

# metacom: an R package for the analysis of metacommunity structure

Tad Dallas

*T. Dallas (tdallas@uga.edu), Odum School of Ecology, Univ. of Georgia, Athens, GA 30606-4288, USA.*

Metacommunity theory is an extension of metapopulation theory with the goal of understanding how ecological communities vary through space and time. One off-shoot of metacommunity theory deals with understanding how community structure varies along biotic or environmental gradients. The Elements of Metacommunity Structure framework is a three-tiered analysis of metacommunity structure that enables the user to identify metacommunity properties that arise in site-by-species incidence matrices. These properties can then be related to underlying variables that influence species distributions. The EMS framework is now implemented in metacom, an open source R package that allows for the analysis and plotting of metacommunities.

Metacommunity theory is an attempt to identify the mechanisms underlying species distributions as they relate to community structure, a central focus of community ecology (Holyoak et al. 2005, Soininen et al. 2007, De Meester 2011). The framework proposed by Leibold and Mikkelsen (2002), and later expanded by Presley (2010), offers an approach to determine metacommunity ‘pattern’, and the environmental or spatial variables underlying this structure. The Elements of Metacommunity Structure (EMS) framework is based on the evaluation of three metrics calculated from a presence–absence interaction matrix, with sites as rows and species occurrences at sites as columns. To determine the metacommunity structure, the interaction matrix is ordinated via reciprocal averaging, a method that groups interactions along the matrix diagonal in a fashion that results in species with similar ranges and sites with similar species compositions to be placed together (Gauch 1982). The weights (or ‘scores’) obtained from ordination then represent a structuring gradient, which can then be related to environmental or spatial variables (Presley and Willig 2010). The three metrics of the EMS framework are coherence, turnover, and boundary clumping (Leibold and Mikkelsen 2002), which can be visualized in three-dimensional space (Fig. 1).

Coherence is measured by counting the number of embedded absences in the ordinated matrix. Statistical significance of coherence is determined by comparing the observed absences to the number of embedded absences observed in many randomized null matrices using a z-test. Many different randomization algorithms exist (Gotelli and Graves 1996), but the best performing algorithms typically hold row (site) totals constant, and

either fill occurrences among sites probabilistically based on the marginal column totals (fixed-proportional null) or by maintaining column sums (fixed-fixed null) (Ulrich and Gotelli 2007). To qualify for further analysis, community matrices must contain statistically significantly fewer embedded absences than expected under null models. Negative community matrix coherence (more embedded absences than expected under null model) indicates a checkerboard pattern, in which species occurrences among sites are mutually exclusive. Positive coherence indicates that species ranges have fewer embedded absences than expected under null model simulations, which allows for subsequent analysis of turnover and boundary clumping.

Turnover is quantified by calculating the number of times one species replaced another between sites, after species distributions are made completely coherent. Therefore, this does not include gaps within species ranges, instead considering only instances where species replace each other from site to site. This metric is compared to the distribution of turnover values obtained through many ordinated null simulations using a z-test. Low turnover is indicative of nested subsets, whereas significantly high turnover is associated with an environmentally-driven community structure (Leibold and Mikkelsen 2002). Boundary clumping is quantified using the Morisita’s index, a measure of the dispersion of species occurrences among sites (Morisita 1971). A Morisita’s index ( $I$ ) of one indicates boundaries are not clumped, while values greater than one ( $I > 1.0$ ) or less than one ( $I < 1.0$ ) indicate clumped or hyperdispersed boundaries, respectively. Statistical significance of the Morisita’s index is determined using a chi-squared test.

