ANALYZING OR EXPLAINING BETA DIVERSITY? COMMENT

Pierre Legendre,[1,3] Daniel Borcard,[1] and Pedro R. Peres-Neto[2]

In 2005, we published a paper (Legendre et al. 2005; hereafter referred to as LBP) explaining the ecological and statistical basis for the analysis of the variation in species composition among sites (this is one operational definition of beta diversity), a subject of great importance for the understanding of the generation and maintenance of beta diversity and the establishment of ecologically sound ecosystem conservation policies. Tuomisto and Ruokolainen (2006) (hereafter referred to as TR) pointed out several issues in our 2005 paper. As a response, this note seeks to clarify the issues behind this contention.

Our major point of disagreement concerns the links between (1) the ecological predictions derived from neutral theories of beta diversity, (2) the statistical hypotheses derived from these predictions, and (3) the statistical methods used to test these hypotheses. In LBP, TR, and the present *Comment*, distance-based methods refer to statistical methods where the geographic relationships among sites are represented by a distance matrix. These methods include the Mantel test and the derived method of regression on distance matrices. The raw-data approach refers to multiple regression and canonical analysis, where the spatial relationships among sites are represented by a rectangular table containing geographic coordinates, a polynomial of the geographic coordinates, principal coordinates of a neighbor matrix (PCNM) (Borcard and Legendre 2002), or derived forms.

The purpose of this *Comment* is to show that (1) some predictions of Hubbell's neutral theory, especially the presence of positive autocorrelation, can be stated and tested using the raw-data approach, (2) the distance approach as proposed by TR (partitioning on distance matrices) is statistically flawed, and (3) when a raw-data hypothesis is translated into distances, the corresponding statistical test lacks power; therefore, whenever a hypothesis can be formulated in terms of raw data

[1] Département de Sciences Biologiques, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec Canada H3C 3J7.
[2] Département des Sciences Biologiques, Université du Québec à Montréal, C.P. 8888, Succursale Centre-ville, Montréal, Québec Canada H3C 3P8.
[3] E-mail: Pierre.Legendre@umontreal.ca

instead of distances, it should be tested using the raw-data approach.

TR (p. 2698) devoted a large section to three levels of abstraction that were proposed and defined in LBP (pp. 437–438). Level-1 questions concern alpha diversity, or the variation in the species identity of organisms at individual sites. For example: What are the soil conditions associated with the presence and identity of 10 tree species in a 1-ha plot of temperate mixed-wood forest? Level-1 questions are irrelevant for the present *Comment*. For levels 2 and 3, which are the concern of LBP and TR, the questions asked are the same in the two papers. Level 2 concerns the variation in community composition among sites in a region of interest, which is beta diversity (community composition encompasses species composition and species abundances); canonical variation partitioning is adequate to address questions pertaining to that level (LBP, p. 438; and TR, p. 2705). For example: Is the variation in community composition among sites due to variation in environmental conditions or to neutral community dynamics? Level-3 questions concern the variation in beta diversity among groups of sites; questions related to that level may be addressed by the distance-based approach (LBP, p. 438; and TR, p. 2705), provided that the technical and statistical uncertainties raised below are settled; see also point 2 in the section *Other points*. Note that TR restrict the level-3 questions to pairs of sites, whereas LBP use level-3 questions to investigate the differences among larger groups of sites; for example, adjacent geographic regions with several sites in each. For example: The public garden in town A is surrounded by identical flower beds, each one containing seven species. The beta diversity (variation in species composition among beds) in that garden is zero. In town B, the public garden was designed with flower beds that are all different in species compositions. Beta diversity is high among the flower beds. One may wonder why there is such a big difference in flower bed composition between the two towns. Is it cultural? Financial? Or is it due to the limited availability of suitable flowers in town A due to soil or climate? This question does not focus on the identities of the flower species, but on the variance in community composition between the two towns: Are the multispecies dispersion matrices homogeneous when comparing the two towns, and if not, why? This would be a level-3 question.

