The Chord-Normalized Expected Species Shared (CNESS)-distance represents a superior measure of species turnover patterns

Running title: *Measuring species turnover by CNESS*

Yi Zou[1*], Jan Christoph Axmacher[2]

1. Department of Health and Environmental Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, China
2. UCL Department of Geography, University College London, London, UK

Correspondence:

Yi Zou (yi.zou@xjtlu.edu.cn)

Abstract

1. Measures of β-diversity characterizing the difference in species composition between samples are commonly used in ecological studies. Nonetheless, commonly used dissimilarity measures require high sample completeness, or at least similar sample sizes between samples. In contrast, the Chord-Normalized Expected Species Shared (CNESS) dissimilarity measure calculates the probability of collecting the same set of species in random samples of a standardized size, and hence is not sensitive to completeness or size of compared samples. To date, this index has enjoyed limited use due to difficulties in its calculation and scarcity of studies systematically comparing it with other measures.
2. Here, we developed a novel R function that enables users to calculate ESS (Expected Species Shared)-associated measures. We evaluate the performance of the CNESS index based on simulated datasets of known species distribution structure, and compared CNESS with more widespread dissimilarity measures (Bray-Curtis index, Chao-Sørensen index, and proportionality based Euclidean distances) for varying sample completeness and sample sizes.
3. Simulation results indicated that for small sample size (*m*) values, CNESS chiefly reflects similarities in dominant species, while selecting large *m* values emphasizes differences in the overall species assemblages. Permutation tests revealed that CNESS has a consistently low CV (coefficient of variation) even where sample completeness varies, while the Chao-Sørensen index has a high CV particularly for low sampling completeness. CNESS distances are also more robust than other indices with regards to undersampling, particularly when chiefly rare species are shared between two assemblages.

4. Our results emphasize the superiority of CNESS for comparisons of samples diverging in sample completeness and size, which is particular important in studies of highly mobile and species-rich taxa where sample completeness is often low. Via changes in the sample size parameter *m*, CNESS furthermore cannot only provide insights into the similarity of the overall distribution structure of shared species, but also into the differences in dominant and rare species, hence allowing additional, valuable insights beyond the capability of more widespread measures.

Key words

β-diversity, CNESS, dissimilarity, species turnover, R function

Introduction

Reliable measurements of biodiversity are crucial in ecological studies, both with regards to species richness (α-diversity) and assemblage composition (β-diversity). Whittaker (1960) defined β-diversity as the species turnover across spatial scale, with α-diversity as the species richness at a sampling unit and γ-diversity as the total number of species over a large geographic area. Assessments of species turnover between samples as a key measure of β-diversity are commonly based on dissimilarity measures using mathematical descriptions of differences between pairs of samples (Legendre & Gallagher 2001; Tuomisto 2010; Mori, Isbell & Seidl 2018). These approaches are generally based on plot × species matrices, often also including information on species' abundances, as basis for the calculation of the (dis)similarity or relative distance between pairs of samples.

The sampling effort for assemblages of diverse, mobile organisms, such as most insect assemblages, is difficult to standardize. The number of species in a sample generally correlates positively with the overall sample size and sampling effort, while sample completeness with regards to the local species pool is often unachievable in species-rich groups and biomes. Therefore, directly comparing the species records between two samples or sites with measures not accounting for the relative sampling effort and completeness creating a potential 'undersampling bias' that will result in highly unstable and unreliable outcomes (Coddington *et al.* 2009; Beck, Holloway & Schwanghart 2013; Iknayan *et al.* 2014). With regards to alpha-diversity, standardization can be achieved for example via the use of rarefaction (Hurlbert 1971) and extrapolation techniques (Chao & Jost 2012; Chao *et al.* 2014), the use of species richness estimators (Hortal, Borges & Gaspar 2006) or by using parametric

diversity indices such as Fisher's α (Beck & Schwanghart 2010). Nonetheless, most widespread measures of species turnover between assemblages are not appropriately accounting for the 'undersampling bias', with results potentially only poorly representing the "true" dissimilarity in the underlying populations (Beck, Holloway & Schwanghart 2013). For example, results are often heavily influenced by dominant species, or by widespread species of low abundance that by chance appear in only a subset of samples (Legendre & Gallagher 2001). Such problems are inherent in results gained by virtually all commonly used techniques to assess changes in species' assemblages, with incidence-based indices more sensitive to sample size than abundance-based ones (Beck, Holloway & Schwanghart 2013). Some efforts have been made to address the influence of incomplete sampling on beta-diversity measures, both by developing indices regarded as less sensitive to sample size (Cardoso, Borges & Veech 2009; Schroeder & Jenkins 2018), or by trying to adjust existing indices (Chao *et al.* 2005; Yue & Clayton 2005) or using rarefaction techniques (Stier, Bolker & Osenberg 2016; Brocklehurst, Day & Fröbisch 2018) that account for sample size-related variations in dissimilarity values. While these have yielded some interesting insights, they were often either plagued by very high levels of uncertainty or by low predictability power, making the interpretation of resulting values very difficult.

