Community Ecology 8(1): 41-46, 2007
1585-8553/$20.00 © Akadémiai Kiadó, Budapest
DOI: 10.1556/ComEc.8.2007.1.6

A spatially explicit measure of beta diversity

G. Bacaro[1] and C. Ricotta[2]

[1] *Department of Environmental Science "G. Sarfatti", University of Siena, Via P. A. Mattioli 4, 53100 Siena, Italy. Fax: +390577232860, E-mail: bacaro@unisi.it*
[2] *Department of Plant Biology, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185 Rome, Italy. Fax: +39064457540, E-mail: carlo.ricotta@uniroma1.it*

Keywords: Multivariate semivariogram, Semivariance, Spatial autocorrelation, Species complementarity.

Abstract. Plant communities are generally spatially structured. Therefore, in order to enhance the interpretation of distance-dependent community patterns, spatially explicit measures of β-diversity are needed that, besides simple species turnover, are able to account for the rate at which biological similarity decays with increasing distance. We show that a multivariate semivariogram computed from species presence and absence data can be considered as a space-dependent alternative to existing definitions of β-diversity. To illustrate how the proposed method works, we used a classical data set from a second-growth piedmont hardwood forest.

Nomenclature: Kartesz (1994).

Introduction

Ecologists traditionally recognize three different components of species diversity: alpha or within-community diversity (α), beta or between-community diversity (β) and gamma or total species diversity in a region (γ) (Whittaker 1972). In particular β-diversity can be defined also as the amount of turnover in species composition from one location to another. In this sense, β-diversity is essentially the same as McArthur's (1965) between habitat diversity.

A common approach to compute β-diversity is to look at the degree to which the species composition of sample plots differs. For presence and absence data, Whittaker (1960) proposed to measure β-diversity as:

$$\beta = \gamma / \overline{\alpha} \qquad (1)$$

where γ is the species richness of the pooled set of plots, and $\overline{\alpha}$ is average species richness in the sample plots. With some rearrangements of Eq. (1), it can be shown that β reflects the inverse of the average frequency of species: $\beta = N / \overline{N}_i$, where N is the total number of plots, and $\overline{N}_i$ is the average number of plots out of N plots that contain species i (Vellend 2001). That is, rare species increase the values of β more than frequent species.

Additional measures of β-diversity can be found in Wilson and Shmida (1984), Tóthmérész (1998), Koleff et al. (2003), Veech et al. (2002), and Legendre et al. (2005). Most of these measures are essentially summary statistics that quantify plot-to-plot variability in species composition independently of the position of individual plots on environmental gradients (Vellend 2001). Nonetheless, plant communities are generally spatially structured. Therefore, in order to enhance the interpretation of distance-dependent community patterns, that is, the rate at which biological similarity decays with increasing distance (Poulin 2003), spatially explicit measures of β-diversity are needed.

In recent years, many authors have looked at distance decay of species composition similarity among sites. Examples from this recent literature include: Nekola and White (1999), Wagner (2003), and Palmer (2005). Among these authors, Wagner (2003) used geostatistical analysis to extend multiscale ordination to nonsystematic spatial samples.

In this paper, we will put the work of Wagner (2003) into the context of diversity theory. We will show that the autocorrelation structure in species composition within a given set of plots provides a meaningful measure of spatially dependent β-diversity. Additionally, by decomposing this measure in single-species contributions to overall β-diversity, we will illustrate the effects of species frequencies and species spatial distribution (i.e., clumpedness) in shaping community structure. To illustrate how the proposed method works, we will show an example using a classical data set from a second-growth piedmont hardwood forest (Reed et al. 1993).

Multivariate variogram modeling

Geostatistical methods deal with the question of how the variance of a given variable depends on the distance between observations (sample plots). For species presences and absences, the occurrence of a species i in a single plot can be described as a binary variable x_i that takes the value 1 if species i is present in a plot, and 0 if it is absent. Accordingly, the autocorrelation structure or distance-dependence of a given set of plots can be quantitatively described in terms of its semivariance $\gamma_i(h)$ for a range of distance classes h:

$$\gamma_i(h) = \frac{1}{2n(h)} \sum_{n(h)} (x_{ia} - x_{ib})^2 = \frac{M}{2n(h)} \tag{2}$$

where x_{ia} and x_{ib} are the values of the variable x_i at plots a and b, respectively, $n(h)$ is the number of pairs of plots a and b separated by the lag distance h, and M is the number of mismatches (i.e., species present in one plot and absent in the next plot, Wagner 2003; see also Podani and Csontos 2006). $\gamma_i(h)$ is thus a measure of the autocorrelation structure of x_i as a function of h. The larger $\gamma_i(h)$, the less similar are the plots. The experimental semivariogram relates the average semivariance of the differences in the data to the intervals between sample plots and provides a concise description of the pattern of spatial variability.