Spatial autocorrelation in community composition data

"We argue that S. P. Hubbell's neutral theory can only be tested using the distance approach, because its testable predictions are stated in terms of distances, not in terms of raw data" (TR, *Abstract*: details on p. 2703). The section *Can we test Hubbell's neutral theory using the raw-data approach?* will show that spatial autocorrelation can and should be tested using the raw-data

approach. Several other aspects and predictions of neutral theory can also be tested without recourse to distance matrices; see McGill et al. (2006) for a recent review. Here are a few examples. First, spatial variance can be used to test neutrality: Under random walk in species abundances ("ecological drift" sensu Hubbell 2001), variance is expected to increase with time (Clark and McLachlan 2003). Second, the compensatory dynamics hypothesis of neutral theory can be tested by assessing the statistical hypothesis that species covary negatively within communities; Houlahan et al. (2007) tested this hypothesis on 41 natural communities. Third, neutrality can be tested by an evolutionary approach, by studying extinction rates and the ages of species (Ricklefs 2006). Finally, the relative importance of environmental control and neutrality can be tested using variation partitioning by canonical analysis, as shown in LBP.

Tests of autocorrelation in the distance world.—Hubbell's neutral theory predicts the presence of positive autocorrelation in community composition due to dispersal limitation. In this section and the next, we will show that spatial autocorrelation can be tested both in the distance and raw-data worlds, but that the raw-data approach is more powerful and has better-known statistical properties. Let us review the distance approach first.

The procedure proposed by Nekola and White (1999), and used by Hubbell (2001) to detect spatial autocorrelation, is to plot ecological similarity as a function of geographic distance. Tuomisto et al. (2003) developed this idea further by log-transforming the geographic distances to make the relationship linear, and measured the fit using a squared matrix correlation (R_M^2), which is the square of the Mantel correlation between distance matrices. We agree with them that this is one way of describing the relationship in this simplified form of correlogram, and we mentioned it in LBP (*Abstract* and p. 442). The squared matrix correlation is then interpreted as an overall coefficient of spatial autocorrelation.

A Mantel correlogram is another form of analysis based on distance matrices (Legendre and Legendre 1998: section 13.1.5). It provides a more detailed analysis than an overall Mantel correlation. A critical point of logic is that a significant value for a distance class in a Mantel correlogram does not mean that *spatial autocorrelation* is present in the response data; it only means that there is a significant *spatial structure* in the data. Users of both raw-data and distance methods should be aware of that and be cautious about the interpretation of the results of correlogram analysis. A spatial structure can have different origins (Fortin and Dale 2005: chapter 5): It may indicate spatial dependence induced by the environmental factors, or spatial autocorrelation resulting from the stochastic demo-

graphic processes described by Hubbell (2001). The algebraic equation describing the spatial variation of a variable y at sampling locations i is: $y_i = f(\mathbf{X}_i) + SA_i + \varepsilon_i$ (Legendre et al. 2002). Because of their spatial structures, the environmental variables in table $\mathbf{X}$ may induce spatial dependence in the response y; that effect is separate from spatial autocorrelation proper (SA) which results from the spatial dynamics of y; ε_i is the "local innovation," or error term, at location i. For example, if there is an environmental effect in the species data and the forcing environmental variable has a broad-scale spatial structure, this will cause similarity decay plots to show monotonically decreasing similarity as geographic distance increases. A significant negative relationship (R_M) in a similarity decay plot or a significant value in a Mantel correlogram does not allow one to conclude about neutrality because it may be due to unmeasured environmental variables that are spatially autocorrelated and that influence the species distributions; it can be interpreted in terms of neutrality only if other factors were logically excluded or partialed out.

