A comprehensive framework for the study of species co-occurrences, nestedness and turnover

Werner Ulrich, Wojciech Kryszewski, Piotr Sewerniak, Radosław Puchałka, Giovanni Strona and Nicholas J. Gotelli

W. Ulrich (http://orcid.org/0000-0002-8715-6619) (ulrichw@umk.pl) and R. Puchałka, Faculty of Biology and Environmental Protection, Nicolaus Copernicus Univ. in Toruń, Lwowska 1, PL-87-100 Toruń, Poland. – W. Kryszewski, Faculty of Mathematics and Informatics, Nicolaus Copernicus Univ. in Toruń, Toruń, Poland. – P. Sewerniak, Dept of Soil Science and Landscape Management, Nicolaus Copernicus Univ., Toruń, Poland. – G. Strona (http://orcid.org/0000-0003-2294-4013), European Commission Joint Research Centre, Inst. for Environment and Sustainability, Ispra, Italy. – N. J. Gotelli, Dept of Biology, Univ. of Vermont, Burlington, VT, USA.

Binary presence–absence matrices (rows = species, columns = sites) are often used to quantify patterns of species co-occurrence, and to infer possible biotic interactions from these patterns. Previous classifications of co-occurrence patterns as nested, segregated, or modular have led to contradictory results and conclusions. These analyses usually do not incorporate the functional traits of the species or the environmental characteristics of the sites, even though the outcomes of species interactions often depend on trait expression and site quality. Here we address this shortcoming by developing a method that incorporates realized functional and environmental niches, and relates them to species co-occurrence patterns. These niches are defined from n-dimensional ellipsoids, and calculated from the n eigenvectors and eigenvalues of the variance–covariance matrix of measured environmental or trait variables. Average niche overlap among species and the spatial distribution of niches define a triangle plot with vertices of species segregation (low niche overlap), nestedness (high niche overlap), and modular co-occurrence (clusters of overlapping niches). Applying this framework to temperate understorey plant communities in southwest Poland, we found a consistent modular structure of species occurrences, a pattern not detected by conventional presence–absence analysis. These results suggest that, in our case study, habitat filtering is the most important process structuring understorey plant communities. Furthermore, they demonstrate how incorporating trait and environmental data into co-occurrence analysis improves pattern detection and provides a stronger theoretical framework for understanding community structure.

Since the pioneering work of Diamond (1975), the assembly of species communities has remained a central research focus (Weiher and Keddy 1999, Stokstad 2009, Götzenberger et al. 2012). The core data structure – a species × sites binary presence–absence matrix (Connor and Simberloff 1979) – has been used to infer species interactions (Fariňa et al. 2009), the distribution of species functional traits (Kraft et al. 2008), and stochastic colonisation and extinction (Hubbell 2001).

Diamond (1975) and subsequent authors (reviewed by Weiher and Keddy 1999, Chesson, 2000, Götzenberger et al. 2012) focused on non-random patterns of compositional change across replicate communities. They interpreted a pattern of reciprocal species occurrence (species segregation) as evidence for competitive exclusion (Gause 1934). But, unless habitats are identical in the replicate patches and there are no limits to dispersal, a pattern of species segregation cannot be distinguished from a change in species composition driven by a gradient in habitat quality or by limited dispersal (species turnover; Ulrich and Gotelli 2013, Blois et al. 2014). Subsequent work (reviewed by

Ulrich and Gotelli 2013) showed that a pattern of reciprocal exclusion in a presence–absence matrix is statistically indistinguishable from a pattern of regular change in species composition (species turnover). Thus, species segregation and turnover appeared to be two sides of the same coin. Any highly segregated matrix can be reordered to show its intrinsic high degree of species turnover among sites (Fig. 1). Recent meta-analyses of animal and plant communities consistently reported a tendency towards species segregation among modern assemblages sampled at a variety of spatial scales (Gotelli and McCabe 2002, Ulrich and Gotelli 2007, 2013), but not among pre-Holocene assemblages (Lyons et al. 2016).

An alternative, biogeographic concept of community assembly focuses on nested patterns of species co-occurrences (Hultén 1937), in which species-poor communities are perfect nested subsets of species-rich communities (Patterson and Atmar 1986, reviewed in Ulrich et al. 2009). A nested community pattern (Fig. 1) is expected from stochastic colonization and extinction in the equilibrium model of island biogeography (MacArthur and Wilson 1963) as well
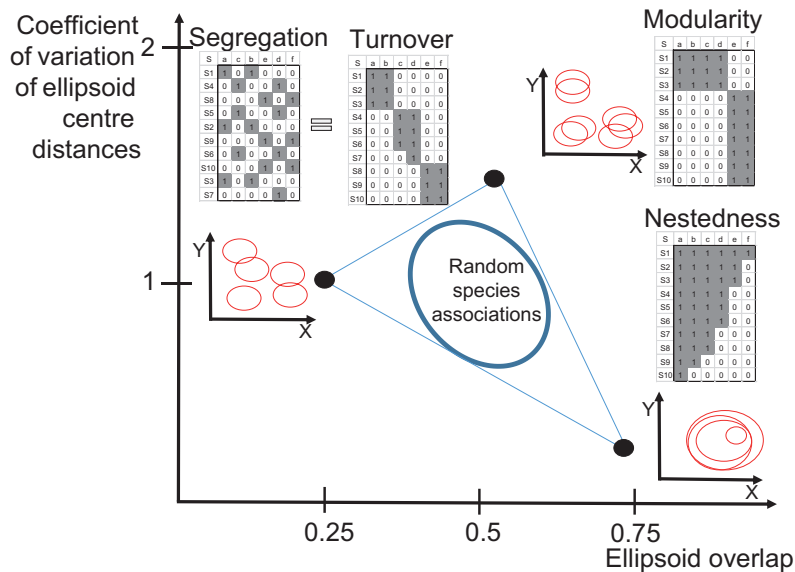
Figure 1. The distribution of eigenvector ellipsoids within the coordinate system of environmental variables allows for the calculation of average ellipsoid overlap and the coefficient of variation of ellipsoid centroid distances. This enables the identification of basic patterns in species niche overlap and co-occurrence (nested, modular, segregated). Using three environmental variables increasing or decreasing along the ordered sites (Supplementary material Appendix 3) the black dots show the observed positions of the example inset matrices (cf. also Supplementary material Appendix 3 Fig. 3.2). Multi-species segregated matrices are increasingly characterized by a Poisson random distribution of ellipsoid centroid distances as identified by the coefficient of variation $CV_{env} \approx 1$. Modular matrices have an aggregated ellipsoid distribution in space ($CV_{env} > 1$), nested ellipsoids have $CV_{env} < 1$. The inset graphs show the position of ellipsoids in the two-dimensional case in which x and y denote the respective environmental/trait variables. A modular pattern might occur at high or low ellipsoid overlap depending on matrix fill. Depending of the structure of the environmental or trait data, a nested pattern might be realised with high variation in ellipsoid centre distances.

as from stochastic speciation, dispersal, and extinction in ecological drift models (Hubbell 2001). Furthermore, nested patterns might also emerge in the presence of ecological gradients promoting the co-existence of habitat generalists and habitat specialists. For this reason, Ulrich et al. (2009) recommended nestedness analysis as a major tool to identify environmental gradients along which local communities are assembled in an orderly manner.

