On the reliability of the Elements of Metacommunity Structure framework for separating idealized metacommunity patterns

Dénes Schmera1,2,*, János Podani3,4, Zoltán Botta-Dukát2,5 and Tibor Erős1,2

1 MTA Centre for Ecological Research, Balaton Limnological Institute, Klebelsberg K. u. 3, H-8392 Tihany, Hungary
2 MTA Centre for Ecological Research, GINOP Sustainable Ecosystem Group, Klebelsberg K. u. 3, H-8392 Tihany, Hungary
3 Department of Plant Systematics, Ecology and Theoretical Biology, Institute of Biology, L. Eötvös University, Pázmány P. s. 1/c, H-1117 Budapest, Hungary
4 Ecology Research Group of the Hungarian Academy of Sciences, Budapest, Hungary
5 MTA Centre for Ecological Research, Institute of Ecology and Botany, Alkotmány u. 2-4, H-2163 Vácrátót, Hungary

*Correspondence: schmera.denes@okologia.mta.hu
Abstract

The Elements of Metacommunity Structure (EMS) framework originally suggested by Leibold and Mikkelson (2002) in Oikos is a popular approach to identify idealized metacommunity patterns (i.e. checkerboard, nested, evenly spaced, Clementsian, Gleasonian), and hereby to infer the existence of structuring processes in metacommunities. Essentially, the EMS framework consists of the rearrangement of the sites-by-species incidence matrix followed by a series of tests for coherence, turnover and boundary clumping in species distributions. Here, we give a critical evaluation of the EMS framework based on theoretical considerations and simulations. We found that user defined site ordering may influence the coherence test (number of embedded absences) depending also on the ordering of species, and therefore we argue that the application of user-defined matrix rearrangement has strong limitations. The recommended ordering by correspondence analysis is sensitive to matrix structure and may even include arbitrary decisions in special cases. Further, we revealed different meanings of the checkerboard pattern and showed that negative coherence is not necessarily associated with this as assumed in the EMS framework. Also, the turnover test cannot always detect nested pattern, because turnover and nestedness are not necessarily the opposite endpoints of a continuum. We argue that the boundary clumping test can only be used for separating Clementsian, Gleasonian and evenly spaced patterns if sites are ordered along a real environmental gradient rather than a latent one identified by correspondence analysis. We found that the series of tests in the EMS framework are burdened by anomalies and that the detection of some metacommunity patterns is sensitive to type II error. In sum, our findings suggest that the analytical methodology of the EMS framework, as well as the conclusions drawn from its application to metacommunity studies require careful reconsideration.

Keywords

biodiversity; community pattern; pattern analysis; idealized metacommunity patterns
1. Introduction

Detecting and understanding drivers of metacommunity structure are key issues in community ecology with significant legacy (Mittelbach 2012). Early ecologists have already inferred the existence of structuring forces from the community patterns observed. For instance, Clements (1916), the pioneer of North American plant ecology, viewed plant communities as coherent units with discrete boundaries formed in response to environmental factors (Clementsian pattern). In contrast, Gleason (1926) argued that species have distinct ecological characteristics and therefore individualistic responses to underlying environmental gradients (Gleasonian pattern). Evenly spaced pattern occurs in systems with trade-offs in fitness in different environments, resulting in a spatial distribution with evenly dispersed populations (Tilman 1982). Intense interspecific competition may generate checkerboard pattern where pairs of species are mutually exclusive (Diamond 1975). Finally, nested pattern occurs when species poor communities consist of subsets of species occurring in richer communities (Patterson and Atmar 1989). These cases have been regarded as idealized types of metacommunity pattern (Ulrich and Gotelli 2013, Heino et al. 2015) and have received increasing attention due to their theoretical interpretation (Carvalho et al. 2013, Ulrich and Gotelli 2013).

The development of metacommunity theory provided a conceptual framework for ecologists to disentangle underlying drivers (niche based species sorting, dispersal, drift, see Vellend 2010, Shipley et al. 2012) of multisite communities (Leibold et al. 2004). Some of the approaches use multispecies distribution patterns for inferring the existence of structuring ecological forces. No doubt that the “elements of metacommunity structure” approach suggested by Leibold and Mikkelsen (2002) and its upgrade (Presley et al 2010, hereafter referred to as EMS framework) provide a very popular methodology developed for this purpose.

