A temporal beta-diversity index to identify exceptional sites in space-time surveys

Pierre Legendre*

Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada H3C 3J7.

* Corresponding author: Pierre Legendre, Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. Email: Pierre.Legendre@umontreal.ca

Running headline: Temporal beta diversity analysis

Abstract

Aim This paper presents the foundations and statistical bases for Temporal Beta diversity analysis, a method for comparison of repeated multi-species surveys at the same sites. Surveys of that type are presently done by ecologists around the world. In particular, the paper describes a method (TBI) to test the differences between community data matrices corresponding to observations made at times T1 and T2 in space-time ecological surveys involving several sites. The objective is to identify the sites that have changed in an exceptional way in species composition between T1 and T2.

Innovation The null hypothesis of the TBI test of significance is that a species assemblage is not exceptionally different between T1 and T2. The problem: testing the significance of dissimilarity coefficients is usually not possible because the values in a dissimilarity matrix are interrelated. However, the dissimilarity between T1 and T2 for a site is independent of the dissimilarities that concern T1–T2 data at other sites. The paper shows that it is possible to compute a valid test of significance in that case. The method also allows users to examine the processes of biodiversity losses and gains through time at the different sites in space-time surveys.

Main conclusion Three applications of the method to different ecological communities are presented. This method is applicable worldwide to all types of ecological communities, marine and terrestrial. It will be of value to identify exceptional sites in space-time ecological surveys carried out to study anthropogenic impacts, including climate change. R software is available implementing the method.

Keywords Beta diversity, B-C plots, space-time analysis, statistical power, temporal beta diversity, temporal beta diversity index (TBI), type I error.

Introduction

In several application fields, researchers want to compare observations made at several sites and at two different times. The question of interest is: are there sites where the difference is so important that they do not seem to belong to the same statistical population as the other sites? If the difference is so exceptional as to lead to a statistically significant result, these sites are worth examining in more detail to identify the cause of the differences. The exceptional character of the difference indicates that some atypical process may be occurring there. Here are some examples. In palaeoecology, comparison of ancient and modern diatom communities preserved in lake sediment cores may indicate areas where acute anthropogenic processes have singularly changed the surrounding land use (e.g., Winegardner et al. 2017). When a strong environmental impact has taken place at a known point in time and an ecological community had been surveyed ahead of the impact, ecologists may survey that community again to determine how it was affected by the impact, and then how it may have recovered in later surveys (e.g., Legendre & Salvat 2015). In community ecology, when studying a permanent stem-mapped forest plot divided into regular quadrats, examining surveys made at two different times may indicate quadrats that have been exceptionally affected by a disturbance, e.g. a climatic or anthropogenic event (e.g., Legendre & Condit in prep.). In population genetics, comparing several local populations of a species observed at two different moments separated by an event of interest may indicate the locations

where the event may have had exceptionally strong effects by changing the population genetic structure. Other examples can be found in other fields of biological and biomedical research.

This paper describes a method to test, for several sampling units (objects), the differences between data vectors corresponding to observations made at times T1 and T2. To fix ideas, I will refer to these objects as sites in this paper, although they may be of different natures. The observed data, assembled in matrices Mat.1 for time T1 and Mat.2 for T2, may be of different kinds; in landscape ecology and genetics, the data are community composition or population gene frequencies observed at different sites. The hypothesis (H_0) to be tested is that a site is not exceptionally different between T1 and T2, compared to other sites in the study that have been observed at the same two times, and belongs to the same statistical population as the other sites.

Tests of significance for dissimilarity coefficients (D) are usually not possible because the D values in a dissimilarity matrix are obtained from the computation of an index between all pairs of objects, e.g. sites in ecology (their number is n), and are thus interrelated, each site contributing to (n – 1) of the dissimilarities in the half-matrix of dissimilarities. In T1–T2 comparisons for individual sites, however, the dissimilarity between T1 and T2 for a site is independent of the dissimilarities computed for T1–T2 data at other sites. So it may be possible to work out a valid test of significance in that case. That test would be of value to identify exceptional sites, which may have a large dissimilarity for different reasons; these reasons may be worth investigating. If the number of sites is large, investigators may want to focus only onto the sites that produce exceptionally (in the statistical sense) large dissimilarity values in T1–T2 comparisons. A dissimilarity D computed between times T1 and T2 for a site, using community composition or gene frequency data for example, is called a *Temporal Beta-diversity Index* (TBI); it measures the change in community composition (or temporal beta diversity) from T1 to T2. A change through time is directional; something (e.g. species, species abundances, gene frequencies) has been gained and/or lost between T1 and T2, and these two components are both of interest to understand the change.

Methods

TBI computation and testing

The proposed method consists basically in the following steps: a dissimilarity index is computed for each site between the data vectors corresponding to T1 and T2, then the indices are tested for significance using a permutational procedure. Two of the dissimilarity indices that can be used in this type of analysis also allow the computation of species losses and gains at each site between T1 and T2. These data provide users with detailed information, at the site level, about the response of the community to the events that occurred between T1 and T2.

1. Compute Temporal Beta-diversity Indices (TBI)

Consider two data matrices, **Mat.1** and **Mat.2**, about the same objects, each one with n sites as rows and the same p variables as columns (Fig. 1). Individual values may be noted $y_{ij}.1$ and $y_{ij}.2$. Compute the dissimilarity $D(\mathbf{y}_i.1, \mathbf{y}_i.2)$ between the two vectors of values, $\mathbf{y}_i.1$ and $\mathbf{y}_i.2$, for each site i. These n dissimilarities form a vector of length n.

The percentage difference dissimilarity ($D_{\%diff}$; method "%difference" in the R function TBI.R, also known as the Bray-Curtis index in other computer packages), and the Ružička dissimilarity (D_{Ruz}; method "ruzicka" in the R function) can be used for beta diversity assessment.

They are obtained by computing a dissimilarity function (equations shown below). With presence-absence data, the percentage difference produces ($1 - S_{Sørensen}$) dissimilarity whereas the Ružička dissimilarity produces ($1 - S_{Jaccard}$), where S designates similarity.

The chord, Hellinger, and log-chord distances are members of the Box-Cox family of distances (Legendre & Borcard 2018). They are also classical indices for beta diversity studies (Legendre & De Cáceres 2013). These indices, as well as the Euclidean distance, are also implemented in the TBI.R function and will be used in the simulations and ecological applications below.

When the percentage difference or the Ružička dissimilarity are used as TBI indices, one can compute two derived indices to study the directionality of the change through time at each site, as proposed by Legendre and Salvat (2015). Consider data vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ corresponding to the multi-species observations at T1 and T2 for a site. The following calculations can be done:

• a_j is the part of the abundance of species j that is common to the two survey vectors: $a_j = \min(y_{1j}, y_{2j})$. A is the sum of the a_j values for all species. It represents the unscaled *similarity* between two surveys.

• b_j is the part of the abundance of species j that is higher in survey 1 than in survey 2: $b_j = y_{1j} - y_{2j}$. B is the sum of the b_j values for all species. It is the unscaled sum of *species losses* between T1 and T2.

• c_j is the part of the abundance of species j that is higher in survey 2 than in survey 1: $c_j = y_{2j} - y_{1j}$. C is the sum of the c_j values for all species. It is the unscaled sum of *species gains* between T1 and T2.

(B+C) represent the unscaled dissimilarity. The values A, B and C are the building elements of the percentage difference, $D_{\%diff} = (B+C)/(2A+B+C)$, and the Ružička dissimilarity, $D_{Ruz} = (B+C)/(A+B+C)$ (Legendre 2014). ($B - C$) indicates the directionality of the process of losses and gains of individuals of the different species between the two surveys. B and C can be scaled by division by a denominator *den*, which is ($2A+B+C$) for $D_{\%diff}$ and ($A+B+C$) for D_{Ruz} case. The $D_{\%diff}$ and D_{Ruz} dissimilarities measure the temporal beta diversity for a site. The scaled B and C statistics can be called D_{loss} and D_{gain}, where $D_{loss} = B/den$ and $D_{gain} = C/den$. An interesting relationship is that $D_{loss} + D_{gain} = D_{\%diff}$ or D_{Ruz}, depending on the denominator *den* that is used. In other words, D_{loss} and D_{gain} partition the $D_{\%diff}$ and D_{Ruz} dissimilarities into *loss* and *gain* components. Values of these indices are in the [0,1] range and are thus directly comparable. The loss and gain statistics can be computed for occurrence (i.e. presence-absence) data as well, because $D_{\%diff}$ becomes the Sørensen dissimilarity with occurrence data and D_{Ruz} becomes the Jaccard dissimilarity, as mentioned above.

What are the ecological applications of D_{loss} and D_{gain}? For each site, one can explore which process, between D_{loss} and D_{gain}, presents the largest contribution to the temporal $D_{\%diff}$ dissimilarity; in other words, which process is dominant at each site. The means of the D_{loss} and D_{gain} components across the sites express the dynamics of the community over all sites. For observations across a large number of sites within a region, or in all quadrats of a stem-mapped dynamics forest plot, the B/den and C/den statistics can be mapped, subjected to canonical analysis (see Ecological application 2), plotted as B-C plots (see subsection 4 below and Ecological application 3), or studied in other ways to understand the differences among the study sites.

2. Testing procedure

To test the significance of TBI indices, the data are permuted at random in both matrices and the indices are recomputed; this procedure is repeated a large number of times and a p-value is computed for the difference between T1 and T2 at each site. Permutations can be done in several ways. A simulation study will compare three permutation methods.

Permutation of species independently of one another is the method used to assess the significance of *Local Contributions to Beta Diversity* (LCBD indices) in the Legendre & De Cáceres (2013) paper. The same logic is followed here in permutation methods 1 and 2. Simulation results (below) will show that this method produces tests with higher power than method 3, the permutation of entire data rows.

2.1. Permutation method 1 –

Permute the raw abundance data at random within each column separately, but in the same way in the two matrices corresponding to T1 and T2.

2.1.1. In each matrix, the original values (e.g. species abundances) are permuted at random, independently in each column. Permutation of the two matrices is started with the same random seed, so that the values in each column (e.g. species) are permuted in the same way in **Mat1.perm** and **Mat2.perm**. With this method, it is the *differences* in values between T1 and T2, for each species, that are permuted at random among the sites. The justification is that we are testing dissimilarities, obtained by combining the species differences between T1 and T2.

2.1.2. The transformation, if any (in the case of chord, Hellinger or log-chord dissimilarities), is recomputed on the permuted data matrices. This is necessary to make sure that the permuted data are transformed in the same way as the initial data, with row sums or row norms of 1. In this way, the D_i of the permuted data will remain comparable to the reference D_i.

2.1.3. The TBI distances between T1 and T2 are recomputed, for each site separately.

2.1.4. After a large number of permutations, a p-value is computed for site i (hence for each $D(\mathbf{y}_{i\cdot 1}, \mathbf{y}_{i\cdot 2})$ index), in the same way as in any permutation test. A correction for multiple testing is applied to obtain a correct experimentwise error rate.

Some technical aspects of permutation method 1 are discussed in Appendix S2.

2.2. Permutation method 2 –

A variant over method 1 is to permute each species independently, as in method 1, without worrying about using the same permutation for species j in matrices **Mat.1** and **Mat.2**. If that method has the same power as method 1, or better, it would lead to simpler code.

2.3. Permutation method 3 –

Another possible method is to permute entire rows of **Mat.1** and **Mat.2**, independently in these two matrices, as it is done in several permutational statistical procedures. The statistical hypothesis under test and the permutation set differ from those in methods 1 and 2 where each species is permuted independently. It will be included in the simulation study only because this method is widely used in multivariate data analysis.

2.4. Permutation method 4 –

• If the sites are part of a geographic broad-scale gradient on a map and spatial autocorrelation is considered to be a salient property of the data, each species could be permuted in a toroidal manner to preserve the spatial autocorrelation of the data. This option is not implemented at the moment in the TBI.R calculation function.

3. BCD computation

When the percentage difference or the Ružička dissimilarity are used as TBI indices, B is the unscaled sum of *species losses* and C is the unscaled sum of the *species gains* between T1 and T2. The unscaled statistics can be scaled to values in the [0,1] range by division by the percentage difference denominator $den = (2A+B+C)$ or by the Ružička denominator $den = (A+B+C)$. The *dissimilarity* D is $(B/den + C/den) = (B+C)/den$. If the TBI dissimilarity is either the percentage difference or the Ružička dissimilarity, one can take advantage of that decomposition of D by listing the B/den and C/den components of TBI indices for each site in the study. These basic statistics can be used in two different ways:

3.1. We can compute summary statistics: the mean of (B/den), the mean of (C/den) and the mean of $D = (B+C)/den$. The following relationship holds: $\text{mean}(B/den) + \text{mean}(C/den) = \text{mean}(D)$. From this decomposition of D, we can derive the contribution of the species losses to the total dissimilarity, $B/(B+C)$, and similarly the contribution of the species gains to the total dissimilarity, $C/(B+C)$. These two ratios sum to 1, providing the relative importance of the species losses and gains phenomena. The result is the same for calculation without a denominator *den*, or with either the percentage difference or the Ružička denominator.

3.2. For each site, we can also obtain the sign of the difference (gains – losses), or $(C - B)$: if B > C, we note a minus sign (–), and if B < C we note a plus (+) sign. This notation allows users to quickly identify the sites where losses or gains dominate. Similarly, the difference $\text{mean}(C/den) - \text{mean}(B/den)$ is computed; its sign tells us if gains (+ sign) or losses (– sign) dominate across all sites. The significance of the difference between the two vectors of statistics B/den and C/den can be computed using a parametric or permutational paired t-test; the R function mentioned in subsection 5 below computes both forms. These tests provide overall indications of the direction of change in community composition over all sites. They help confirm the asymmetry between abundance or occurrence losses (B/den) and abundance or occurrence gains (C/den). In Ecological application 2 (Tikus Island coral communities), the two forms of calculation provided complementary information.

4. B-C plot

We can also use the B/den and C/den statistics as coordinates of points (representing sites) in bivariate graphs with B/den in the ordinate and C/den in the abscissa. We call these graphs *B-C plots*. They display visually the relative importance of the loss and gain processes across the study sites, informing researchers about the detailed and global structure of the species losses and gains.

A B-C plot is presented in Ecological application 3 (Chesapeake Bay benthos data). In that B-C plot, a diagonal green line, with slope of 1, was drawn through the origin; it represents the theoretical positions of sites where D_{gain} would be equal to D_{loss}. A red line was also drawn parallel to the green line, passing through the centroid of all points. When the red line is below

the green line, it indicates that the survey interval was dominated by species losses across the sites, and the opposite if the red line is above the green line. Points found higher in the plot towards the upper-right corner represent higher temporal beta diversity than points found lower in the direction of the lower-left corner.

Sites found at the highest diagonal margin of the distribution of points, in the direction of the upper-right corner of the plot, have high *D* values (beta differentiation). In most cases, this happens because communities have undergone great changes from T1 to T2. High *D* values may also be found at sites that contain very few species and individuals. This situation is discussed in Ecological application 3, where four such sites are found. Users should check the number of species and individuals involved in the dissimilarity calculation of these sites before drawing ecological conclusions.