One measure specifically designed to account for the issues relating to sample standardization is the 'Chord-Normalized Expected Species Shared' (CNESS)-distance. The CNESS index was introduced by Trueblood, Gallagher and Gould (1994), and it is based on the calculation of the 'Normalized Expected (number of) Species Shared' (NESS) between two samples as proposed by Grassle and Smith (1976). Both CNESS and NESS are in turn derived from the 'Expected Species

97 Shared' (ESS)-index that reflects the probability of obtaining the same set of species
98 when randomly drawing a specific number of individuals from a community (Morisita
99 1959; Grassle & Smith 1976). In other words, CNESS has been developed to cater
100 for the effect that two samples of equal size randomly drawn from the same
101 underlying community will by chance vary in their exact composition of species, and
102 in the distribution of individuals across the different species. High CNESS
103 dissimilarity values in this context reflect a low probability that two samples are
104 drawn from the same community. Additionally, CNESS calculations allow for the
105 sample size compared between two samples to be varied by adjusting the sample
106 size parameter, *m*. This allows for a direct comparison of assemblages represented
107 by two samples of varying sample size, by estimating their similarity for a
108 standardized sample size common to both samples. In this context, small values of
109 *m* are believed to emphasize the similarity specifically in dominant species, whereas
110 for large values of *m*, results are assumed to be increasingly affected by the
111 composition of the entire species assemblage (Trueblood, Gallagher & Gould 1994).
112 Calculating dissimilarities for different *m* values therefore generates unique insights
113 into the similarity patterns between samples with regards to their different
114 components (Trueblood, Gallagher & Gould 1994). CNESS has already been used
115 particularly in studies of insect biodiversity, where samples are commonly showing
116 large differences in the number of specimens caught at individual sampling events
117 and in their sample completeness (Axmacher *et al.* 2004; Beck & Vun Khen 2007;
118 Zou *et al.* 2014).

119 In spite of its theoretical advantages over other, commonly used dissimilarity metrics,
120 the uptake of CNESS has been limited. For example, CNESS was excluded in a
121 recent study by Schroeder and Jenkins (2018) who evaluated the sensitivity of

several dissimilarity indices to the 'undersampling bias', with the authors recommending measures such as the Bray-Curtis index due to their relative robustness to this effect. One of the reasons for the low profile of CNESS distances might relate to problems in calculating these dissimilarity values, with no suitable software tools available to date. The Compah96 software used in previous studies that is programmed in FORTRAN for MS DOS-based systems (Gallagher 1998) has become unavailable. In contrast, commonly used dissimilarity measures such as the Sørensen or Bray-Curtis indices can be calculated already by a number of standard packages in the open source R programming language (Oksanen *et al.* 2014). In addition, dissimilarity value of CNESS range between 0 and √2 (see details in the method section), which makes direct comparisons with other dissimilarity measures whose values usually range between 0 (samples are the same) and 1 (samples are 100% different) problematic.

Here, we provided scripts for a function to conveniently calculate the entire family of ESS (Expected Species Shared) measures using the R language (see Appendix 1) to make these dissimilarity measures more easily and widely available. We additionally introduced a slightly amended version of the CNESS measure adjusted so that values now range between 0 and 1. We used this function to explore how CNESS performs for assemblages of different species distribution structures for different sample size parameters, *m*. In addition, we evaluated the sensitivity of the CNESS measure in comparison to other, commonly used dissimilarly measures, with regards both to incomplete samples and variations in sample size. We used simulated rather than empirical data-sets to explore patterns and draw conclusions for the general behaviour of the different dissimilarity and distance measures.

Method

The expression of CNESS

CNESS is derived from the Expected Species Shared (ESS) measures introduced by Trueblood (Trueblood, Gallagher & Gould 1994). The ESS value for sites *i* and *j* ($ESS_{ij|m}$) (Grassle & Smith 1976), represents the number of species expected to be shared between two randomly selected samples of a standardized size of *m* individuals, and can mathematically be expressed as:

$$ESS_{ij|m} = \sum_{k=1}^{S}\left[1 - \frac{\binom{N_{i*}-N_{ik}}{m}}{\binom{N_{i*}}{m}}\right] \times \left[1 - \frac{\binom{N_{j*}-N_{jk}}{m}}{\binom{N_{j*}}{m}}\right]$$

where S represents the total number of species, N_{i*} and N_{j*} represent the total number of individuals of site i and j, and N_{ik} and N_{jk} represent the abundance of the k^{th} species at sites *i* and *j*.

While ESS calculations follow logical probability assumptions, the value of $\binom{N_{i*}}{m}$ for a large value of *m* can become almost infinite, leading to potential calculation failures during computation. The function nonetheless can be amended as follows (see mathematical proof in Appendix 2):

$$ESS_{ij|m} = \sum_{k=1}^{S}\left[1 - \prod_{n=0}^{m-1}\frac{(N_{i*} - N_{ik} - n)}{N_{i*} - n}\right] \times \left[1 - \prod_{n=0}^{m-1}\frac{(N_{j*} - N_{jk} - n)}{N_{j*} - n}\right]$$

Although generally creating the same values for ESS, this formula is more robust with regards to the aforementioned calculation problems. The ESS values can in a next step be normalized, leading to the NESS (Normalized Expected Species

Shared) similarity measure between two samples, with values ranging between 0 and 1 (Grassle & Smith 1976):

$$\mathrm{NESS}_{\mathrm{ij|m}} = \frac{2 \times \mathrm{ESS}_{\mathrm{ij|m}}}{\mathrm{ESS}_{\mathrm{ii|m}} + \mathrm{ESS}_{\mathrm{jj|m}}}$$