The EMS framework includes the rearrangement of the sites-by-species incidence matrix followed by three tests (Fig. 1), each related to a given element of metacommunity structure. First, the rows and the columns of the matrix are ordered along the first axis of correspondence analysis (CoA) to discern variation in response to a latent environmental gradient. According to Leibold and Mikkelsen (2002, p. 241), the simultaneous ordering of sites and species has three purposes: (1) it often minimizes the number of interruptions in species' ranges (number of embedded absences), (2) it provides a basis for judging whether a given metacommunity is nested, or dominated by turnover (high number of species replacements), and (3) it defines the boundaries of species' ranges (boundary clumping). Consequently, matrix rearrangement via CoA has strong impact on the assessment of each element of metacommunity structure. Note that although this procedure is recommended for general use, the EMS framework also allows user-defined matrix ordering. Secondly coherence, the first element of metacommunity structure is defined as the number of embedded absences in the matrix and its significance is examined using null model tests. Following the study of Gotelli (2000), species richness of sites is kept constant in the recommended null model (Presley et al. 2010). If coherence is negative (the number of embedded absences is significantly higher than expected by chance) then the EMS
framework detects checkerboard pattern. If the number of embedded absences does not differ significantly from a randomly generated value (coherence is random) then the EMS framework indicates a random pattern. If coherence is positive (the number of embedded absences is lower than expected by chance) then the matrix should be examined for turnover. Turnover, the second element of metacommunity structure, is measured as the number of times one species replaces another between two sites (i.e. number of replacements) for each possible pair of species and for each possible pair of sites. If turnover is negative (the number of replacements is lower than expected by chance) then the EMS framework reveals a nested pattern, if turnover is random the EMS detects quasi pattern (see Presley et al. 2010), and if turnover is positive (the number of replacements is higher than expected by chance) then the EMS framework suggests the existence of Clementsian, Gleasonian or evenly spaced patterns. These latter three are separated from each other by examining the boundary clumping of species ranges, the third element of metacommunity structure, using the Morisita test. If clumping is positive (Morisita I is significantly larger than 1.0) then the EMS framework detects Clementsian pattern; if clumping is negative (Morisita I is significantly lower than 1.0) evenly spaced pattern is indicated, and if clumping is random (Morisita I does not significantly differ from 1.0) then the pattern is thought to be Gleasonian.

There is, however, much controversy about the relative merits of the EMS framework. Gotelli and Ulrich (2012, p. 178), for instance, noted that species segregation and aggregation examined in the coherence test "might be the different sides of the same coin" and that rearranging the matrix (i.e. the reordering of sites by correspondence analysis) "does not alter any of the underlying information on species occurrences in the matrix". By examining the power of different null model algorithms, Gotelli and Ulrich (2012) found that a segregation measure was not exactly opposite in its behavior to a nestedness measure, suggesting that nested and segregated patterns (i.e. evenly spaced, Gleasonian and Clementsian) are not necessarily mutually exclusive as implied by the turnover test in the EMS framework. The same authors repeated this comment later and also argued that "The frameworks proposed by Leibold and Mikkelson (2002), and Presley et al. (2010) implicitly assume that measures of coherence, turnover, and boundary clumping describe orthogonal, independent properties of matrices. But if the measures are strongly correlated, some of the proposed cells in their classification frameworks may be redundant or not achievable. Leibold and Mikkelson (2002) recognized this problem and noted that they were able to identify empirical matrices that fit each of the five different scenarios they described" (Ulrich and Gotelli 2013, p. 3). A more recent paper stated that the efficiency of the EMS framework is heavily dependent on data quality (Mihaljevic et al. 2015, see also Gotelli and Graves 1996, Ulrich and Gotelli 2013) and suggested the use of occupancy models to at least partly overcome this problem. These models allow an estimation of predicted occupancy at each sample site and thus make it possible to distinguish between the probability of a species occurring at a site and the probability of a species being detected at a site in which it does occur (Mihaljevic et al. 2015). These critical comments, however, did not prevent community ecologists from using the methodology even further. The EMS framework has still been used increasingly both in terrestrial and aquatic realms for finding the best fit to idealized

However, the reliability of the method in discerning idealized (meta)community patterns has not been tested as yet.

To fill this methodological gap, this paper examines the performance of the EMS framework. Combining theoretical aspects with simulation approaches we go through this approach step by step and inspect how the rearrangement of the matrix, the output of individual tests as well as their series influence the success of analysis. We examined also the robustness of the methodology to increasing noise in the data, as well as the practice of researchers in revealing the importance of environmental factors structuring metacommunity patterns.