Species turnover and nestedness have been interpreted as the endpoints of a continuum along which meta-communities are organised (Presley et al. 2010, Brown et al. 2017). Here, we define a meta-community as a set of local patches that are connected by dispersal, and in which all species have the potential to occur (Wilson 1992). However, it is still unclear whether meta-communities really can be ordered according to this assumed one-dimensional continuum from turnover to nestedness. For example, Ulrich and Gotelli (2013) illustrated how presence–absence matrices identified by standard metrics can be simultaneously nested and segregated. Similar results were also obtained in analyses of trophic networks: Fortuna et al. (2010) reported both positive and negative correlations of metrics quantifying nestedness and segregation. Moreover, the structure of these correlations changed with the fill of the matrix.

The natural counterpart of segregation is modularity, a pattern in which groups of species frequently occur in the same sites (Fig. 1). In the extreme scenario of a Clementsian 'super-organism', strong positive associations among the species in a community mean that all species occur together in one environment, and are replaced at a sharp ecotone by a different set of species in a different environment. Therefore, this modular pattern of meta-community structure is not equivalent to a nested pattern (Fortuna et al. 2010). Following Leibold and Mikkelson (2002), we argue that it represents instead a third independent pattern of meta-community organisation. In this respect we note that the corresponding counterpart of nestedness – anti-nestedness – (Almeida-Neto et al. 2007), still awaits a formal definition.

We also observe, that, for a meta-community containing only two species, a strong positive aggregation will generate a modular pattern if both species have the same number of occurrences, but a nested pattern if one of the species has more occurrences than the other. Thus, the patterns of nestedness, segregation, and modularity depend not only on the degree of co-occurrence, but also on the occurrence frequencies of individual species. For a multi-species assemblage, each individual species may be associated with other pairs that can be nested, aggregated, or segregated, making it difficult to generate a composite index that applies to the entire matrix.

Consequently, we argue that the simple nestedness – segregation gradient is insufficient to describe the complexity of meta-community organisation. Instead, we propose a triangular framework, with nested, modular and turnover patterns as vertices (Fig. 1). Modular patterns have long been identified as important in both mutualistic networks and food webs (Bascompte 2009), in which groups of species at one trophic or functional level interact with a second group of a different level. A modular community organisation is therefore a pattern of within-module aggregation and

between-module turnover. In contrast, a nested pattern is a single module in which species occurrence frequencies vary, and composition is nested.

Environmentally defined patterns of co-occurrence

The classical approach of analysing species co-occurrence with only a binary presence–absence matrix has three important drawbacks. First, it does not consider environmental constraints (Leibold and Mikkelson 2002, Presley et al. 2010, Ulrich and Gotelli 2013), and identifies patterns on the sole basis of occurrences/absences. Ordination techniques such as correspondence analysis and multidimensional scaling naturally include species abundances and environmental data (Legendre and Legendre 1998). However, these techniques are not designed to infer the precise pattern of species associations (Dallas et al. 2016), and the basic definitions of turnover, modularity, and nestedness do not refer to environmental covariates. Consequently, nearly all studies on co-occurrence analysis focused on the presence–absence matrix only.

Co-occurrence is frequently linked either to species interactions in the case of modular and segregated patterns (Bascompte 2009, Götzenberger et al. 2012), or to environmental gradients (Ulrich et al. 2009) and colonisation trajectories (Gassert et al. 2013, Louy et al. 2014) in the case of nested patterns. We feel that this conceptual mismatch hinders progress in meta-community analysis, and we argue that there is need for an extension of the basic analysis of co-occurrences to incorporate environmental variables and possibly species functional traits. In this paper we develop such an extension to redefine species turnover, modularity and nestedness.

The second major shortcoming of binary matrix analysis lies in the plethora of metrics used to quantify co-occurrence patterns (Ulrich and Gotelli 2013). The most common metric of nestedness, NODF (Almeida-Neto et al. 2008), depends on specific ways of ordering the rows and columns of the matrix, whereas metrics of turnover, segregation, β-diversity (Stone and Roberts 1990, Tuomisto 2010), and modularity (Newman and Girvan 2004) are independent of matrix order. Because these metrics are based on different algorithms and use different information from the matrix structure, results are not directly comparable. For instance, a recent meta-analysis (Ulrich and Gotelli 2013) confirmed that the degrees of turnover and nestedness are significantly negatively correlated. However, this correlation explained less than 20% of variance along the assumed nestedness – turnover gradient, suggesting the existence of additional intermediate patterns elusive to both metrics.

The third problem with the analysis of binary matrices is that raw scores of all these metrics heavily depend on matrix shape and fill and thus cannot be compared directly (Gotelli and Ulrich 2012). Consequently, pattern identification involves a second step, a comparison with an appropriate ecological standard, often in form of a randomised matrix (Gotelli and Graves 1996). Large effect sizes reveal the existence of the focal pattern, while small effect sizes imply a lack thereof. Theoretically, it should be possible to define a gradient (based on effect sizes) spanning, for instance, from strongly nested – to random – to strongly anti-nested.

However, because there are three idealized patterns, it is problematic to place an assemblage on a single gradient from segregated to nested by using effect sizes measured relative to a specific null model. Furthermore, to date, there is no procedure permitting to place a given meta-community unequivocally within the triangle defined by the modular, nested, and turnover vertices (Fig. 1).

There have been several attempts to provide a unifying framework (Leibold and Mikkelson 2002, Presley et al. 2010) of meta-community structure. However, these attempts have not found a solution to the issue, nor they have included additional environmental information aimed at linking the observed patterns to ecological processes. We argue that, in order to make further progress in the classification of communities, we need to incorporate both environmental data, and species functional traits (McGill et al. 2006). Here we propose an expanded and unifying framework for the identification and measurement of nestedness, modularity, and turnover that surpasses the present methods of co-occurrence analysis by incorporating data on environmental characteristics and species traits. We show the simultaneous analysis of co-occurrences and environmental data provides new insight into the mechanisms of community assembly and the spatial distribution of species niches.