This measure is further modified to specifically account for the often large number of rare species that randomly occur in a small number of samples, even if samples are drawn from the same, underlying population. This modification is the CNESS (Chord-Normalized Expected Species Shared)-distance measure (Trueblood, Gallagher & Gould 1994). CNESS values can be calculated as:

$$\mathrm{CNESS}_{\mathrm{ij|m}} = \sqrt{2 \times \left[1 - \frac{\mathrm{ESS}_{\mathrm{ij|m}}}{\sqrt{\mathrm{ESS}_{\mathrm{ii|m}} \times \mathrm{ESS}_{\mathrm{jj|m}}}}\right]}$$

While NESS values vary between 0 and 1, Trueblood, Gallagher and Gould (1994) formulated CNESS in a way that theoretical values range between 0 and √2. This may result in difficulties when comparing its values with other dissimilarity indices that usually range between 0 and 1. We therefore slightly modified the CNESS index by removing the √2 multiplicator from the function, leading to the amended formula for CNESS_a:

$$\mathrm{CNESS}_{\mathrm{a(ij|m)}} = \sqrt{1 - \frac{\mathrm{ESS}_{\mathrm{ij|m}}}{\sqrt{\mathrm{ESS}_{\mathrm{ii|m}} \times \mathrm{ESS}_{\mathrm{jj|m}}}}}$$

We have created an R function (Appendix 1) that conveniently allows us and our readers to calculate CNESS_a, CNESS, NESS and ESS values in the R environment. The function contains three parameters, *x, m*, and *index* (by default, the *index* is set

as $CNESS_a$); where *x* represents the species × sample (as row × column) matrix, and *m* the sample size parameter representing the number of individuals to be randomly drawn from the two samples that are compared. Theoretically, the choice of *m* can be any positive integer that is ≥ 1. However, if the total sample size for a site is < *m*, this site will automatically be excluded from the analysis.

Simulation and analysis

To assess the performance of $CNESS_a$ in comparison to other distance or dissimilarity measures, we first created a theoretical "control" dataset containing 100 species. The abundance of these species was fitted to a logarithmic distribution pattern. The log-mean value of the resulting dataset is 6.5, with a log-sd value of 1, with the resulting dataset representing the trial community therefore containing about 100,000 specimens distributed across the 100 species. For "treatments", we created assemblages of equal size and distribution patterns, but with different amounts of "dominant" (D) and "rare" (R) species shared with the control. Each treatment contained three different populations, sharing 25%, 50% and 75% of their dominant (D) or rare (R) species with our "control". Thus, we created a total of six "treatment" assemblages. The "dominant" species group shares the most abundant species from the control group. For example, the 25% dominant species (D25) group shares the 25 species most abundant in the control group with that group, while randomizing their respective species rank order in the new group. The remaining 75 species in this second group are "new species" when compared with the control group. Likewise, the rare species assemblage shares the least abundant species with the control, with species ranks again randomized. The overall abundance distributions for different datasets are displayed in Appendix 3.

206 The actual analysis of index performances was separated into two parts. In the first
207 part, the relative influence of abundant and rare species on the $CNESS_a$ calculated
208 for different *m* values was evaluated. This was achieved by calculating the pairwise
209 $CNESS_a$ value between the “control” and “treatment” datasets, with *m* values
210 increasing from 1 to 100,000.

211 The second part of the analysis focuses on comparisons of the stability of distance
212 or dissimilarity values of $CNESS_a$ and a selection of other, commonly used
213 abundance-based dissimilarity indices. We selected three indices: i) the Bray-Curtis
214 index, which is the most commonly used abundance-based dissimilarity index that
215 has been argued to also be relatively robust with regards to undersampling bias
216 (Schroeder & Jenkins 2018), ii) The Chao-Sørensen index, which is an abundance
217 based form of the Sørensen index developed by Chao *et al.* (2005) in order to
218 reduce the species distribution bias inherent in incidence-based indices, and iii)
219 proportion-based Euclidean distances. For the $CNESS_a$ index, we selected *m* values
220 of 1, 10, 100 and 1000.

221 We simulated two sampling strategies in order to investigate the effects of
222 incompleteness of samples, and of unequal sample sizes. The first strategy was to
223 have an equal sampling coverage for both “treatment” and “control” datasets, with
224 the coverage varying between 0.01% (~10 individuals), 1%, 10% and 100% (all
225 specimens present in the sample). Our sampling coverage refers to the number of
226 individuals sampled from the overall pool, while we also calculated the sampling
227 completeness that refers to the proportion of species sampled in comparison to the
228 total number of species contained in the pool. Species completeness reach 9%, 54%
229 and 97% for the individual coverage at 0.01%, 0.1% and 1, and reach 100% when

individual coverage is higher than 10%. The second strategy then compared the dissimilarity or distance between two samples that varied in their coverage, again with the coverage in the individual treatment samples varying from 0.01% to 1%, 10% and 100%, but using a constant number of 1000 specimens for the control treatment. We calculated the pairwise distance or dissimilarity values between the "control" and "treatment" samples from these combinations for all the above indices, carrying out permutations with 1000 iterations.