2. Methods

To guarantee unambiguous answers, we first carefully review terms and procedures related to the EMS framework. We discuss possible interpretations of terms and evaluate the performance of different procedures. In case of equivocal use of any term or procedure, we attempt to clarify the situation by suggesting a solution.

We calculated the following indices: the number of embedded absences (the index of coherence test, Leibold and Mikkelsen 2002, Presley et al. 2010), the number of mutually exclusive species pairs (Diamond 1975), turnover (the index of turnover test, Leibold and Mikkelsen 2002, Presley et al. 2010). As nestedness is not defined in the EMS framework, we used two nestedness measures, the relativized nestedness ($N_{\text{rel}}$, Podani and Schmera 2011) and the site-order independent version of NODF (Almeida-Neto et al. 2008) called as $\text{NODF}_{\text{max}}$ (Podani and Schmera 2012, Ulrich and Almeida-Neto 2012).

We examined the behavior of indices themselves as well as the behavior of the indices in null model tests. Indices were examined using toy data sets in series of site-by-species incidence matrices. We examined the relationship between indices in two-site situations using the random parameter approach (Chao et al. 2012, see also Baselga and Leprieur 2015). In the first (Random parameter approach 1), we assumed that the numbers of species present in both sites ($a$), present only in the first site ($b$), and present only in the second site ($c$) are derived from a uniform distribution ranging from 0 to 100. We generated 50,000 triplets of random $a$, $b$ and $c$ values, and removed data records with empty sites. In the second case (Random parameter approach 2), we assumed that 200 species are distributed among the three sets ($a$, $b$ and $c$). We produced all possible combinations and removed data records with empty sites. Furthermore, we simulated all the possible sites-by-species matrices containing 4 sites and 4 species (degenerate matrices were omitted). This procedure resulted in 41,503 binary matrices, called hereafter as 4-by-4 binary matrices.

Although the 4-by-4 binary matrices allow examining the response of indices to all possible situations in the matrix, the null model test of the matrix might be problematic due to the small number of sites and species. We therefore produced 10,000 random matrices with 10 sites and 10 species (degenerate matrices, i.e. those containing empty rows or columns, were omitted). These are referred to as 10-by-10 matrices. We used them in null model tests.
(Gotelli and Graves 1996). For each random matrix, we generated 1000 null matrices. Although there are many algorithms to produce 'random' or 'null' matrices and these algorithms have different statistical properties and ecological meanings (Gotelli and Ulrich 2012, Ulrich and Gotelli 2013, Strona et al. 2017), we selected the null model method that maintained the species richness of every site and filled species ranges based on their marginal probabilities ("r1" method in metacom package, Dallas 2014). The P value (estimated probability of type I error) was calculated as the number of null matrices whose index value was more extreme than or equal to the observed index. We applied a two-tailed test at \( \alpha = 0.05 \). The Jaccard index (Jaccard 1912) was used to measure the similarity of different null model tests: the number of matrices proved to be significantly positive (or negative) in both tests was divided by the number of such matrices plus those that were found significant only in either of the two tests. Positive and negative results in the two tests were not distinguished.

We used a noise test (Gotelli 2000, Podani and Schmera 2012) to examine the sensitivity of the EMS framework to increasing randomness in community data. We started with 20-by-20 perfectly structured nested, Gleasonian, evenly spaced and Clementsian patterns (Electronic Appendix 1). These patterns were regarded as initial patterns (step 0, 0% noise). We then gradually added noise (randomness) to the matrix in the following way: In the first step (5% noise), 20 pairs of randomly chosen values in the matrix were interchanged (referred to the full randomization model in Podani and Schmera 2012). In the second step (10% noise), 40 pairs of randomly chosen values were interchanged. Complete randomness (100% noise) is achieved after 20 steps, with a total of 400 interchanges. Degenerate matrices were omitted. This procedure was repeated 100 times for every step. EMS analysis was performed for each step (21 steps) 100 times. The output of the noise test shows the relative frequency of detected metacommunity patterns in response to increasing noise level (from 0% to 100%). The ideal - and expected - situation is that at low noise level the methodology detects mostly the initial pattern. At intermediate noise level, the initial pattern is detected in a decreasing number of times, while the frequency of random pattern is increasing. At high noise level, the frequency of random pattern should be the largest. If the initial pattern is not detected many times even at low noise level, then the EMS framework is sensitive to type II error. In contrast, if the initial pattern is detected with high frequency even at high noise level, the EMS framework is sensitive to type I error.