Methods

A multidimensional unifying concept for the study of meta-community structure

Eigenvector ellipses and their higher dimensional counterparts, ellipsoids, are commonly used in statistics (Friendly et al. 2013) and ecology (Legendre and Legendre 1998, Doledec et al. 2000) to visualise and quantify the spatial arrangement of sets of variables in two- and multi-dimensional space. Specifically, in principal components analysis (PCA), these ellipsoids are generated from the variance–covariance matrices of a set of equally weighed variables for which the associated eigenvectors and eigenvalues depict, respectively, the direction in space, and the semi-radii. In ecological analysis, these eigenvector ellipsoids were used to identify a functional niche space within an n-dimensional hyperspace defined by a focal set of species characters (Hutchinson 1978, Jackson et al. 2011, Zalewski et al. 2014, Maire et al. 2015, Qiao et al. 2016). When applied to environmental measures collected at each site, our ellipsoid approach shows how species are distributed across multiple environmental dimensions. By identifying the region within this hyperspace occupied by each species, the ellipsoids can be interpreted as the realised range of environmental conditions within which a focal species occurs (Doledec et al. 2000, Janžeković and Novak 2012). More notably, they provide novel ways to derive metrics of species turnover, modularity or nestedness. Building on this idea, here we use eigenvector ellipsoids (as defined in the Supplementary material Appendix 1) to redefine the concepts of nestedness, modularity, and turnover in terms of environmental characteristics or geographic positions of a set of sites hosting a meta-community. The corresponding Fortran software application 'NicheNew' is available from

W. Ulrich by request. We provide the respective source code in the Supplementary material Appendix 3.

Environmental based patterns of co-occurrence

Our framework combines the $m \times n$ presence–absence matrix (m species, n sites) of species occurrences (**M**), and either a $2 \times n$ matrix of the geographical position of sites (typically longitude and latitude) (**D**), or a $k \times n$ matrix of k environmental variables (**V**) characterising each site. These matrices are used to define and to compare species-specific eigenvector ellipsoids.

As shown in the worked example in Supplementary material Appendix 3 Fig. 3.1, for each species, the variables in **D** or **V** are normalised by the common Z-transformation to have a mean of zero and a standard deviation of one in order to ensure comparability between variables. Then, for each species, a **C** matrix of dimensions $k \times n_i$ is obtained by reducing **D** or **V** to a submatrix that includes only the n_i sites of occurrence of the i-th species (Fig. 1). Each **C** matrix is then used to define an ellipsoid E(C), which identifies the realised environmental niche (the range of environmental conditions) (Fig. 1, 2) for each one of the m species as:

$$E(C) = \{x \in R; (x-c)^T \, \Sigma^{-1} (x-c) \leq r^2$$
$$= (x-c)^T \, U^{T-1} U (x-c) \leq r^2 \quad (1)$$

where r^2 is the 99% a quantile of a χ^2-distribution with k degrees of freedom. The vector **c**, indicating the ellipsoid centre, has coordinates $(c_1,...,c_k)$ with $c_j = \frac{1}{n_i} \sum_{s=1}^{n_i} v_{js}$, $j = 1,...,k$. Here n_j is the number of occurrences of the focal species, and v_i is the environmental value associated with each occurrence. The matrix Σ is the $k \times k$ variance – covariance matrix of environmental variables (Supplementary material Appendix 1).

Equation 1 reveals whether or not a vector x representing the environmental features (or the geographic coordinates) at a given location belongs to a given environmental niche. Specifically, our approach quantifies the overall degree of niche overlap between pairs of species as exemplified in the supplementary material (Supplementary material Appendix 3 Fig. A3.1). The directions of axes of the k-dimensional ellipsoid E(c) are given by the unit eigenvectors $u_1,...,u_k$ of the matrix Σ; these vectors are the columns of the **U** matrix. The half-lengths l_j, $j = 1,...,k$, of axes are defined by $l_j = \sqrt{r^2 \lambda_j}$, where λ_j is the eigenvalue of Σ and is contained in the diagonal $k \times k$ matrix Λ (Supplementary material Appendix 1).

We can now redefine perfect nestedness as a pattern in which, for any pair of species, the eigenvector ellipse of one species is fully contained within that of the second species (Fig. 1). This behaviour results in a nested Russian-doll pattern when applied to a species rich meta-community. The common true subset definition of nestedness (Patterson and Atmar 1986) appears as a special case of this generalised definition.

Accordingly, we define perfect species turnover as a pattern in which the eigenvector ellipses of any pair of species do not overlap. Again, this is an extension of the common definition of turnover based on presence–absence data only (Fig. 1). Perfect turnover is the opposite of perfect nestedness, but can be achieved only if $m < n$. At higher matrix fill, the eigenvector ellipses necessarily partly overlap, and turnover is associated with some degree of species aggregation/modularity due to spatially separated groups of species having high ellipsoid overlap (O).

The degree of modularity can be quantified by the coefficient of variation $CV_{env} = \frac{\sigma}{\mu}$, where μ is the average Euclidean distance between ellipsoid centroids within the hyperspace spanned by the Z-transformed environmental variables and σ its standard deviation. High values of CV_{env} are associated with increased modularity. As a reference, a Poisson random distribution is characterized by $CV_{env} \approx 1$. This metric distinguishes turnover and modularity from the placement of the ellipsoid centres. The turnover pattern is characterised by equally distributed ellipses in space (Fig. 1) resulting in comparably low values of CV_{env}.

Therefore, the ellipsoid approach makes it possible to identify three basic patterns in ecological presence–absence
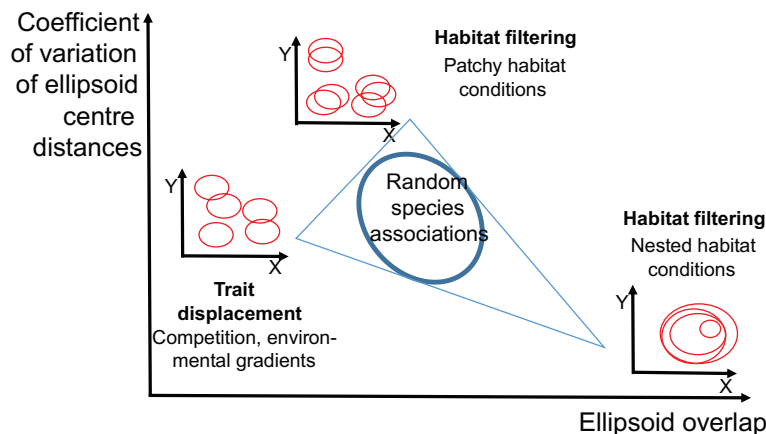

Figure 2. Ellipsoids of the functional traits realised in a given habitat calculated from local communities generate three basic patterns in the distribution of trait space: Habitat filtering, trait displacement, and trait clustering. As in Fig. 1, the distribution of eigenvector ellipsoids with the coordinate system made by ellipsoid overlap and the coefficient of variation of ellipsoid centre distances (CV_{trait}) allows for an identification of these patterns.

matrix. The results can be conveniently summarized in a plot of the coefficient of variation CV_{env} versus ellipse overlap O_{env} (Fig. 1). This representation provides an intuitive way to discriminate between nestedness (high O_{env} and low CV_{env}), turnover (low O_{env} and low CV_{env}), and modularity (intermediate O_{env} and high CV_{env}) (Fig. 1).