Can we test Hubbell's neutral theory using the raw-data approach?—In this section, we will show by simulations that the raw-data approach is the method that should be used to test hypotheses about positive spatial autocorrelation in neutral communities, one of Hubbell's predictions. We conducted a new Monte Carlo study akin to the one in our original publication (LBP). In that paper, canonical variation partitioning had proved to be the most powerful method to detect spatial autocorrelation in simulated community composition data, when compared to Mantel tests, whatever the method used to represent the spatial relationships in the two forms of analysis (see LBP: Table 1, columns with headings [b + c]). This was especially true when the spatial structure was modeled using PCNM variables, which are directly related to spatial autocorrelation functions (Dray et al. 2006). In the new study reported here, we simulated spatially patterned communities along a transect following Hubbell's neutral model with migration. Our Monte Carlo study was based on the following steps: (1) Generate a neutral metacommunity distribution of relative species abundances using Hubbell's algorithm (Hubbell 2001: Fig. 9.1, p. 291). The algorithm is based on two parameters: J_m is the number of individuals in the metacommunity, and θ is the fundamental biodiversity number. The metacommunity starts with a single individual ($j = 1$) of a single species and individuals are added until the metacommunity reaches J_m individuals. As each individual is added to the metacommunity, the individual is assigned to either an already existent species or to a new species (which is then added to the metacommunity). The probability of the jth individual being assigned to a new species is $\theta/(\theta + j - 1)$. If the jth individual was not assigned to a new species, then it was added to a previously existent species with a probability

TABLE 1. Rates of rejection of H_0 at significance level $\alpha = 0.05$ for the raw-data and distance approaches, each based on 1000 simulations, for detecting spatial structures due to migration in simulated neutral communities.

θ	m	lc	$J = 1000$		$J = 10\,000$	
			Raw data	Distance	Raw data	Distance
10	0.1	20	0.940	0.215	0.137	0.069
10	0.1	30	0.992	0.201	0.200	0.043
10	0.1	50	1.000	0.170	0.194	0.053
10	0.2	20	1.000	0.450	0.277	0.077
10	0.2	30	1.000	0.430	0.401	0.077
10	0.2	50	1.000	0.397	0.545	0.066
20	0.1	20	0.998	0.286	0.165	0.055
20	0.1	30	0.999	0.276	0.228	0.066
20	0.1	50	1.000	0.221	0.296	0.047
20	0.2	20	1.000	0.652	0.424	0.101
20	0.2	30	1.000	0.620	0.571	0.094
20	0.2	50	1.000	0.544	0.776	0.088
30	0.1	20	1.000	0.385	0.196	0.075
30	0.1	30	1.000	0.332	0.306	0.065
30	0.1	50	1.000	0.311	0.412	0.057
30	0.2	20	1.000	0.759	0.541	0.088
30	0.2	30	1.000	0.759	0.725	0.101
30	0.2	50	1.000	0.674	0.902	0.099

Note: Here, θ is the fundamental biodiversity number; m is the migration rate; lc is the number of local communities; and J is the number of individuals in the local community.

equal to the species' proportion of individuals in the metacommunity. (2) Once the neutral metacommunity was created, we randomly populated lc local communities with individuals from the metacommunity. Each local community was populated with the exact same number of individuals J (i.e., $J = J_m/lc$), randomly sampled with replacement from the metacommunity. Again, each species was sampled according to the species' proportion of its individuals in the metacommunity. (3) We then set out a death and immigration process, following McGill (2003), in a simulation based on 10 000 time steps as follows: First, local communities in step 2 were spatially distributed along a transect. Then, the simulation started, and at each time step, a randomly chosen individual from each local community was killed (deleted). Next, a random uniform number was generated: If smaller than m (migration rate), a randomly chosen individual was copied (i.e., birth with migration) from one of the two nearest-neighbor local communities along the transect, with equal chances, to replace the deleted individual; if larger than m, an individual from the local community was randomly chosen and duplicated (i.e., birth without migration). The probability of an individual belonging to a particular species was equal to the species' proportion of individuals in the local community (without migration) or in the chosen adjacent local community (with migration). (4) When the migration dynamics was completed after 10 000 time steps, we tested whether

the community distribution was spatially structured using both the raw-data and distance approaches.