To quantitatively describe the way in which $\gamma_i(h)$ changes with distance h, experimentally derived semivariances are commonly used to fit an approved variogram function. The principal parameters of a fitted variogram function are (Cressie 1993): nugget variance (the level of spatial variation in the data at scales smaller than the distance separating adjacent pairs of plots), sill (the amount of spatially dependent structural variance in the data), and range (the lag distance up to the beginning of the sill where the spatial dependence extends).

As suggested by Wagner (2003), to generalize the univariate definition of Eq. (2) to multivariate community data the binary values x_{ia} and x_{ib} are substituted by the vectors X_a and X_b that represent the presence/absence data of S species x_i at plots a and b and the empirical semivariance becomes half the squared dissimilarity between X_a and X_b:

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{n(h)} \left\| X_a - X_b \right\|^2 \tag{3}$$

where the term $\|X_a - X_b\|$ represents the (symmetric) dissimilarity between the binary vectors X_a and X_b. There are in principle many coefficients that can be used to measure plot-to-plot dissimilarity in species composition (for a review, see Podani 2000). If the squared Euclidean distance is used on species presence/absences, Eq. (3) becomes:

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{n(h)} \left\| X_a - X_b \right\|^2 = \sum_{i=1}^{S} \frac{1}{2n(h)} (x_{ia} - x_{ib})^2$$

$$= \sum_{i=1}^{S} \hat{\gamma}_i(h) \tag{4}$$

Therefore, using the squared Euclidean distance for summarizing the dissimilarity between X_a and X_b, the empirical semivariance $\hat{\gamma}_i(h)$ can be decomposed into the sum of the empirical semivariances of the single species variables x_i.

One might argue that the squared Euclidean distance is not always the best distance to use in ecological data; more suitable functions are the chord, Hellinger, chi-square, or Bray-Curtis dissimilarity (Legendre et al. 2005). However, in this specific case, using the squared Euclidean distance for computing plot-to-plot dissimilarity in species composition, the resulting semivariance $\gamma_i(h)$ equals the mean number of species that are present in only one of a pair of observations, regardless of the direction of comparison (see Eq. 4). That is, $\hat{\gamma}(h)$ describes the complementarity of the species composition of two plots, thus representing an ecologically meaningful measure of spatially explicit β-diversity. Wagner (2003) refers to $\hat{\gamma}(h)$ as to the variogram of complementarity.

Worked example

Area and methods

The proposal of using geostatistical methods for summarizing spatial-dependent β-diversity (see Eq. 4) is illustrated with a worked example on a classical data set from a second-growth piedmont hardwood forest located in the Oosting Natural Area of the Duke Forest, North Carolina (Reed et al. 1993, Palmer and White 1994, Wagner 2003).

The study site considered contains an heterogeneous assemblage of forest communities with gradual transition, and the plot design was developed specifically to address questions related to scale and spatial pattern (Reed et al. 1993). The more xeric communities are dominated by *Quercus alba*, *Q. rubra*, and *Carya tomentosa*, while the more mesic ones are mainly composed of *Acer rubrum*, *Liquidambar styraciflua*, and *Fagus grandifolia*. Accord-

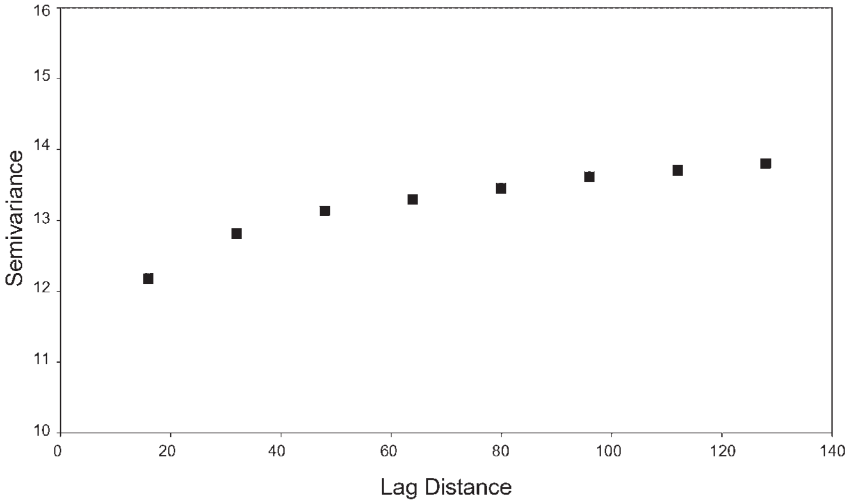

Figure 1. Empirical semivariogram of complementarity for the 206 species found in the 4 m × 4 m modules of the Oosting data set.

ing to Reed et al. (1993), the abundance of typically successional species, such as *Pinus taeda* and *Liriodendron tulipifera* suggests that the site experienced a severe anthropogenic disturbance, probably in the late 1800s.