Functional trait based patterns of co-occurrence

The ellipsoid approach also applies to species trait data (cf. Lamanna et al. 2014 for a related approach), coded in a $m \times t$ species $\times$ trait matrix **T**. As for the environmental variables the trait variables must allow the calculation of variance–covariance matrix, in this case based on interspecific variability. The transposes of the **M** and **T** matrices, with dimensions $n \times m$ and $t \times m$, respectively, define the size of the trait ellipsoids of each site calculated over all occurring species, which constitutes a measure of the realised total trait space per site calculated over all species present there. These trait ellipsoids are calculated from all species occurring at a focal site. The respective quotient of realised trait space (defined by the occurring species) to total trait space (defined by all species of the meta-community) quantifies the proportion of trait space utilized and allows for comparisons among sites. Thus our approach adds to the traditional study of functional trait space (D'Andrea and Ostling 2016) the environmental perspective by calculating habitat specific total traits spaces.

As for environmental ellipsoids, the overlap in trait ellipsoids (O_{trait}) and the variability in trait centroid distances (CV_{trait}) can be used to assess whether species assemble according to similar traits, as predicted by habitat filtering theory, or whether species assemble to equally fill the available trait space, as predicted by classical niche theory (Fig. 2). Nested trait ellipsoids indicate that communities assemble for similar trait centroids and thus point to habitat filtering effects. In contrast, segregated trait ellipsoids point to trait displacement among species (Fig. 2). One of the processes leading to displacement is interspecific competition. A major advantage of our ellipsoid method is the possibility of identifying a third pattern of trait space distribution, namely the displacement of trait clusters (Fig. 2). These are characterised by intermediate average ellipsoid overlap and high variability in ellipsoid centroid distances. These clusters may indicate joint influences of habitat filtering and competitive effects.

Study sites and sampling of empirical data

We studied the Scots pine *Pinus sylvestris* understorey vegetation of 130 plots (400 m² each) from three Forest Divisions (Bolesławiec, Głogów and Oława) in southwestern Poland (Supplementary material Appendix 4 Fig. A4.1). The investigated plots were located in habitats spanning from pine mono-stands to mixed pine forests, and showed a wide variation in both soil moisture (from dry to boggy) and soil fertility (from poor sandy to fertile fine-textured). Respective raw data regarding species composition, plot geographic position, age, and soil fertility are provided in the Supplemenary material Appendix 2.

Vegetation surveys, conducted according to Braun-Blanquet (1964) sampling procedures, took place from June to August 2003 to 2005. From these data, we compiled a single presence–absence data matrix **M**$_{total}$ documenting the occurrence of 90 understorey plant species (matrix rows) across 130 plots (matrix columns; Supplementary material Appendix 2).

In each of the 130 studied plots, standard humus horizon soil samples were collected (Operat 2003, 2004, 2005). For the present study, we estimated soil fertility from the plot average of pH, N content, and the respective carbon–nitrogen (C/N) ratio. Species functional traits were assessed by Ellenberg indicator values (Ellenberg et al. 1992), with light, moisture, and soil nitrogen demands calculated by JUICE 7.0 (Tichý 2002) (Supplementary material Appendix 2). For simplicity we refer to them as functional traits although they rather reflect resource demands. Distances in trait space were assessed using the centroid distances of the respective eigenvalue ellipsoids. Of course the results of our approach depend on trait choice. Therefore, a careful choice of those traits that potentially influence the pattern of species co-occurrence is crucial for a proper interpretation of the results.

Data analysis

To relate environmental conditions to the pattern of species co-occurrence and environmental ellipsoid overlap (O_{env}), we used a sliding window approach, moving a window of eight columns along the 130 forest plots (with overlap of four columns). Prior to this operation, the plots were sorted according to the first eigenvector of the geographic distance matrix. This sorting ensured that the 31 resulting windows contained nearby plots with comparably similar environmental characteristics. This procedure ensured sufficient environmental variability among the windows to obtain reliable regression results, whereas a comparable bootstrap resampling approach would have combined plots regardless of their respective spatial and environmental distances. We used the same window approach for species traits and moved a sliding window (width eight, overlap four species) along the rows of species. This resulted in 15 windows combining species of similar numbers of occurrences.

Intuitively, raw scores of ellipsoid overlap and centroid distance should be affected by matrix size and fill. To test for these effects, we first applied a general linear model (with identity link function and equiprobable error structure) to assess (using all species pairs) the dependence of pairwise ellipsoid overlap and centroid distances on: the numbers of occurrences of the two species; fill of the matrix; and the respective pair-wise Sørensen similarity index. This was computed as $2n_{joint}/(n_1 + n_2)$, where n_{joint} is the number of joint occurrences and n_1 and n_2 the numbers of occurrences of both species. Additionally, we calculated for each sliding window the C-score as a measure of matrix-wide species segregation (Stone and Roberts 1990) and the NODF score (Almeida-Neto et al. 2008), which quantifies the degree of ordered species loss among plots (the degree of nestedness).

In species co-occurrence analysis, statistical inference is commonly based on a null model approach (Gotelli 2000, Gotelli and Ulrich 2012) to eliminate influences of matrix fill

and size. We tested the performance of three well-established null models (Gotelli 2000, Gotelli and Ulrich 2012), and compared the above described metric scores with those obtained from 19 the fixed – fixed (FF) randomization based on the independent swap algorithm, 2) the equiprobable – equiprobable (EE) randomization based on an equiprobable random reshuffling of matrix occurrences, and 3) an equiprobable reshuffling (SH) of each environmental variable among columns. We choose these three algorithms because they span from very permissive (EE, which does not account for differential occurrence probabilities that are not linked to the focal pattern) to very conservative (FF and SH, which assume observed numbers of occurrences to be independent of the focal pattern).

Each analysis was based on 200 null matrices. We used raw scores of O_{env} and CV_{env} and respective null model effect sizes ($\Delta = Obs - Exp$; Obs and Exp: observed and expected scores) as response variables in a general linear model (identity link function, equiprobable error structure) to infer possible dependencies on soil characteristics (ln-transformed averages per window and the respective standard deviations). The degrees of segregation and nestedness based on presence–absence data were assessed from standardized effect sizes (SES = Δ/σ; σ is the standard deviation of the null model distribution) and similarly linked to soil variables. Effect sizes of predictors were estimated by partial η^2-values. Because the results obtained with FF and SH were qualitatively similar to those obtained with EE, we show the EE results only and provide the respective results from FF and SH in the electronic supplement D.

