Examining residual spatial correlation in variation partitioning of beta diversity in a subtropical forest

Ke Cao[1,2], Xiangcheng Mi[2], Liwen Zhang[3], Haibao Ren[2], Mingjian Yu[4], Jianhua Chen[5], Jintun Zhang[1,]* and Keping Ma[2,]

[1] Key Laboratory of Biodiversity Sciences and Ecological Engineering, Ministry of Education, College of Life Sciences, Beijing Normal University, Beijing 100875, China
[2] State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, 20 Nanxingcun, Xiangshan, Beijing 100093, China
[3] Key Laboratory of Coastal Zone Environmental Processes, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003, China
[4] College of Life Sciences, Zhejiang University, Hangzhou 310058, China
[5] College of Chemistry and Life Sciences, Zhejiang Normal University, Jinhua 321004, China
*Correspondence address. College of Life Sciences, Beijing Normal University, Beijing 100875, China.
Tel: 010-58807647, Fax: 010-58807721, E-mail: zhangjt@bnu.edu.cn

Abstract

Aims

The relative roles of ecological processes in structuring beta diversity are usually quantified by variation partitioning of beta diversity with respect to environmental and spatial variables or gamma diversity. However, if important environmental or spatial factors are omitted, or a scale mismatch occurs in the analysis, unaccounted spatial correlation will appear in the residual errors and lead to residual spatial correlation and problematic inferences.

Methods

Multi-scale ordination (MSO) partitions the canonical ordination results by distance into a set of empirical variograms which characterize the spatial structures of explanatory, conditional and residual variance against distance. Then these variance components can be used to diagnose residual spatial correlation by checking assumptions related to geostatistics or regression analysis. In this paper, we first illustrate the performance of MSO using a simulated data set with known properties, thus making statistical issues explicit. We then test for significant residual spatial correlation in beta diversity analyses of the Gutianshan (GTS) 24-ha subtropical forest plot in eastern China.

Important Findings

Even though we used up to 24 topographic and edaphic variables mapped at high resolution and spatial variables representing spatial structures at all scales, we still found significant residual spatial correlation at the 10 m × 10 m quadrat scale. This invalidated the analysis and inferences at this scale. We also show that MSO provides a complementary tool to test for significant residual spatial correlation in beta diversity analyses. Our results provided a strong argument supporting the need to test for significant residual spatial correlation before interpreting the results of beta diversity analyses.

Keywords: beta analysis, residual spatial correlation, spatial scale, canonical ordination, multi-scale ordination, variation partitioning

INTRODUCTION

Beta diversity, generally defined as species compositional difference among communities, is a central concept to understanding mechanisms of community assembly (Anderson *et al.* 2011; Condit *et al.* 2002; Kraft *et al.* 2011). Mechanistic hypotheses explaining beta diversity patterns includes dispersal limitation, deterministic processes such as habitat filtering or stochastic processes generating ecological drift, or larger-scale processes creating differences in the size of species

pools (Hubbell 2001; Chase and Myers 2011; Ricklefs and He 2016). However, inferences of the importance of underlying processes strongly depend on the robustness of methods for evaluating the contribution of these processes to beta diversity. For example, it has been well recognized that the strong dependence of some beta diversity metrics on gamma diversity may bias toward an underestimation of the importance of local processes and an overestimation of the importance of larger-scale processes (Kraft *et al.* 2011; Bennett and Gilbert 2016; Ulrich *et al.* 2017). One important but neglected issue is the residual spatial correlation in beta diversity analysis.

Patterns of beta diversity among communities may exhibit spatial correlation induced by community spatial dynamic processes (e.g. dispersal, local plant–plant interaction) or spatial dependence of communities on underlying spatially structured environmental variables (Legendre *et al.* 2002; Wagner 2004). Spatial correlation of beta diversity among communities may be left unaccounted for beta diversity analysis due to the following oversights: First, some important environmental or spatial variables may be omitted from the analysis. For example, some soil nutrients may not have been measured. Or, spatial variables generated by trend-surface polynomials based on geographic coordinates may be used in the analysis instead of spatial eigenfunctions because a trend-surface polynomial cannot capture the spatial structures of communities at local scales (Borcard *et al.* 1992). Second, a scale mismatch may occur between the scale of ecological processes of interest and the scale of environmental or spatial data (Holland *et al.* 2004; Knegt *et al.* 2010; Wagner and Fortin 2005). For example, a scale mismatch will likely arise between community data observed at fine scale and fine-scale environmental variables that have been spatially interpolated from coarse-scale observations. Finally, the standard forward selection procedure used to select spatial variables is not necessarily appropriate to detect all community-level spatial structures because the principle of parsimony results in omission of spatial variables that do not account for the spatial structure of a minimum number of species (Peres-Neto and Legendre 2010). Therefore, the unaccounted spatial correlation may appear in the residual errors, resulting in residual spatial correlation. The presence of residual spatial correlation in the analysis violates the assumption for most statistical models that the residual errors are independent (Legendre 1993) and will lead to problematic inferences in beta diversity studies (Legendre *et al.* 2002; Peres-Neto and Legendre 2010; Wagner 2004). For example, Dormann (2007) compared spatial and non-spatial models in 24 studies of species distribution data, and found that spatial correlation leads to mis-estimation of model coefficient on average 25%.