The two analytical approaches were identical to those followed in LBP. The raw-data analyses used canonical redundancy analysis (RDA) of the Hellinger-transformed response species table by a table of principal coordinates of neighbor matrices (PCNM). For simplicity, $n/2$ PCNM variables with positive eigenvalues were used (i.e., no selection of the best PCNMs), providing a conservative test since a large number of degrees of freedoms are lost to the PCNMs. The PCNMs used in the tests only modeled patterns related to positive spatial autocorrelation, which was the kind predicted by Hubbell's theory. The distance approach was based on a Mantel test of the correlation between a Hellinger distance matrix among sites, representing the species data, and a geographic distance matrix $\mathbf{D}(XY)$ computed from the X and Y geographic coordinates (i.e., positions along the transect); this is the way spatial relationships are represented in most instances in Mantel tests. Significance tests were based on 999 random permutations. Using combinations of θ, J, lc, and m, we generated 1000 metacommunities for each combination, and each was tested using the raw-data and distance approaches. Results are presented in Table 1; they clearly show that the raw-data approach is a far more powerful method than the distance approach for detecting spatial autocorrelation in neutral communities.

In the last paragraph of page 2703, TR argued that

An existing spatial pattern in community composition can be described a posteriori, especially by such powerful methods as PCNM [. . .]. However, doing so does not test the neutral model, because the neutral model did not predict that this was the particular spatial pattern that was expected to emerge in this particular case. Any specific spatial pattern in community composition is just as much in accordance with the neutral model as any other, as long as the degree of spatial autocorrelation is similar.

The last statement is true for distance as well as raw-data methods and does by no means invalidate the latter in favor of the former. This is why we issued a warning about the various origins of spatial structure in the section *Tests of autocorrelation in the distance world* above. This being said, *if* other sources of variation have been logically or technically excluded, PCNM analysis (which is closely related to autocorrelation functions; Dray et al. 2006) is far more powerful at detecting raw-data structures emerging as a consequence of autocorrelation, than the distance approach is at detecting distance patterns emerging from autocorrelation. Therefore, we advocate the raw distance (PCNM) approach.

TR argued on page 2705 that, because Hubbell's neutral theory's "testable predictions are stated in terms

of distances, not in terms of raw data," only a method involving distances would be valid for testing the theory. They went on and wrote that "attempting to test this ecological hypothesis using the raw-data approach may give quite misleading results." Contrary to the simulations reported by LBP, autocorrelation in all communities analyzed in the present simulations was strictly generated under Hubbell's neutral model. Table 1 shows that the distance decay method was not nearly as powerful as the raw-data approach in detecting spatial autocorrelation in the data. Our new simulations show, therefore, that the raw-data approach is actually the most appropriate for testing this aspect of the theory.

Partitioning on distance matrices

Tuomisto et al. (2003) proposed to partition the variation of the ecological resemblance matrix using multiple regression on distance matrices: "More recently, variation partitioning has been extended to the distance approach by using multiple regression on distance matrices (Duivenvoorden et al. 2002, Tuomisto et al. 2003)" (TR, p. 2698). We developed that regression method for phylogenetic analysis (Legendre et al. 1994) and implemented it in the computer program (Casgrain 2001) that was used by Duivenvoorden et al. (2002), Tuomisto et al. (2003), and other authors who followed in their footsteps. A technical problem arises with the extension that they proposed. They combined by subtraction the coefficients of determination of three matrix regressions (R_M^2) to compute linearly independent fractions of variation allegedly corresponding to: (a) the nonspatially structured variation explained by the environmental distance matrix, (b) the spatially structured species variation explained by the environmental distance matrix, and (c) the spatially structured species variation not explained by the environmental distance matrix. We argue that variation partitioning cannot be computed in that way from distance matrices even if the ecological hypotheses are stated in the distance world. The problem here is not ecological in nature but mathematical (application of variation partitioning to distance matrices) and statistical (low power of the distance approach when compared to the raw-data approach). These points are discussed below.

TR have indeed proposed to carry out variation partitioning using distance-based R_M^2, but for this method to be demonstrated to be valid, one should first show (1) that it produces approximately correct estimates of the fractions of variation; (2) that the fractions of variation thus isolated are additive; (3) whether families of variables (e.g., environment) should be represented by a single synthetic distance matrix or by one matrix per variable, two procedures which lead to very different results; and (4) how the R_M^2 coefficients and the fractions of variation should be interpreted. The rationale presented by TR for their methodological

extension of variation partitioning solely relies on the Legendre et al. (1994) paper. This is insufficient: The validity of multiple regression on phylogenetic distance matrices does not warrant the extension to variation partitioning on ecological distance matrices. TR would provide an extremely constructive element to this debate by producing simulations giving clear answers to the four questions above.