Results

Ellipsoid overlap and species co-occurrences

Presence–absence based co-occurrence analysis revealed for the majority of shifting windows a significantly aggregated pattern of co-occurrence in comparison to the EE null model (average SES C-score = –12.9 ± 0.53, average SES NODF = 1.29 ± 0.29; mean ± SE, Fig. 3c) and a segregated pattern in comparison to the FF null model (average SES C-score = 4.16 ± 0.62, average SES NODF = –2.34 ± 0.71).
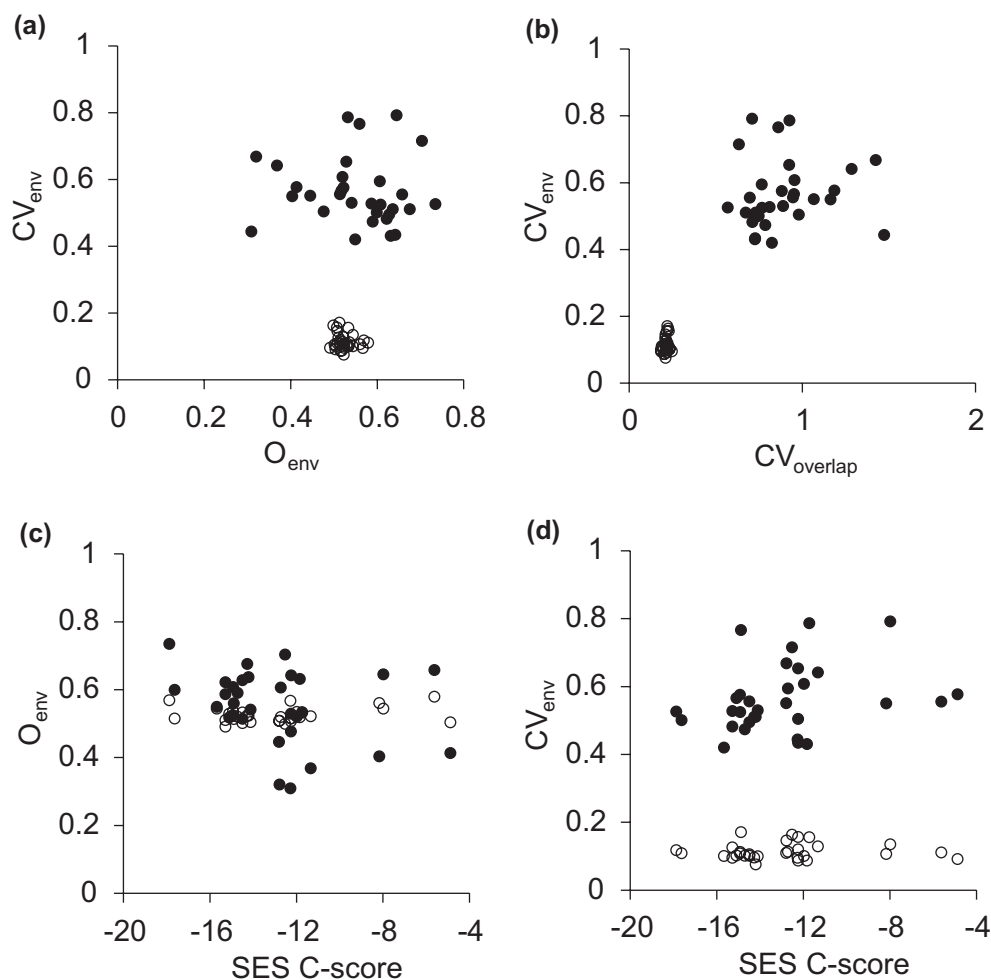

Figure 3. (a) Observed variability (black circles) in environmental centroid distances (CV_{env}) but not ellipsoid overlap (O_{env}) was larger than expected from the equiprobable (EE) null model (open circles, equivalent to the random space in Fig. 1 and 2) in all 31 shifting windows moved across plots of the matrix of forest understorey plants. Observed variability in ellipsoid overlap ($CV_{overlap}$) was higher than expected from EE (b). Observed O_{env} (c) and CV_{env} (d) did not depend on the standardised effect size (SES) of the C-score (EE null model).

Table 1. General linear modelling (normal error structure, identity link function) using all pairs of understorey forest species and 12 sliding windows (N = 10699). S1, S2, V1, V2: numbers of occurrences and environmental ellipsoid volumes of the two species compared; J: pair-wise Sørensen similarity; fill: matrix fill. Given are partial η^2 values. Raw scores and effect sizes Δ of the equiprobable null model of the environmental ellipsoid overlap (O_{env}) and the respective centroid distances (D_{env}) served as dependent variables. Significances: *: p < 0.05, **: p < 0.01, ***: p < 0.001. (+): positive effect; (–): negative effect.

Variable	Ellipsoid overlap		Centroid distance	
	O_{env}	$\Delta\,O_{env}$	D_{env}	$\Delta\,D_{env}$
S1	< 0.01	0.01	< 0.01	0.10
S2	0.01	0.02	< 0.01	0.06
V1	< 0.01	< 0.01	< 0.01	0.01
V2	< 0.01	< 0.01	< 0.01	0.02
J	(+) 0.34**	(+) 0.56***	0.02	(–) 0.46***
Fill	0.06	0.04	< 0.01	< 0.01
r^2 (model)	0.37**	0.58***	0.02	0.49***

Pairwise species environmental ellipsoid overlap of the understorey plant communities increased with the respective pairwise Sørensen similarity of species occurrences (r = 0.53, P(r = 0) < 0.0001 and Table 1). Average overlap varied between 0.3 and 0.8 and did not differ from random expectation (Figs. 3a, Supplementary material Appendix 4 Fig. A4.2a–A4.3a). The C-score was not related to the average ellipsoid overlap (Fig. 3c, Supplementary material Appendix 4 Fig. A4.2c). For the EE and FF null models, observed but not expected variability in ellipsoid centroid distances increased with spatial segregation (Fig. 3d, Supplementary material Appendix 4 Fig. A4.2d). Distances among ellipsoid centroids decreased with increasing similarity (r = 0.75, P(r = 0) < 0.0001 and Table 1).

Observed variability in overlap, calculated from all pairwise species comparisons, was in all sliding windows higher than expected from the EE null model (Fig. 3c) and from the FF and SH null models (Supplementary material Appendix 4 Fig. A4.2a, A4.3a). Observed variability in overlap was always higher than expected from the EE null model (Fig. 3b) but did not differ from FF and SH expectations (Supplementary material Appendix 4 Fig. A4.2b, A4.3b).