However, we are unaware of any previous empirical studies investigating residual spatial correlation in beta diversity research. Perhaps testing for significant residual spatial correlation in beta diversity analysis, particularly in canonical analysis such as redundancy analysis (RDA) or canonical correspondence analysis (CCA), is methodically more

difficult than in single-species studies of species-environment regression models. Multi-scale ordination (MSO) with RDA or CCA integrates the ordination results into a framework of geostatistics and diagnoses the residual spatial correlation by checking assumptions related to geostatistics and regression analysis (Couteron and Ollier 2005; Wagner 2004). This method complements the beta diversity analysis using canonical analysis, and will much improve our understanding of the degree to which environmental and spatial processes influence community structure.

In this study, we employed MSO as a diagnostic tool to assess the validity of beta diversity analysis and the inferences from the models by examining the significance of residual spatial correlation across scales. We first used a simulated data set of known properties to illustrate how MSO examines residual spatial correlation in canonical ordination and thus making statistical issues explicit. We then performed the assessment of validity on variation partitioning models by using data from the Gutianshan plot (GTS), a 24-ha plot of a subtropical forest in east China. This plot is ideal for this assessment because up to 24 topographic and edaphic variables and species distribution have been mapped at high resolution throughout the study site (Zhang *et al.* 2011). Our investigation provides a strong argument for the examination of significance of residual spatial correlation for beta diversity analyses when quantifying relative contributions of environmental or spatial processes.

MATERIALS AND METHODS

Simulated data

In this work, we first used a simulated data set with known properties to illustrate how MSO examines residual spatial correlation across scales, and thus reveals statistical issues. The simulated data set was generated using a regular square grid with 100 (10 m × 10 m) quadrats. Three groups of artificial variables were examined:

(1) Three spatial variables respectively representing spatial structures at broad, medium and fine scales induced by dispersal limitation and neighborhood competition. A total of 48 dbMEM (distance-based Moran's eigenvector maps) eigenfunctions were created following the MEM method to characterize spatial structures at all scales based on distances between quadrats in the grid (Borcard and Legendre 2002; Borcard *et al.* 2004; Dray *et al.* 2006). MEM variables 3, 13 and 33 were chosen as the three spatial variables to represent spatial processes at broad, medium and fine scales in the species data (online supplementary Fig. S1a–c).

(2) Three environmental variables consist of two components: spatially structured and non-spatially structured components. Spatially structured components respectively have spatial structure at broad, medium and fine scales. The three environmental variables were gener-

ated with uniformly distributed random values plus MEM variables 5, 15 and 35 for broad, medium and fine scales, respectively (as illustrated in online supplementary Fig. S1d).

(3) Nine species were simulated to be equally controlled by one environmental variable and one spatial variable in 100 quadrats (as illustrated in online supplementary Fig. S1e). Species j at quadrat i (S_{ij}) was generated by implementing a model that was modified from Legendre et al. (2005) and Legendre et al. (2002):

$$S_{ij} = \alpha_j E_{ik} + \beta_j S_{il} + \varepsilon_{ij}$$

where α_j is the effect of environmental variable (E_{ik}) k at quadrat i on species j, β_j is the effect of spatial variable (S_{il}) l at quadrat i on species j and ε_{ij} is a spatially unstructured random error for species j at quadrat i. We set α_j and β_j to 0.5 to simulate equal dependence of species on both environmental and spatial processes. The R code for simulation is available in online supplementary Appendix B.

We used three environmental variables as explanatory variables, and six spatial variables (MEM variables 5, 15 and 35 representing spatial structures of environmental variables, and MEM variables 3, 13 and 33 representing spatial processes) as covariates to explain variation of simulated communities using pRDA. Subsequently, we conducted the following analyses: (i) we omitted either one of the environmental variables (e.g. environmental variable 1), the spatial variable representing the spatial structure of the environmental variable (e.g. MEM variable 5), or both simultaneously, then examined residual spatial correlation with MSO. (ii) We omitted one of the MEM variables 3, 13 or 33 from the spatial data respectively; then examined residual spatial correlation with MSO. These analyses allowed us to illustrate how MSO can be used to examine missing spatially structured environmental variables, missing spatial variables or both at different scales.