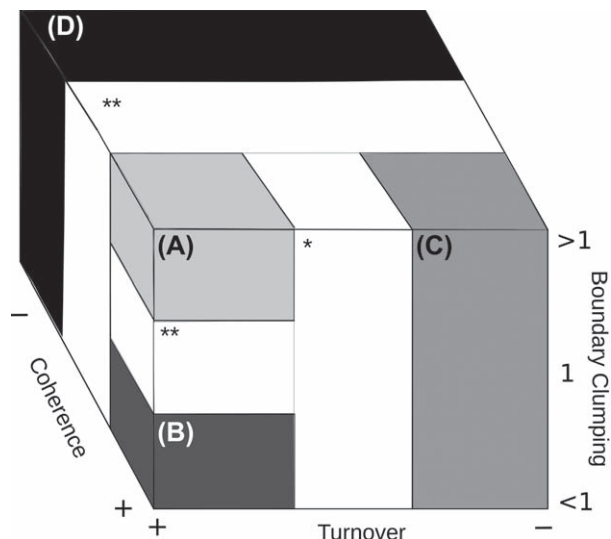


Figure 1. A three-dimensional space formed by the three attributes used to identify metacommunity structure (coherence, turnover and boundary clumping). Metacommunity patterns exist as regions within the space, and specific results can be plotted as points. This allows for a more continuous view of the patterns of metacommunity structure (A: Clementsian, B: evenly spaced gradients, C: nested subsets, D: checkerboard). Metacommunity structures unable to be distinguished by the current EMS framework (i.e. Gleasonian and random structures) are indicated by double asterisks (\*\*). A single asterisk (\*) indicates the 3-dimensional space in which the quasi-structures proposed by Presley et al. 2010 would occupy. However, we do not consider these structures in the current analysis, and treat metacommunities within this area as areas in which pattern is unknown.

It is important to note that my interpretation of the EMS analysis differs slightly from that of Leibold and Mikkelsen (2002) and Presley et al. (2010). Some studies have interpreted non-significant results to be evidence that the null hypothesis ( $H_0$ ) is true. However, a non-significant result ( $p > \text{significance level } \alpha$ ) does not mean that  $H_0$  is 'accepted'. Instead, a non-significant result means that the data do not provide enough evidence to determine if  $H_0$  is true or false (see Wackerly et al. 2008 or Gelman 2013 for further discussion). To interpret non-significant results as indicators (e.g. non-significant coherence used to indicate 'random' metacommunity structure) is to commit the error of accepting the null hypothesis. In light of this, I do not consider the EMS analysis able to distinguish two structures (Fig. 1) (i.e. 'random' and 'Gleasonian') that were part of the original formulation of Leibold and Mikkelsen (2002). This is a subtle point, and the interpretation of the analysis is left to the package user.

However, the EMS framework remains one of the best methods for the determining metacommunity structure, and subsequently relating structure to underlying biotic and abiotic gradients (see Ulrich and Gotelli 2013 for both a critique and validation of components of the EMS framework). Currently, the *vegan* package (Oksanen et al. 2012) in the R statistical programming environment (R Core Team) contains many functions for the analysis of community-level data. However, *vegan* does not include

functions necessary for the EMS analysis and the only available implementation of the EMS framework is Chris Higgins's MatLab code (available at <http://faculty.tarleton.edu/higgins/metacommunity-structure.html>). In the present software note, I present R package *metacom* ver. 1.3 (Dallas 2013), a simplified expansion of Higgins's Matlab code, providing a freely-available platform to evaluate the three statistics of the EMS framework and to explore the biotic and abiotic gradients along which species distributions are structured.

## Basic functions

The *metacom* package is composed of functions that calculate the 3 metrics of the EMS framework independently, the *Metacommunity()* function that performs the analysis of all 3 metrics using the same null matrices (reducing computation time), and a plotting function.

The functions used to determine the coherence and turnover of an interaction matrix (*Coherence()* and *Turnover()*, respectively) are based on null model simulations. Null community matrices are produced with the *NullMaker()* function, which combines methods of community simulation from functions *commsimulator()* and *permatfull()* in the R package *vegan* (Oksanen et al. 2012) with an additional argument that either allows row (site) or column (species) totals to contain no occurrences. Biologically, this is an important consideration, as the null expectation of a site containing no species or a species occurring in none of observed sites may be unrealistic, depending on the endemism of species and the suitability of sites to occupancy.

The *Imagine()* function is a flexible plotting function for interaction matrices that produces a publication-quality image plot of an interaction matrix. The function includes arguments concerning whether or not to ordinate the community matrix before plotting, whether to make species ranges completely coherent (a common way to plot interaction matrices (Presley and Willig 2010, Willig et al. 2011)), and other cosmetic alterations such as changing color or whether to print the names of species and sites.