Ellipsoid overlap in relation to environmental conditions

The relationship between environmental ellipsoid overlap and distances among ellipsoid centroids of understorey

plants was partly determined by soil factors (Table 2, Supplementary material Appendix 4 Table A4.1). We found that nitrogen content and the respective variability among plots increased ellipsoid overlap, but did not influence distances among ellipsoid centroids (Table 2). In turn, pH and variability in pH were positively correlated with the distances among ellipsoid centroids. Neither of these patterns was detected by traditional species based co-occurrence analysis (Table 2). High C/N ratios and consequently low nitrogen availability decreased average distances among ellipsoid centroid (Table 2) and therefore the degree of species aggregation. Again, this pattern was not detected by the occurrence-based analysis. Compared to EE and FF, the SH null model proved to be conservative, and did not point to environmental influence of ellipsoid overlap and distances (Supplementary material Appendix 4 Table A4.2). High pH (r = 0.54, P(r = 0) < 0.01) and the high C/N ratio (r = 0.40, P(r = 0) < 0.05), but not soil nitrogen content (r < 0.05, P(r = 0) > 0.50) increased species richness. Variability in soil characteristics was not detectably linked to species richness (all pairwise r < 0.1, P(r = 0) > 0.1, not shown).

Average variability in distance but not in trait overlap was in all 15 sliding windows higher than expected from the EE null model (Fig. 4a). Distance in environmental space was significantly negatively correlated with trait overlap (Fig. 4b) and with variability in trait centroid distances (Fig. 4c). In 11 of the 15 sliding windows, average trait overlap among species was higher than expected from the EE null model (Fig. 4b).

Discussion

We have shown that species occurrence ellipsoids based on sets of environmental variables are able to discern between basic patterns of species co-occurrences (Fig. 1, 2). Importantly, our approach goes beyond the traditional one-dimensional view of co-occurrence centred upon the nestedness – turnover gradient (Leibold and Mikkelson 2002, Presley et al. 2010). Because it is based on environmental and functional trait data, our approach offers a straight-forward method to assess how environmental characteristics determine patterns of species co-occurrences. Finally, it allows for a direct comparison of environmental and functional niche space (Fig. 1, 2).

Crucially, our method shifts the analysis of co-occurrence from an occurrence-based perspective to a niche-occupancy

Table 2. General linear modelling (normal error structure, identity link function), with log-transformed environmental predictor variables in rows and community response variables in columns (environmental ellipsoid overlap O_{env}, respective centroid distances D_{env}, and the coefficients of variation of overlap $CV_{overlap}$, and centroid distance CV_{env} (effect sizes Δ of the equiprobable null model), and standardised effect sizes of the C-score and NODF). Data from 31 sliding windows moved upon the matrix of understorey forest plants. Given are partial η^2 values. Significances: *: p < 0.05, **: p < 0.01, ***: p < 0.001. (+): positive effect; (–): negative effect.

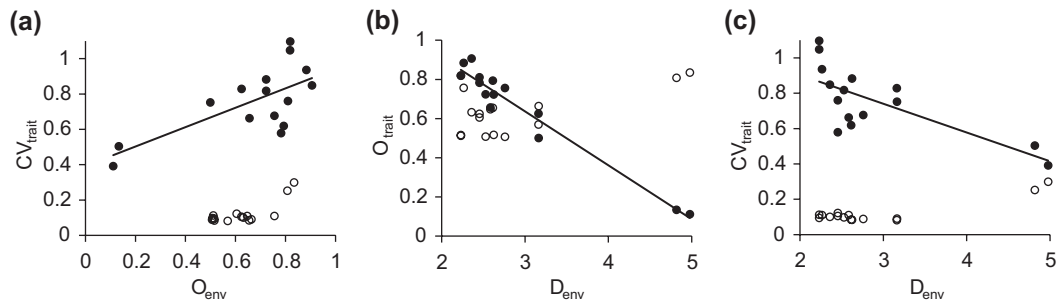| Variable	$\Delta\,O_{env}$	$\Delta\,CV_{overlap}$	$\Delta\,D_{env}$	$\Delta\,CV_{env}$	SES C-score	SES NODF
ln pH	0.01	0.01	(–) 0.11	(–) 0.31*	0.08	0.01
ln N	(+) 0.20*	(+) 0.27*	0	0.03	0.01	< 0.01
ln C/N	0.01	0	0.03	(+) 0.20	0.02	(–) 0.19*
ln StDev pH	0	0	(+) 0.19*	(+) 0.26*	0.02	0.03
ln StDev N	0.02	0.03	0.05	0.01	0.07	0.06
ln StDev C/N	0	0.01	0.09	(–) 0.33**	< 0.01	(+) 0.28*
r^2 (whole model)	0.45***	0.55***	0.37**	0.34**	0.03	0.26*

Figure 4. (a) Observed variability (black circles) in trait centroid distances (CV_{trait}) was in all shifting windows moved across species of the matrix of forest understorey plants larger than expected from the equiprobable (EE) null model (open circles, equivalent to the random space in Fig. 1 and 2) and increased with environmental ellipsoid overlap O_{env} ($r^2 = 0.48$, permutation $p < 0.01$). Trait ellipsoid overlap O_{trait} (b: $r^2 = 0.95$, $p < 0.001$) and CV_{trait} (c: $r^2 = 0.52$, $p < 0.001$) decreased with environmental centroid distance D_{env}. O_{trait} was in 12 of the 15 windows larger than expected from the EE null model.

perspective. The geometry of species co-occurrences has been historically interpreted in terms of interspecific competition (Diamond 1975, Gotelli 2000). Thus, mutual species exclusions served as a surrogate to study competitive exclusion according to Gauses' principle (1934), based on niche overlap and limiting similarity (MacArthur and Levins 1967, Chesson 2000). However, evidence for mutual exclusion is scarce (Abrams and Rueffler 2009, van Leeuwen and Etienne 2013) and evidence is mounting for other mechanisms leading to segregated species co-occurrences (Blois et al. 2014).

A major reason for this lack of corroboration might be that the use of occurrences as a proxy for niche segregation is only justified if the sites where species exclude one another have similar environmental characteristics (Blois et al. 2014). Consequently, a common implicit assumption of co-occurrence analysis is that sites are environmentally similar. However, such an assumption is only rarely justified, and observed patterns of species co-occurrence might be more often the outcome of subtle environmental differences among sites rather than competitive interactions among species. Environmental differences induce filter effects, leading to the development of associations of species with similar resource requirements, and therefore ecological niches. Consequently, observed patterns of species segregation might indicate species interactions or habitat filtering (Blois et al. 2014). Classical co-occurrence analysis is unable to discern between these two drivers in community assembly. The ellipsoid approach widens the analysis of species co-occurrences in both environmental and trait dimensions, improving the assessment of overlap in resource use.