Field data

Description of study site and plot

The Gutianshan Forest Dynamic Plot, an old-growth evergreen broad-leaved subtropical forest, is located in Gutianshan National Nature Reserve (29°10′19′′-29°17′41′′N, 118°03′49′′-118°11′12′′E), Zhejiang Province, East China. Detailed descriptions of climate, topography and flora can be found in Legendre et al. (2009) and Zhu et al. (2010).

We sampled soil in the Gutianshan (GTS) plot following BCI soil sample protocols (http://ctfs.si.edu/datasets/bci/soil-maps/BCIsoil.html) (John et al. 2007). We divided the 24 ha plot into 30 × 30 m² cells. All cell intersection points were taken as base points, and two extending points within the boundary of the plot were selected at 2, 5 or 15 m from each base point along a random compass direction (N, E, S, W, NE, NW, SE or SW). All base and extending points were sampled. Overall, we took 892 soil samples from the whole plot

(Zhang et al. 2011). We measured soil moisture, bulk density, nitrogen mineralization rate (Nmin), pH and 16 soil nutrients [including total carbon (TC), total nitrogen (TN), total phosphorus (TP), available Fe, Mn, Zn, Cu, K, P, Ca, Mg, Na, B, Si, N (including $NH4^+$ and $NO3^-$) and Al] following BCI soil sampling protocols (John et al. 2007; Zhang et al. 2011).

Community data

We grouped the trees into territory units of four grain sizes (i.e. size of each observational unit) of 10 × 10 m², 20 × 20 m², 40 × 40 m² and 50 × 50 m² then counted the abundance of the 159 tree species in each unit.

Environmental data

Mean elevation, convexity, slope and aspect using elevational data at all grain sizes were calculated following the method of Legendre et al. (2009). Aspect is a circular variable, and sin (aspect) and cos (aspect) were used in linear models. So there were five variables in total in the topographic data matrix. For edaphic variables, we obtained spatial predictions for 5 × 5 m² blocks using ordinary kriging following the method of John et al. (2007). Then we calculated block averages at all grains from 10 × 10 m² to 50 × 50 m². Twenty edaphic factors were used to create third-degree polynomial equations including a total of 60 monomials. The combination of the original variables and their high order functions allowed modeling of nonlinear relationships between species composition and edaphic predictors (Jones et al. 2006). Then forward selection (with permutation tests, at the 5% significance level, of the increase in R^2) was employed to select parsimonious subsets of the edaphic and topographic variables at different scales.

Spatial data

At all four grain sizes, the spatial relationships among the grid units were represented by MEM variables. The MEM variables were computed across all points of the spatial grid with different truncation distances for the different grain sizes (Legendre et al. 2009). In the present study, forward selection (with permutation tests, at the 5% significance level, of the increase in R^2) was employed to select parsimonious subsets of the MEM variables before variation partitioning and MSO.

Statistical methods

Variation partitioning

Variation partitioning of community composition data using environmental variables and spatial eigenfunctions decomposes the beta diversity into four fractions of variation: pure environmental [a], pure spatial [c], joint effect of environmental and spatial [b] and unexplained [d] (Borcard and Legendre 2002; Legendre et al. 2009). Here, we aim to examine the significance of residual spatial correlation in beta diversity analysis based on spatial coordination, such as redundancy analysis (RDA) and canonical correspondence analysis (CCA), which is methodologically more difficult than other beta diversity analysis.

Multi-scale ordination

MSO integrates direct ordinations such as simple or partial redundancy analysis (pRDA) into a geostatistical framework to partition the canonical ordination results by distance into a set of empirical variograms (Wagner 2004). These empirical variograms estimate the explained and residual variance versus distance to characterize the spatial structure of these variance components (Wagner 2004).

In geostatistics, data stationarity is an important property that refers to the approximately constant variance or semi-variance of data properties over the study area (Rossi *et al.* 1992). Community composition data may display prominent trends along environmental gradients. The community composition data could be detrended by environmental data and spatial variables in pRDA before performing MSO analysis.

We employed MSO as a diagnostic tool to check the following three assumptions related to geostatistics and regression analysis:

(1) Stationarity: For a stationary process, spatial correlation among community compositions results in reduced variance within short distances, whereas at larger distances beyond the range of spatial correlation, community compositions are spatially independent and thus the variance reaches a constant level, which is termed the sill. If the variogram of the residual variance does not reach a sill, it violates the assumption of intrinsic stationarity required in geostatistics, indicating that important explanatory factors were not included into analysis (Wagner 2004).