It need to be noticed that the main aim of our study was to test the 'stability' or 'robustness' of distance measures based on $CNESS_a$ and the other indices for differences in sampling coverage and unequal sample size scenarios, rather than evaluating how each index specifically reflects the underlying differences between samples and assemblages. The applicability of individual indices may partly depend on the actual sample patterns, with some measure comparisons in this regard provided in earlier studies (Chao *et al.* 2005; Beck, Holloway & Schwanghart 2013; Barwell, Isaac & Kunin 2015). In order to evaluate the stability of the different indices under the different sampling strategies, we then compared the coefficient of variation (CV = SD / mean) of the permutations results. In order to check the change of dissimilarity under different levels of sampling coverage, we computed the change rate ($D_{c,n}$) between the undersampled dataset (D_n) and the final, full sample dataset (D_1 , i.e. representing either the full dataset in sampling approach 1, or the 1% control dataset in approach 2) using the formula:

$$D_{c,n} = \frac{|D_n - D_1|}{D_1}$$

All calculations were conducted in R V3.1.2 (R Core Team 2014), and we used the "CommEcol" package (Melo 2014) to calculate the Chao-Sørensen index, while the package "plyr" (Wickham 2011) was used for the data sorting during the simulation. The simulation scripts can be found in Appendix 4.

Results

$CNESS_a$ distances between control samples and samples taken from assemblages sharing rare species with the control were generally larger than distances between control samples and samples sharing dominant species with the control. Nonetheless, the difference between these scenarios decreased with an increase in the sample size parameter *m*, with the shared rare species sample distances decreasing and the distances for samples sharing dominant species initially decreasing, but then increasing (Figure 1). For very large *m* – values, distances for "rare" and "dominant" treatments converged towards a common value, representing the value when ESS accounts for the actual number of species for site *i* (ESS_{ii}) and site *j* (ESS_{jj}), and the shared total number of species between two sites (ESS_{ij}).

Comparisons of the different dissimilarity metrics show that the CV values generally increase with a decrease in sample coverage across all indices and for both, equal and unequal sampling strategies, as well as across both, the rare and the dominant shared species scenarios. Only the Bray-Curtis measures shows an exceptional peak in CV at a sampling coverage of 1% for the unequal sampling strategy (i.e. both samples have the same coverage). In all scenarios, the CV of $CNESS_a$, Bray-Curtis

and proportion-based Euclidean distances never exceeded 0.1 (<0.05 for $CNESS_a$ in most cases), while the CV of Chao-Sørensen exceeded 0.1 in several scenarios, for example for samples sharing dominant species with the control for a coverage <0.1%, reaching a maximum value of 0.69 (Figure 2). With regards to variations in the standardized sample size in $CNESS_a$ (*m*), an increase in its value resulted in a lower CV in the scenario of shared dominant species, but in a higher CV in the "rare species" shared scenario (Figure 2).

Where rare species were shared between control and treatment samples, $CNESS_a$ showed a stable performance across different sampling coverages and sampling strategies, as the change rate in comparison to the full coverage value never exceeded 0.1 and remained <0.05 for the majority of cases. In comparison, the changes of all other three indices exceeded 0.1 in some cases, for example in scenarios where sampling coverage <0.1% (Figure 3). Where dominant species were shared between control and treatment samples, all indices showed high change values >0.1 under a sampling coverage < 0.1% (this value could not be calculated for $CNESS_a$ *m* =1000), expect for Bray-Curtis distances under the unequal sampling strategy, but the change for this index exceed 0.1 when sampling coverage reached 10% and 0.1% (Figure 3).

Discussion

The R function we developed for this study and present in the appendix enables users to calculate the entire family of ESS-related distance measures. It allowed us to simulate and compare the performance of these widely neglected dissimilarity measures with more widespread measures for communities across a wide range of shared species and sample completeness scenarios. The values of the amended

298 $CNESS_a$ range from 0 to 1, which enables users to compare results directly with
299 common dissimilarity measures. The sample size, *m*, which by default is set to 1,
300 can be changed according to the users' requirements. In the simulation we selected
301 a low sampling coverage of 0.01%, (~ 10 individuals) as our lower margin. This
302 coverage, equivalent to 9% in species richness-based sample completeness, is
303 much lower than that used in previous simulation studies dealing with the
304 undersampling issue, for example ~ 40% by Brocklehurst, Day and Fröbisch (2018),
305 ~ or 30% by Beck, Holloway and Schwanghart (2013). Such a low number of
306 individuals in a sample is actually not uncommon in real-life arthropod studies (e.g.
307 Beck & Kitching 2009; Duan *et al.* 2016), although we are commonly unable to
308 assert the correct number of species in a sampling plot given the associated effort
309 that would be required to completely sample such communities. This is also one
310 reason that simulated groups with known species and abundance distributions were
311 used in this study.

312 Our first simulation confirms that pair-wised results based on CNESS distances are
313 strongly influenced by the distribution of shared species. Previous, empirical studies
314 often calculated species turnover for different values of *m*, following the assumption
315 that a smaller value of *m* emphasizes the similarity of samples with regards to their
316 dominant species (Brehm, Homeier & Fiedler 2003; Axmacher *et al.* 2004; Hilt &
317 Fiedler 2005), while here, we for the fist time analyse in detail the implications of
318 changes in its value across a wide variety of values up to the entire generated
319 species pool. For a small sample size parameter *m*, CNESS distances between
320 treatment assemblages sharing rare species with the control assemblage are much
321 higher than the ones sharing dominant species with the control. This is reflecting the
322 basic probability calculations on which the measure is based, since when taking a

relatively small sample (i.e. a small value of *m*), the probability of any of the *m* individuals belonging to shared species is higher when assemblages share their dominant species rather than their rare ones. Nonetheless, for large values of *m*, CNESS approaches a constant value, i.e. the (chord-normalized) proportion of the shared number of species between two samples in both scenarios (shared rare or common species), explaining the convergence of CNESS dissimilarity values for large values of *m*. This confirms that for small *m* values, results chiefly reflect similarities in the dominant species (e.g. Hilt & Fiedler 2005), while for large *m*-values, the dissimilarity reflects the overall turnover between samples in their underlying species pool, irrespective of the abundance of the individual species within that pool. Altering values of *m* therefore enables researchers to shift the focus from the share of abundant species to the overall species pool. This ability in our view makes CNESS already a superior measure of species turnover patterns, since other, widespread beta-diversity indices only generate one fixed value that is strongly influenced by the underlying species abundance distribution pattern (Beck, Holloway & Schwanghart 2013).