The vegetation of Oosting Natural Area is sampled in a square grid consisting of 256 modules of 16×16 m^2. A nested series of square quadrats (with linear dimensions of 0.125 m, 0.25 m, 0.5 m, 1 m, 2 m, and 4 m) is located in the southwestern corner of each module. Presence was recorded for all vascular plant species at each quadrat size (for further details see Reed et al. 1993, Palmer and White 1994).

In this paper, we applied Eq. (4) to calculate the spatially-explicit β-diversity for the 256 modules of 4×4 m^2. Unlike Wagner (2003), who used only the 25 most abundant species, we used all 206 species present in at least one 4×4 m^2 quadrat. All calculations were performed with GSTAT (Pebesma 2004; see also gstat web site).

The maximal distance (h) at which semivariance is computed was set equal to 128 m (i.e., half the extent of the study area). This is because the central quadrats of the sampling grid cannot contribute to semivariance estimate for distances greater than the selected threshold (Wagner 2003).

The effects of single species patterns on community structure

Figure 1 plots species complementarity vs. lag for the analyzed data set. Due to coarse-scale heterogeneity, the empirical variogram of Figure 1 is not bounded by a sill such that β-diversity shows a continuous increase without reaching a range within the usable distance h (Cressie 1993).

As for Whittaker's β-diversity (see Eq. 1), the overall pattern of species complementarity $\gamma(h)$ can be decomposed into the single-species contributions $\gamma_i(h)$. One advantage of this additive property of $\gamma_i(h)$ is that we can evaluate the role of each species in shaping the overall β-diversity pattern. For instance, there are two major parameters that affect the contribution of each species to $\gamma_i(h)$: species frequency and the degree of species overall dispersion in the sampling grid, or species 'clumpedness' (Podani and Czárán 1997, Podani et al. 1998).

For binary data, such as species presences/absences sampled in independent quadrats, the expected semivariance $\hat{\gamma}_i(\infty)$ is a function of the mean p_i (i.e., the probability of occurrence of species i in a given quadrat): $\hat{\gamma}_i(\infty) = p_i(1 - p_i)/2$. As a result, the highest semivariance is reached for these species that are found in 50% of the quadrats sampled. On the other hand, if the data show some degree of autocorrelation, species semivariance becomes spatially dependent. In this case, the relation between $\hat{\gamma}_i(h)$ and lag distance is expressed by Eq. (2).

From Eq. (2) it is easily shown that very rare and very abundant species contribute only marginally to overall species complementarity, while subdominant species with intermediate abundance values are much more important in shaping the overall pattern of β-diversity. For instance, for occasional species, nearly all terms in the summation on the right-hand side of Eq. (2) become $(x_{ia} - x_{ib})^2 = (0 - 0)^2 = 0$, while for dominant species most of these terms become $(x_{ia} - x_{ib})^2 = (1 - 1)^2 = 0$.

To show this effect in practice, for each species in the Oosting data set, we analyzed the behavior of $\hat{\gamma}_i(h)$ as a function of the number of presences in the sampling quadrats N_i. In Figure 2, the plot of vs. N_i is shown for a lag