In the present case study, we found clear indication of a modular organisation of understorey forest plant communities (indicated by the higher variability in centroid distances than expected from the null model) with respect to environmental characteristics (Fig. 3). This was not visible from co-occurrence analysis alone, which gave contrasting results mostly driven by null model's choice (EE versus FF). Trait ellipsoid analysis also revealed a significant modular structure (Fig. 4), caused by co-occurring groups of species within similar environmental space (Fig. 3a). Importantly, our new approach allowed for the identification of the major drivers behind this modular pattern in environmental niche

space (Table 2). High pH and soil nitrogen content, both linked to nutrient rich soils and productivity, increased environmental overlap and decreased the degree of modularity (Table 2). Consequently, we speculate that a modular community organisation might be particularly associated with variable and resource poor environments. These relationships did not depend on species richness and total ellipsoid volume (Table 1), and were not evident in traditional co-occurrence analysis (Table 2). To our knowledge the relationship between niche modularity and environmental factors has not been studied so far. Interestingly, prior studies on the connection between plant species turnover and habitat productivity have returned inconsistent results (Paoli et al. 2006, Fernández-Going et al. 2013, Ulrich et al. 2014), perhaps because the pattern of niche modularity changes with soil characteristics.

Soil variables were correlated with both species richness and increasing modularity. Recent theoretical work (Ulrich et al. 2017a) on variation in beta diversity (species turnover) has shown that richness and the degree of turnover are intimately linked and cannot be disentangled by simple null model analysis. Traditional statistical inference of species co-occurrences has tried to control for richness effects when inferring the degree of species turnover (Gotelli 2000). Our results here suggest that to a certain degree modularity itself in either traits or environmental niches might be a consequence of increased species richness. We note that previous studies of food web structure also found a positive correlation between functional modularity and community species richness (Sebastián-González et al. 2015, Montoya et al. 2015). Further, a modular organisation can stabilise communities (Grilli et al. 2016). However, it remains an open question whether these results also hold for environmentally- and trait-defined niches.

We did not find evidence for a nested or segregated environmental niche structure. Instead, species occurrences were aggregated but not nested in comparison to an equiprobable null expectation. Nested patterns of co-occurrences often result from neutral community assembly (Ulrich et al. 2009), whereas segregation occurs in heterogeneous habitats (Boeye et al. 2014), or in competitively organised communities (Diamond 1975) or in communities with high rates of temporal species turnover (Thuiller et al.

2007). Consequently, our results do not point to competition as a major driver of community organisation. We also have no evidence for high colonisation rates from a regional species pool and a resulting neutral community organisation. However, because community structure is known to change with spatial scale (Cottenie 2005), our interpretation does not preclude strong within-plot competitive interactions at a spatial scale below the resolution of our data.

Our approach allows for a straightforward and readily interpretable link between trait and environmental niche spaces. Previously, such links were provided by (mostly) linear ordination techniques like correspondence analysis or non-metric multidimensional scaling. However, these methods did not differentiate between the specific patterns of species co-occurrence. They did also not directly refer to species functional niches. Because the degree of trait modularity was negatively correlated with the segregation of environmental ellipsoids (and hence niche space; Fig. 4c), we argue that environmental filtering sorted species according to environmental factors, which led to segregation in trait space. Such a pattern is expected if species with similar environmental demands have also similar functional traits. In this respect, we found a clear indication of modularity in the distribution of functional traits (Fig. 4a).

Interestingly, the degree of modularity was positively linked to average ellipsoid overlap. Thus, high overlap in environmental niche space, and therefore high species packing, might be associated with the appearance of species groups with similar functional traits. In turn, this might increase community stability, as reported by Bastolla et al. (2009) and Rohr et al. (2014) for mutualistic networks. However, specific studies aimed at linking trait distribution to species richness are lacking. In line with this finding, we observed decreased trait overlap with increasing environmental distance (Fig. 4b–c). Such a pattern is expected if environmental filtering also filters for traits resulting in a distance decay of trait overlap along environmental gradients (Kraft et al. 2015).

Here we focused on presence–absence data. However, our approach can be easily extended to incorporate abundance data by weighting the environmental variables by abundances in the second mapping step of analysis. Furthermore, our method can be applied to dataset including phylogenetic information. In fact, because the ellipsoid method is able to directly link phylogenetic distances to environmental niche overlap, it has also the potential to identify those environmental or geographical variables that contribute most to phylogenetic differentiation.

References

Abrams, P. A. and Rueffler, C. 2009. Coexistence and limiting similarity of consumer species competing for a linear array of resources. – Ecology 90: 812–822.

Almeida-Neto, M. et al. 2007. On nestedness analyses: rethinking matrix temperature and anti-nestedness. – Oikos 116: 716–722.

Almeida-Neto, M. et al. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and quantification. – Oikos 117: 1227–1239.

Bascompte, J. 2009. Disentangling the web of life. – Science 325: 416–419.

Bastolla, U. et al. 2009. The architecture of mutualistic networks minimizes competition and increases biodiversity. – Nature 458: 1018–1020.

Blois, J. L. et al. 2014. A framework for evaluating the influence of climate, dispersal limitation, and biotic interactions using fossil pollen associations across the late Quaternary. – Ecography 37: 1095–1108.

Boeye, J. et al. 2014. Habitat structure mediates spatial segregation and therefore co-existence. – Landscape Ecol. 29: 593–604.

Braun-Blanquet, J. 1964. Pflanzensoziologie – Grundzüge der Vegetationskunde, 3rd edn. – Springer.

Brown, B. L. et al. 2017. Making sense of metacommunities: dispelling the mythology of a metacommunity typology. – Oecologia 183: 643–652.

Chesson, P. 2000. Mechanisms of maintenance of species diversity. – Annu. Rev. Ecol. Syst. 31: 343–366.

Connor, E. H. and Simberloff, D. 1979. The assembly of species communities: chance or competition? – Ecology 60: 1132–1140.

Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. – Ecol. Lett. 8: 1175–1182.

D'Andrea, R. and Ostling, A. 2016. Challenges in linking trait patterns to niche differentiation. – Oikos 125: 1369–1385.

Dallas, T. A. et al. 2016. Ordination obscures the influence of environment on plankton metacommunity structure. – Limnol. Oceanogr. 1: 54–61.

Diamond, J. M. 1975. Assembly of species communities. – In: Cody, M. L. and Diamond, J. M. (eds), Ecology and evolution of communities. Harvard Univ. Press, pp. 342–444.

Doledec, S. et al. 2000. Niche separation in community analysis: a new method. – Ecology 81: 2914–2927.

Ellenberg, H. et al. 1992. Zeigerwerte von Pflanzen in Mitteleuropa. – Scripta Geobotanica XVIII, Erich Goltze KG, Göttingen.

Fariňa, J. M. et al. 2009. Can conservation biologists rely on established community structure rules to manage novel systems? … Not in salt marshes. – Ecol. Appl. 19: 413–422.

Fernández-Going, B. M. et al. 2013. Climate interacts with soil to produce beta diversity in Californian plant communities. – Ecology 94: 2007–2018.

Fortuna, M. A. et al. 2010. Nestedness versus modularity in ecological networks: two sides of the same coin? – J. Anim. Ecol. 79: 811–817.

Friendly, M. et al. 2013. Elliptical insights: understanding statistical methods through elliptical geometry. – Stat. Sci. 28: 1–39.