Comparisons of the CV values confirms the robustness of the CNESS measure of compositional dissimilarity across a wide range of scenarios, including in cases where two communities share rare species. In contrast, the high variance particularly of the Chao-Sørensen dissimilarity measure for a low sampling completeness suggests that this index is not suitable to measure compositional dissimilarity in such scenario. Where communities share chiefly their dominant species, most indices show a high change ratio under a low sampling coverage, which means they all do not provide strong representations of the actual dissimilarity between the two samples. Nonetheless, even under this condition, CNESS still performs much better

than Chao-Sørensen and Euclidean distance measures. It needs to be noticed that the performance of Bray-Curtis is likely influenced by the sampling strategy, i.e. its changing ratio showed the same increasing trend with the decrease of sampling coverage under equal sampling strategy (i.e. undersampling for both assemblages, or only for one of the assemblages). However, Bray-Curtis reached a peak in similarity for the 1% coverage under the unequal sampling strategy, i.e. it behaves in a more unstable and unpredictable way across these scenarios when compared to CNESS that shows similar performances under the two sampling strategies.

In this study, we calculated CNESS for different sample size parameters *m* and three widespread beta diversity indices based on simulated datasets of known dissimilarity and using different sampling scenarios, to compare the difference between the different dissimilarity measures. It needs to be stressed that we did not assess how close resulting values were to the “true dissimilarity”. Instead, we focused on the variance and change ratio observed in the indices, since in the ordination or in other visualization approaches used to present the data, plots are commonly grouped by their relative distance or dissimilarity values. A robust prediction of dissimilarities across the different scenarios and under repeat extraction of random samples from underlying assemblages in this context is seen as an absolutely crucial basic criterion (Brehm & Fiedler 2004). In this regard, our results clearly emphasize the suitability and superiority of CNESS in samples of diverging sample sizes. The value of CNESS is sensitive to the distribution structure of shared species, which can be reflected by the changing of the sample size parameter *m*. While being highly useful in studying the compositional difference for overall species assemblage, in many real-life cases, setting *m* to large values comes at the cost of having to remove a number of samples whose overall sample sizes are smaller than *m*. Nonetheless, the

CNESS uniquely allows to address this problem via variations in the sample size parameter according to the respective underlying data structure of the samples that are being compared. We generally recommend researchers to calculate the CNESS (or similar measures such as NESS) dissimilarity for a number of different *m* values to obtain insights both into the share in dominant species and across the overall species pool (see e.g. Brehm, Homeier & Fiedler 2003; or Axmacher *et al.* 2004).

Acknowledges

This study was financially supported by the National Natural Science Foundation of China (31700363) and Jiangsu Science and Technology Programme (BK20181191).

Authors' contributions

YZ and JCA conceived the idea. YZ wrote the script and did the analysis. YZ and J JCA wrote the manuscript.

References

Axmacher, J.C., Holtmann, G., Scheuermann, L., Brehm, G., Müller-Hohenstein, K. & Fiedler, K. (2004) Diversity of geometrid moths (Lepidoptera: Geometridae) along an Afrotropical elevational rainforest transect. *Diversity and Distributions,* **10,** 293-302.

Barwell, L.J., Isaac, N.J. & Kunin, W.E. (2015) Measuring β-diversity with species abundance data. *Journal of Animal Ecology,* **84,** 1112-1122.

Beck, J., Holloway, J.D. & Schwanghart, A. (2013) Undersampling and the measurement of beta diversity. *Methods in Ecology and Evolution,* **4,** 370-382.

Beck, J. & Kitching, I.J. (2009) Drivers of moth species richness on tropical altitudinal gradients: a cross-regional comparison. *Global Ecology and Biogeography,* **18,** 361-371.

Beck, J. & Schwanghart, W. (2010) Comparing measures of species diversity from incomplete inventories: an update. *Methods in Ecology and Evolution,* **1,** 38-44.

Beck, J. & Vun Khen, C. (2007) Beta-diversity of geometrid moths from northern Borneo: effects of habitat, time and space. *Journal of Animal Ecology,* **76,** 230-237.

Brehm, G. & Fiedler, K. (2004) Ordinating tropical moth ensembles from an elevational gradient: a comparison of common methods. *Journal of Tropical Ecology,* **20,** 165-172.

Brehm, G., Homeier, J. & Fiedler, K. (2003) Beta diversity of geometrid moths (Lepidoptera: Geometridae) in an Andean montane rainforest. *Diversity and Distributions,* **9,** 351-366.

Brocklehurst, N., Day, M.O. & Fröbisch, J. (2018) Accounting for differences in species frequency distributions when calculating beta diversity in the fossil record. *Methods in Ecology and Evolution,* **9,** 1409-1420.