(2) Residual spatial correlation: The residual variances were tested against distance classes for significant correlation using multivariate Mantel correlograms with permutation test (Borcard and Legendre 2002; Legendre and Legendre 2012). Spatial correlation of the residual variances indicates that some significant spatial structures in communities remain unaccounted for in some distance classes. The violation of assumptions one and two suggests that some important explanatory variables were not incorporated in the analysis or that the spatial structure was not the same in the spatial and environmental variables (Wagner 2004). Therefore, the statistical significance of fractions explained by environmental and spatial variables as well as regression coefficients of species-habitat relationships in RDA or pRDA cannot be determined using parametric or ordinary permutation tests.

(3) Scale-dependent correlation structure: Containment of the empirical variogram of the sum of the explained and residual variance entirely within a point-wise envelope indicates the relationships between community compositions and explanatory variables do not vary significantly with scale. Else, if the sum of the explained and residual variance exceeds the point-wise envelope for the sum of explained and residual variance, the re-lationship between community composition and the explanatory variables may be dependent on scale, and thus, relationships between community composition and explanatory variables derived from RDA or pRDA, which are based on global analysis of all examined scales, should not be interpreted (Wagner 2004).

The RDA analyses, variation partitioning, tests of significance of the fractions and MSO were computed using the R package 'vegan' (Oksanen *et al.* 2015). Geostatistical analyses were carried out using the R package 'gstat' (Pebesma 2004). MEM variables were created with the R package 'PCNM'; forward selection was computed using the 'packfor' package (Both packages are available on website: http://www.bio.umontreal.ca/legendre/indexEn.html).

RESULTS

In the simulation, environmental variable 1 was synthesized by addition of a uniform distributed random variable and MEM variable 5. When using all environmental and spatial variables as explanatory factors or omitting either environmental variable 1 or MEM variable 5 from explanatory variables, all three assumptions were satisfied: residual variance reached a sill, residual variance showed no significant spatial correlation and sum of explained and residual variance fell into the point-wise envelope (Fig. 1a–c). This indicated that no significant residual spatial correlation was found in canonical analysis. Only when both environmental variable 1 and MEM variable 5 were omitted, residual spatial correlation could be found; although the sum of residual and explained variance fell into the point-wise envelope, residual variance did not reach a sill and significant residual correlation were found in most distance classes (Fig. 1d). This significant residual spatial correlation occurs when the community compositional difference in species 1, 4 and 7 remains unexplained, thus the results are problematic. Similarly, significant residual correlations and other parts of community compositional differences that are not fully explained could also be found when omitting both environmental variable 2 and MEM variable 15 or environmental variable 3 and MEM variable 35 (Fig. 1e and f) or any of the MEM variables 3, 13 and 33 (data not shown).

After the illustration of the performance of MSO, we then examined the significance of residual spatial correlation in the beta diversity of the Gutianshan plot. As shown in Table 1, all testable fractions [a], [a+b], [c] and [b+c] explained by environmental or spatial factors were significant across all grains.