## Example metacommunity structure analysis

Leibold and Mikkelsen (2002) created artificial matrices to test their proposed EMS framework, many of which are provided in the list object 'TestMatrices'. This is accessible in R by typing 'data(TestMatrices)' after loading the *metacom* package. The help file (*help(TestMatrices)* or *?TestMatrices*) provides details on the size and origin of the matrices. To demonstrate the utility of the EMS analysis, I provide an example of metacommunity analysis and determination of important variables associated with the structuring gradient utilizing a well-studied dataset available within the *vegan* R package (Oksanen et al. 2012). The data are from a floristic and environmental survey of 44 species among 24 sites in eastern Fennoscandia (Väre et al. 1995).

```
#loads the vegan package and imports the interaction
matrix ('varespec') and site-specific #environmental
variables ('varechem').
```

```

> require(vegan)
> data(varespec, varechem)

#determines the 3 metrics of the EMS analysis
# Coherence

> Coherence(comm = varespec, scores = 1, method =
'fixedfixed', sims = 1000, order = TRUE, allow.
empty = FALSE)

# Turnover

> Turnover(comm = varespec, scores = 1, method =
'fixedfixed', sims = 1000, order = TRUE, allow.
empty = FALSE)

# Boundary clumping

> BoundaryClump(comm = varespec, order = TRUE,
scores = 1)

#...or all three metrics can be computed at once using the
Metacommunity() function

> Metacommunity(comm = varespec, scores = 1,
method = 'fixedfixed', sims = 1000, order = TRUE,
allow.empty = FALSE)

```

The `Metacommunity()` function produces a list element composed of the ordinated empirical matrix ('Comm'), and the results of the analysis of coherence ('Coherence'), turnover ('Turnover') and boundary clumping ('Boundary'). The community matrix can then be visualized utilizing the `Imagine()` function, plotted with the 'varespec' data in Fig. 2. The vegetation metacommunity contained fewer embedded absences than expected ( $z = 3.65$ ,  $p < 0.001$ ), more species replacements than expected ( $z = -3.37$ ,  $p < 0.0001$ ), and species range boundaries were not

significantly different from the null expectation (Morisita's index = 1.24,  $p = 0.151$ ). Based on these results, the vegetation metacommunity would traditionally be considered 'Gleasonian' (but see discussion above on inference based on non-significance). The results of the analysis of coherence and turnover suggest that the 'varespec' metacommunity responds to a structuring gradient. Along this gradient, species ranges contain fewer embedded absences and species tend to replace each other more often than expected under null model simulations that maintain both site and species totals.

### Environmental variables associated with structuring gradient

The site scores obtained from reciprocal averaging can be related to environmental variables (site and species scores can be obtained using `OrderMatrix()`) to provide evidence of the importance of environmental variables in structuring species distributions. This may be performed using a variety of methods such as regression methods (Elith et al. 2008, Meynard et al. 2013), correspondence analyses (López-González et al. 2012), or permutation tests (as implemented in `vegan`). For simplicity, I use non-parametric Spearman's correlations to investigate the associations between the site scores obtained from reciprocal averaging and a small subset of environmental variables, including nitrogen, phosphorous, pH and depth of the humus layer. Nitrogen ( $R_s = 0.13$ ,  $p = 0.538$ ) and phosphorous ( $R_s = 0.332$ ,  $p = 0.113$ ) were both uncorrelated to site scores, while pH ( $R_s = -0.465$ ,  $p = 0.022$ ) and the depth of the humus layer ( $R_s = 0.747$ ,  $p < 0.0001$ ) were both significantly correlated, suggesting that pH and humus layer

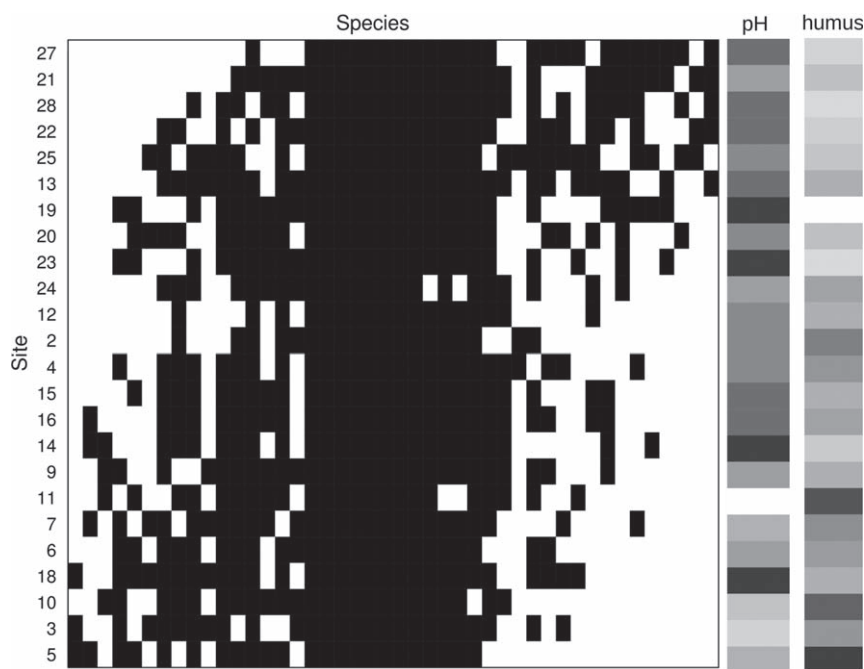


Figure 2. A visualization of the 'varespec' data, with numbered sites as rows and species as columns. Black rectangles indicate a species occurrence at a site. pH and humus depth, two variables associated with the structuring gradient species distributions respond to, are show on the right with darker colors indicating lower (more acidic) pH and larger humus depth.

depth (Fig. 2) are associated with plant community composition, while nitrogen and phosphorous are potentially less important in determining species composition at a site.