Cardoso, P., Borges, P.A.V. & Veech, J.A. (2009) Testing the performance of beta diversity measures based on incidence data: the robustness to undersampling. *Diversity and Distributions,* **15,** 1081-1090.

Chao, A., Chazdon, R.L., Colwell, R.K. & Shen, T.-J. (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters,* **8,** 148-159.

Chao, A., Gotelli, N., Hsieh, T.C., Sander, E., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs,* **84,** 45-67.

Chao, A. & Jost, L. (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology,* **93,** 2533-2547.

Coddington, J.A., Agnarsson, I., Miller, J.A., Kuntner, M. & Hormiga, G. (2009) Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. *Journal of Animal Ecology,* **78,** 573-584.

Duan, M., Liu, Y., Yu, Z., Baudry, J., Li, L., Wang, C. & Axmacher, J.C. (2016) Disentangling effects of abiotic factors and biotic interactions on cross-taxon congruence in species turnover patterns of plants, moths and beetles. *Scientific Reports,* **6,** 23511.

Gallagher, E.D. (1998) Compah96.

Grassle, J.F. & Smith, W. (1976) A similarity measure sensitive to the contribution of rare species and its use in investigation of variation in marine benthic communities. *Oecologia,* **25,** 13-22.

Hilt, N. & Fiedler, K. (2005) Diversity and composition of Arctiidae moth ensembles along a successional gradient in the Ecuadorian Andes. *Diversity and Distributions,* **11,** 387-398.

Hortal, J., Borges, P.A.V. & Gaspar, C. (2006) Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology,* **75,** 274-287.

Hurlbert, S.H. (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology,* **52,** 577-586.

Iknayan, K.J., Tingley, M.W., Furnas, B.J. & Beissinger, S.R. (2014) Detecting diversity: emerging methods to estimate species diversity. *Trends in Ecology & Evolution,* **29,** 97-106.

Legendre, P. & Gallagher, E. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia,* **129,** 271-280.

Melo, A.S. (2014) CommEcol: Community Ecology Analyses. R package version 1.6.5.

Mori, A.S., Isbell, F. & Seidl, R. (2018) β-Diversity, Community Assembly, and Ecosystem Functioning. *Trends in Ecology & Evolution,* **33,** 549-564.

Morisita, M. (1959) Measuring of interspecific association and similarity between communities. *Memoirs of the Faculty of Science, Kyushu University, Series E (Biology),* **3,** 65-80.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Simpson, G., Solymos, P., Stevens, M. & Wagner, H. (2014) vegan: Community Ecology Package. R package version 2.2-0.

R Core Team (2014) *R: A language and environment for statistical computing. Version 3.1.2*. R Foundation for Statistical Computing, Vienna, Austria.

Schroeder, P.J. & Jenkins, D.G. (2018) How robust are popular beta diversity indices to sampling error? *Ecosphere,* **9**.

Stier, A.C., Bolker, B.M. & Osenberg, C.W. (2016) Using rarefaction to isolate the effects of patch size and sampling effort on beta diversity. *Ecosphere,* **7**.

Trueblood, D.D., Gallagher, E.D. & Gould, D.M. (1994) Three stages of seasonal succession on the Savin Hill Cove mudflat, Boston Harbor. *Limnology and Oceanography,* **39,** 1440-1454.

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.

Whittaker, R.H. (1960) Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs,* **30,** 279-338.

Wickham, H. (2011) The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software,* **40,** 1-29.

Yue, J.C. & Clayton, M.K. (2005) A Similarity Measure Based on Species Proportions. *Communications in Statistics - Theory and Methods,* **34,** 2123-2131.

Zou, Y., Sang, W., Zhou, H., Huang, L. & Axmacher, J.C. (2014) Altitudinal diversity patterns of ground beetles (Coleoptera: Carabidae) in the forests of Changbai Mountain, Northeast China. *Insect Conservation and Diversity,* **7,** 161-171.