For MSO(Y~E|S) (denoting MSO analysis for pRDA of community composition data with environmental variables as explanatory variables and MEM variables as covariates) and MSO(Y~S|E) at the 20×20 m^2 grain size, the sum of explained and residual variance reached a sill and thus the intrinsic stationarity assumption was satisfied (Fig. 1c and d). At the same time, the sum of the explained and residual variance
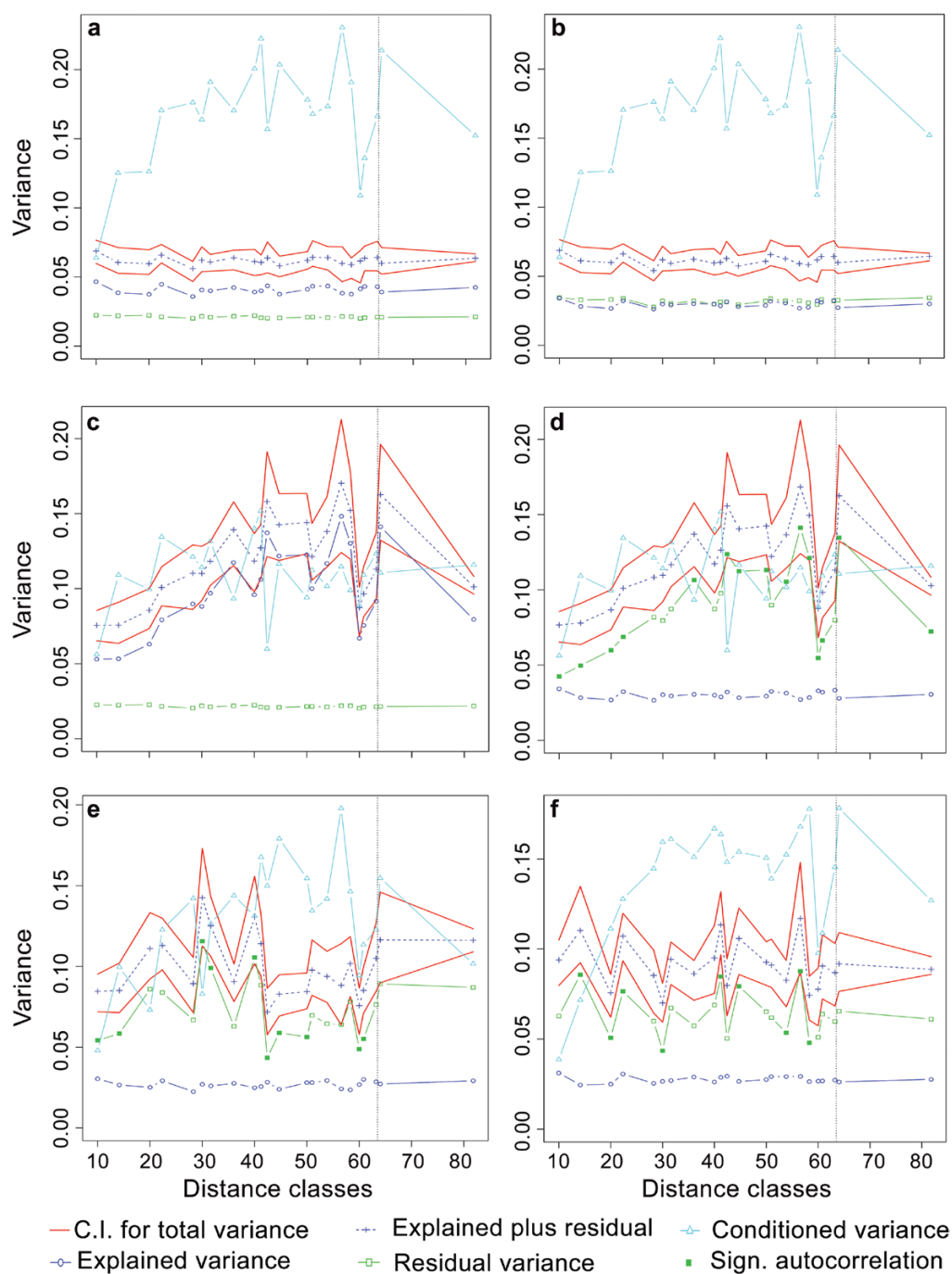
Figure 1: multi-scale ordination with partial redundancy analysis of a simulated data set using environmental variables as explanatory variables and spatial variables as covariates (MSO(Y~E|S)) when (a) all environmental and spatial variables were used; (b) environmental variable 1 was omitted from the data set; (c) MEM variable 5 representing spatial structure of environmental variable 1 was omitted; (d) both environmental variable 1 and MEM variable 5 were omitted; (e) both environmental variable 2 and MEM variable 13 were omitted; (f) both environmental variable 3 and MEM variable 35 were omitted. 'CI for total variance' is the confidence interval envelope for the variogram of the sum of explained and residual variance of community composition. 'Explained variance' is the variogram of the variance constrained by the explanatory variables after accounting for the conditioning variables, which is pure environment-explained variation (fraction [a]) for MSO(Y~E|S), or pure space-explained variance (fraction [c]) for MSO(Y~S|E). 'Residual variance' is the variogram of the residual variance. 'Explained plus residual' is the sum of 'explained variance' and 'residual variance'. 'Conditioned variance' is the variogram of conditioned variance, which is the variance explained by the joint effect of spatial and environmental variables (fraction [b]). Filled box symbols indicate a significant correlation of the residual variance at distance classes. The dotted line indicates half the maximum distance between observations, beyond which the variogram should not be interpreted.

in most of the distance classes between quadrats fell within the envelope of the sum of explained and residual variance, indicating that correlations between the species-explaining variable were scale-independent at most distances (Fig. 1c and d). The residuals were not spatially autocorrelated beyond distance classes 70 m (Fig. 1c and d). These results indicate that almost no significant residual spatial correlation was found in the variation partitioning model at the 20×20 m² grain size (Fig. 2c and d). We also found similar results at the 40×40 m² and 50×50 m² grains (Fig. 2e–h).

We did find residual correlation at the 10×10 m² grain size (Fig. 2a and b). Although the intrinsic stationarity assumption was satisfied and the species-explained variation was scale-independent in most of the distance classes for both MSO(Y~E|S) and MSO(Y~S|E) (Fig. 1a and b), the residual variance exhibited significant correlation in both analyses of MSO(Y~E|S) and MSO(Y~S|E) (Fig. 1a and b). This indicates that significant residual spatial correlation occurred in the analysis at grain size 10×10 m².