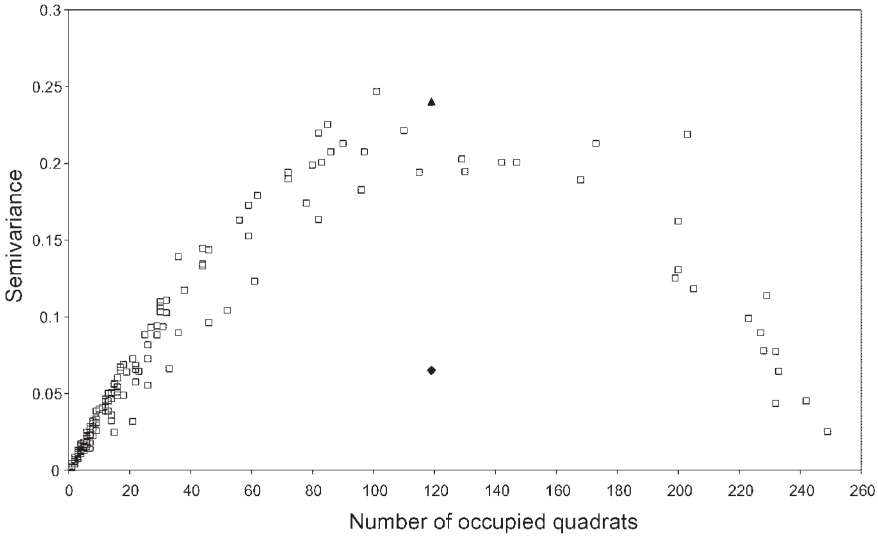

Figure 2. Plot of specific semivariances at lag distance h = 16 m vs. the number of presences in the sampling quadrats for the 206 species of the Oosting data set. The triangle shows the semivariance of *Quercus rubra* (with the spatial pattern of Fig. 3); the diamond shows the semivariance of an artificial clumped distribution of very high aggregation with the same abundance (see Figure 4).

	X	X		X		X	X			X	X			X	X
X	X	X	X			X	X			X		X			X
	X		X		X			X	X		X	X	X		X
X				X		X	X						X		
X						X				X					X
X	X					X			X				X	X	X
X		X		X	X		X			X			X		X
	X		X			X			X				X	X	X
	X				X									X	X
X			X	X			X	X		X	X	X		X	X
X	X		X			X		X		X		X			X
		X			X						X			X	X
X	X		X			X	X	X			X		X	X	X
X		X	X	X		X						X	X		X
X	X	X		X	X	X	X		X						X
X	X	X			X		X	X	X						

Figure 3. Spatial distribution of *Quercus rubra* in the 4 m × 4 m modules of the Oosting data set. The number of presences in the sampling grid $N_i = 119$.

distance $h = 16$ m. The plots of $\hat{\gamma}_i(h)$ vs. N_i for larger lag distances provide analogous results and are not shown here.

As shown in Figure 2, in good agreement with the theoretical expectation, the highest semivariance values are associated to species with intermediate number of presences, while rare and dominant species both possess very low values of $\hat{\gamma}_i$. This general argument that species at near 50% frequency influence the values of the most is mathematically trivial. Nonetheless, this constitutes a major difference with respect to Whittaker's β that is worth pointing out.

In addition, species with a comparable number of presences in the sampling grid may have very different semivariance values depending on their degree of clumpedness. To illustrate the influence of species clumpedness on $\hat{\gamma}_i$, we compared the semivariance obtained from the actual distribution of *Quercus rubra* (Fig-

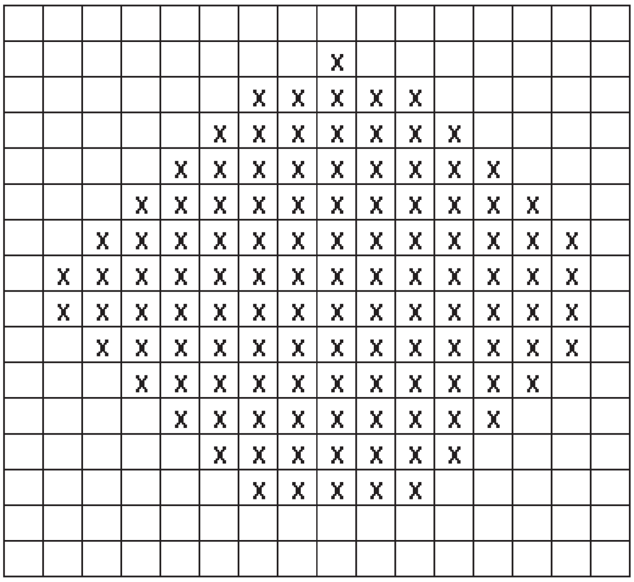

Figure 4. Artificial spatial distribution of very high clumpedness with the same number of presences in the Oosting sampling grid ($N_i = 119$) as *Quercus rubra*.

ure 3; 119 presences in the Oosting sampling grid) with the equal-sized artificial distribution of Figure 4. In Figure 2, the semivariance value of *Quercus rubra* is denoted by a triangle, while a diamond indicates the semivariance of its artificial counterpart. Comparing both values, it becomes apparent that clumped species contribute less to overall β-diversity than more dispersed (i.e., more evenly distributed) species.

This is an obvious consequence of the spatially-explicit character of Eq. (2). For instance, for species that are confined to a limited portion of the sampling grid, two adjacent quadrats are most likely to be either both empty or both filled such that most of the terms in the summation on the right-hand side of Eq. (2) go to zero.