Finally, we examined how researchers use the EMS framework and handle the importance of environmental factors in shaping metacommunity patterns. To reveal this, first we made a search using ISI Web of Science (access date: 28 July 2015) on the number of papers citing Presley et al. (2010). In the second step, we searched for papers applying the EMS framework. We divided these papers into two groups: those applying user defined matrix ordering and articles using CoA for site and species ordering. Then, we searched for papers that reported the variance explained by CoA axes. In our view, this information is essential, and should be obligatorily added to EMS analysis as an expression of the reliability of the method. No doubt that the amount of community variation explained must be used for
assigning the studied metacommunity to an idealized pattern. Finally, we examined whether
the axes of CoA (EMS framework) were related to any environmental variables.

All calculations were performed in R (R Core Team 2016). All possible matrices containing 4
sites and 4 species were produced by the gtools package (Warnes et al. 2015). Null matrices
were produced by the metacom package (Dallas 2014). Correspondence analysis (CoA) was
performed by the ca package (Nenadic and Greenacre 2007), the number of mutually
exclusive species pairs, number of embedded absences, turnover, relativized nestedness and
NODF_{max} were calculated by R-scripts developed by the authors (Electronic Appendix 2).

3. Site and species ordering

By definition, site and species orderings influence the number of embedded absences
(order-dependent measure) in the data matrix, but they have no impact on the number of
replacements (order-independent measure). In addition, site ordering also affects patterns
in boundary clumping (order-dependent measure). That site and species ordering both
influence coherence can be explained by the definition of embedded absence: "an
interruption in a range or community" (p. 242 in Leibold and Mikkelson 2002).

Studying communities along an environmental gradient is a typical situation for user-defined
site-ordering. The EMS framework allows user-defined matrix ordering without emphasizing
the importance of species ordering. Since coherence is influenced not only by the order of
sites but also by species ordering, as said, user-defined matrix ordering has strong
limitations. Therefore, if the data matrix is ordered by the user, we recommend a clear
definition of species ordering, if it is possible at all.

Alternatively, the recommended matrix-ordering uses the first axis of CoA to define the
order of sites (and species) for the coherence test. In this case, we disclaim real
environmental gradients and focus on the "within-matrix data structure". In complex data
structures, however, the first axis of CoA does not necessarily explain considerably more
variation than the subsequent axes. In other words, the first axis of CoA might identify one
dominant but not necessarily the only dominant axis of community variation. This means
that analyses of the same data matrix reordered along different axes might reveal
contrasting aspects of data structure. We by no means state that the use of the first axis of
CoA is a bad decision but emphasize that further studies are needed to reveal the effect of
choosing among similarly important axes, and to merge alternative results into a consensus.

Finally, although CoA has been one of the most popular ordination methods in numerical
ecology, it has some limitations. One of these is that sites with single and unique species
cannot be ordered due to the lack of overlap with other sites and species. If software
packages do order such matrices "in a way" then the result is based on an arbitrary decision
(Electronic Appendix 3). Accordingly, CoA performed by different software packages may
provide differently ordered matrices and thus it cannot be regarded as "a standardized
approach to order sites and species" as stated by Presley et al. (2010, p. 910). Although field
ecologists might argue that actual data sets rarely contain sites with unique species, we have
three arguments in favor of discussing this situation. First, a methodology should work under all circumstances, or at least its users should be aware of any limitations. Second, the chance of observing sites with unique species cannot be excluded completely in actual data sets. Finally, checkerboard pattern, a key term of the EMS framework, has a strong theoretical connection to sites with unique species (see next paragraph).