DISCUSSION

The relative roles of ecological processes in structuring beta diversity are of considerable interest to ecologists (Anderson *et al.* 2011; Kraft *et al.* 2011; Ricklefs and He 2016). However, inferences in some studies of beta diversity are problematic because of oversights, such as failure to include some important environmental or spatial factors in the analysis or scale mismatches, resulting in residual autocorrelation (de Knegt *et al.* 2010; Wagner and Fortin 2005). In this study, we demonstrated that residual spatial autocorrelation likely occurred even when analyses were performed using up to 24 topographic and edaphic variables mapped with high resolution as well as a set of spatial variables with structures at all scales created using MEM method.

Our simulation showed that significant residual spatial correlation occurred when both spatially structured environmental variables and the spatial variables representing their spatial structures were omitted from the analyses (Fig. 1b–d), thus important components of beta diversity

remained unexplained. This suggests the omission of both important spatially structured environmental variables and related spatial variables at particular scales will lead to significant residual spatial correlation (Fig. 1d–f), and underestimation of environmental or spatial processes. Thus, in order to avoid residual spatial correlation, important spatial structures of communities at all scales should be accounted for in beta diversity analyses.

The fractions explained by environmental and spatial variables in Table 1 are similar to those found by Legendre *et al.* (2009). After edaphic variables were included into this analysis, the joint effect of environmental and spatial processes ([b]) increased by 10% compared with results from Legendre *et al.* (2009) (Table 1).

At 20×20 m² to 50×50 m² grain sizes, we found almost no significant residual autocorrelation in the analyses. This validated variation partitioning results and significance test of fractions at these grains. These results demonstrate that environmental and spatial variables at 20×20 m² to 50×50 m² grains explained the beta diversity in Gutianshan plot quite well.

However, we indeed found significant residual autocorrelation at 10×10 m² grain size, which invalidated the results of the variation partitioning model and significance tests of fractions explained by environmental and spatial variables at this scale (Table 1). Residual spatial autocorrelation was found in both MSO(Y~E|S) and MSO(Y~S|E) analyses, indicating that the spatial structures of communities were not fully accounted for by either spatially structured environmental variables or spatial variables. On the one hand, the missing spatially structured environmental variables were likely caused by: (i) failure to include some important spatially structured environmental variables such as some local soil nutrients or light in our analysis or (ii) a scale mismatch arising from interpolation of the environmental variables from measurements on a 30×30 m² grid, despite using a multi-scale systematic soil sampling scheme. On the other hand, the standard forward selection procedure of MEM variables likely could not detect spatial structures of communities at local scales. Therefore, some important spatial variables,

Table 1: variation partitioning results for different grain sizes

Scale	[a]	[b]	[c]	[d]	[a+b]	[b+c]	Number of significant MEM variables
10×10 m²	0.0093	0.3074	0.3147	0.3686	0.3167	0.6221	187
20×20 m²	0.0728	0.3871	0.1738	0.3664	0.4599	0.5608	151
40×40 m²	0.1813	0.3720	0.0982	0.3486	0.5533	0.4702	29
50×50 m²	0.1479	0.4085	0.0593	0.3843	0.5563	0.4678	19