Assuming that TR could answer the four questions about their partitioning method, the question of power remains acutely important. It motivated our comparative simulation study of the two partitioning methods. In both LBP and the present paper (previous section), we showed that the Mantel test had extremely low power in spatially explicit simulations. This means that it was unlikely to detect a species–environment relationship or a spatial structure when such an effect was present in the data. Considering the high cost of good ecological data, we felt it was our responsibility to explain to ecologists, who sought variation partitioning results to support their theories, what the most appropriate statistical method was to achieve their aims.

There are several statistical reasons for the low power of the Mantel test. The reasons revolve around (1) the fact that the R_M^2 statistic is inappropriate for questions about variation of community composition among sites in a region of interest (level-2 questions in LBP, TR, and in the section *Other points* below); (2) the difficulty of computing an adjusted form of R_M^2; and (3) the lack of additivity of R_M^2, which is crucial for variation partitioning.

Incorrect statistic.—The recent literature shows that many researchers still use the distance approach to solve level-2 questions, which is incorrect. LBP and TR agree on this point (TR: Fig. 2). Some examples are listed in LBP (pp. 438–439). It is thus urgent to warn researchers against this confusion. The technical reason is that the quantity which is partitioned in partitioning on distance matrices is the sum-of-squares of the distances, SS(**D**), and this quantity is not equivalent and cannot be reduced to the total sum-of-squares of the response data matrix SS(**Y**) (LBP: Eq. 2). Hence, the Mantel statistic R_M^2 is not equivalent to the canonical R^2. The simulation results reported in LBP were crucial in showing that this makes a big difference for interpretation of the results.

Adjusted coefficient of determination.—We now know that in regression, the R^2 statistic is a biased estimate of the true population R^2. Adjustments for this bias exist: Ohtani (2000) showed that, under certain assumptions, Ezekiel's (1930) adjusted coefficient of determination (R_a^2) is an unbiased estimator of the contribution of a set of explanatory variables **X** to the explanation of the variance of a single response variable **y**: $R_a^2 = 1 - a(1 - R^2)$, where $a = (n - 1)/(n - m - 1)$ for models with an intercept; n is the number of observations, and m is the number of explanatory

variables in the model. Peres-Neto et al. (2006) have shown that the adjusted canonical R_a^2, obtained by applying Ezekiel's correction to the canonical R^2 obtained in RDA, also produces unbiased estimates of the real contribution of the variables in $\mathbf{X}$ to the explanation of a response matrix $\mathbf{Y}$, and that it is these values of R_a^2 that must be used to obtain unbiased estimates of the fractions in variation partitioning. This conclusion is important for variation partitioning based upon distance matrices. No equation has been proposed to compute an adjusted R-square (R_{Ma}^2) in Mantel-type regression. In Ezekiel's equation, should we use $m = 1$ for a single explanatory matrix $\mathbf{X}$ in the regression, or should we make m equal to the number of original variables that were included in the calculation of the distances leading to $\mathbf{X}$? Should n be the number of original objects (sites) or the number of distances in half or the whole distance matrix? In any case, the very idea of an adjusted R^2 is suspicious in matrix regression because R_M^2 should be interpreted as a measure of fit of a model to pairwise distances and not in terms of the proportion of variation of a response matrix $\mathbf{D_1}$ explained by an explanatory matrix $\mathbf{D_2}$.

Additivity.—One final point concerns the additivity of the fractions resulting from variation partitioning. We now know how to partition the variation of a response matrix $\mathbf{Y}$ with respect to several explanatory matrices $\mathbf{X}$ using RDA. In raw-data partitioning, an identical total fraction of explained variation is obtained, whether all explanatory variables are put in a single table $\mathbf{X}$ or they are divided into any number of sub-tables (environmental, spatial, and so on). The effects of the explanatory variables are thus additive. This is not the case in partitioning on distance matrices: Different total amounts of explained variation for the response $\mathbf{Y}$ are obtained if one includes all explanatory variables in a single distance matrix or if separate distance matrices are computed for the various explanatory variables. This clearly shows that variation partitioning based on distances lacks the essential property of additivity, which is the basis for interpretation of variation partitioning results. TR are well aware of this fact, as they mention that in the distance approach "R^2 values will change depending on [. . .] whether all environmental variables are combined into a single distance matrix or used in separate matrices" (p. 2707).