4. The multiple meaning of checkerboard pattern

Since the coherence test of the EMS framework is supposed to separate checkerboard, random and other data structures from each other, first we review the meaning of checkerboard pattern and then identify its most conspicuous realization. In examining the co-occurrence of bird species on islands, Diamond (1975) proposed the term "checkerboard distribution" for competing pairs of species with mutually exclusive island-by island distributions. In this metaphor, the distribution of the competing pair of species reflects the alternating squares of dark and light colors in a checkerboard pattern (of which the chessboard is a special case). Accordingly, we can visualize this checkerboard distribution for a given pair of species by the following matrix ($M_1$), where sites are rows and species are columns:

$$
M_1 = \begin{bmatrix}
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1
\end{bmatrix}
$$

It is very important to emphasize that Diamond's original checkerboard distribution reflects the distribution of pairs of species, and the "checkerboard character" of a community has been mostly (but not always, see Stone and Roberts 1990, 1992; Gotelli 2000) characterized by the number of checkerboard species pairs (i.e. the number of species pairs with non-overlapping occurrence). In agreement with this, and for compatibility with the EMS framework (Presley et al. 2010), we will use the term checkerboard pattern for binary matrices in which the number of checkerboard species pairs (Stone and Roberts 1990) is high. This definition means that the checkerboard pattern may be identified in binary (presence-absence) matrices without any restriction as to the order of sites and species.

Almeida-Neto et al. (2008) used the term "checkerboard" for any binary matrix in which every cell containing the value of 1 has the same value in all diagonally neighboring cells and 0 in the remaining neighboring cells, for example:
We refer to this as *visual checkerboard pattern*, which has been used mostly for visualization purposes. Since its first appearance (Almeida-Neto et al. 2008), the visual checkerboard pattern has been used frequently as an example matrix for developing pattern analysis procedures (Podani and Schmera 2011, Ulrich and Gotelli 2013). However, many analytical tools in metacommunity ecology are insensitive to the order of sites and species in the matrix and therefore are unable to detect and test this unique property. A noted example is the compartment pattern, which differs from the visual checkerboard pattern only in the ordering of sites and species (Podani and Schmera 2011, Ulrich and Gotelli 2013).

Recently, Connor et al. (2013) have contributed by two very important points to the proper interpretation of the checkerboard pattern in situations where ordering is fixed by geographical constraints. First, they argued that the checkerboard metaphor reflects the mutually exclusive distribution of two species on a set of islands, where only one of the two species (denoted by the letters A or B) is present on a single island, while the position of the letters corresponds to the explicit geographic position of the islands. Consequently, a set of 16 islands arranged spatially in a regular 4-by-4 grid and occupied by two species (A and B) shown below indicates a checkerboard pattern:

\[
\begin{bmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1
\end{bmatrix}
\]

Connor et al. (2013) argued also that the metaphor does not intend to visualize the presence-absence of the species in a sites-by-species incidence matrix (for example, \( M_2 \) as given above) and thus this incidence matrix, in disagreement with other studies (Almeida-Neto et al. 2008, Podani and Schmera 2011, Ulrich and Gotelli 2013), should not be regarded as a "real" spatial checkerboard pattern.

As a second contribution, Connor et al. (2013) pointed out that the idea of Diamond's checkerboard distribution includes not only mutually exclusive island-by-island distribution of species pairs, but also the overlapping geographic ranges of the species. According to this argument and without any information on the distribution ranges of species, the lack of co-occurrence due to competition cannot be separated from lack of co-occurrence owing to non-overlapping ranges (spatial turnover). All of these suggest that presence-absence
matrices fail to address Diamond's (1975) original idea because a binary data matrix by itself has "no explicit geography" and provides no information on the distribution ranges of species. To clarify the situation, Connor et al. (2013) suggested the term true checkerboard pattern for a pair of species which never co-occur on the same island and the islands occupied by these two species are geographically alternating.

In sum, we will use the term checkerboard pattern when the binary matrix contains a large number of checkerboard species pairs. Visual checkerboard pattern refers to a binary matrix, in which zeros regularly alternate with 1-s over columns and rows. Finally, true checkerboard pattern refers to a pair of species which never co-occur on the same island and the islands occupied by these two species geographically alternate.

The EMS framework (Presley et al. 2010) refers to Diamond's definition (Diamond 1975), and thus to checkerboard pattern. Some studies using the EMS framework, however, apparently have to do with the visual checkerboard pattern (see Fig. 2 in Tonkin et al. 2017) to which this methodology does not apply.

Of the 41,503 4-by-4 binary matrices, the largest number of mutually exclusive species pairs (6) was observed when all sites contained only a unique species (Electronic Appendix 4). This suggests that if the checkerboard pattern is quantified by the number of checkerboard species pairs (as in the EMS framework), then sites with single and unique species will contribute the most to the checkerboard character. Unfortunately, ordering of such matrices is arbitrary in correspondence analysis (Electronic Appendix 3), and thus the application of CoA and exclusive species pairs within the same approach may not be optimal.