Fraction [a] is variation explained only by the environmental variables, fraction [c] is variation explained only by the MEM variables, fraction [b] is the intersection of the amount of variation explained by linear models of the two sets of explanatory tables, fraction [d] is the unexplained variation. Environmental variables used to compute [a+b]: edaphic and topographic variables described in methods section. MEM variables were the explanatory variables used to calculate fraction [b+c]. Fraction [a], [a+b], [b+c] and [c] are testable, whereas fraction [b] and [d] are not testable.
The reported values are adjusted R-squares.
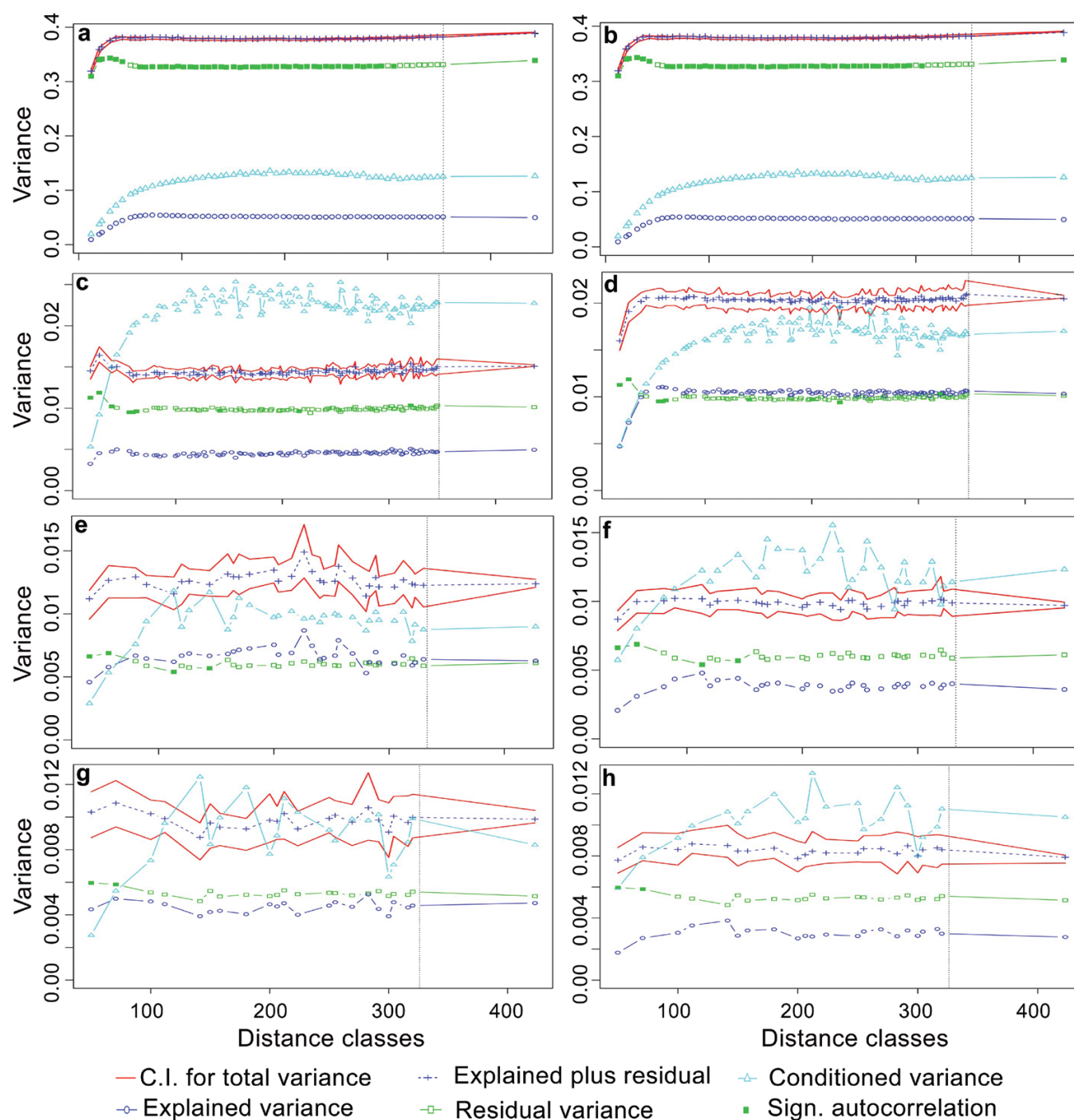
Figure 2: multi-scale ordination with partial redundancy analysis using environmental variables as explanatory variables and spatial factors as covariates (MSO(Y~E|S)) at grains of (a) 10×10 m^2, (c) 20×20 m^2, (e) 40×40 m^2, (g) 50×50 m^2 and with spatial variables as explanatory variables and environmental variables as covariates (MSO(Y~S|E)), at grains of (b) 10×10 m^2, (d) 20×20 m^2, (f) 40×40 m^2, (h) 50×50 m^2. Symbols in Fig. 2 have the same meaning as in Fig. 1.

which represent local community dynamics, such as inter-specific or intraspecific interactions, were not selected by the forward selection procedure. These results demonstrate that residual spatial autocorrelation arose even when up to 20 soil nutrients were mapped with a high resolution and a set of

spatial variables with spatial structures at all the scales were created using the MEM method. This suggests that some important components of community composition especially at local scales induced by neighborhood competition, likely remain unexplained. The unexplained components will lead to

problematic inferences, i.e. underestimation of either environmental or spatial processes. This provided strong empirical justification that ecologists should examine the significance of residual spatial autocorrelation before interpreting the results of variation partitioning of beta diversity.

When significant residual autocorrelation is found in canonical analysis, changes to the experimental design and hypothesis testing are required in order to discern whether residual spatial autocorrelation is due to the omission of important spatially structured environmental variables or spatial variables or caused by a scale mismatch. In this study, further soil nutrient sampling at a finer scale (e.g. a 20 × 20 m² grid based multi-scale sampling scheme) and measurement of more soil nutrients to account for residual spatial autocorrelation in the analysis at scale of 10 × 10 m² would be required. On the other hand, the individual species selection model, which applies the forward selection procedure to each species separately, provides an easier solution to pick up the missing spatial variables or the missing spatial structures of environmental variables (Peres-Neto and Legendre 2010). However, the individual species selection model is likely computationally infeasible due to large number of MEM variables at 10 × 10 m² quadrats (1327 MEM variables × 2400 quadrats) and species in the Gutianshan plot (159 species). This may inform future attempts to refine and optimize individual species selection models for large ecological data set.