In B-C plots, the points representing sites can be labelled with colours or symbols representing the types of environment, the geographic areas where they come from, or any other independent classifier of interest. Separate B-C plots can be drawn for sites surveyed in different types of environment, although all sites have been analysed in the same TBI analysis. Comparison of these plots will immediately show which types of environment have produced mostly losses or gains in species occurrences or abundances. Ecological application 3 shows a B-C plot with the sites separated in two classes of a temperature classifier.

5. Software

These calculations are implemented in the TBI() function in R, presently available on the Web page http://adn.biol.umontreal.ca/~numericalecology/FonctionsR/. Examples of output files of the TBI function are shown in Appendices S3 and S4.

Numerical simulations

Numerical simulations were used to check the type I error rate and power of the permutation methods described in subsection 2 above. The data simulation methods and results are described in detail in Appendix S1. A summary of these results is presented here, with recommendations to users of the method.

Simulation to estimate type I error rates

The simulation results reported in Appendix S1 show that the TBI tests had correct rates of type I error with the three permutational testing methods, for the two community-like data generation methods (Poisson and lognormal) and all dissimilarity indices used, and this for all significance levels (α) considered, from $\alpha = 0.01$ to $\alpha = 0.50$.

Simulations to compare power of *D* indices and permutation methods

For the analysis of community composition data, permutation methods 1 and 2 are equally appropriate (Figs. S1.5 and S1.6, Appendix S1). The percentage difference and Ružička indices produced tests with the highest power, followed by the indices in the Box-Cox family: the chord, Hellinger and log-chord distances. The Euclidean distance alone produced TBI tests with extremely low power. It should not be used for TBI tests of community composition data.

The best combination to obtain TBI tests of community data with maximum power is to use the percentage difference or the Ružička indices with permutation methods 1 or 2. These two dissimilarities can also be decomposed into species losses (*B*/*den*) and gains (*C*/*den*), which can be used to examine the processes of losses and gains at the site level and to produce B-C plots.

For standardized environmental variables, only the Euclidean distance was tested in the simulation study because this is the only one that makes sense with this type of data. The simulation results clearly showed that permutation method 1 produced the highest power with simulated quantitative environmental data. It would be the testing method of choice for this type of data.

Additional simulations involving different numbers of sites with an effect and different total numbers of sites showed that power of the test with permutation method 1 was high as long as the proportion of sites with an effect was smaller than $n/2$, independently of the total number of sites n in the study (Fig. S1.7, Appendix S1).

Warning – In real ecological studies, when the TBI test is applied to data where some sites are highly impoverished due to pollution or other extreme environmental situations, whereas other sites have higher species richness, this may result in sites with very few species and no species in common in the T1–T2 comparisons due to sampling variation at these impoverished sites. The TBI test will indicate a significant difference between T1 and T2 for these sites and this is a legitimate statistical outcome. When users of the method identify sites showing significant TBI tests in real data, they should check the species composition of these sites at T1 and T2. Interpretation of the test results should be done with caution when high and significant TBI indices are associated with community composition vectors with low richness and no species in common between T1 and T2. Examples are found in Ecological application 3, Chesapeake Bay data.

Application to physical environmental or community trait data

It could be interesting to determine in what sites the changes in environmental data (e.g. land use) were the most important. One could then determine if these sites are also those for which the community has changed the most. Functional trait matrices could also be analysed in that way (Laliberté & Legendre 2010) in order to determine at which sites the trait composition of the community has been altered the most.

Use the TBI method to compare two matrices containing the same environmental variables observed at T1 and T2. This is a situation where the Euclidean distance would be appropriate as a basis for computing a TBI index. Data preparation:

• If all environmental variables are quantitative, they should be standardized before they are used in TBI analyses to make sure that all variables have the same weight (i.e. the same variance) in the calculation of TBI indices. The correct way of standardizing the variables is to put them in a single data table, **Y** = rbind(Y.T1, Y.T2); standardize **Y** by columns [Y.stand = scale(Y)]; then separate the two tables before TBI analysis. In that way, the differences in values of each variable for all pairs of sites in the two tables will remain comparable to the original differences in unstandardized values and the distances computed between sites in T1 and T2 will be meaningful. Appendix S6 contains an R function to carry out this special standardization.

298 • If the environmental data contain a mixture of quantitative and qualitative data, one could put
299 the two data tables together as above, then compute the Gower dissimilarity using the gowdis(**Y**)
300 function of package FD, which can handle mixtures of quantitative and qualitative variables, and
301 finally apply principal coordinate analysis (PCoA) to the square-rooted Gower dissimilarities.
302 Square-rooting should make a Gower **D** matrix Euclidean before PCoA; see Legendre &
303 Legendre (2012, Table 7.2). These operations will produce a table of principal coordinates, which
304 can be split in two matrices and used as input into TBI analysis. No standardization of these data
305 matrices will be required.

306 • For community trait matrices, use the same method: Gower dissimilarity, PCoA of the square-
307 rooted dissimilarities, split the principal coordinates in two matrices, compute TBI using the
308 Euclidean distance.

309 No application of TBI analysis to environmental or trait data is presented in this paper to
310 save space.

311

312 ## Ecological applications

313 ### Ecological application 1 – Insecticide treatments in mesocosms

314 The invertebrate insecticide treatment data, from van den Brink & ter Braak (1999), are described
315 in Appendix S3. We will compare data of surveys #4 and #11. Survey #4 was done one week
316 after the insecticide treatment; then, the fauna of the mesocosms was considered by the authors to
317 have fully recovered from treatment at the time of survey #11. For example, in the two
318 mesocosms that had received the highest insecticide doses, species richness increased by 9 and 19
319 species from survey #4 to #11.

320 All TBI dissimilarities showed that in the mesocosms with the highest insecticide doses,
321 community compositions was the most different between T1 and T2 (Table 1, upper panel). The
322 p-values were identical for the percentage difference and Ružička dissimilarities (Table 1, lower
323 panel); the two mesocosms that had received the highest doses of the insecticide, M11 and M12,
324 showed significant differences in community composition between surveys #4 and 11. The chord,
325 Hellinger and log-chord distances led to the same conclusion. These five distances are deemed
326 appropriate for beta diversity study (Legendre & De Cáceres, 2013). On the contrary, the
327 Euclidean distance is known to be inappropriate for such studies and, indeed, tests based on that
328 distance did not show significant differences in community composition between surveys #4 and
329 #11 in any of the mesocosms.

330 Detailed analysis of the species losses (*B/den*) and gains (*C/den*), obtained from TBI
331 analysis computed with the percentage difference (Appendix S3), showed that in the 8 treated
332 mesocosms, the changes in community composition always consisted of species gains; that is,
333 statistic *C/den* (gains) was always larger than *B/den* (losses). Analysis of the mean values of
334 *B/den* and *C/den* for these 8 mesocosms showed that *C/den* represented 58% of the
335 dissimilarities, as expected in a study of recovery after an insecticide treatment. The
336 permutational paired *t*-test showed a significant difference ($p = 0.0074$) between T1 and T2
337 across the 8 mesocosms (additional calculations, not shown in Appendix S3).

TBI calculations using the Sørensen dissimilarity (Appendix S3) indicated that, in addition to mesocosms #11 and 12, mesocosm #10 (treated with 6 µg/L of insecticide) also displayed a significant difference between T1 and T2 in species occurrence data.

Further analyses were run with the three permutation methods proposed in this paper, using the percentage difference dissimilarity and 999 random permutations for the tests of significance. Only the results for mesocosms M10 to M12 are reported because all other p-values were 1.0 after correction for multiple testing.

• Permutation method 1 – The last three corrected p-values were 0.230, 0.012 and 0.012; M11 and M12 were significant at alpha=0.05.

• Permutation method 2 – The last three corrected p-values were 0.470, 0.012 and 0.044; M11 and M12 were significant at alpha=0.05.

• Permutation method 3 – The last three corrected p-values were 1.000, 0.120 and 0.759; M10, M11 and M12 were **not significant** at alpha=0.05.

Permutation method 1 had the highest power to detect changes in species composition whereas method 3 lacked power; it did not detect the changes between T1 and T2 at any site. These observations are in agreement with the simulation results reported in a previous section and in Appendix S1 of this paper, which showed that method 3 had very low power with community composition data.

Ecological application 2 – South Tikus Island coral communities

Data on the abundances of 75 coral species at 10 sites in the island of South Tikus, Indonesia, are described in Appendix S4. We will examine the changes in community composition between the 1981 survey and all five following surveys: 1983, 1984, 1985, 1987 and 1988. This study is not meant to identify sites that were exceptionally different between two years or test specific hypotheses about them because specific environmental conditions at each site have not been reported for each year. Instead of testing the TBI statistics for individual sites, we will carry out a detailed study of the species loss (*B*/*den*) and gain (*C*/*den*) statistics, as described in the Methods. These statistics were computed with the denominator (*den*) of the percentage difference index; they decompose the percentage difference into additive components.

First, we will plot the mean values of *B*/*den*, *C*/*den* and *D* statistics across the sites, in comparisons of the 1981 survey with all successive surveys in turn (1983, 1984, 1985, 1987 and 1988) (Fig. 2). This method of analysis was used by Legendre and Salvat (2015, Fig. 3), who described the effects of a nuclear test on the mollusc communities of an atoll in the Pacific. Here, we are studying the effect of an El Niño event on coral communities.

Fig. 2a shows the changes in *D* between years, and its components *B*/*den* and *C*/*den*. We observe that after El Niño, species losses (*B*/*den*) dominated the changes, accounting for 96% of the dissimilarities (*D*) between 1981 and 1983; species gains (*C*/*den*) represented only 4% of *D*. In later years, the species losses decreased. The asymmetry between *B*/*den* and *C*/*den*, with dominance of *B/den* (losses) over *C*/*den* (gains), was significant for all year pairs in Figs. 2a and 2b, as shown by the overall paired *t*-tests of the asymmetry, described in the Methods, computed for each year pair over the 10 study sites, which were all significant.

Does that mean that some of the species that had disappeared had recovered, or that only the species that remained had increased their abundances-per-species? The answer is found in Fig. 2b, which displays the same statistics, computed for species occurrence data. That second graph shows that many species disappeared at first from the surveyed sites after El Niño (*B*/*den* was 77% for the 1981-1983 comparison), then some of the original species recovered on the surveyed reefs (*B*/*den* decreased to 62% for 1981-1984 and to 45% for 1981-1985). Some of the coral species reestablished themselves at the surveyed sites during the following years, possibly by budding from colonies that had survived at nearby sites, or by dispersion of larvae from elsewhere. During that time, new species that were not present in 1981 occupied the depleted reefs, starting in the 1981-1983 comparison (*C*/*den* = 6%) and increasing in the following years (17% for 1981-1984 and 19% for 1981-1985). Observed changes in abundance-per-species and in occurrence became small, possibly caused by sampling variation.

The overall similarity in community composition between the years can be appreciated in a RDA biplot, where the centroid of each year is shown surrounded by the 10 site observations of that year (Fig. 3). This biplot was produced as follows: first, a percentage difference matrix was computed among all years and sites, square-rooted to make the dissimilarity matrix Euclidean, and subjected to principal coordinate analysis (Gower 1966). The entire matrix of principal coordinates was used as the response data in a RDA against a factor representing the 6 survey years of the study. This form of canonical ordination is called distance-based redundancy analysis (dbRDA, Legendre and Anderson 1999). The figure shows that the sites in 1981 had quite different species composition than in surveys after El Niño. The communities moved to a position in the ordination quite distant from 1981 after heavy species losses, then it moved to a new position in 1984 after it recuperated some of its former species, plus some new species that were not present in 1981 and 1983. It moved again to a new position in 1985. From then on, the changes observed in 1987 and 1988 seem to represent random variation due to observed random losses and gains of species, which may be due in part to sampling variation and in part to random species losses and gains.

As in the Legendre and Salvat (2015) study, where the effect on communities was due to a man-made disturbance, the communities found in South Tikus after the natural El Niño event differed in species composition from the structure they had in 1981 and they kept changing, apparently randomly, in later years. These observations are compatible with the neutral theory of evolution of communities.

Ecological application 3 – Chesapeake Bay data

The data of the Chesapeake Bay Benthic Monitoring Program are described in Appendix 5. They concern 205 benthic species caught at 27 sites in the Chesapeake Bay, sampled spring and fall during 13 years from 1996 to 2008. For the present example, we will concentrate on the faunal data of the 25 brackish sites observed during the fall surveys conducted in 2005 and 2008. 52 species were observed in these two years: 38 in 2005 and 45 in 2008, with an overlap of 31 species found in both years.

This example offers the opportunity to build a B-C plot described in section 4 of the Methods section. The percentage difference index was used; the Ružička index would have produced similar results. These data will be used to demonstrate how to draw B-C plots and how to interpret them.

For the pair of years 2005 and 2008, the B-C plot is shown in Fig. 4. In the plot, the red line is *above* the green line. This indicates that *gains* in benthic abundances-per-species dominated *losses* in the Chesapeake (fall surveys) from 2005 to 2008.

A simple classification of the sites by an environmental factor, water temperature, was used to separate the sites in two groups, providing an example of the kind of information that can be derived from displaying different habitat groups as symbols or colours in B-C plots. In the present example, the dispersion of the sites shows a strong relationship between water temperature and the gains and losses of species. The analysis and B-C plots could have been repeated on species occurrence (presence-absence) data.

In addition to the computation of the *B*/*den* and *C*/*den* components at each site, the R function also computed TBI tests of significance of the difference between years at each site. Although this is not the prime purpose of this example, let us mention that four sites were significant at the 0.05 level after Holm correction for multiple testing (25 simultaneous tests). They are sites S22, S23, S52 and S71. These 4 sites, shown in Fig. 4, all had a TBI dissimilarity $D = 1$, no species in common between T1 and T2, and very few species present: 7 species at T1 and 1 species at T2 for site S22, 3 and 3 for S23, 0 and 3 for S52, 4 and 7 for S71. Due to the small numbers of species and individuals at these sites, the test results should not be taken to represent strong evidence of an important change in community composition. These $D = 1$ results could be due to sampling variation.

A map of the 25 brackish sites on the Chesapeake Bay, plotted with the RgoogleMaps package, is shown in Fig. S5.1 (Appendix S5) of the present paper. In the map, signs on the symbols indicate the sites dominated by abundance-per-species gains and losses between 2005 and 2008. The site identification numbers are those found in the data base.

Discussion

TBI analysis and B-C plots are useful to identify exceptional sites in space-time ecological surveys carried out to study the effects of natural and anthropogenic impacts, including the effects of climate change on natural communities and other types of biodiversity data.

The method was elaborated while different applications involving parts of the method were being developed. Some of these applications have already been published in papers that offered opportunities to develop the TBI theory and software, providing pertinent application questions and data.

• Impact of a field experiment – The loss (*B*/*den*) and gain (*C*/*den*) statistics were first analysed by Legendre & Salvat (2015) to compare community composition data (marine molluscs) during 30 years, before and after a man-made disturbance on an atoll in the Pacific. This disturbance to the mollusc community was the atmospheric test of a Hydrogen bomb in 1968.