The debate about the R^2 does not concern the statistical tests themselves (Mantel test for two matrices, or global test in a multiple regression on distance matrices). It only concerns the use of the squared Mantel coefficient, or the R^2 of the multiple regression on distance matrices R_M^2, as a measure of the fraction of explained variation, and following that as the basis for computation of the fractions in variation partitioning.

Question levels

We agree with TR that Mantel tests should be restricted to level-3 questions, described in the *Introduction*, whereas canonical variation partitioning addresses level-2 questions: We spelled that point out in LBP. However, TR's last paragraph of their section *Testing ecological hypothesis C* (p. 2704) claims that the hypothesis of environmental control of species distributions (a level-2 question) "is testable with the distance approach." They simply restate a level-2 question (the effect of environmental variables on species distributions) in level-3 terms (correlation between environmental distances and community distances). TR's recommendation is thus in blatant contradiction to our simulation results; distance-based methods must be avoided for level-2 questions. The simulations reported in Table 1 (sections B–E, column [a + b]) of LBP have clearly shown that Mantel tests were highly inefficient at detecting species–environment relationships when such relationships were present in data. It is therefore counterproductive to restate a level-2 question in distance terms.

Canonical variation partitioning is perfectly adequate to test predictions under TR's hypotheses A, B, and C (p. 2703). The low power of the Mantel approach to detect environmental relationships or spatial structures in community composition data where these relationships were present, as demonstrated by the simulations of LBP (Table 1, A–C), is a sufficient reason to avoid the distance approach to answer questions related to these hypotheses. Neutral processes generate spatial autocorrelation in community data, and PCNM analysis is very efficient at detecting it. This is shown by the simulation results reported in Borcard and Legendre (2002), in LBP, and in Table 1 of the present paper. These simulations also demonstrate that the Mantel test is very inefficient at detecting spatial autocorrelation in data. In all cases investigated here, the Mantel test is inadequate for level-2 questions reformulated in terms of distances, due to its extremely low statistical power; hence, the ecological hypotheses related to level-2 questions must be tested in the world of the raw data whenever that is possible.

Other points

1) In their section *The difference between "analyzing" and "explaining" beta diversity,"* TR (p. 2701) propose that any pairwise comparison of sites is a level-3 question. This is an incorrect statement: Pairwise comparisons can be made in the raw-data as well as in the distance world. It is therefore not surprising that, from this incorrect premise, they derived the incorrect conclusion that all the examples we cited as misuses of the Mantel test were in fact legitimate.

2) TR agreed with LBP that level-3 questions concern the variation in beta diversity among pairs or larger

groups of sites. Let us concentrate on larger groups of sites found in different sub-areas of the region under study: Establishing that there is variation among sub-areas can be translated into testing a hypothesis of homogeneity/heterogeneity of dispersion matrices among regions, each containing a group of sites. This can be done by the classical test of homogeneity of multivariate dispersion matrices (which compares the full within-group multivariate variance-covariance matrices; Kullback 1959), or the new test of homogeneity of within-group dispersions (ANOVA of distances of individual multivariate observations to their group centroid) described by Anderson (2006). More than two sites per group should be available for these analyses in order to obtain reliable estimates of the mean and variance parameters. The Mantel-based approach should be compared, by numerical simulations, to homogeneity analysis in order to determine which of the two approaches is the most powerful to detect heterogeneity in data.

Conclusion

The main argument presented by TR is that the predictions of Hubbell's model are formulated in terms of distances; hence, tests of significance should be conducted in the distance world. We have shown (1) that the presence of autocorrelation predicted by the neutral model is a level-2 question, which can and should be tested by canonical analysis using raw-data tables; (2) that the Mantel test should not be extended to variation partitioning on distance matrices until the very serious technical issues about that extension have been settled; and (3) that important level-2 hypotheses, for instance, those about environmental control determining community structure, should imperatively be tested in the world of raw data due to the extremely low power of the Mantel test.