In summary, our main findings were: (i) ecologists should check for significant residual autocorrelation before interpreting results of variation partitioning of beta diversity and (ii) MSO provides a complementary tool to test for significant residual spatial autocorrelation to avoid problematic inferences, i.e. underestimation of either environmental or spatial processes. We believe testing for significant residual spatial autocorrelation is critical for understanding the relative roles of different ecological processes in shaping beta diversity among communities.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of Plant Ecology* online.

FUNDING

ACKNOWLEDGEMENTS

REFERENCES

Anderson MJ, Crist TO, Chase JM, *et al*. (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecol Lett* **14**:19–28.

Bennett JR, Gilbert B (2016) Contrasting beta diversity among regions: how do classical and multivariate approaches compare? *Glob Ecol Biogeogr* **25**:368–77.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Modell* **153**:51–68.

Borcard D, Legendre P, Avois-Jacquet C, *et al*. (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology* **85**:1826–32.

Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. *Ecology* **73**:1045–55.

Chase JM, Myers JA (2011) Disentangling the importance of ecological niches from stochastic processes across scales. *Philos Trans R Soc Lond B Biol Sci* **366**:2351–63.

Condit R, Pitman N, Leigh EG Jr, *et al*. (2002) Beta-diversity in tropical forest trees. *Science* **295**:666–9.

Couteron P, Ollier S (2005) A generalized, variogram-based framework for multi-scale ordination. *Ecology* **86**:828–34.

de Knegt HJ, Langevelde FV, Coughenour MB, *et al*. (2010) Spatial autocorrelation and the scaling of species–environment relationships. *Ecology* **91**:2455–65.

Dormann CF (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Glob Ecol Biogeogr* **16**:129–38.

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Modell* **196**:483–93.

Holland H, Bert D, Fahrig L (2004) Determining the spatial scale of species' response to habitat. *Bioscience* **54**:227–33.

Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography (MPB-32). (monographs in population biology).

John R, Dalling JW, Harms KE, *et al*. (2007) Soil nutrients influence spatial distributions of tropical tree species. *Proc Natl Acad Sci USA* **104**:864–9.

Jones MM, Tuomisto H, Clark DB, *et al*. (2006) Effects of mesoscale environmental heterogeneity and dispersal limitation on floristic variation in rain forest ferns. *J Ecol* **94**:181–95.

Kraft NJ, Comita LS, Chase JM, *et al*. (2011) Disentangling the drivers of β diversity along latitudinal and elevational gradients. *Science* **333**:1755–8.

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**:1659–73.

Legendre P, Borcard D, Peres-Neto PR (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecol Monogr* **75**:435–50.

Legendre P, Dale MRT, Fortin MJ, *et al.* (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* **25**:601–15.

Legendre P, Legendre L (2012) *Numerical Ecology*, 3rd edn. Amsterdam, The Netherlands: Elsevier Science.

Legendre P, Mi X, Ren H, *et al.* (2009) Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology*, **90**:663–74.

Oksanen J, Blanchet FG, Kindt R, *et al.* (2015) *vegan: Community Ecology Package. R Package Version 2.3-0.*

Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. *Comput Geosci* **30**:683–91.

Peres-Neto PR, Legendre P (2010) Estimating and controlling for spatial structure in the study of ecological communities. *Glob Ecol Biogeogr* **19**:174–84.

Ricklefs RE, He F (2016) Region effects influence local tree species diversity. *Proc Natl Acad Sci USA* **113**:674–9.

Rossi RE, Mulla DJ, Journel AG, *et al.* (1992) Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecol Monogr* **62**:277–314.

Ulrich W, Baselga A, Kusumoto B, *et al.* (2017) The tangled link between β- and γ-diversity: a Narcissus effect weakens statistical inferences in null model analyses of diversity patterns. *Glob Ecol Biogeogr* **26**:1–5.

Wagner HH (2004) Direct multi-scale ordination with canonical correspondence analysis. *Ecology* **85**:342–51.

Wagner HH, Fortin MJ (2005) Spatial analysis of landscapes: concepts and statistics. *Ecology* **86**:1975–87.

Zhang L, Mi X, Shao H, *et al.* (2011) Strong plant-soil associations in a heterogeneous subtropical broad-leaved forest. *Plant Soil* **347**: 211–20.

Zhu Y, Mi XC, Ren HB, *et al.* (2010) Density dependence is prevalent in a heterogeneous subtropical forest. *Oikos* **119**:109–19.