Figures

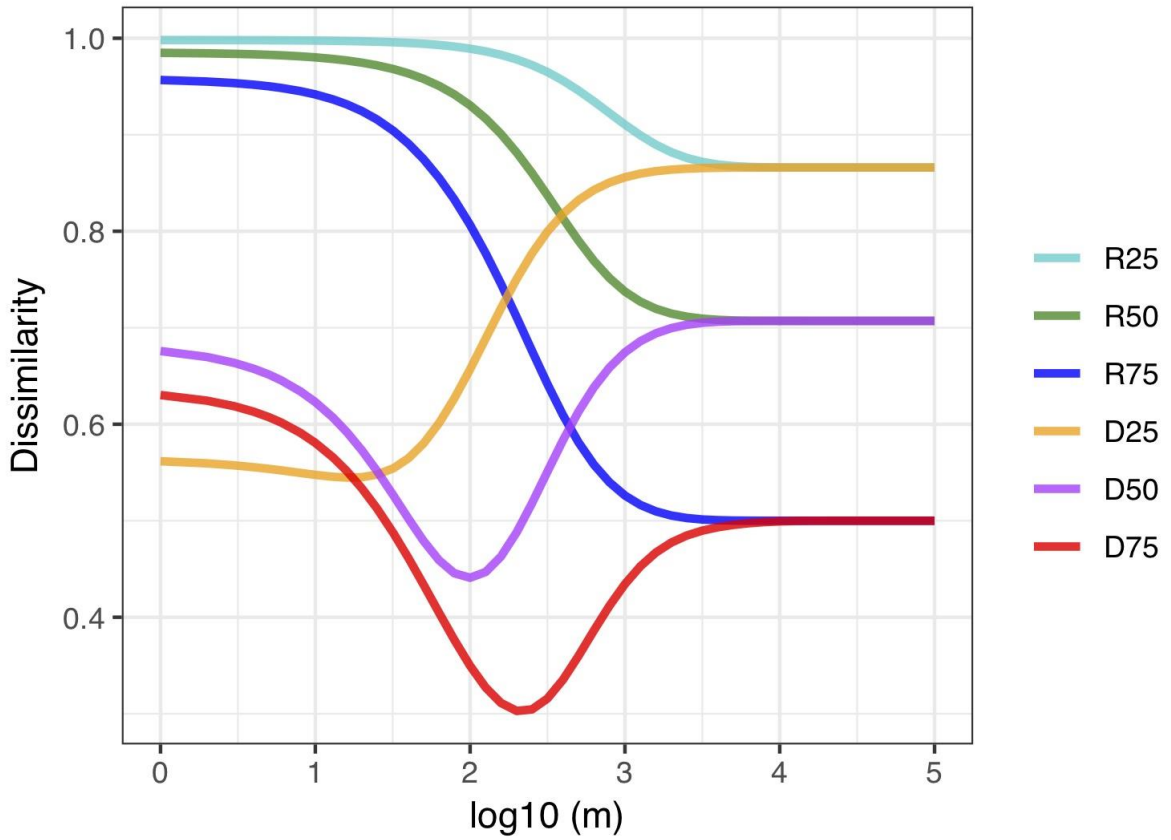
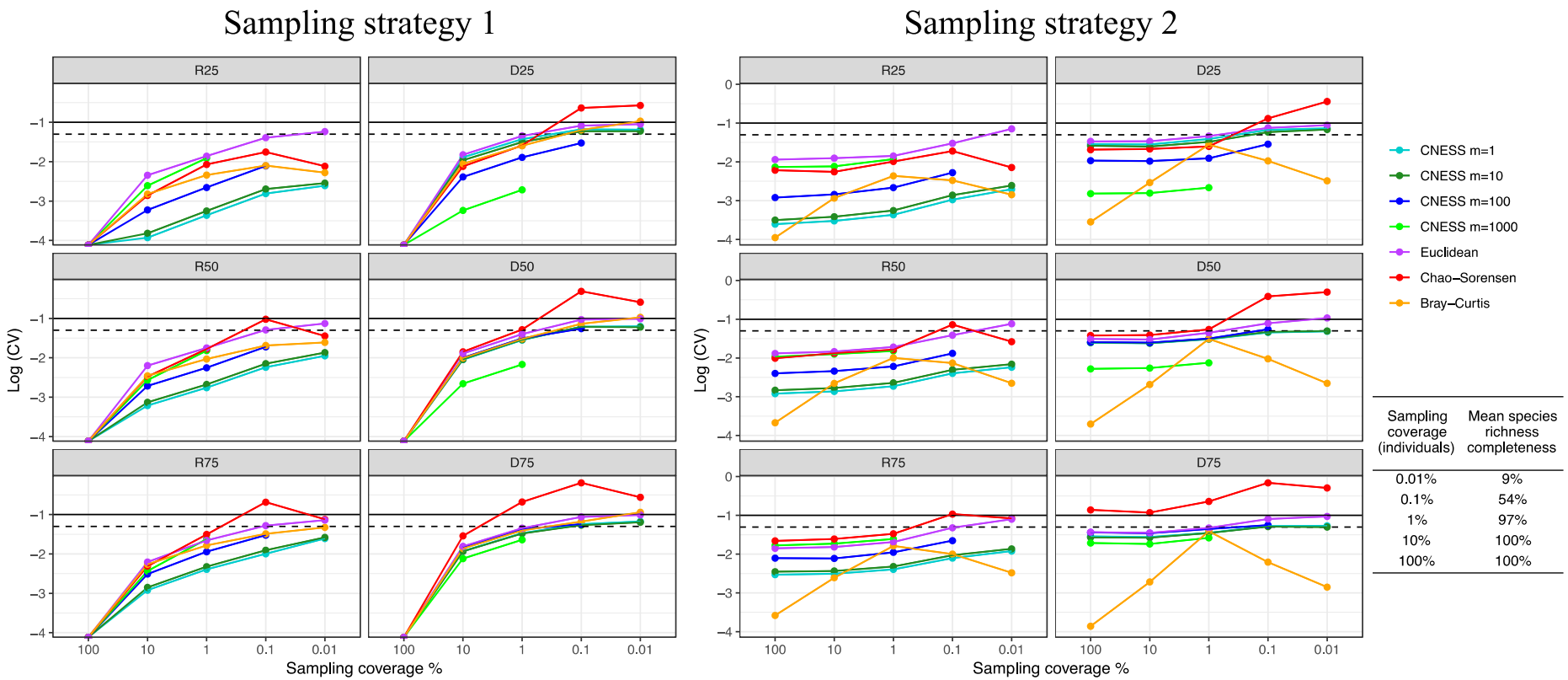
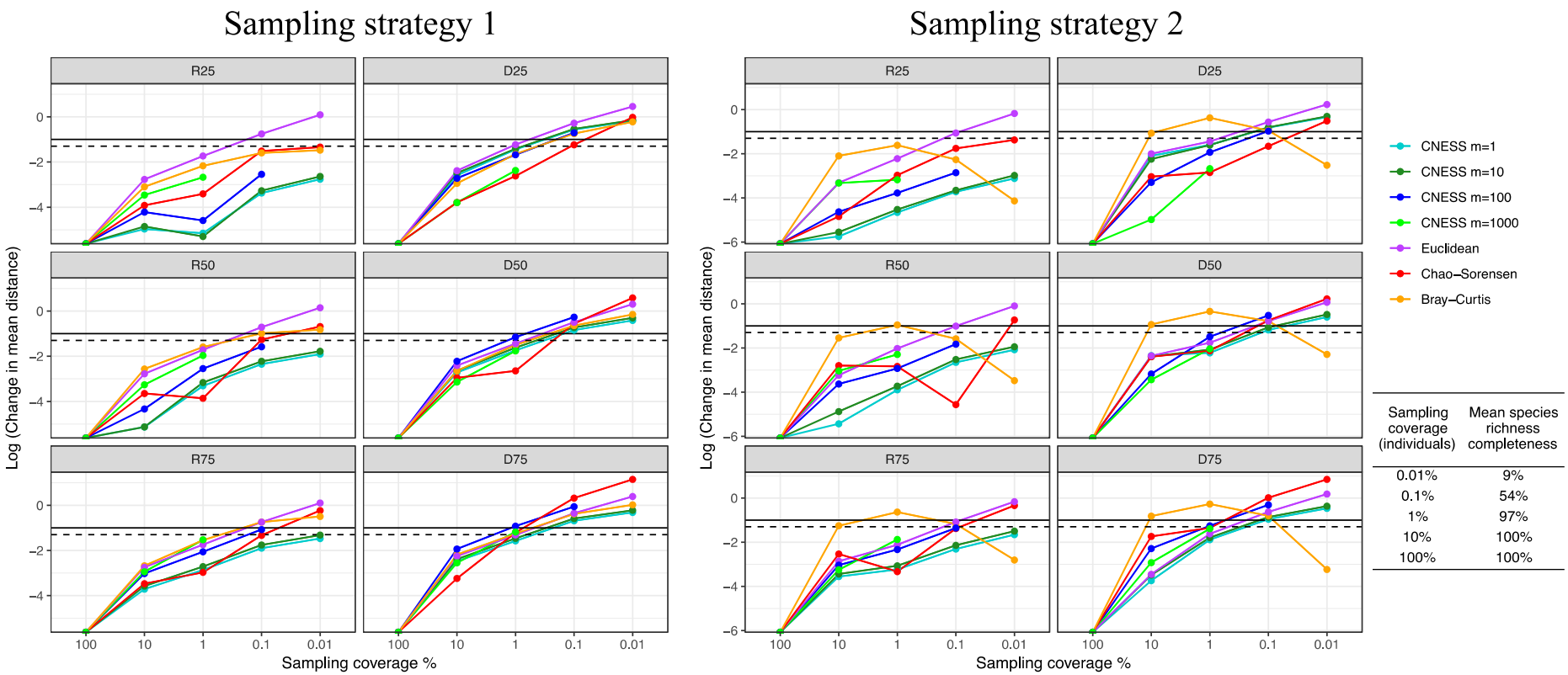