5. Coherence test

The first promise of the coherence test is that a high number of embedded absences (negative coherence) indicates checkerboard pattern (high number of checkerboard species pairs). The examination of 4-by-4 binary matrices shows that sites with single and unique species (matrices with the highest number of checkerboard species pairs) have no embedded absences. To get a deeper insight into this relationship, we plotted the number of embedded absences (quantifying negative coherence) against the number of checkerboard species pairs (quantifying checkerboard pattern) for the 41,503 4-by-4 incidence matrices (Fig. 2). As seen, a high number of embedded absences is not necessarily associated with low number of checkerboard species pairs and vice versa. Moreover, 4-by-4 matrices with the highest number of checkerboard species pairs (sites with single and unique species) contain no embedded absences, while matrices with the largest number of embedded absences (not shown) do not contain checkerboard species pairs.

The analyses of 10-by-10 matrices revealed that 303 matrices showed a significantly higher number of embedded absences (negative coherence) than expected and thus exhibited checkerboard pattern. The null model test detected 66 matrices with significantly large number of checkerboard species pairs, of which 15 matrices were selected also by the coherence test. This suggests that 4.24% (Jaccard similarity = 15/354) is the agreement.
between null model tests using the number of embedded absences (coherence test of the EMS framework) and the number of checkerboard species pairs. Consequently, the number of embedded absences does not necessarily indicate checkerboard pattern and thus cannot be used alone as its indicator.

6. Definition of turnover and its test

The EMS framework assumes that turnover and nestedness are opposing patterns (Leibold and Mikkelson 2002, p. 239). This means that if we observe low nestedness then turnover should be high and vice versa. In an extreme situation, zero nestedness should yield maximum turnover. To illuminate the relationship between turnover and nestedness, let us examine the following example (rows are sites and species are columns):

\[
\begin{array}{cccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{array}
\]

Here turnover (number of times one species replaces another between two sites: in a two-site situation it means \(b \times c\), where \(b\) is the number of species present only in the first, while \(c\) is the number of species present only in the second site, Presley et al. 2010) equals to 9.

Note that in a 2-by-10 incidence matrix the maximum possible value of turnover is 25 (obtained when \(b = 5\) and \(c = 5\)). Consequently, this turnover value is 64% lower than the theoretical maximum. Although several nestedness indices do exist (the EMS framework does not define any measure), all of them agree that if two sites do not share any species then nestedness should be zero (Ulrich et al. 2009, Podani and Schmera 2012). Thus, this example demonstrates a situation with relatively low turnover and zero nestedness.

Before discussing the relationship between turnover and nestedness, we should note that the turnover definition applied by Presley et al. (2010) and used here differs from many existing definitions of turnover (see Tuomisto 2010, Anderson et al. 2011, Gotelli and Ulrich 2012). We by no means state that this measure does not quantify the concept of turnover, but emphasize its uniqueness in community ecology and therefore further studies are needed to clarify its performance.

We examined the relationship between turnover and nestedness in two-site situations using both random parameter approaches (Fig. 3). All combinations of nestedness measures and random parameter approaches showed that high turnover associates mostly with low nestedness. However, low turnover values can be associated with a wide range of nestedness values, suggesting that turnover and nestedness are not necessarily opposing patterns. Although under specific conditions we can assume that high turnover predicts low nestedness, this is not always the case (see Random parameter approach 2). On the other hand, low turnover does not necessarily predict high nestedness.
We studied the relationship between turnover and nestedness using all possible 4-by-4 matrices. When nestedness was quantified by the relativized nestedness measure, we found a relatively strong negative relationship between the two variables (r = -0.860, Fig. 4).

Although low turnover values indicate high relative nestedness, high turnover does not necessarily indicate low relativized nestedness. When nestedness was quantified by NODF\textsubscript{max}, the negative relationship with turnover was lower than with relativized nestedness (r = -0.641, Fig 5), and a low turnover value may be indicative of low nestedness.

We used null model tests on 10-by-10 matrices to examine whether significantly high turnover is associated with significantly low nestedness, and whether significantly low turnover with high nestedness. Null model tests indicated 421 matrices with high turnover and 433 matrices with low nestedness when the latter is measured by the relativized measure. The agreement between the two assessments was 23.59% (i.e. Jaccard similarity = 163/691). When nestedness was quantified by NODF\textsubscript{max}, 296 matrices showed low nestedness. The agreement between high turnover and nestedness (NODF\textsubscript{max}) was only 5.60% (Jaccard similarity = 38/679). None of our null model tests indicated significantly low turnover, high relativized nestedness and high NODF\textsubscript{max}. These results suggest that high turnover is not necessarily associated with low nestedness in the null model tests. In agreement with these findings, Ulrich and Gotelli (2013) and Ulrich et al. (2017) have already published similar results.