• A palaeoecological study – Winegardner et al. (2017) compared diatom communities in lake sediment surveyed 150 years apart across the USA. Temporal beta diversity varied significantly as a function of forest cover, with higher temporal beta in watersheds with contemporary lower forest cover.

• Space-time freshwater ecology – Kuczynski et al. (2018) compared freshwater fish surveys 20 years apart in rivers throughout France. They observed biotic homogenization over time in fish

communities. Changes in community composition mainly resulted from population declines and were favoured by an increase in temperature seasonality and in non-native species density.

• Forest ecology – Legendre & Condit (to be submitted) computed and analysed B-C plots for six habitat types, comparing tree community composition (abundance data) from the surveys conducted 30 years apart, in 1985 and in 2015, in the Barro Colorado Island Forest Dynamics Plot in Panama, divided into 1250 (20 m × 20 m) quadrats.

Other methods can be used to further our understanding of the difference between surveys conducted at T1 and T2.

• The space-time interaction (STI) can be tested in repeated surveys without replication, using the STI method (Legendre et al. 2010). When no significant interaction is found between space and time in multivariate community data, (a) we should not expect to identify sites that have exceptional values of TBI except for type I error cases. The test of space-time interaction can thus be considered as a global test of the STI indices in the comparison of two surveys carried out at times 1 and 2. (b) In that case, STI analysis allows us to test the overall difference between times, using MEM eigenfunctions as covariables. For two surveys only (T1, T2), testing the interaction with the STI method requires, however, that the coordinates of the sites be known. If they are not, the interaction cannot be tested.

• One can plot maps of the D_i dissimilarities, computed at all sites i, to assess their variation. Analysis of the spatial variation of the D_i values can be conducted, using for example a correlogram or variogram, kriging, or Moran's eigenvector map analysis described for example in Legendre & Legendre (2012, Chapter 14 and references therein).

• For field experiments with multiple treatments, as shown in Ecological application 1, TBI analysis is complementary to principal response curves (PRC), a method developed by van den Brink & ter Braak (1999) to analyse the results of experiments conducted over time involving multivariate response data (e.g. community composition data) and multiple treatments. The data in Ecological application 1 were those used by van den Brink & ter Braak (1999) to illustrate the PRC method.

• To determine what are the species responsible for the main changes in community composition, carry out RDA on chord, Hellinger, or log-chord-transformed community data on the two data matrices combined, **Y** = rbind(Y.T1, Y.T2), with the binary factor "T1, T2" as the explanatory variable, and a factor representing the sites as covariable. The RDA plot will have a single canonical axis (abscissa) and the species more abundant at T1 and T2 will have long arrows pointing left or right.

• Other methods for space-time analysis have been described in the Legendre & Gauthier (2014) review paper.

The simulation study reported in this paper could be extended to include other aspects not covered in the simulations reported in Appendix S1; for example simulations of community composition data forming spatial gradients, which should be tested using torus permutations. Simulation ventures of that kind would represent nice projects for honour or M.Sc. students in statistical ecology or in statistics.

Acknowledgements

I am thankful to Amanda Winegardner and Lucie Kuczynski who experimented with this testing procedure and computed it on real data during the development phase of this paper, and to Daniel Borcard and Marie-Hélène Brice who provided useful comments on the manuscript before submission. Lucie Kuczynski suggested the acronym TBI for the method. This research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) research grant no. 7738 to P. Legendre.

References

Gower, J. C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338.

Kuczynski, L., P. Legendre, P. & Grenouillet, G. (2018) Concomitant impacts of climate change, fragmentation and non-native species have led to reorganization of fish communities since the 1980s. *Global Ecology and Biogeography* **27**, 213–222.

Laliberté, E. & Legendre, P. (2010) A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **91**, 299–305.

Legendre, P. (2014) Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography* **23**, 1324–1334.

Legendre, P. & Anderson, M.-J. (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* **69**, 1–24.

Legendre, P. & Borcard, D. (2018) Box-Cox-chord transformations for community composition data prior to beta diversity analysis. *Ecography* **41**, 1–5. http://dx.doi.org/10.1111/ecog.03498.

Legendre, P. & Condit, R. Spatial and temporal analysis of beta diversity in the Barro Colorado Island forest dynamics plot, Panama. *Forest Ecosystems* (to be submitted).

Legendre, P. & De Cáceres, M. (2013) Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters* **16**, 951–963.

Legendre, P., De Cáceres, M. & Borcard, D. (2010) Community surveys through space and time: testing the space-time interaction in the absence of replication. *Ecology* **91**, 262–272.

Legendre, P. & Gauthier, O. (2014) Statistical methods for temporal and space-time analysis of community composition data. *Proceedings of the Royal Society B* **281**, 20132728.

Legendre, P. & Legendre, L. (2012) *Numerical ecology, 3rd English edition*. Elsevier Science BV, Amsterdam.

Legendre, P. & B. Salvat, B. (2015) Thirty-year recovery of mollusc communities after nuclear experimentations on Fangataufa atoll (Tuamotu, French Polynesia). *Proceedings of the Royal Society B* 2**82**, 20150750.

van den Brink, P. J. & ter Braak, C. J. F. (1999) Principal response curves: analysis of time-dependent multivariate responses of biological community to stress. *Environmental Toxicology and Chemistry* **18**, 138–148.

Winegardner, A. K., Legendre, P., Beisner, B. E. & Gregory-Eaves, I. (2017) Diatom diversity patterns over the past c. 150 years across the conterminous United States of America: identifying mechanisms behind beta diversity. *Global Ecology and Biogeography* **26**, 1303–1315.

Biosketch

PIERRE LEGENDRE is a specialist in numerical ecology, a Fellow of the Royal Society of Canada (Academy of Science) and a Web of Science Highly Cited Researcher (2014, 2015, 2016 and 2017) in Environment/Ecology. He is interested in the ecological and biogeographical processes that organize biodiversity spatially and temporally.

Supporting Information

Additional supporting information found on-line.

Appendix S1. Simulations involving artificial survey data at times T1 and T2.

Appendix S2. Discussion of some aspects of permutation method 1.

Appendix S3. Insecticide treatments in mesocosms.

Appendix S4. South Tikus Island coral communities.

Appendix S5. The Chesapeake Bay benthos data.

Appendix S6. An R function to standardize environmental data prior to TBI analysis (text file).

Table 1. The dissimilarities (top panel) and p-values (lower panel) associated with the tests of significance of the distances between T1 (survey #4) and T2 (survey #11), for 12 mesocosms (M1 to M12) shown in order of increased insecticide doses. The p-values were corrected for multiple testing (Holm correction); corrected significant values at the 0.05 level are marked with an asterisk (*). Each test involved 9999 random permutations. The maximum possible value is 1 for the %difference and Ružička dissimilarities, and $\sqrt{2}$ = 1.4142 for the chord, Hellinger and log.chord distances. The Euclidean distance does not have an upper bound. Permutation method 1 was used in these tests.

Mesocosms	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Treatment, µg/L	0	0	0	0	0.1	0.1	0.9	0.9	6	6	44	44
TBI dissimilarity values												
%difference *D*	0.433	0.449	0.405	0.459	0.496	0.439	0.488	0.485	0.474	0.621	0.735	0.672
Ružička *D*	0.604	0.620	0.576	0.630	0.663	0.610	0.656	0.653	0.643	0.766	0.847	0.804
Chord *D*	0.806	0.836	0.688	0.774	0.872	0.777	0.884	0.895	0.807	0.981	1.144	1.075
Hellinger *D*	0.859	0.863	0.807	0.860	0.922	0.848	0.930	0.895	0.887	1.072	1.194	1.101
Log.chord *D*	0.862	0.865	0.814	0.865	0.929	0.854	0.933	0.899	0.899	1.078	1.200	1.106
Euclidean *D*	26.33	26.55	22.63	22.44	27.24	24.36	26.43	27.12	23.27	24.82	29.79	27.72
p-values corrected for multiple testing												
%difference *D*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.342	0.001*	0.001*
Ružička *D*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.342	0.001*	0.001*
Chord *D*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.001*	0.001*
Hellinger *D*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.173	0.001*	0.001*
Log.chord *D*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.174	0.001*	0.001*
Euclidean *D*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.414	1.000

Appendices to:

Legendre, P. (2018) A temporal beta-diversity index to identify exceptional sites in space-time surveys. (Submitted).

Appendix S1

SIMULATIONS INVOLVING ARTIFICIAL SURVEY DATA AT TIMES T1 AND T2

Introduction

This section reports the results of a simulation study carried out to determine if the new test of significance of the temporal beta change at individual sites (TBI) has correct type I error rates and is able to detect sites for which the response data had exceptionally high dissimilarities between time 1 (T1) and time 2 (T2). Simulations were done with random data generated in three different ways; six dissimilarity coefficients were used as TBI indices. Three different permutation methods were also compared.

Data generation methods

Two methods were used to generate random community-like data in matrices Mat1 and Mat2. The first one (gen.method=1) was to draw values at random from the Poisson distribution. The second method (gen.method=2) was to use random lognormal data.

The random Poisson deviates were generated with a probability of occurrence (lambda parameter of the distribution) of 0.8. The generated data were skewed and contained approximately 45% zeros. For the lognormal data, the normal distribution generating the deviates that are exponentiated to produce random lognormal data had mean = 0 and standard deviation = 2.0. The values were rounded to the nearest integer. The generated data were highly skewed and contained approximately 36.5% zeros. Random lognormal data are much more highly skewed than random Poisson data.

It may be of interest to ecologists to identify sites that are exceptional in the changes to their community structure on the one hand, and sites that are exceptional in the changes to their environmental conditions on the other hand. So, in a third series of simulations, quantitative environmental data will be generated through random normal deviates; they will represent environmental variables in simulations. Real quantitative environmental data are often not normally distributed, but in many cases then can be normalized using data transformations. Qualitative variables (factors) will not be used in the simulations. How to handle them is described in section "Application to physical environmental or community trait data" of the main paper.

Dissimilarity coefficients

Six dissimilarity coefficients were used in the simulation study. Among the coefficients that are often used to analyse community composition data, we used the chord, Hellinger and log-chord distances that have the Euclidean property (a useful property for PCoA ordination), and the percentage difference and Ružička dissimilarities, which are non-Euclidean dissimilarities and can produce negative eigenvalues and complex axes in PCoA. The Euclidean distance was also used because it is the most widely used coefficient to analyse environmental data tables transformed by

standardization or ranging. Its behaviour for the analysis of community composition data will be assessed against the coefficients specialized for this type of data.

Permutation methods

The three permutation methods are described in the main paper. Briefly, they are:

• Method 1 (permute.sp = 1) — Permute data separately in each column, but in the same way in both matrices.

• Method 2 (permute.sp = 2) — Permute data separately in each column. Do not force the permutations to start at the same point in the two matrices.

• Method 3 (permute.sp = 3) — Permute entire rows in each matrix separately.

Simulations to estimate type I error rates

These simulations will provide an assessment of the validity of the three testing methods. "A statistical testing procedure is valid if the probability of a type I error (rejecting H_0 when true) is no greater than α, the level of significance, for any α." (Edgington 1995, p. 37).

Simulation methods

Simulation series 1, community composition data, random Poisson deviates

Two subseries of simulations were carried out:

1.1. In the first subseries, the two matrices (mat.1 and mat.2, of size $n \times p$) contained random Poisson deviates, as described in the introduction of this Appendix; Fig. S1.1. Three permutation methods were used with 999 random permutations. See results in Table S1.1.

1.2. In this subseries of simulations, in addition to the random rata in mat.1a and mat.2a (called mat.1 and mat.2 in Fig. S1.1), a submatrix mat1.d containing zeros was added to mat.1a and a matrix mat2.d containing random Poisson deviates was added to mat.2a; see Fig. S1.2. Results with *p3* = 6 are shown in Table S1.2.

The objective of the second subseries was to show the effect on the tests of significance of having extra species in the data matrices showing strong difference between T1 and T2 (absent in **Mat.1** and present in **Mat.2**) but with only random differences among the sites. These extra species should have no effect on the TBI tests.

Simulation series 2, community composition data, random lognormal deviates

Two subseries of simulations were carried out:

2.1. In the first subseries, the two matrices (mat.1 and mat.2, of size $n \times p$) contained random lognormal deviates, as described in the introduction of this Appendix; Fig. S1.1. Three permutation methods were used with 999 random permutations. See results in Table S1.3.

2.2. In this subseries of simulations, in addition to the random rata in mat.1a and mat.2a (called mat.1 and mat.2 in Fig. S1.1), a submatrix mat1.d containing zeros was added to mat.1a and a matrix mat2.d containing random Poisson deviates was added to mat.2a; see Fig. S1.2. Results with *p3* = 6 are shown in Table S1.4.

The objective of the second subseries was the same as for subseries 1.2, with a different way of generating community composition-like data.

Simulation series 3, environmental data, random normal deviates

3. In this last series, quantitative environmental data were simulated using random normal deviates instead of species-like data. There were $n = 20$ sites and $p = 20$ variables in matrices **Mat.1** and **Mat.2**, as in Fig. S1.1. TBI tests were only computed with the Euclidean distance. The other distances investigated in the previous simulation series only make sense for community composition and other frequency-like data (Legendre & Legendre 2012, Chapter 7). The three permutation methods were used with 999 random permutations.

The data vectors were standardized as described in Appendix S5. Explanation: (a) the two data tables are joined into a single data matrix, **Y** = rbind(**Mat.1**, **Mat.2**), before standardization. In this way, the differences in values of each variable for a given pair of sites in the two tables will remain comparable to the differences of computed from the original unstandardized values; in this way, the distances computed between sites in T1 and T2 will be meaningful. This precaution is important when there are differences in means between T1 and T2. (b) Standardizing the variables insures that all variables will contribute the same variance to the calculation of the TBI indices; the variances will not depend on the physical units of the variables or other contingencies that make the variances unequal.

Results, type I error study

Results are presented in Tables S1.1 and S1.2 for random Poisson deviates, in Tables S1.3 and S1.4 for random lognormal deviates deviates, and in Table S1.5 for normal deviates.

The simulations produced the expected result that type I error was always correct. The three testing methods are thus valid following the definition shown above (Edgington 1995).

1. Examination of the tables of rejection rates of the null hypothesis (Table S1.1-S1.4) showed that the TBI tests had correct rates of type I error with the three permutational testing methods, for the two community-like data generation methods (Poisson and lognormal) and all dissimilarity indices used, and this for all significance levels (α) considered, from $\alpha = 0.01$ to $\alpha = 0.50$. Deviations from the nominal significance levels, shown at the top of each table, were very small. Simulations involving random environmental-like quantitative data (Table S1.5) also showed correct levels of type I error with the Euclidean distance used in the computation of TBI indices.

2. The tables of rejection rates were divided into separate matrices per data generation and permutation methods, and transformed into squared differences (or squared errors) between the rejection rates and the nominal significance levels α. The sum of the squared differences was computed for each matrix. Examination of the results (not shown in detail in this Appendix) showed no significant difference (Friedman's test) in type I error rejection rates between the data generation methods (Poisson or lognormal) or among the permutation methods.

