and/or once spatial component is included (Wagner 2003). It remains to be assessed how Z-scores perform under the wide selection of null models available for these kinds of data and representing variety of ecological hypotheses.

Another factor affecting generality of my results is whether one studies species associations or beta diversity. I set this study to be in

R-mode; that is, it examined the indices as measures of between-species associations (R-mode), and not beta diversity (Q-mode). The reason was that the null models are better developed for the R-mode (but see Chase et al. 2011, Legendre 2019) and it was more straightforward to simulate spatially explicit two-species associations than patterns of beta diversity. This is because of the interpretation of the d component: In the case of species associations, d is the number of sites where none of the two species has been found, and it is bounded by the finite (bounded) space. In the case of beta diversity, however, d is the number of unobserved species from the species pool, for which the upper boundary may be elusive. In spite of this, I consider my results to be valid for beta diversity, given that one can make a reasonable assumption about the size of the species pool, since the mathematics of the raw beta diversity is identical to the association metrics, only the indices are applied to a transposed site-by-species matrix (Arita 2017).

To conclude, in this paper I have demonstrated that most Z-scores converge to the same value, although the original indices may have different values and meanings. Z-scores also, on average, improve the ability to capture meaningful association between two species, and I suggest that they may account for confounding effects of spatial resolution of the data, uneven prevalences of the species, or intra-specific spatial aggregation, although this needs to be explored more thoroughly. I thus argue that null models and null expectations, perhaps more than particular indices, should be the central focus of future methodological research, particularly in the R-mode analyses of beta diversity for which the null models seem to be underdeveloped.

Literature Cited

Araújo, M. B., and A. Rozenfeld. 2014. The geographic scaling of biotic interactions. Ecography 37:406–415.

Arita, H. T. 2017. Multisite and multispecies measures of overlap, co-occurrence, and co-diversity. Ecography 40:709–718.

Chase, J. M., N. J. B. Kraft, K. G. Smith, M. Vellend, and B. D. Inouye. 2011. Using null models to disentangle variation in community dissimilarity from variation in α-diversity. Ecosphere 2:1–11.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. Ecology 81:2606–2621.

Gotelli, N. J., and A. M. Ellison. 2013. EcoSimR 1.00. http://www.uvm.edu/~ngotelli/EcoSim/EcoSim.html

Gotelli, N. J., and D. J. McCabe. 2002. Species co-occurrence: a meta analysis of J. M. Diamond's assembly rules model. Ecology 83:2091–2096.

Gurevitch, J., et al. 1992. A meta-analysis of competition in field experiments. American Naturalist 140:539–572.

Hubálek, Z. 1982. Coefficients of association and similarity, based on binary (presence–absence) data: an evaluation. Biological Reviews 57:669–689.

Koleff, P., et al. 2003. Measuring beta diversity for presence–absence data. Journal of Animal Ecology 72:367–382.

Legendre, P. 2019. A temporal beta-diversity index to identify sites that have changed in exceptional ways in space–time surveys. Ecology and Evolution 9:3500–3514.

Legendre, P., and L. Legendre. 2012. Numerical ecology. Elsevier, Amsterdam, The Netherlands.

Rajagopalan, S., and R. Robb. 2005. Assessment of similarity indices to quantify segmentation accuracy of scaffold images for tissue engineering. Pages 1636–1647 in J. M. Fitzpatrick and J. M. Reinhardt, editors. Medical imaging 2005: Image processing, SPIE 5747. SPIE, Bellingham, Washington, USA.

Stevens, W. L. 1938. The distribution of entries in a contingency table with fixed marginal totals. Annals of Human Genetics 8:238–244.

Ulrich, W., M. Almeida-Neto, and N. J. Gotelli. 2009. A consumer's guide to nestedness analysis. Oikos 118:3–17.

Ulrich, W., A. Baselga, B. Kusumoto, T. Shiono, H. Tuomisto, and Y. Kubota. 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., and N. J. Gotelli. 2010. Null model analysis of species associations using abundance data. Ecology 91:3384–3397.

Ulrich, W., and N. J. Gotelli. 2013. Pattern detection in null model analysis. Oikos 122:2–18.

Ulrich, W., Y. Kubota, B. Kusumoto, A. Baselga, H. Tuomisto, and N. J. Gotelli. 2018. Species richness correlates of raw and standardized co-occurrence metrics. Global Ecology Biogeography 27:395–399.

Wagner, H. H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. Ecology 84:1045–1057.

Wiegand, T., and K. A. Moloney. 2014. Handbook of spatial point-pattern analysis in ecology. CRC Press, Boca Raton, Florida, USA.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.2933/full