Figure 1. The $CNESS_a$ values calculated between the control and different treatment datasets with different m values. R25, R50 and R75 refer to treatments that share 25%, 50% and 75% of the rare species in the theoretical population, while D25, D50 and D75 refer to the respective share in dominant species with the control.



Sampling coverage (individuals)	Mean species richness completeness
0.01%	9%
0.1%	54%
1%	97%
10%	100%
100%	100%

Figure 2. The coefficient of variation (CV, log10 transformed) based on 1000 permutations for different dissimilarity or distance measures calculated between the different treatments and the control sample for equal sampling (sampling strategy 1) and unequal sampling (sampling strategy 2) for different sampling coverage. Solid and dashed vertical lines refer to 0.1 (log10 value of -1) and 0.05 (log10 value of -1.3) CV values. R25, R50 and R75 refer to treatments that share 25%, 50% and 75% of the rare species in the theoretical population, while D25, D50 and D75 refer to the respective share in dominant species with the control. The table refers to the mean species richness completeness for different sampling coverages calculated based on the control group.



Sampling coverage (individuals)	Mean species richness completeness
0.01%	9%
0.1%	54%
1%	97%
10%	100%
100%	100%

Figure 3: Change in the mean value (log10-transformed) based on 1000 permutations for different indices between treatment and control group for equal sampling (Sampling strategy 1) and unequal sampling (Sampling strategy 2) under different sampling coverage. Solid and dashed vertical lines refer to 10% (log10 value of -1) and 5% (log10 value of -1.3) change. R25, R50 and R75 refer to treatments that share 25%, 50% and 75% of the rare species in the theoretical population, while D25, D50 and D75 refer to the respective share in dominant species with the control. The table refers to the mean species richness completeness for different sampling coverages calculated based on the control group.

Electronic Supplementary Materials

Appendix 1. R scripts to calculate Expected Species Shared (ESS) family

Appendix 2, Mathematical proof for the transformation of ESS formula

Appendix 3. The abundance distribution of simulated species for different “treatment” groups

Appendix 4. Simulation R scripts