3. The additional species that were present in **Mat.2** but not in **Mat.1**, did not affect the type I error rates of the TBI tests. The rejection rates in Tables S1.1 and S1.2 (random Poisson deviates), and those in Tables S1.3 and S1.4 (random lognormal deviates), were very similar. In Tables S1.2 and S1.4 (simulation subseries 2), the rejection rates produced by permutation method 2 were the closest to the nominal significance levels, as it was the case in Tables S1.1 and S1.3 (subseries 1).

Note — The percentage difference and Ružička dissimilarity indices differ only by their denominators. The tests of significance of these two indices produce the same p-value if they are run with the same series of permuted vectors. Since the random number generator was started at the same value at the beginning of all simulation runs, it is normal that the rejection rates found for these two dissimilarities in TBI simulations be the same in the report tables (Tables S1.1 to S1.4).

Simulations to compare power of *D* indices and permutation methods

Simulation methods

For the power simulations, some of the sites were generated with a strong difference between T1 and T2 whereas other sites only had random differences. The objective of these simulations was to determine (a) if some permutation methods produced more powerful tests than other methods for the two types of data (community composition and environmental), and (b) if some dissimilarity functions were better suited to identify sites with strong differences in community composition data between T1 and T2 than other dissimilarities.

Data generation proceeded as follows for community composition data. *n1* sites were assigned to a first group that differed in composition between T1 and T2, whereas *n2* sites only had random variation between T1 and T2. Refer to Fig. S1.3.

1. The *n1* sites received strong differences between T1 and T2, as follows:

• A first group of *p1* species received random abundance values in mat.1a.

• A second group of *p2* species received random abundance values in mat.2b

• In **Mat.1**, the *p2* species received the values of mat.2b times a contribution constant (parameter “contr") with values between 0 and 1; contr = 0.2 was used in the simulations with random Poisson deviates reported below. For the simulations with random lognormal deviates, contr was either 0.01 or 0.02, as described below.

• In **Mat.2**, the *p1* species received the values of mat.1a times the same contribution constant "contr" as in the previous paragraph.

The generated abundance-like values were either random Poisson or random lognormal deviates, as described at the beginning of this Appendix.

2. The *n2* sites received random abundance-like values, either Poisson or lognormal. Hence the differences between T1 and T2 were random for these sites (Fig. S1.3).

In the results reported below, $n1 = 5$, $n2 = 15$, $p1 = p2 = 10$. For random Poisson deviates, contr = 0.2. For the simulations with random lognormal deviates, contr was either 0.01 or 0.02, as described below. 1000 data sets were independently generated and analysed for each reported rejection rate (0.01, 0.05, 0.10) in Tables S1.6 and S1.7 reporting the results.

Simulation series 1, community composition data, random Poisson deviates

Two subseries of simulations were carried out:

1. In the first subseries, all matrices simply random Poisson deviates, structured following Fig. S1.3.

2. In addition, a submatrix mat1.d containing zeros was added to **Mat.1** and a matrix mat2.d was added to **Mat.2**, as in the subseries 2 simulations for type I error. See Fig. S1.4.

The objective of the second subseries was to show the effect on the tests of significance of having extra species in the data matrices showing strong difference between T1 and T2 (absent in **Mat.1** and present in **Mat.2**) but with only random differences among the sites. These extra species should have little effect, if any, on the TBI tests of significance.

Simulation series 2, community composition data, random lognormal deviates

Two subseries of simulations were carried out.

1. Random lognormal deviates. The contributions of the mat.2b data to the mat.1b data, and of the mat.1a data to the mat.2a data, was determined by contr = 0.01.

2. The contribution parameter (contr) had the value 0.02.

The contribution parameter was adjusted to produce different rejection rates among the three permutation methods. That was obtained with the contribution values contr = [0.01, 0.02] that were used to generate the values in mat1b and mat2a. With contr = 0, the differences between **Mat.1** and **Mat.2** for the first 5 sites were so strong that all methods rejected H_0 with very high rates and identified the first 5 sites as exceptional; differences among the permutation methods were not clearly shown.

Simulation series 3, environmental data, random normal deviates

For simulation involving data simulating pseudo-environmental variables, the construction of the data matrices followed the same method as in simulation subseries 1 above, except that the data were random normal deviates standardized as described in Appendix S5.

In the results reported below, there were $n1 = 5$ and $n2 = 15$ sites, $p1 = p2 = 10$ variables. The contribution parameter was chosen to be contr = 0.05. 1000 data sets were independently generated and analysed for each reported rejection rate (0.01, 0.05, 0.10) in Tables S1.8 reporting the results.

Results, power study

Results are presented in Tables S1.6 and S1.7 for simulated community composition data and in Table S1.8 for simulated environmental data.

Differences among D and permutation methods, community composition data

Power is the ability to detect an effect when one is present in the data. In the simulations reported here, we know that rather large effects were present in all data sets because the simulation function had generated it in the data.

From the simulation results for community composition data (simulation series 1: Table S1.6, Fig. S1.5; simulation series 2: Table S1.7, Fig. S1.6), we can make the following observations, working from the bottom of the graphs up:

1. All simulations involving the Euclidean distance for the computation of TBI indices had dismally low power (triangles pointing down). TBI indices computed with the Euclidean distance hardly ever detected the presence of exceptional sites in the species-like data files simulated with Poisson or lognormal deviates. The Euclidean distance should not be used for TBI tests of community composition data.

2. The tests of significance of TBI indices carried out with permutation method 3 (black dashed lines) had much lower power than the tests produced with permutation methods 1 and 2 (blue and red lines). The power of this permutation method improved at nominal significance level 0.10, but even then power was lower than that of permutation methods 1 and 2. For the significance levels (0.01 and 0.05) routinely used in tests of significance, the power for detecting exceptional sites in data that contained an effect was much lower than with permutation methods 1 and 2. Permutation method 3 should not be used with community composition data.

3. For data generated with Poisson or lognormal distributions, the most powerful TBI tests were computed with the percentage difference (aka Bray-Curtis) and Ružička dissimilarities, followed by the group {Hellinger, log-chord} distances which produced very similar results with species-like data simulated with Poisson random deviates, and log-chord distances for community data simulated with lognormal deviates. TBI tests based on the chord dissimilarity had the lowest power among the distances that produced usable tests. The log-chord distance was expected to be the most appropriate (and thus produce more powerful tests that the chord or Hellinger distances) with lognormally distributed data since the log transformation, which is the first transformation in the calculation of that distance, makes the random data normal before the chord transformation is computed.

4. Permutation methods 1 (blue lines) and 2 (red lines) can both be used, with a small advantage for method 1 in results of Poisson simulations, and equivalent powers for methods 1 and 2 in simulations based on lognormal deviates.

5. When sites had entirely different species compositions between T1 and T2, the TBI test had maximum power: it always rejected H_0 at significance levels of 0.05 and 0.10, and in 98.7% of the cases at level 0.01. This situation was produced by setting the contribution parameter to the value contr=0 for the generation of data in submatrices mat1b and mat2b (see Fig. S1.3). These rejection rates were obtained with all dissimilarity indices tested: percentage difference, Ruzicka, chord, Hellinger and log-chord distances.

The best combination for TBI tests of community composition data with maximum power is to use the percentage difference or the Ružička indices with permutation methods 1 or 2. These two dissimilarities can also be decomposed into species losses (*B*/*den*) and gains (*C*/*den*), which can be used to examine the processes of losses and gains at the site level and to produce B-C plots.

Differences in power for simulated environmental data

For standardized environmental variables (simulation series 3, Table S1.8), the only distance tested in the simulation study was the Euclidean distance. The results show a marked advantage of

permutation method 1, which had higher power than the other two methods to detect effects that were present in simulated data. It would be the method of choice for this type of data.

Additional simulations, differences in power associated with different values of n1 and n2

Power to detect an effect in statistical tests is well-known to be a function of three parameters: the importance of the effect to be detected, the significance level α, and the number of observations n. Additional simulations were conducted to detect the effect on power of the number of exceptional sites (*n1*) and the number of sites with random variation (*n2*) in the study. Community composition was simulated using random lognormal deviates.

1. The number of sites n was 20 in all simulations, with the number of affected sites *n1* = varying from 1 to 19; *n2* = (n – *n1*). The simulations used the percentage difference index, permutation method 1, and contr = 0.02, 1000 independent simulations and permutation method 1 with 999 permutations.

The results show that for tests of significance at level alpha = 0.05, optimal power was obtained with *n1* = 1 to 9 in simulations (Fig. S1.7a). *n1* should not be larger than n/2. Tests carried out with *n1* equal to or larger than n/2 can still be used but they have lower power.

2. For *n1* = n/4, simulations were repeated for different values, with n = {8,16,24,32,40,48,56,64} and *n1* = {2,4,6,8,10,12,14,16}. Again, the simulations used random lognormal data, the percentage difference index, contr = 0.02, 1000 independent simulations and permutation method 1 with 999 permutations.

Rejection rates are reported in Fig. S1.7b for three significance levels alpha: 0.01, 0.05 and 0.10. Power remained constant over all values of n investigated.

In summary, power of the test performed with permutation method 1 was high when the effect was strong, and as long as the proportion of sites with an effect was smaller than n/2 (Fig. S1.7a). For a fixed proportion of affected sites, power did not increase when the total number of sites n in the study was larger (Fig. S1.7b).

References

Edgington, E. S. (1995) *Randomization Tests, 3rd edn.* Marcel Dekker, New York.

Legendre, P. & L. Legendre. (2012) *Numerical ecology, 3rd English edn*. Elsevier Science BV, Amsterdam.

Table S1.1. Type I error rates of the test of the TBI *D* indices shown in the first column: rejection rates (i.e. number of rejections of H_0 divided by the number of simulations, which was 1000) of the TBI test when there were no exceptional sites in the simulated data. The data were drawn from a random Poisson distribution; n = 20 sites, p = 20 species (Fig. S1.1). Three permutational methods were used with 999 random permutations. Simulation series 1: data at all sites and both times came from the same statistical population, hence H_0 was true.

	Permutation	Nominal significance levels						
	methods	0.01	0.05	0.10	0.20	0.30	0.40	0.50
% difference	1	0.000	0.034	0.108	0.208	0.275	0.335	0.491
% difference	2	0.009	0.048	0.094	0.192	0.298	0.398	0.497
% difference	3	0.008	0.049	0.097	0.195	0.289	0.385	0.493
Ružička	1	0.000	0.033	0.108	0.208	0.275	0.333	0.492
Ružička	2	0.009	0.047	0.093	0.190	0.296	0.397	0.498
Ružička	3	0.008	0.050	0.099	0.199	0.291	0.388	0.497
Chord	1	0.000	0.042	0.075	0.200	0.292	0.383	0.475
Chord	2	0.011	0.052	0.099	0.196	0.298	0.404	0.503
Chord	3	0.009	0.053	0.100	0.195	0.297	0.398	0.492
Hellinger	1	0.000	0.042	0.100	0.175	0.300	0.367	0.517
Hellinger	2	0.011	0.050	0.101	0.199	0.301	0.411	0.505
Hellinger	3	0.010	0.050	0.101	0.197	0.295	0.400	0.500
Log-chord	1	0.000	0.050	0.117	0.183	0.300	0.375	0.483
Log-chord	2	0.011	0.052	0.101	0.198	0.301	0.408	0.506
Log-chord	3	0.010	0.052	0.102	0.196	0.297	0.395	0.503
Euclidean	1	0.000	0.025	0.094	0.200	0.319	0.400	0.494
Euclidean	2	0.007	0.046	0.092	0.194	0.290	0.390	0.488
Euclidean	3	0.007	0.046	0.091	0.195	0.289	0.381	0.480

Table S1.2. Type I error rates of the test of the TBI *D* indices shown in the first column. See caption of Table S1.1. The data were drawn from a random Poisson distribution; n = 20 sites. Simulation series 2: for the basic *p1* = 20 species, data at all sites and both times came from the same statistical population. In addition, T2 had *p3* = 6 species more than T1 (Fig. S1.2). For these 6 species, there were no differences among the sites besides random variation; hence H_0 was still true.

	Permutation	Nominal significance levels						
	methods	0.01	0.05	0.10	0.20	0.30	0.40	0.50
% difference	1	0.017	0.042	0.092	0.183	0.267	0.425	0.541
% difference	2	0.008	0.054	0.086	0.176	0.269	0.383	0.494
% difference	3	0.007	0.043	0.091	0.192	0.292	0.388	0.496
Ružička	1	0.017	0.042	0.091	0.183	0.267	0.425	0.541
Ružička	2	0.009	0.048	0.095	0.199	0.302	0.401	0.505
Ružička	3	0.007	0.043	0.090	0.191	0.290	0.386	0.494
Chord	1	0.000	0.062	0.112	0.212	0.312	0.450	0.525
Chord	2	0.010	0.051	0.100	0.205	0.307	0.406	0.509
Chord	3	0.009	0.045	0.093	0.197	0.299	0.399	0.497
Hellinger	1	0.000	0.075	0.112	0.213	0.312	0.375	0.550
Hellinger	2	0.010	0.049	0.099	0.202	0.305	0.403	0.504
Hellinger	3	0.009	0.043	0.091	0.194	0.292	0.392	0.497
Log-chord	1	0.000	0.075	0.100	0.238	0.312	0.400	0.550
Log-chord	2	0.010	0.050	0.100	0.203	0.304	0.406	0.503
Log-chord	3	0.009	0.045	0.094	0.194	0.294	0.397	0.497
Euclidean	1	0.008	0.050	0.100	0.208	0.300	0.383	0.508
Euclidean	2	0.011	0.050	0.095	0.189	0.289	0.384	0.484
Euclidean	3	0.006	0.043	0.097	0.191	0.293	0.391	0.491

Table S1.3. Type I error rates of the test of the TBI *D* indices shown in the first column: rejection rates (i.e. number of rejections of H_0 divided by the number of simulations, which was 1000) of the TBI test when there were no exceptional sites in the simulated data. The data were drawn from a random lognormal distribution; n = 20 sites, p = 20 species (Fig. S1.1). Three permutational methods were used with 999 random permutations. Simulation series 1: data at all sites and both times came from the same statistical population, hence H_0 was true.