## Conclusion

The EMS framework, as it is implemented in the metacom R package, provides a method for examining metacommunity structure from pattern (i.e. distinguishing between the four metacommunity structures) to process, through the identification of variables associated with the axis scores representative of a gradient to which communities respond. In the vegetation ('varespec') metacommunity examined, I determined that the metacommunity followed what is traditionally referred to as a 'Gleasonian' pattern, and was structured along a gradient that was related to humus depth and pH.

## Package installation

The current implementation (ver. 1.3) of the metacom package is freely-available (license GPL-2) from the CRAN repository <http://cran.r-project.org/web/packages/metacom/index.html>, and can be installed by downloading from CRAN or within the R environment using the command `install.packages('metacom', dependencies = TRUE)`. Further, the project is hosted on Github (url: <https://github.com/taddallas/metacom>), which allows users to openly contribute to the package.

To cite metacom or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for 'version 0'.

Dallas, T. 2014. metacom: an R package for the analysis of metacommunity structure. – *Ecography* 37: 000–000 (ver. 0).

## References

Dallas, T. 2013. metacom: Analysis of the 'elements of metacommunity structure'. – R package ver. 1.3, <http://CRAN.R-project.org/package=metacom>.

- De Meester, L. 2011. A metacommunity perspective on the phylo- and biogeography of small organisms. – In: De Meester, L. (ed.), *Biogeography of microscopic organisms: is everything small everywhere?* Cambridge Univ. Press, pp. 324–334.
- Elith, J. et al. 2008. A working guide to boosted regression trees. – *J. Anim. Ecol.* 77: 802–813.
- Gauch, H. G. 1982. *Multivariate analysis in community ecology*. – Cambridge Univ. Press.
- Gelman, A. 2013. Commentary: p values and statistical practice. – *Epidemiology* 24: 69–72.
- Gotelli, N. J. and Graves, G. R. 1996. *Null models in ecology*. – Smithsonian Inst. Press.
- Holyoak, M. et al. 2005. *Metacommunities: spatial dynamics and ecological communities*. – Univ. of Chicago Press.
- Leibold, M. A. and Mikkelsen, G. M. 2002. Coherence, species turnover, and boundary clumping: elements of metacommunity structure. – *Oikos* 97: 237–250.
- López-González, C. et al. 2012. Metacommunity analysis of Mexican bats: environmentally mediated structure in an area of high geographic and environmental complexity. – *J. Biogeogr.* 39: 177–192.
- Meynard, C. N. et al. 2013. Disentangling the drivers of metacommunity structure across spatial scales. – *J. Biogeogr.* 40: 1560–1571.
- Morisita, M. 1971. Composition of the I-index. – *Res. Popul. Ecol.* 13: 1–27.
- Oksanen, J. et al. 2012. *vegan: community ecology package*. – R package ver. 2.0-4, <http://CRAN.R-project.org/package=vegan>.
- Presley, S. J. and Willig, M. R. 2010. Bat metacommunity structure on Caribbean islands and the role of endemics. – *Global Ecol. Biogeogr.* 19: 185–199.
- Presley, S. J. et al. 2010. A comprehensive framework for the evaluation of metacommunity structure. – *Oikos* 119: 908–917.
- Soininen, J. et al. 2007. Neutrality, niches, and determinants of plankton metacommunity structure across boreal wetland ponds. – *Ecoscience* 14: 146–154.
- 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.
- Väre, H. et al. 1995. Effects of reindeer grazing on understorey vegetation in dry *Pinus sylvestris* forests. – *J. Veg. Sci.* 6: 523–530.
- Wackerly, D. D. et al. 2008. *Mathematical statistics with applications*. – Thomson Learning, pp. 488–512.
- Willig, M. R. et al. 2011. Tropical metacommunities along elevational gradients: effects of forest type and other environmental factors. – *Oikos* 120: 1497–1508.