The method of partitioning the variation of multivariate community composition data tables between environmental and spatial components (Borcard et al. 1992, Borcard and Legendre 1994), based on RDA, has now been generalized to several explanatory data matrices; a statistical function is freely available in the R language "vegan" library (Oksanen et al. 2007) to compute the results for up to four such matrices. Its domain of application as a method for spatial analysis was greatly improved by the development of PCNM analysis (Borcard and Legendre 2002, Borcard et al. 2004). We did the theoretical and the simulation work necessary to demonstrate the statistical correctness and usefulness of our method. Likewise, we developed regression on distance matrices for phylogenetic studies (Legendre et al. 1994), but we did not interpret R^2_{M} as anything but the adjustment of a model to data, for the reasons developed in the section *Partitioning on distance matrices* above. The proponents of partitioning on distance matrices are

invited to provide the same in-depth work to demonstrate the soundness of their approach.

Literature cited

Anderson, M. J. 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics 62:245–253.

Borcard, D., and P. Legendre. 1994. Environmental control and spatial structure in ecological communities: an example using Oribatid mites (Acari, Oribatei). Environmental and Ecological Statistics 1:37–61.

Borcard, D., and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling 153:51–68.

Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. Ecology 85:1826–1832.

Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. Ecology 73:1045–1055.

Casgrain, P. 2001. Permute! Version 3.4. User's manual. Département de sciences biologiques, Université de Montréal, Montreal, Canada.

Clark, J. S., and J. S. McLachlan. 2003. Stability of forest biodiversity. Nature 423:635–638.

Dray, S., P. Legendre, and P. R. Peres-Neto. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling 196:483–493.

Duivenvoorden, J. F., J.-C. Svenning, and S. J. Wright. 2002. Beta diversity in tropical forests. Science 295:636–637.

Ezekiel, M. 1930. Methods of correlation analysis. John Wiley and Sons, New York, New York, USA.

Fortin, M.-J., and M. R. T. Dale. 2005. Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, UK.

Houlahan, J. E., et al. 2007. Compensatory dynamics are rare in natural ecological communities. Proceedings of the National Academy of Sciences (USA) 104:3273–3277.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey, USA.

Kullback, S. 1959. Information theory and statistics. John Wiley and Sons, New York, New York, USA.

Legendre, P., D. Borcard, and P. R. Peres-Neto. 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. Ecological Monographs 75:435–450.

Legendre, P., M. R. T. Dale, M.-J. Fortin, J. Gurevitch, M. Hohn, and D. Myers. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography 25:601–615.

Legendre, P., F.-J. Lapointe, and P. Casgrain. 1994. Modeling brain evolution from behavior: a permutational regression approach. Evolution 48:1487–1499.

Legendre, P., and L. Legendre. 1998. Numerical ecology, 2nd English edition. Elsevier Science BV, Amsterdam, the Netherlands.

McGill, B. J. 2003. A test of the unified neutral theory of biodiversity. Nature 422:881–885.

McGill, B. J., B. A. Maurer, and M. D. Weiser. 2006. Empirical evalution of neutral theory. Ecology 87:1411–1423.

Nekola, J. C., and P. S. White. 1999. The distance decay of similarity in biogeography and ecology. Journal of Biogeography 26:867–878.

Ohtani, K. 2000. Bootstrapping R^2 and adjusted R^2 in regression analysis. Economic Modelling 17:473–483.

Oksanen, J., R. Kindt, P. Legendre, and R. B. O'Hara. 2007. vegan: community ecology package version 1.8-5. ⟨http://cran.r-project.org/⟩

Peres-Neto, P., P. Legendre, S. Dray, and D. Borcard. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87:2614–2625.

Ricklefs, R. E. 2006. The unified neutral theory of biodiversity: Do the numbers add up? Ecology 87:1424–1431.

Tuomisto, H., and K. Ruokolainen. 2006. Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. Ecology 87:2697–2708.

Tuomisto, H., K. Ruokolainen, and M. Yli-Halla. 2003. Dispersal, environment, and floristic variation of western Amazonian forests. Science 299:241–244.