	Permutation	Nominal significance levels						
	methods	0.01	0.05	0.10	0.20	0.30	0.40	0.50
%difference *D*	1	0.014	0.064	0.121	0.193	0.293	0.386	0.486
	2	0.010	0.051	0.101	0.199	0.300	0.403	0.504
	3	0.007	0.045	0.079	0.173	0.254	0.362	0.476
Ružička *D*	1	0.014	0.064	0.121	0.193	0.293	0.386	0.486
	2	0.010	0.051	0.101	0.199	0.300	0.403	0.504
	3	0.007	0.045	0.079	0.173	0.254	0.362	0.476
Chord *D*	1	0.000	0.038	0.100	0.188	0.251	0.388	0.513
	2	0.012	0.051	0.101	0.195	0.298	0.401	0.498
	3	0.008	0.031	0.085	0.176	0.278	0.370	0.468
Hellinger *D*	1	0.000	0.050	0.100	0.175	0.263	0.375	0.562
	2	0.012	0.052	0.102	0.201	0.303	0.398	0.498
	3	0.007	0.036	0.072	0.165	0.270	0.381	0.496
Log-chord *D*	1	0.000	0.025	0.076	0.213	0.313	0.425	0.512
	2	0.011	0.054	0.103	0.197	0.299	0.399	0.497
	3	0.008	0.036	0.081	0.163	0.264	0.374	0.479
Euclidean *D*	1	0.000	0.050	0.112	0.188	0.312	0.413	0.500
	2	0.009	0.051	0.100	0.202	0.303	0.403	0.502
	3	0.018	0.048	0.100	0.183	0.273	0.390	0.503

Table S1.4. Type I error rates of the test of the TBI *D* indices shown in the first column. See caption of Table S1.3. The data were drawn from a random lognormal distribution; $n = 20$ sites. Simulation series 2: for the basic $p1 = 20$ species, data at all sites and both times came from the same statistical population. In addition, T2 had $p3 = 6$ species more than T1 (Fig. S1.2). For these 6 species, there were no differences among the sites besides random variation; hence H_0 was still true.

	Permutation methods	Nominal significance levels						
		0.01	0.05	0.10	0.20	0.30	0.40	0.50
%difference *D*	1	0.022	0.061	0.105	0.210	0.294	0.389	0.500
	2	0.009	0.055	0.100	0.195	0.299	0.404	0.501
	3	0.008	0.046	0.114	0.193	0.275	0.386	0.462
Ružička *D*	1	0.022	0.061	0.105	0.210	0.294	0.389	0.500
	2	0.009	0.055	0.100	0.195	0.299	0.404	0.501
	3	0.008	0.046	0.114	0.193	0.275	0.386	0.462
Chord *D*	1	0.017	0.061	0.100	0.222	0.300	0.411	0.517
	2	0.009	0.048	0.098	0.197	0.295	0.395	0.497
	3	0.014	0.059	0.106	0.179	0.269	0.359	0.485
Hellinger *D*	1	0.017	0.055	0.111	0.194	0.273	0.368	0.501
	2	0.012	0.049	0.095	0.196	0.295	0.399	0.491
	3	0.005	0.061	0.094	0.184	0.280	0.374	0.501
Log-chord *D*	1	0.022	0.072	0.116	0.194	0.273	0.401	0.506
	2	0.009	0.048	0.096	0.192	0.290	0.398	0.491
	3	0.005	0.043	0.103	0.198	0.295	0.391	0.506
Euclidean *D*	1	0.017	0.056	0.133	0.194	0.322	0.405	0.505
	2	0.011	0.050	0.100	0.197	0.299	0.393	0.501
	3	0.004	0.050	0.107	0.186	0.303	0.400	0.480

Table S1.5. Type I error rates of the test of TBI indices computed using the Euclidean distance. Simulation series 3: the data were drawn from a random normal distribution; n = 20 sites, p = 20 variables. Three permutation methods were used with 999 random permutations. Data at all sites and both times came from the same statistical population, hence H_0 was true.

	Permutation	Nominal significance levels						
	methods	0.01	0.05	0.10	0.20	0.30	0.40	0.50
Euclidean *D*	1	0.007	0.050	0.100	0.193	0.321	0.421	0.528
	2	0.009	0.049	0.096	0.204	0.304	0.410	0.500
	3	0.010	0.050	0.112	0.209	0.304	0.392	0.495

Table S1.6. Power analysis of TBI *D* indices shown in the first column, random Poisson data. There were *n1* = 5 exceptional sites in the simulated data. The results in the table are the mean rejection rates of the TBI test for these 5 sites, computed as the number of rejections of H_0 divided by the number of simulations, which was 1000. (A) Data with *p1* and *p2* species only (Fig. S1.3). (B) *p3* = 6 extra species with values of 0 in Mat.1 and random Poisson deviates in all sites of Mat.2 (Fig. S1.4); H_0 is true for these *p3* species.

		(A) p1 and p2 species			(B) p1, p2 and p3 species		
	Permutation	Significance levels			Significance levels		
	methods	0.01	0.05	0.10	0.01	0.05	0.10
%difference *D*	1	0.341	0.787	0.859	0.478	0.858	0.900
	2	0.370	0.699	0.897	0.377	0.765	0.878
	3	0.000	0.256	0.686	0.000	0.212	0.716
Ružička *D*	1	0.341	0.787	0.859	0.478	0.858	0.900
	2	0.370	0.699	0.897	0.377	0.765	0.878
	3	0.000	0.256	0.686	0.000	0.212	0.716
Chord *D*	1	0.355	0.511	0.600	0.381	0.580	0.720
	2	0.350	0.400	0.493	0.355	0.425	0.527
	3	0.000	0.211	0.418	0.000	0.190	0.396
Hellinger *D*	1	0.377	0.577	0.711	0.421	0.660	0.840
	2	0.350	0.541	0.699	0.357	0.574	0.722
	3	0.000	0.214	0.529	0.000	0.191	0.526
Log-chord *D*	1	0.355	0.600	0.711	0.401	0.620	0.800
	2	0.350	0.492	0.660	0.355	0.523	0.694
	3	0.000	0.212	0.497	0.000	0.191	0.477
Euclidean *D*	1	0.000	0.000	0.000	0.000	0.000	0.020
	2	0.000	0.001	0.006	0.000	0.002	0.012
	3	0.000	0.000	0.002	0.000	0.002	0.009

Table S1.7. Power analysis of TBI *D* indices shown in the first column, random lognormal data. There were 5 exceptional sites in the simulated data. The results in the table are the mean rejection rates of the TBI test for these 5 sites, computed as the number of rejections of H_0 divided by the number of simulations, which was 1000. Data with *p1* and *p2* species only (Fig. S1.3). There were no extra species in these simulations (*p*3 = 0). The contribution parameter of these simulations varied: (A) contr = 0.01, (B) contr = 0.02.

		(A) contr = 0.01			(B) contr = 0.02		
	Permutation	Significance levels			Significance levels		
	methods	0.01	0.05	0.10	0.01	0.05	0.10
%difference *D*	1	0.875	0.975	1.000	0.463	0.849	0.962
	2	0.872	0.997	1.000	0.472	0.865	0.982
	3	0.000	0.387	0.971	0.002	0.354	0.770
Ružička *D*	1	0.875	0.975	1.000	0.463	0.849	0.962
	2	0.872	0.997	1.000	0.472	0.865	0.982
	3	0.000	0.387	0.971	0.002	0.354	0.770
Chord *D*	1	0.625	0.675	0.713	0.301	0.313	0.338
	2	0.668	0.668	0.681	0.360	0.360	0.360
	3	0.000	0.301	0.601	0.002	0.284	0.345
Hellinger *D*	1	0.612	0.737	0.850	0.315	0.401	0.529
	2	0.657	0.714	0.824	0.363	0.419	0.586
	3	0.000	0.303	0.639	0.003	0.285	0.381
Log-chord *D*	1	0.676	0.887	0.912	0.401	0.638	0.750
	2	0.714	0.902	0.939	0.403	0.659	0.766
	3	0.000	0.306	0.756	0.002	0.296	0.532
Euclidean *D*	1	0.012	0.025	0.062	0.012	0.025	0.062
	2	0.000	0.003	0.038	0.000	0.003	0.018
	3	0.002	0.022	0.048	0.002	0.018	0.044

Table S1.8. Power analysis of TBI indices computed using the Euclidean distance, random normal data. There were 5 exceptional sites in the simulated data. The results in the table are the mean rejection rates of the TBI test for these 5 sites, computed as the number of rejections of H_0 divided by the number of simulations, which was 1000. Data with p1 and p2 species only. There were no extra species in these simulations ($p3 = 0$). contr = 0.05.

		contr = 0.05		
	Permutation	Significance levels		
	methods	0.01	0.05	0.10
Euclidean *D*	1	0.627	0.772	0.843
	2	0.251	0.449	0.657
	3	0.033	0.182	0.310

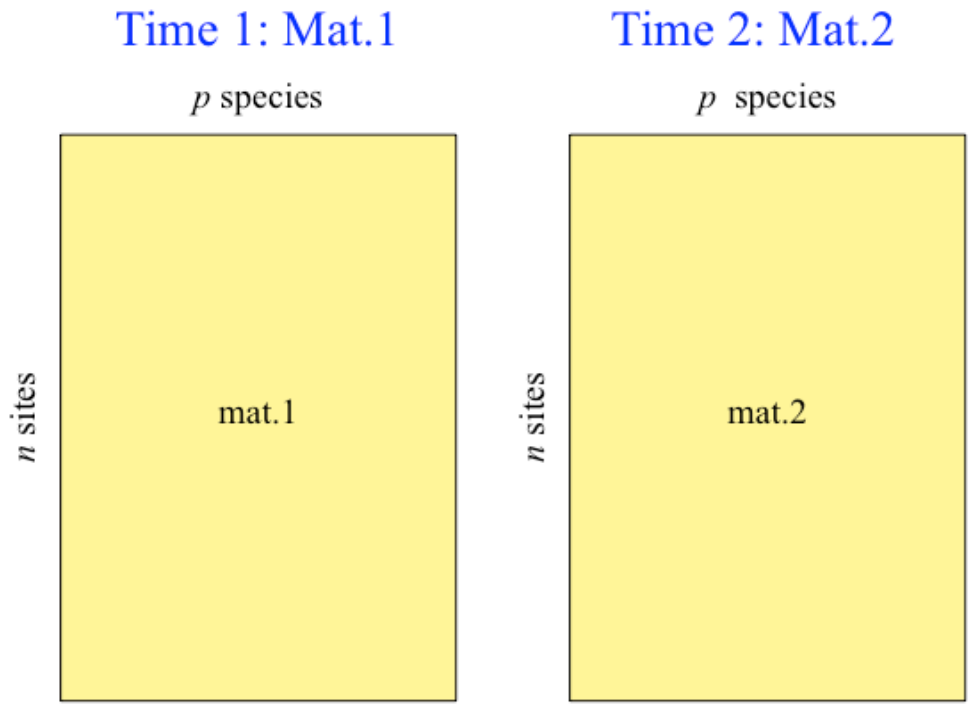

Fig. S1.1. Two data matrices (**Mat.1** and **Mat.2**) used in the type I error simulations. Sections mat.1 and mat.2 were filled with random numbers, so that H_0 was true. Note: **Mat.1** and **Mat.2** contain only one section each here; they will contain more sections in the next figures.

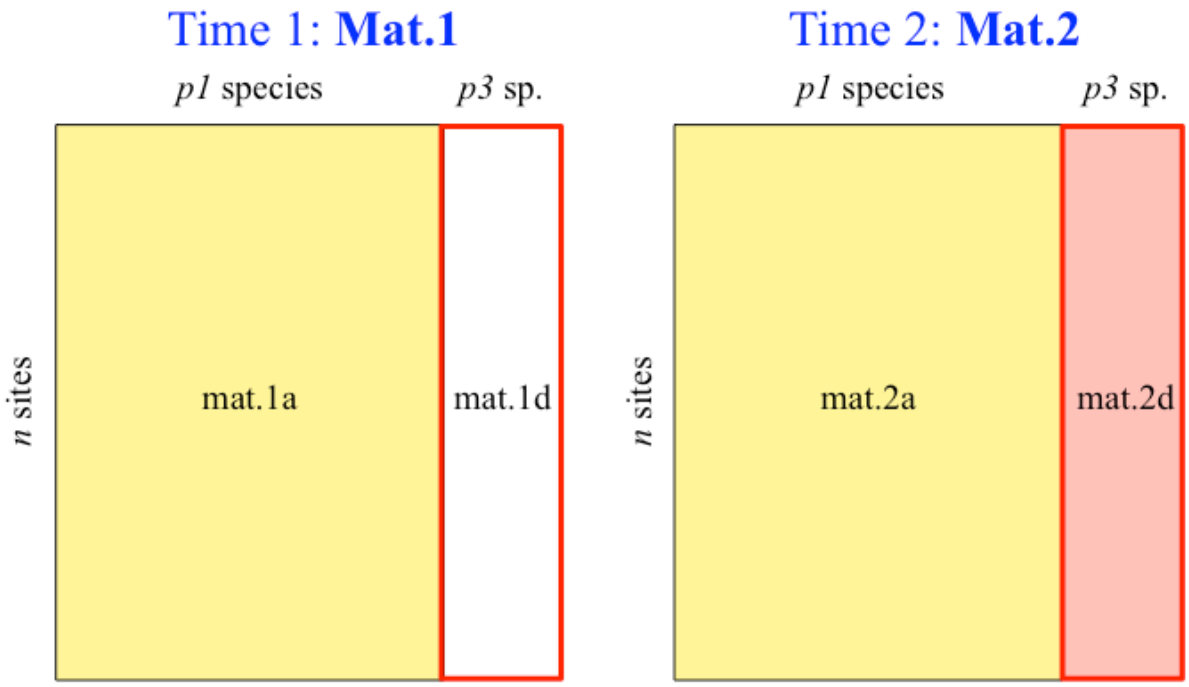

Fig. S1.2. Two data matrices (**Mat.1** and **Mat.2**) used in the type I error simulations; mat.1a and mat.2a contained random numbers, as in Fig. S1.1. In the simulations of subseries 2, mat.1d (white, containing zeros) and mat.2d (pink, containing random numbers) were joined to mat.1a and mat.2a.

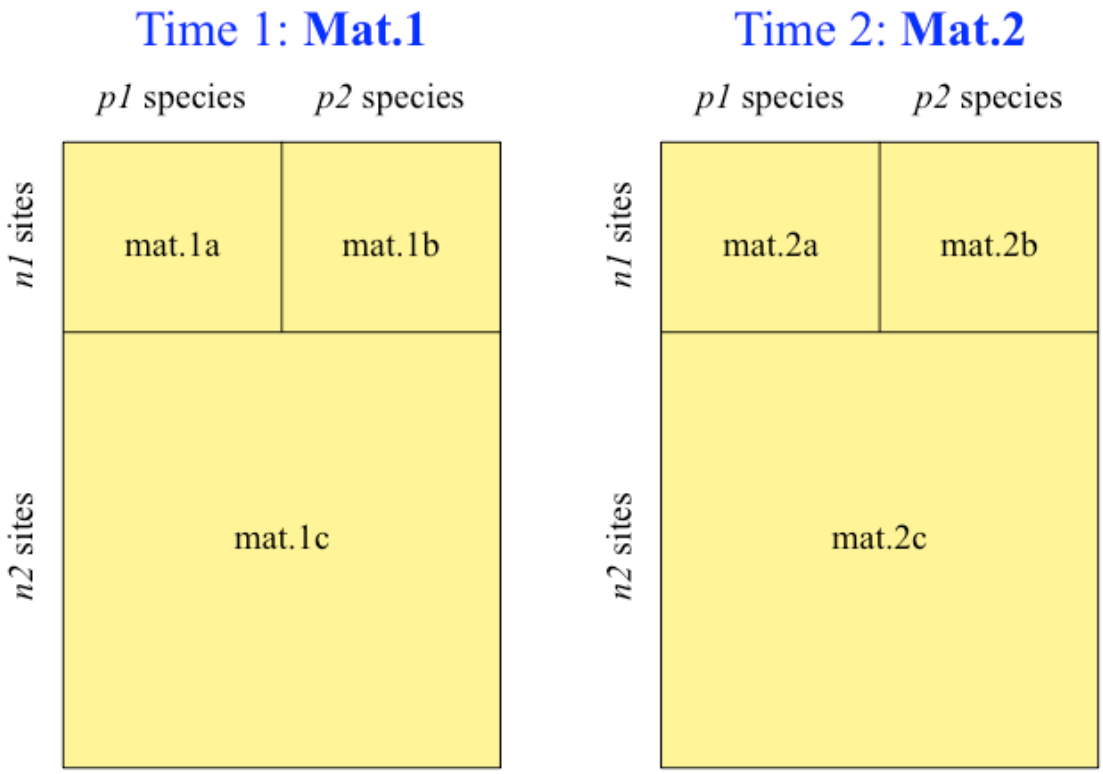

Fig. S1.3. Two matrices (**Mat.1** and **Mat.2**) used in the power simulations. Abundances in mat.1a and mat.2b were generated independently using either random Poisson or random lognormal deviates. Submatrix mat.1b received a fraction of the abundances in mat.2b and mat 2a received a contribution of the abundances in mat 1a. The values of these contributions are described in the text and depend of the random data generator used. Submatrices mat.1c and mat.2c received random deviates drawn from the same statistical population.