From an ecological viewpoint, species clumpedness is an important factor for explaining the spatial heterogeneity of plant communities as it is directly related to major biotic and abiotic constraints like dispersal strategies or the distribution of local disturbances and limiting environmental factors.

Conclusions

Community structure is spatially organized. This spatial pattern need to be taken into account when developing strategies for sustainable management of natural resources in order to preserve the spatial organization of the species-environment relationships necessary for nature to recreate and maintain β-diversity (Legendre et al. 2005).

In this short paper, we point out the relationship between geostatistics and β-diversity. By contrast, the analysis of the causes that create species turnover is behind the scope of our paper. As shown by Nekola and White (1999), some of it is due to strong, local environmental gradients, while other turnover is due to geographic replacement among similar sites; disentangling the observed turnover into ecological vs. geographic drivers would deserve a much longer paper.

The advantage of using variogram modeling for the analysis of plant community structure lies in its ability to integrate a basic ecological concept as species diversity within an advanced statistical method like geostatistics. In this view, the multivariate semivariogram of Figure 1 represents a summary statistics at the community level expressing overall spatial relationships among the constituent species without any direct reference to the actual details of their spatial positions and their interspecific dynamical interactions. To obtain information on the spatial pattern of single species, or on the combined pattern of any particular multiplet of species, β-diversity can be decomposed into additive species-specific terms according to Eq. (4). We hope, this additive property will prove fruitful in future ecological research.

Acknowledgements. We wish to thank M. Palmer for the data of the Oosting Natural Area. Comments by M. Palmer, J. Nekola and an anonymous referee on previous drafts of this paper are greatly appreciated.

References

Cressie, N.A.C. 1993. *Statistics for Spatial Data*. Wiley Interscience, New York.

Kartesz, J.T. 1994. *A Synonymized Checklist of the Vascular Flora of the United States, Canada, and Greenland.* 2nd ed. Timber Press, Portland, OR.

Koleff, P., K.J. Gaston, and J.J. Lennon. 2003. Measuring beta diversity for presence-absence data. *J. Anim. Ecol.* 72: 367-382.

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

McArthur, R.H. 1965. Patterns of species diversity**.** *Biol. Rev.* 40: 510-533.

Nekola, J.C. and P.S. White. 1999. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* 26: 867-878.

Palmer, M.W. 2005. Distance decay in an old-growth neotropical forest. *J. Veg. Sci.* 16: 161-166.

Palmer, M.W. and P.S. White. 1994. Scale dependence and the species-area relationship. *Am. Nat.* 144: 717-740.

Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30: 683-691.

Podani, J. 2000. *Introduction to the Exploration of Multivariate Biological Data.* Backhuys Publishers, Leiden, NL.

Podani, J. and P. Csontos. 2006. Quadrat size dependence, spatial autocorrelation and the classification of community data. *Comm. Ecol.* 7: 117-127.

Podani, J. and T. Czárán. 1997. Individual-centered analysis of mapped point patterns representing multi-species species assemblages. *J. Veg. Sci.* 8: 259-270.

Podani, J., T. Czárán, and I. Scheuring. 1998. Individual-centered analysis of community pattern: some case studies. *Abstr. Bot.* 22: 101-112.

Poulin, R. 2003. The decay of similarity with geographical distance in parasite communities of vertebrate hosts. *J. Biogeogr.* 30: 1609-1615.

Reed, R.A., R.K. Peet, M.W., Palmer, and P.S. White. 1993. Scale dependence of vegetation-environment correlations: a case study of a North Carolina piedmont woodland. *J. Veg. Sci.* 4: 329-340.

Tóthmérész, B. 1998. On the characterization of scale-dependent diversity. *Abstr. Bot.* 22: 149-156.

Veech, J.A., K.S. Summerville, T.O. Crist, and J.C. Gering. 2002. The additive partitioning of species diversity: recent revival of an old idea. *Oikos* 99: 3-9.

Vellend, M. 2001. Do commonly used indices of β-diversity measure species turnover? *J. Veg. Sci.* 12: 545-552.

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

Whittaker, R.H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* 30: 279-338.

Whittaker, R.H. 1972. Evolution and measurement of species diversity. *Taxon* 21: 213-251.

Wilson, M.W. and A. Shmida. 1984. Measuring beta diversity with presence-absence data. *J. Ecol.* 72: 1055-1064.

Web site

GSTAT – www.gstat.org.

Received June 6, 2006
Revised November 22, 2006
Accpted February 22, 2007