Gassert, F. et al. 2013. From southern refugia to the northern range margin: genetic population structure of the common wall lizard *Podarcis muralis*. – J. Biogeogr. 40: 1475–1498.

Gause, G. F. 1934. The struggle for existence. – Williams and Wilkins, Baltimore.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. – Ecology 81: 2606–2621.

Gotelli, N. J. and Graves, G. R. 1996. Null models in ecology. – Smithsonian Inst. Press.

Gotelli, N. J. and McCabe, D. J. 2002. Species co-occurrence: meta-analysis of J. M. Diamond's assembly rules model. – Ecology 83: 2091–2096.

Gotelli, N. J. and Ulrich, W. 2012. Statistical challenges in null model analysis. – Oikos 121: 171–180.

Götzenberger, L. et al. 2012. Ecological assembly rules in plant communities – approaches, patterns and prospects. – Biol. Rev. 87: 111–127.

Grilli, J. et al. 2016. Modularity and stability in ecological communities. – Nature Comm. 7: 12031.

Hubbell, S. P. 2001. The unified theory of biogeography and biodiversity. – Princetion Univ. Press.

Hultén, E. 1937. Outline of the history of Arctic and Boreal biota during the Quaternary period. – Thule, Stockholm.

Hutchinson, G. E. 1978. An introduction to population ecology. – Yale Univ. Press.

Jackson, A. L. et al. 2011. Comparing isotopic niche widths among and within communities: SIBER – stable isotope Bayesian ellipses in R. – J. Animal Ecol. 80: 595–602.

Janžeković, F. and Novak, T. 2012. PCA – a powerful method for analysing ecological niches. – In: Sanguansat, P. (ed.), Principal component analysis – multidisciplinary applications. InTech, Rijeka, pp. 127–142.

Kraft, N. J. B. et al. 2008. Functional traits and niche-based tree community assembly in an Amazonian forest. – Science 322: 580–582.

Kraft, N. J. et al. 2015. Community assembly, coexistence and the environmental filtering metaphor. – Funct. Ecol. 29: 592–599.

Lamanna, C. et al. 2014. Functional trait space and the latitudinal diversity gradient. – Proc. Natl Acad. Sci. USA 111: 13745–13750.

Legendre, P. and Legendre, L. 1998. Numerical ecology. – Elsevier Science.

Leibold, M. A. and Mikkelson, G. M. 2002. Coherence, species turnover, and boundary clumping: elements of metacommunity structure. – Oikos 97: 237–250.

Louy, D. et al. 2014. Out of the Alps: the biogeography of a disjunctly distributed mountain butterfly, the almond-eyed ringlet *Erebia alberganus* (Lepidoptera, Satyrinae). – J. Heredity 105: 28–38.

Lyons, S. K. et al. 2016. Holocene shifts in the assembly of terrestrial plant and animal communities implicate increasing human impacts. – Nature 529: 80–83.

MacArthur, R. H. and Wilson, E. O. 1963. An equilibrium theory of insular zoogeography. – Evolution 17: 373–387.

MacArthur, R. and Levins, R. 1967. The limiting similarity, convergence and divergence of coexisting species. – Am. Nat. 101: 377–385.

Maire, E. et al. 2015. How many dimensions are needed to accurately assess functional diversity? A pragmatic approach for assessing the quality of functional spaces. – Global Ecol. Biogeogr. 24: 728–740.

McGill, B. et al. 2006. Rebuilding community ecology from functional traits. – Trends Ecol. Evol. 21: 178–185.

Montoya, D. et al. 2015. Functional group diversity increases with modularity in complex food webs. – Nature Comm. 6: 8379.

Newman, M. E. J. and Girvan, M. 2004. Finding and evaluating community structure in networks. – Phys. Rev. E 69: 026113.

Operat 2003. Elaboration of forest sites for the Oława Forest Division. – Operat S.C., Toruń, unpublished data for forest administration use.

Operat 2004. Elaboration of forest sites for the Bolesławiec Forest Division. – Operat S.C., Toruń, unpublished data for forest administration use.

Operat 2005. Elaboration of forest sites for the Głogów Forest Division. – Operat S.C., Toruń, unpublished data for forest administration use.

Paoli, G. D. et al. 2006. Soil nutrients and beta diversity in the Bornean Dipterocarpaceae: evidence for niche partitioning by tropical rain forest trees. – J. Ecol. 94: 157–170.

Patterson, B. D. and Atmar, W. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. – Biol. J. Linn. Soc. 28: 65–82.

Presley, S. J. et al. 2010. A comprehensive framework for the evaluation of metacommunity structure. – Oikos 119: 908–917.

Qiao, H. et al. 2016. NicheA: creating virtual species and ecological niches in multivariate environmental scenarios. – Ecography 39: 805–813.

Rohr, R. P. et al. 2014. On the structural stability of mutualistic systems. – Science 345: e1253497.

Stokstad, E. 2009. On the origin of ecological structure. – Science 326: 33–35.

Sebastián-González, E. et al. 2015. Macroecological trends in nestedness and modularity of seed-dispersal networks: human impact matters. – Global Ecol. Biogeogr. 24: 293–303.

Stone, L. and Roberts, A. 1990. The checkerboard score and species distributions. – Ocologia 85: 74–79.

Tichý, L. 2002. Juice, software for vegetation classification. – J. Veg. Sci. 13: 451–453.

Thuiller, W. et al. 2007. Stochastic species turnover and stable coexistence in a species-rich, fire-prone plant community. – PloS One 2: e938.

Tuomisto, H. 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. – Ecography 33: 2–22.

Ulrich, W. and Gotelli, N. J. 2007. Disentangling community patterns of nestedness and species co-occurrence. – Oikos 116: 2053–2061.

Ulrich, W. and Gotelli. N. J. 2013. Pattern detection in null model analysis. – Oikos: 122: 2–18.

Ulrich, W. et al. 2009. A consumer's guide to nestedness analysis. – Oikos 118: 3–17.

Ulrich, W. et al. 2014. Climate and soil attributes control plant species turnover in global drylands. – J. Biogeogr. 41: 2307–2319.

Ulrich, W. et al. 2017a. The tangled link between β- and γ-diversity: a Narcissus effect weakens statistical inferences in null model analyses of diversity patterns. – Global Ecol. Biogeogr. 26: 1–5.

van Leeuwen, E. and Etienne, R. S. 2013. Caught in the middle: asymmetric competition causes high variance in intermediate trait abundances. – Theor. Popul. Biol. 85: 26–37.

Weiher, E. and Keddy. P. (eds) 1999. Ecological assembly rules: perspectives, advances, retreats. – Cambridge Univ. Press.

Wilson, D. S. 1992. Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection. – Ecology 73: 1984–2000.

Zalewski, M. et al. 2014. High niche overlap in stable isotope space of ground beetles. – Ann. Zool. Fenn. 51: 301–312.