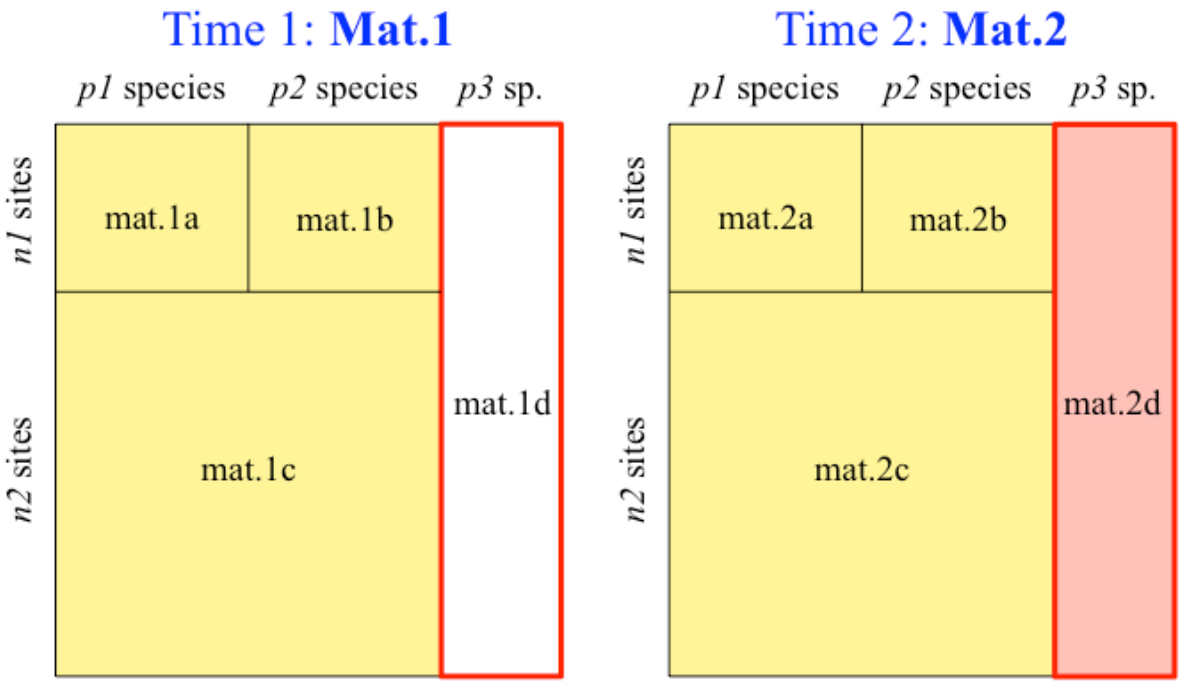

Fig. S1.4. Two matrices (**Mat.1** and **Mat.2**) used in additional power simulations with random Poisson error. In these simulations, mat.1d (white, containing zeros) and mat.2d (pink, containing random numbers) were joined to mat.1a-b-c and mat.2a-b-c. H_0 was true for the *p3* species in mat.1d and mat.2d.

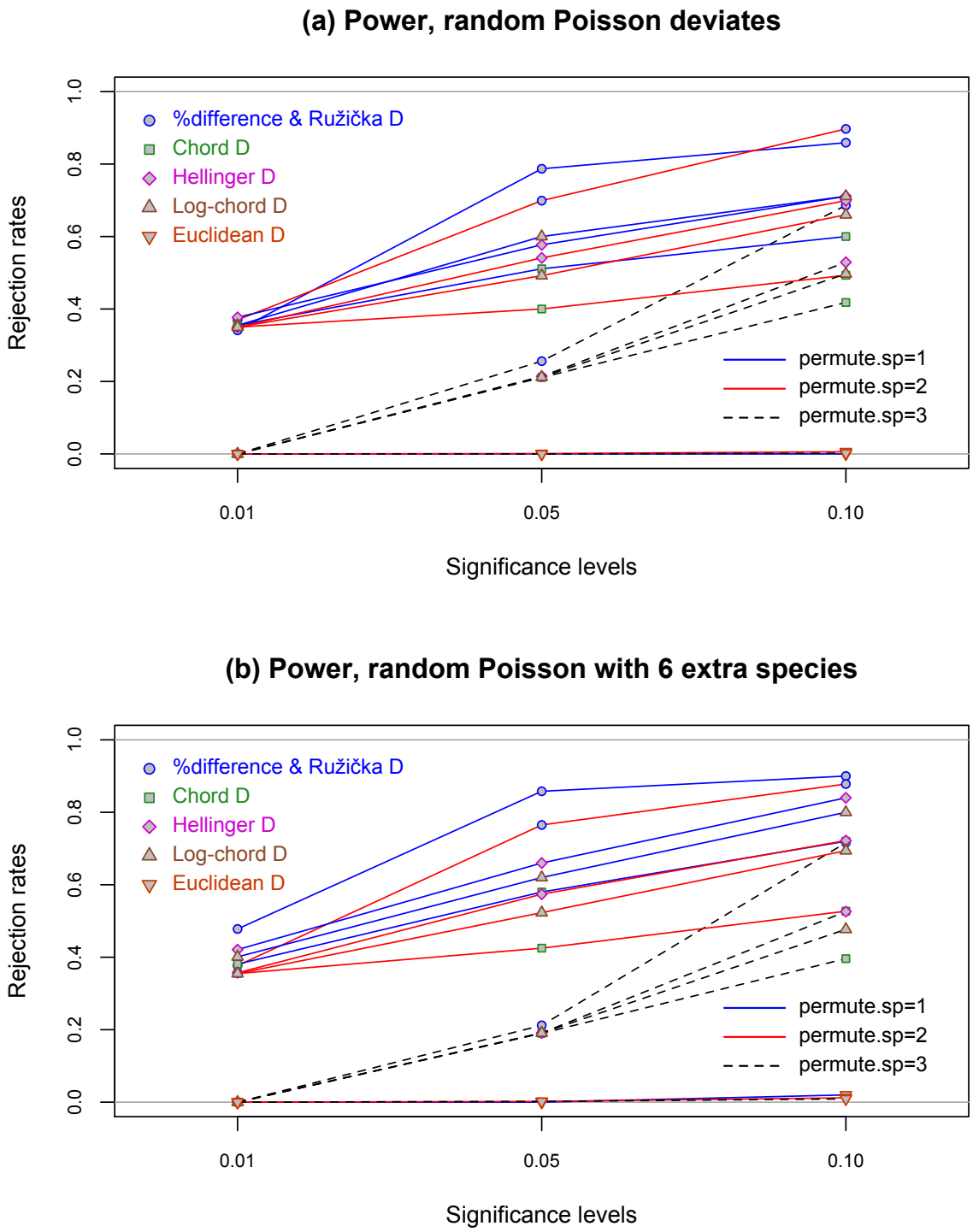

Fig. S1.5. Power study, random Poisson deviates. Rejection rates obtained with five dissimilarity coefficients and three permutation methods. Rejection rates are reported for three significance levels alpha: 0.01, 0.05 and 0.10 (abscissa).

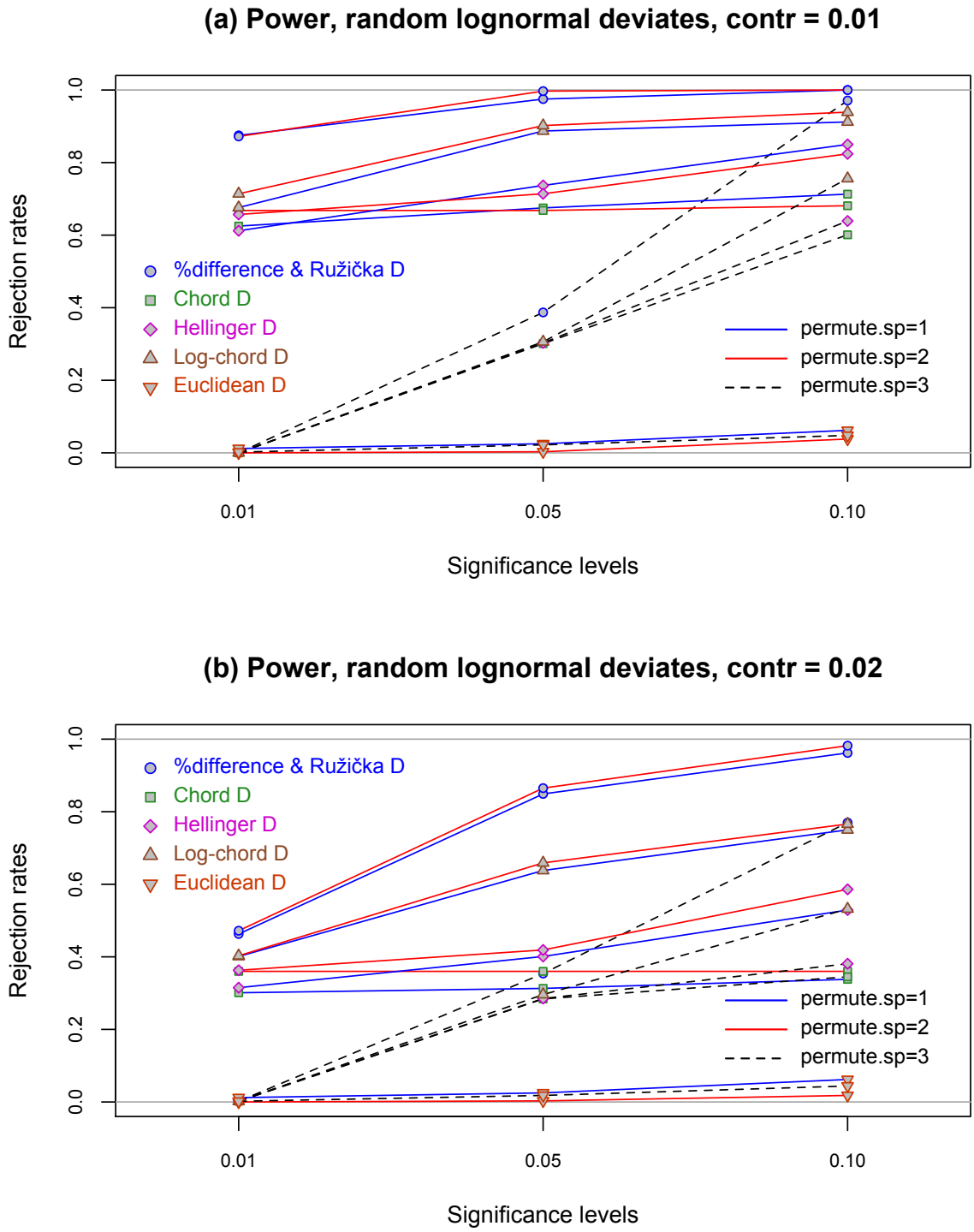

Fig. S1.6. Power study, random lognormal deviates. Rejection rates obtained with five dissimilarity coefficients and three permutation methods. Rejection rates are reported for three significance levels alpha: 0.01, 0.05 and 0.10 (abscissa).

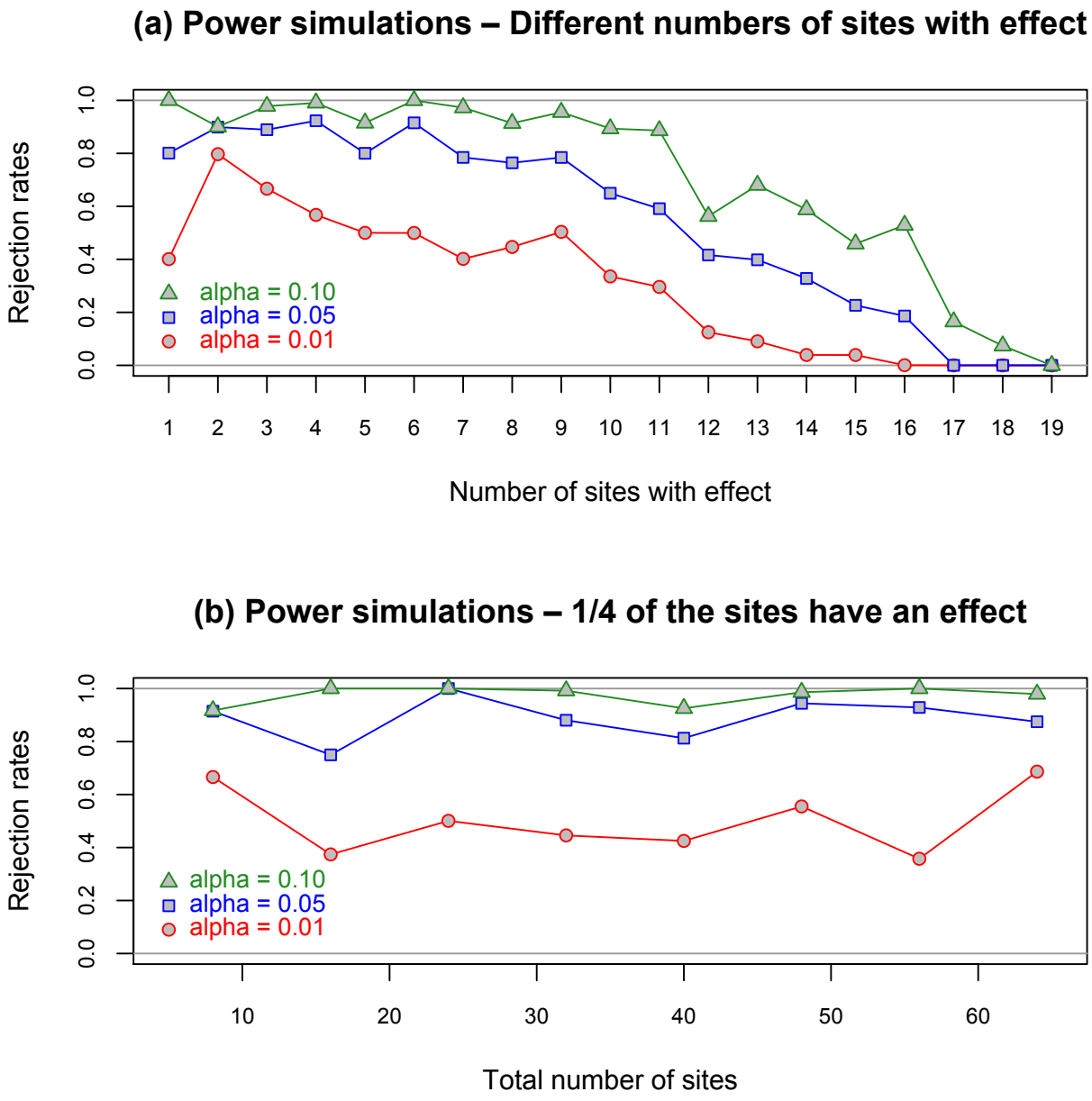

Fig. S1.7. Rejection rates of the TBI tests in power simulations. Rejection rates are reported for three significance levels alpha: 0.01, 0.05 and 0.10. (a) Different numbers of sites received an effect (abscissa, *n1* = 1 to 19), i.e. a difference in community composition between T1 and T2. There were 20 sites in total in each simulation. (b) The proportion of sites with an effect was kept constant, here *n1* = *n*/4, for different values of *n* (abscissa).