7. Boundary clumping test

Our starting point is that Clementsian, Gleasonian and evenly spaced patterns can only be interpreted along an actual (real) environmental gradient (Clements, 1916, Gleason 1926, Tilman 1982, Shipley and Keddy 1987). We argue that "within matrix data structure" revealed by CoA is inadequate for this purpose. Although user defined site ordering might allow testing real environmental gradients, difficulties associated with the coherence test (number of embedded absences is influenced by species ordering, unclear interpretation of coherence) strongly limit this possibility. All of these suggest that no boundary clumping test can be performed within the context of the EMS framework.

8. Series of tests

The EMS framework includes a well-defined sequence of three tests (coherence, turnover and boundary clumping). If we assume that these tests indicate orthogonal and independent properties of matrices then all these tests could be performed independently from the results of tests made earlier in the series. Although Leibold and Mikkelsen (2002, p. 239) argue that "turnover and clumping are most meaningful in the context of reasonably coherent ranges", the application of a series of tests has strong consequences. First, a test performed in a series or alone has different statistical and ecological meanings. For instance, the turnover test alone indicates the concept where species are replaced by one another, while within the EMS framework it indicates the existence of replacement in positively
coherent metacommunity patterns. Second, some patterns should be more frequently
detected than others because earlier tests restrict the possible outputs (i.e. turnover test
can only be performed if coherence is high and cannot be performed when coherence is
random or negative). Consequently, if we perform a two-tailed statistical test with 5%
significance level, then about 2.5% of the examined random matrices should show
checkerboard pattern, 95% random pattern, 0.0625% (2.5% × 2.5%) nested pattern, 2.375%
(2.5% × 95%) quasi pattern, 0.0015625% (2.5% × 2.5% × 2.5%) evenly spaced and
Clementsian pattern, and 0.05937% (2.5% × 2.5% × 95%) Gleasonian pattern in a series of
tests suggested by the EMS framework.

If we assume that the tests are not orthogonal and not independent then a series of tests
may have a clear ecological meaning. In this case, however, the output of an earlier test
should predict the output of a later test, or the ecological meaning of the output of an
earlier test suggests that there is no need for further ecological information. The argument
of Leibold and Mikkelson (2002, p. 239) that "turnover and clumping are most meaningful in
the context of reasonably coherent ranges" suggests that the EMS framework considers
coherence as primary feature of metacommunity organization. However, we see no strong
theoretical support for the priority of coherence in metacommunity structuring.

9. Noise test

The noise test showed that the reliability of the method to identify idealized structures is
different at the same level of noise (Fig. 6, Electronic Appendix 5). For example, the noise
level at which the idealized structure was detected at least with 50% reliability was below
only 5%, 5%, 10% and 20% for the evenly spaced, Gleasonian, nested and Clementsian
gradients, respectively. These results may explain why Clementsian (or quasi-Clementsian)
and nested patterns are identified most frequently in actual data sets and why Gleasonian
pattern is identified relatively infrequently. Further, identification of the evenly spaced
gradient was not possible in all cases even at zero noise. As low as 5% noise in the data
already yielded that the EMS method identified either Clementsian or Gleasonian structure.
Above 15% (nested), 20% (evenly spaced), 25% (Gleasonian) or 30% (Clementsian) noise
levels, the EMS method identified random metacommunity structure in more than 50% of
cases, which further proves the sensitivity of the method to the characterization of idealized
structures at different noise levels.