Appendix S2

DISCUSSION OF SOME ASPECTS OF PERMUTATION METHOD 1

Permutation method 1 is described in the Methods section of the paper. Some technical aspects of this permutation method are discussed here.

• By permuting the columns of each matrix separately, we produce realizations of H_0 corresponding to the idea that all sites come from the same statistical population and the variation observed among sites for each species is due to random sampling of the same statistical population. This type of permutation does not deny that species may be correlated to one another. If they are and all sites are drawn from a single populations, the variation observed within each species will still fluctuate as a result of random sampling.

• The permutations are done separately in Mat1 and Mat2, i.e. the values of a species at T1 and T2 are not combined and then permuted at random, because we want to preserve the differences per species, if any, between T1 and T2. The test is designed to identify sites where the difference between T1 and T2 is noticeably larger than that of most other sites in the study. If, on the contrary, one only wanted to identify a general difference between T1 and T2 in multivariate species [or other type of] data, i.e. a difference in the positions of the centroids of the T1 versus T2 data, one could use redundancy analysis (RDA) or Manova to achieve that; it is an entirely different question.

• In function TBI() with argument permute.sp = 1, each random permutation of a species is done in the same way in T1 and T2. The seed fed to the permutation function (which is sample() in R) used to permute the values in the columns of Mat1 is noted and used again to perform the permutations of Mat2.

• The data in both Mat1 and Mat2 are permuted. This is done because, if only one matrix (say, Mat1) was permuted, the dissimilarities under permutation, used to assess the reference dissimilarity for object *i*, would be computed by comparison with a fixed vector of values for that object at T2. The test would then not have access to the entire set of possible permutations, but this is necessary condition for the permutations to correctly represent H_0. As a consequence, the test would have reduced power.

Appendix S3

INSECTICIDE TREATMENTS IN MESOCOSMS

The Insecticide treatment invertebrate data

Observations on the abundances of 178 invertebrate species (macroinvertebrates and zooplankton) subjected to insecticide treatments in aquatic mesocosms (called "ditches") were used by van den Brink & ter Braak (1999) as an application example in their paper describing Principal Response Curves (PRC) analysis. The authors agreed to make the data available to researchers in the CANOCO program documentation and in the R package vegan (Oksanen et al. 2017).

The experiment involved twelve mesocosms, which were surveyed on eleven occasions. Four mesocosms served as controls (dose = 0) and the remaining eight were treated once with the insecticide chlorpyrifos, with dose levels of 0.1, 0.9, 6.0 and 44.0 µg/L in two mesocosms each. The data are log-transformed species abundances, $y_{tr} = \log_e(10y + 1)$. In their paper, the authors used the log-transformed invertebrate data in PRC analysis; PRC preserved the Euclidean distance among the observations.

The 12 mesocosms had been attributed at random to the treatments. However, to facilitate presentation of the results, they will be presented here in order of increased insecticide doses: {0, 0, 0, 0, 0.1, 0.1, 0.9, 0.9, 6.0, 6.0, 44.0, 44.0} µg/L. Results of the calculations with the R function TBI() are shown for the species abundance and occurrence data of this ecological application.

References

Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. & Wagner, H. (2017) *vegan: Community ecology package.* R package version 2.4-4. https://cran.r-project.org/package=vegan.

van den Brink, P. J. & ter Braak, C. J. F. (1999) Principal response curves: analysis of time-dependent multivariate responses of biological community to stress. *Environmental Toxicology and Chemistry* **18**, 138–148.

Results of calculations with R function TBI()

Pyrifos insect treatment data: compare survey #4 (one week after the insecticide treatment) to survey #11 (after full recovered from treatment). # Comments added to the output files.

```
library(vegan)
data(pyrifos)
survey4.order = c(38,39,41,47,37,44,40,46,43,48,42,45)
survey11.order = c(122,123,125,131,121,128,124,130,127,132,126,129)
```

1. Comparison based upon species abundance data, percentage difference D

```
( res1 <- TBI(pyrifos[survey4.order,], pyrifos[survey11.order,], method="%diff",
nperm=9999, permute.sp=1, BCD=TRUE, test.t.perm=TRUE, clock=TRUE) )
# Computation time = 51.634000  sec

-----
$TBI
 [1] 0.4332125 0.4490831 0.4048151 0.4593321 0.4958159 0.4392330 0.4884889 0.4851041
0.4740264 0.6205484 0.7345825 0.6721440

$p.TBI
 [1] 0.9305 0.8618 0.9827 0.7249 0.4664 0.8448 0.5350 0.6370 0.7404 0.0342 0.0001 0.0001

$p.adj
 [1] 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 0.3420 0.0012 0.0012

$BCD.mat
        B/(2A+B+C) C/(2A+B+C) D=(B+C)/(2A+B+C) Change
Site.1   0.1616465  0.2715660        0.4332125    +   # Untreated
Site.2   0.1973186  0.2517645        0.4490831    +   # Untreated
Site.3   0.2305092  0.1743059        0.4048151    -   # Untreated
Site.4   0.2643243  0.1950077        0.4593321    -   # Untreated
Site.5   0.2303800  0.2654359        0.4958159    +   # Treated, 0.1 microgram/L
Site.6   0.1980843  0.2411487        0.4392330    +   # Treated, 0.1 microgram/L
Site.7   0.2425404  0.2459484        0.4884889    +   # Treated, 0.9 microgram/L
Site.8   0.1854199  0.2996843        0.4851041    +   # Treated, 0.9 microgram/L
Site.9   0.1901665  0.2838599        0.4740264    +   # Treated, 6 micrograms/L
Site.10  0.3094316  0.3111168        0.6205484    +   # Treated, 6 micrograms/L
Site.11  0.3232546  0.4113279        0.7345825    +   # Treated, 44 micrograms/L
Site.12  0.1829121  0.4892319        0.6721440    +   # Treated, 44 micrograms/L

$BCD.summary   # Here the BCD summary is computed for the 12 mesocosms, not the 8 treated
 mean(B/den) mean(C/den)   mean(D)  B/(B+C)  C/(B+C) Change
   0.2263323   0.2866998 0.5130322 0.441166 0.558834    +

$t.test_B.C    # Here the tests is computed for the 12 mesocosms, not for the 8 treated
                mean(B-C)      Stat    p.param p.perm   p<=0.05
Paired t.test -0.06036748 -2.132286 0.05635548 0.0364          # Permutation test signif.

$BC
[1] NA
-----
```

2. Comparison based upon species occurrence (i.e. presence-absence) data, Sørensen D

```
( res2 <- TBI(pyrifos[survey4.order,], pyrifos[survey11.order,], method="sorensen",
nperm=9999, permute.sp=1, BCD=TRUE, test.t.perm=TRUE, clock=TRUE) )
# Computation time = 38.387000  sec

-----
$TBI
 [1] 0.4390244 0.4324324 0.4457831 0.4705882 0.4666667 0.4358974 0.5000000 0.4153846
0.4545455
[10] 0.6800000 0.7551020 0.6595745

$p.TBI
 [1] 0.8788 0.9134 0.7933 0.6390 0.6892 0.7769 0.4826 0.8964 0.9113 0.0001 0.0001 0.0001

$p.adj
 [1] 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 0.0012 0.0012 0.0012

$BCD.mat
        B/(2A+B+C) C/(2A+B+C) D=(B+C)/(2A+B+C) Change
Site.1   0.1463415  0.2926829        0.4390244    +   # Untreated
Site.2   0.1891892  0.2432432        0.4324324    +   # Untreated
Site.3   0.2048193  0.2409639        0.4457831    +   # Untreated
Site.4   0.2205882  0.2500000        0.4705882    +   # Untreated
Site.5   0.1733333  0.2933333        0.4666667    +   # Treated, 0.1 microgram/L
Site.6   0.1666667  0.2692308        0.4358974    +   # Treated, 0.1 microgram/L
Site.7   0.2105263  0.2894737        0.5000000    +   # Treated, 0.9 microgram/L
Site.8   0.1384615  0.2769231        0.4153846    +   # Treated, 0.9 microgram/L
Site.9   0.1363636  0.3181818        0.4545455    +   # Treated, 6 micrograms/L
Site.10  0.2800000  0.4000000        0.6800000    +   # Treated, 6 micrograms/L
Site.11  0.2857143  0.4693878        0.7551020    +   # Treated, 44 micrograms/L
Site.12  0.1276596  0.5319149        0.6595745    +   # Treated, 44 micrograms/L

$BCD.summary   # Here the BCD summary is computed for the 12 mesocosms, not the 8 treated
 mean(B/den) mean(C/den)   mean(D)   B/(B+C)   C/(B+C) Change
    0.189972   0.3229446 0.5129166 0.3703759 0.6296241    +

$t.test_B.C    # Here the test is computed for the 12 mesocosms, not for the 8 treated
               mean(B-C)      Stat      p.param p.perm   p<=0.05
Paired t.test  -0.1329727 -4.621706 0.0007383173  7e-04         *  # Both tests signif.

$BC
[1] NA
```

Appendix S4

SOUTH TIKUS ISLAND CORAL COMMUNITIES

South Tikus Island coral data

Brown and Suharsono (1990) surveyed coral communities (75 species) at 10 sites in the island of South Tikus, Indonesia, in the years 1981, 1983, 1984, 1985, 1987 and 1988. An El Niño event occurred in 1982–1983, which caused coral bleaching and death of coral colonies, and triggered changes in the composition of coral communities. They reported that "as many as 80-90% of corals died on the reef flats at the study sites, with the major casualties being branching species in the genera *Acropora* and *Pocillopora*".

Coral forms colonies which occupy surfaces, so that the data are not in numbers of individuals but in areal cover of each species. The sum of the species areal covers at a site may exceed 100% because coral colonies may overlap one another vertically. The Brown and Suharsono (1990) data have been used in several papers to demonstrate the application of multivariate methods for the analysis of beta diversity and the comparison of surveys across time, e.g. by Warwick et al. (1990), Anderson et al. (2011) and Chao and Chiu (1996). Following these papers, the data in the present application were treated as if they were species abundances. They were obtained from Appendix S1 of the Anderson et al. (2011) article.

References

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L. et al. (2011) Navigating the multiple meanings of b diversity: a roadmap for the practicing ecologist. *Ecology Letters* **14**, 19–28.

Brown, B. E. & Suharsono. (1990) Damage and recovery of coral reefs affected by El Niño related seawater warming, in the Thousand Islands, Indonesia. *Coral Reefs* **8**, 163–170.

Chao, A. & Chiu, C.-H. (2016) Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. *Methods in Ecology and Evolution* **7**, 919–928.

Warwick, R.M., Clarke, K.R. & Suharsono. (1990) A statistical analysis of coral community responses to the 1982–83 El Niño in the Thousand Islands, Indonesia. *Coral Reefs* **8**, 171–179.

Appendix S5

THE CHESAPEAKE BAY BENTHOS DATA

The Chesapeake Bay Benthic Monitoring Program data

The data set used in this example was extracted from the Maryland Data Sets of the Chesapeake Bay Benthic Monitoring Program (http://www.baybenthos.versar.com/data.htm), which is a portion of the Chesapeake Bay Program (http://www.chesapeakebay.net/). Detailed information about the sampling protocol is found on that web page. The data, available online, come in the form of numerous text files, one per group of variables and per year. Legendre & Gauthier (2014) compiled and formatted these files in a *Rdata* file for immediate analysis in R. The <ChesapeakeBay.Maryland.RData> data are available in a zipped file found in S5 of their paper. The file contains faunal data collected during surveys at 27 sites, spring and fall, during 13 years, i.e. from 1996 to 2008, for a total of 702 data rows. It also contains data frames describing sampling information, sediment, water quality, and geographic coordinates of the sites; these files all have corresponding data rows. These authors used the Chesapeake data in a series of practical exercises in R for ecologists who want to learn how to analyse space-time ecological data. The exercises are detailed in Appendices S2 and S3 of their paper.

The faunal data for all sites, years and seasons consist of the abundances of 205 benthic macrofaunal taxa (203 invertebrates and 2 chordates) identified in the sediment of 25 brackish-water sites of the bay. Table S5.1 shows how the species are split between seasons and salinity groups. The spring survey data contain 181 species and the fall data 142 species. Two freshwater sites (#36 and #79) were present in the database; they contained 105 species. These two sites were excluded from the present example, which focussed on 25 brackish sites where 155 species were identified.

A map of the 25 brackish sites used in Ecological application 3 of the main paper, plotted with the RgoogleMaps package, is shown in Fig. S5.1. Also shown are the results of calculations with the R function TBI(), for the species abundance and occurrence data of this ecological application.

Reference

Legendre, P. & Gauthier, O. (2014) Statistical methods for temporal and space-time analysis of community composition data. *Proceedings of the Royal Society B* **281**, 20132728.

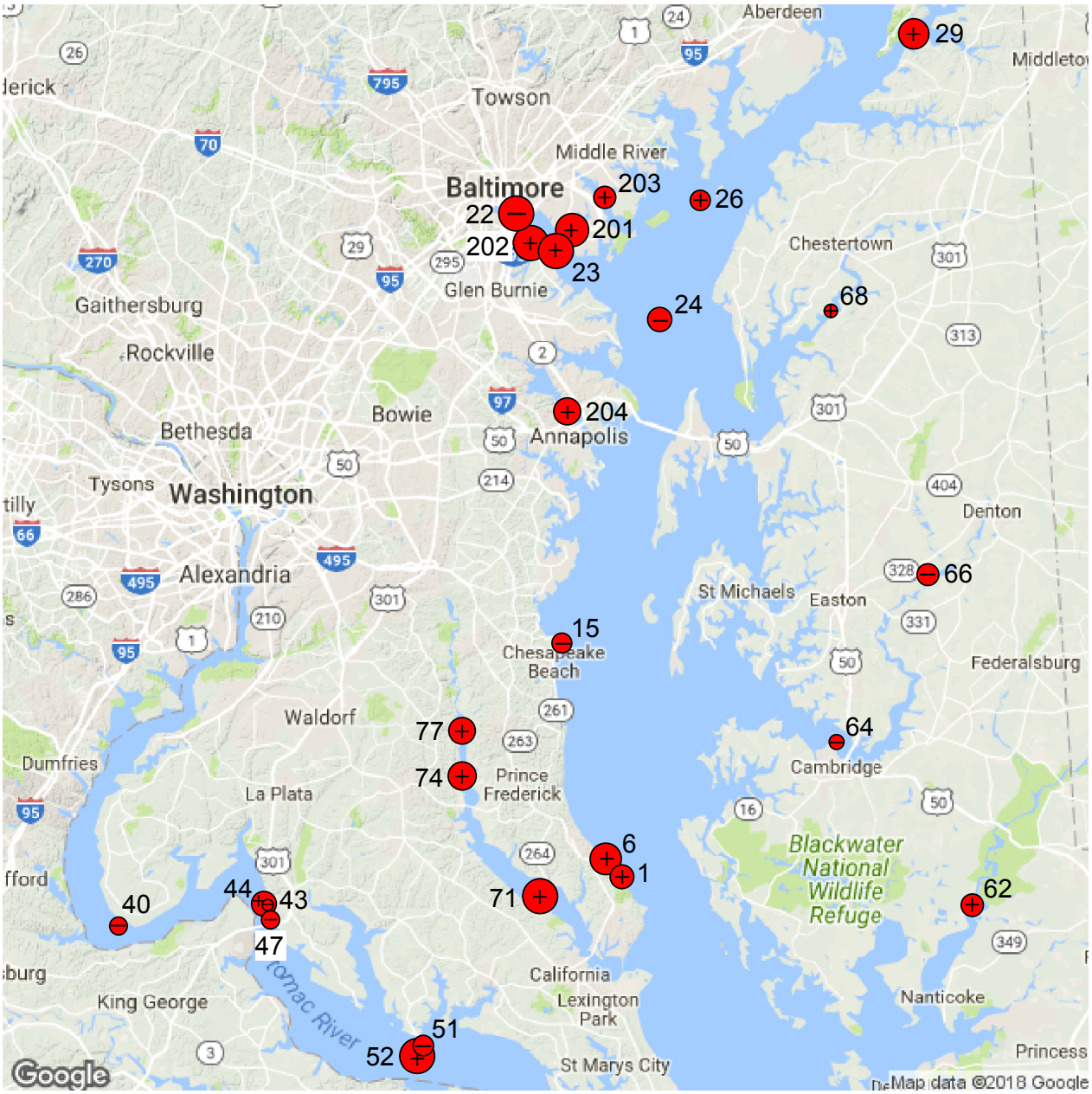

Figure S5.1. Map of the 25 brackish sites (red symbols) of the Chesapeake Bay ecological survey, produced with the RgoogleMaps package in R. Comparison of survey in years 2005 and 2008: point sizes are proportional to the TBI indices (percentage difference *D*). + signs indicate the 17 sites where *gains* in abundances-per-species dominated; – signs, the 8 sites where *losses* dominated.

Table S5.1. Number of species in subsets of the Chesapeake fauna data surveyed during 13 years, spring and fall. In total, 205 benthic species were found at the 27 survey sites.

	Spring	Fall	Spring and fall
Freshwater (2 sites)	93	58	105
Brackish (25 sites)	128	121	155
All survey sites (27 sites)	181	142	205

Results of calculations with R function TBI()

Compare Chesapeake Bay benthic fauna, 25 brackish sites, years 2005 and 2008, fall survey data.

1. Comparison based upon species abundance data, percentage difference D

```
( res.fauna.05.08.pcdiff = TBI(Y1, Y2, "%diff", pa.tr=FALSE, permute.sp=1, nperm=99999,
BCD=TRUE, test.BC=TRUE, test.t.perm=TRUE, clock=TRUE) )
# Computation time = 792.754000  sec

-----
$TBI
 [1] 0.6766467 0.5704698 0.9411765 1.0000000 0.6309524 0.7685950 1.0000000 1.0000000
 [9] 0.6960000 0.5777778 0.8632812 0.5081967 0.3229572 0.7083333 0.5164835 0.5843137
[17] 1.0000000 0.8983051 0.6385965 0.4244604 0.6256158 0.3846154 1.0000000 0.8020833
[25] 0.7611940

$p.TBI, nperm=9999
 [1] 0.3929 0.6419 0.0032 0.0033 0.4416 0.1919 0.0001 0.0001 0.3223 0.5884 0.0705
[12] 0.7248 0.9681 0.3324 0.7594 0.6169 0.0001 0.0381 0.4727 0.9135 0.5359 0.9165
[23] 0.0001 0.1564 0.2146

$p.adj, nperm=9999
 [1] 1.0000 1.0000 0.0672* 0.0672* 1.0000 1.0000 0.0025* 0.0025* 1.0000 1.0000 1.0000
[12] 1.0000 1.0000 1.0000 1.0000 1.0000 0.0025* 0.7239 1.0000 1.0000 1.0000 1.0000
[23] 0.0025* 1.0000 1.0000

$p.TBI, nperm=99999
 [1] 0.39289 0.64205 0.00322 0.00319 0.44089 0.19191 0.00002 0.00002 0.32282 0.58789
[11] 0.07023 0.72529 0.96808 0.33231 0.76022 0.61646 0.00001 0.03828 0.47287 0.91356
[21] 0.53649 0.91707 0.00001 0.15645 0.21396

$p.adj, nperm=99999
 [1] 1.00000 1.00000 0.06699 0.06699 1.00000 1.00000 0.00046 0.00046 1.00000 1.00000
[11] 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 0.00025 0.72732 1.00000 1.00000
[21] 1.00000 1.00000 0.00025 1.00000 1.00000
```

```
$BCD.mat
         B/(2A+B+C) C/(2A+B+C) D=(B+C)/(2A+B+C) Change
Site.1  0.167664671 0.50898204        0.6766467    +
Site.2  0.436241611 0.13422819        0.5704698    -
Site.3  0.029411765 0.91176471        0.9411765    +
Site.4  0.250000000 0.75000000        1.0000000    +
Site.5  0.053571429 0.57738095        0.6309524    +
Site.6  0.347107438 0.42148760        0.7685950    +
Site.7  0.950000000 0.05000000        1.0000000    -  * Abundances-per-species losses
Site.8  0.400000000 0.60000000        1.0000000    +  * 40% Ab.-per-sp. losses, 60% gains
Site.9  0.576000000 0.12000000        0.6960000    -
Site.10 0.100000000 0.47777778        0.5777778    +
Site.11 0.224609375 0.63867188        0.8632812    +
Site.12 0.483606557 0.02459016        0.5081967    -
Site.13 0.190661479 0.13229572        0.3229572    -
Site.14 0.229166667 0.47916667        0.7083333    +
Site.15 0.340659341 0.17582418        0.5164835    -
Site.16 0.482352941 0.10196078        0.5843137    -
Site.17 0.000000000 1.00000000        1.0000000    +  * All species gains
Site.18 0.135593220 0.76271186        0.8983051    +
Site.19 0.057894737 0.58070175        0.6385965    +
Site.20 0.316546763 0.10791367        0.4244604    -
Site.21 0.571428571 0.05418719        0.6256158    -
Site.22 0.161538462 0.22307692        0.3846154    +
Site.23 0.098039216 0.90196078        1.0000000    +  * All species gains
Site.24 0.005208333 0.79687500        0.8020833    +
Site.25 0.008955224 0.75223881        0.7611940    +

$BCD.summary
 mean(B/den) mean(C/den)   mean(D)   B/(B+C)   C/(B+C) Change
   0.2646503   0.4513519 0.7160022 0.3696222 0.6303778    +

$t.test_B.C, nperm=99999
               mean(B-C)      Stat    p.param  p.perm   p<=0.05
Paired t.test  -0.1867016 -1.826046 0.08031088 0.08221

$BC
[1] NA
-----
```

Note – The site names, Site.1 to Site.25, in the function output files correspond to the following site names on the map:

```
site.names
 [1] "S1"   "S15"  "S201" "S202" "S203" "S204" "S22"  "S23"  "S24"  "S26"
[11] "S29"  "S40"  "S43"  "S44"  "S47"  "S51"  "S52"  "S6"   "S62"  "S64"
[21] "S66"  "S68"  "S71"  "S74"  "S77"
```

2. Comparison based upon species occurrence (i.e. presence-absence) data, Sørensen D

```
( res.fauna.05.08.sor = TBI(Y1, Y2, "sorensen", pa.tr=FALSE, permute.sp=1, nperm=9999,
BCD=TRUE, test.BC=TRUE, test.t.perm=TRUE, clock=TRUE) )
# Computation time = 63.074000  sec

-----
$TBI
 [1] 0.4838710 0.4166667 0.6666667 1.0000000 0.2727273 0.1578947 1.0000000 1.0000000
 [9] 0.3846154 0.1818182 0.3600000 0.3333333 0.2500000 0.4000000 0.2500000 0.2857143
[17] 1.0000000 0.5238095 0.2592593 0.2592593 0.3333333 0.1538462 1.0000000 0.1304348
[25] 0.2592593

$p.TBI
 [1] 0.1533 0.3568 0.0259 0.0032 0.7316 0.9651 0.0001 0.0001 0.4021 0.9175 0.4696
[12] 0.5722 0.8212 0.3680 0.8236 0.7183 0.0001 0.1347 0.7819 0.8123 0.6038 0.9808
[23] 0.0001 0.9646 0.7098

$p.adj
 [1] 1.0000 1.0000 0.5180 0.0672? 1.0000 1.0000 0.0025* 0.0025* 1.0000 1.0000 1.0000
[12] 1.0000 1.0000 1.0000 1.0000 1.0000 0.0025* 1.0000 1.0000 1.0000 1.0000 1.0000
[23] 0.0025* 1.0000 1.0000

$BCD.mat
        B/(2A+B+C) C/(2A+B+C) D=(B+C)/(2A+B+C) Change
Site.1  0.06451613 0.41935484        0.4838710    +
Site.2  0.25000000 0.16666667        0.4166667    -
Site.3  0.16666667 0.50000000        0.6666667    +
Site.4  0.50000000 0.50000000        1.0000000    +
Site.5  0.09090909 0.18181818        0.2727273    +
Site.6  0.15789474 0.00000000        0.1578947    -
Site.7  0.87500000 0.12500000        1.0000000    -  * 87.5% species losses, 22.5% gains
Site.8  0.50000000 0.50000000        1.0000000    +  * 50% losses, 50% gains
Site.9  0.15384615 0.23076923        0.3846154    +
Site.10 0.09090909 0.09090909        0.1818182    +
Site.11 0.16000000 0.20000000        0.3600000    +
Site.12 0.22222222 0.11111111        0.3333333    -
Site.13 0.08333333 0.16666667        0.2500000    +
Site.14 0.20000000 0.20000000        0.4000000    +
Site.15 0.12500000 0.12500000        0.2500000    +
Site.16 0.14285714 0.14285714        0.2857143    +
Site.17 0.00000000 1.00000000        1.0000000    +  * All species gains
Site.18 0.38095238 0.14285714        0.5238095    -
Site.19 0.11111111 0.14814815        0.2592593    +
Site.20 0.11111111 0.14814815        0.2592593    +
Site.21 0.28571429 0.04761905        0.3333333    -
Site.22 0.03846154 0.11538462        0.1538462    +
Site.23 0.36363636 0.63636364        1.0000000    +  * All species gains
Site.24 0.04347826 0.08695652        0.1304348    +
Site.25 0.11111111 0.14814815        0.2592593    +

$BCD.summary
 mean(B/den) mean(C/den)   mean(D)  B/(B+C)  C/(B+C) Change
   0.2091492   0.2453511 0.4545004 0.460174 0.539826    +

$t.test_B.C
               mean(B-C)       Stat   p.param p.perm   p<=0.05
Paired t.test  -0.0362019 -0.6218665 0.5398928 0.5533

$BC
[1] NA
```

```
#                                  Appendix S6, R function
#
# An R function to standardize environmental data prior to TBI analysis.


#' Special standardization for environmental data prior to TBI analysis.
#'
#' After standardization, all variables will have the same weight (i.e. they will
#' all contribute the same variance) in the calculation of TBI indices.
#'
#' @param mat1 First data matrix, class matrix or data.frame.
#' @param mat2 Second data matrix, class matrix or data.frame.
#' @param non.neg=TRUE : make the data non-negative before scaling (recommended).
#'        non.neg=FALSE: keep standardized data with signs (due to centring).
#'
#' @return A list with the two matrices standardized as dexscribed above.
#'
#' @details
#' The two data sets are joined into a single data matrix, Y = rbind(Y.T1, Y.T2).
#' Y is standardized [Y.stand = scale(Y)], then it is separated into two tables of
#' the sizes of the original data matrices, before analysis with function TBI().
#'
#' Explanation:
#' (a) the two data tables are joined into a single data matrix, Y = rbind(Y.T1,
#' Y.T2), before standardization. In this way, the differences in values of each
#' variable for a given pair of sites in the two tables will remain comparable
#' to the differences ofcomputed from the original unstandardized values; in this
#' way, the distances computedbetween sites in T1 and T2 will be meaningful. This
#' precaution is important when there are differences in means between T1 and T2.
#' (b) Standardizing the variables insures that all variables will contribute the
#' same variance to the calculation of the TBI indices; the variances will not
#' depend on thephysical units of the variables or other contingencies that make
#' the variances unequal.
#'
#' Argument non.neg=TRUE makes the values non-negative to produce data without
#' negative signs. It does not change the results of the TBI tests.
#'
#' @author  Pierre Legendre \email{pierre.legendre@umontreal.ca}, 2018
'scale.for.TBI' <-
      function(mat1,mat2,
                  non.neg=TRUE)
      {
      mat1 <- as.matrix(mat1)
      mat2 <- as.matrix(mat2)
      dim.1 <- dim(mat1)
      dim.2 <- dim(mat2)
      if(!is.numeric(mat1)) stop("First data matrix not numeric")
      if(!is.numeric(mat2)) stop("Second data matrix not numeric")
      if(dim.1[1] != dim.2[1]) stop("Data sets have different numbers of rows")
      if(dim.1[2] != dim.2[2]) stop("Data sets have different numbers of columns")
      #
      tmp <- scale(rbind(mat1,mat2))
      if(non.neg) tmp <- tmp - min(tmp)
      mat1 <- tmp[1:n12,]
      mat2 <- tmp[(n12+1):(2*n12),]
      list(mat1=mat1, mat2=mat2)
      }
```