10. The importance of environmental factors in shaping metacommunity patterns: current
practice

We found fifty papers citing Presley et al. (2010). Twenty-six papers, each of them ordering
sites and species by correspondence analysis, applied the EMS framework. We found that
only four papers out of these 26 attempted to provide information on the variance explained
by CoA in some way. Three of these 4 papers provided eigenvalues of the first two axes of
CoA. However, these two eigenvalues by themselves do not quantify the percentage of
variance they explained. There was a single paper of the 26 (3.8%) that provided information on the variance explained by CoA. This paper showed also that the first axis of CoA accounted for 17.7% to 24.0% community variation depending on the metacommunity studied and that environmental variables explained 47.9% to 77.4% variance of the first axis of CoA (Erős et al. 2014). Although this single study does not allow general conclusions to be drawn, it implies that metacommunity patterns are detected based on a limited amount of community variation and that this limited community variation is correlated only at an intermediate-level with multiple environmental variables. In most studies, the variance explained by CoA is not given at all, only the relationship between the site position in ordination axis and environmental variables (80.8%). In some cases, CoA is used in the EMS framework, but environmental variables are related to canonical correspondence analysis by the reasoning that canonical correspondence analysis is related to CoA (de la Sancha et al. 2014), or can be regarded as a constrained extension of CoA (Heino et al. 2015). These studies ignore the fact that CoA and canonical correspondence analysis need not result in the same ordering of sites along any ordination axis. Overall, our literature survey shows that essential information, including at least some hints on the reliability of the identification of idealized metacommunity structures remains completely hidden in almost all studies which used the EMS framework.

11. Conclusions

Our theoretical and statistical considerations show that the EMS framework has to be used with caution for the identification of idealized metacommunity patterns. While it is appealing to identify the best-fit metacommunity structure under a single analytical framework, the reliability of the test to distinguish among the idealized structures is strongly case dependent.

We showed that although user-defined site-ordering allows testing the response of community to an actual environmental gradient, its application is problematic due to the dependence of the coherence test upon the order of species. Unfortunately, this dependence strongly limits the performance of the EMS framework in testing the response of communities to real environmental gradients. Even if CoA is used for the ordering of sites, the EMS framework is relatively unreliable for separating evenly spaced, Gleasonian and Clementsian patterns. Our results demonstrate that the coherence test is the most critical step of the EMS framework. We found that it is not necessarily adequate for separating checkerboard pattern and its ecological meaning is not clearly defined. Our observations are in strong agreement with the findings of Gotelli and Ulrich (2012) in that the turnover test is not necessarily adequate for detecting a nested pattern.

We concluded that the application of a series of tests requires further considerations and that the detection of some idealized patterns is prone to type II error. Our literature survey clearly indicated that the documentation of the results of the EMS framework analysis is insufficient and thus information is extremely limited on the amount of community variation used for detecting idealized metacommunity patterns and also on the relationship between
this variation and environmental drivers. These findings call for reconsidering the analytical steps of the EMS framework, and for careful interpretation of its results.

Acknowledgements

This work was supported by the OTKA K104279 and the GINOP 2.3.3-15-2016-00019 grants.

References


Warnes, GR et al. 2015. gtools: Various R programming tools. R package version 3.5.0. https://CRAN.R-project.org/package=gtools

FIGURES
Fig. 1: Diagrammatic representation of the Elements of Metacommunity Structure (EMS) framework following Leibold and Mikkelson (2002) and Presley et al. (2010).
Fig. 2: The relationship of the number of embedded absences and the number of checkerboard species pairs when incidence matrices with 4 sites and 4 species were examined.
Fig. 3: The relationship between turnover (horizontal axes) and nestedness (vertical axes) when pairs of sites were examined. Upper subfigures show when nestedness was quantified as $N_{rel}$, while lower subfigures show when nestedness was quantified as $NODF_{max}$. Left subfigures show the results of the Random parameter approach 1, while right subfigures those of the Random parameter approach 2.
Fig. 4: The relationship between turnover and nestedness ($N_{\text{rel}}$) when incidence matrices with 4 sites and 4 species were examined.
Fig. 5: Relationship between turnover and nestedness (NODF_{max}) when incidence matrices with 4 sites and 4 species were examined.
Fig. 6: Bar plot showing the frequency of idealized metacommunity patterns (vertical axis) detected by the Elements of Metacommunity Structure (EMS) framework when evenly spaced pattern was exposed to increasing noise (horizontal axis).
THE MANUSCRIPT CONTAINS ALSO THE FOLLOWING ELECTRONIC APPENDICES

Electronic Appendix 1: Idealized metacommunity patterns used in the noise test.

Electronic Appendix 2: R script used for calculating indices.

Electronic Appendix 3: The ordering of an incidence matrix with sites with single and unique species. R scripts.

Electronic Appendix 4: Visualization of 4-by-4 incidence matrices with the largest number of checkerboard species pairs.

Electronic Appendix 5: The results of the noise tests on nested, Gleasonian and Clementsian patterns.