# LETTER

# Integrating environmental and spatial processes in ecological community dynamics

**Karl Cottenie**

*National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, 735 State Street Suite 300, Santa Barbara, CA 930101, USA*

*Correspondence and Present address: Department of Integrative Biology, University of Guelph Guelph, ON, Canada N1G 2W1. E-mail: cottenie@uoguelph.ca*

## Abstract

The processes controlling the abundances of species across multiple sites form the cornerstone of modern ecology. In these metacommunities, the relative importance of local environmental and regional spatial processes is currently hotly debated, especially in terms of the validity of neutral model. I collected 158 published data sets with information on community structure, environmental and spatial variables. I showed that approximately 50% of the variation in community composition is explained by both environmental and spatial variables. The majority of the data sets were structured by species-sorting dynamics (SS), followed by a combination of SS and mass-effect dynamics. While neutral processes were the only structuring process in 8% of the collected natural communities, disregarding neutral dispersal processes would result in missing important patterns in 37% of the studied communities. Moreover, metacommunity characteristics such as dispersal type, habitat type and spatial scale predicted part of the detected variation in metacommunity structure.

## Keywords

Body size, dispersal, environmental heterogeneity, mass effects, meta-analysis, meta-community theory, neutral theory, patch dynamics, species sorting, variation decomposition.

## INTRODUCTION

Metacommunities are local communities connected by the dispersal of multiple species (Hubbell 2001). While the neutral metacommunity model presented by Hubbell was important in stressing the importance of spatial processes in structuring community dynamics, it asserted that species–environment relations were not important in establishing some well-known ecological relationships such as the distribution of abundance or the species–area law. A recent synthesis of metacommunity dynamics by Leibold *et al.* (2004) integrated both environmental heterogeneity with associated niche explanations and dispersal processes. Depending on the relative importance of both processes, they suggested four types of metacommunities: the species-sorting (SS), the mass-effect (ME), the neutral model (NM) and patch dynamics (PD) type. The NM assumes ecological equivalence of species with dispersal-limited communities, resulting in strong spatial structures (Hubbell 2001; Bell 2001). Sufficient dispersal in a heterogeneous environment with associated niche differences will promote SS along resource gradients (Chase & Leibold 2003). Increasing dispersal in heterogeneous environments will result in local populations being quantitatively affected by dispersal through source–sink relations, independent of resource gradients (ME; Holt 1993; Mouquet & Loreau 2002, 2003). Patch dynamics are another form of niche differentiation that does not revolve around habitat. Instead, trade-offs lead to spatiotemporal niches (Hastings 1980; Tilman 1994).

Although experiments have illustrated several aspects of metacommunities (Gonzalez *et al.* 1998; Holyoak 2000; Forbes & Chase 2002), comprehensive tests with observational data are limited in information used and extent of the tests. Only species abundance distributions are checked for zero-sum multinomial distribution in the context of NM for a limited number of data sets (Condit *et al.* 2002; McGill 2003; Etienne & Olff 2005). However, no information on environmental or dispersal characteristic of the studied system are used in this method, while conceptually these processes distinguish the different metacommunity types. This clearly illustrates the limited power of using species abundance distributions.

In a recent review of neutral theory, Chave (2004) advocated the use of databases and well-designed statistical methods to test neutral theory, and by extension the other metacommunity models. Recently, a few studies incorporated relevant environmental and spatial information (Duivenvoorden *et al.* 2002; Cottenie *et al.* 2003; Tuomisto *et al.* 2003; Gilbert & Lechowicz 2004) using powerful multivariate methods that can actually detect the effects of these processes. However, the link to the different metacommunity types is not made explicit in these studies.

This paper is the first to evaluate NM, SS and ME types from many different natural metacommunities within a unified framework using standardized statistical methods. This is a very powerful and encompassing test as (i) it includes variables related to the processes actually structuring metacommunities, (ii) it provides a test with specific predictions for three of the four different metacommunity types, (iii) it assesses the generality or distribution of the different metacommunity types in observational data sets by applying it to 158 different data sets, (iv) it provides a continuous metacommunity score that captures the dynamic aspect of the different metacommunity types (i.e. they occur in a continuum with transition zones between the different types), and (v) it provides a starting point for explaining the occurrence of the different metacommunity types in certain systems.

The first objective was to determine the variation in community composition explained by environmental and spatial variables. The second objective was to relate these variation components to metacommunity dynamics. The third objective was to determine the importance of local and dispersal processes to four characteristics of a metacommunity. This would provide potential causal mechanisms for explaining a variety of metacommunity models in nature. For instance, increase in body size is expected to decrease local population densities and decrease dispersal rates (Finlay 2002; Cohen *et al.* 2003). Increasing the spatial scale of a system results in both an increase in habitat heterogeneity and dispersal limitation between the different sites. This would be expected to increase SS and decrease ME dynamics. Habitats with different degrees of contiguity offer different contexts for dispersal limitation, for example, compare continuous marine environments with discrete lakes in an uninhabitable terrestrial matrix. Finally, organisms have qualitatively different dispersal types that will influence spatial scale.

## MATERIAL AND METHODS

### Data set requirements and acquisition

In order to be included in the analyses, a data set needed: (i) community composition as either absolute abundances,

relative abundances, or presence–absence of its constituent taxa; (ii) environmental variables relevant for structuring that particular community; (iii) spatial configuration of the sampled sites to be collected within one growing season. This ensured only direct dispersal interactions between sites, with no confounding in the data through multiple dispersal events by, for instance, lumping sites measured for several seasons or in different seasons.

Data were located by database and literature searches, and directly contacting other scientists. Authors of relevant studies were contacted directly for the raw data. Publication bias is probably not an important problem, as the majority of these data sets were not collected for testing metacommunity theory. However, all the studies did include environmental data, indicating that the authors expected at least some sort of environmental structuring. (For description of the different data sets, see Table S1 in Supplementary Material.)

### Variation decomposition

I determined the importance of local environmental and spatial processes (the different competing hypotheses) by decomposing the total variation in the community matrix into unique environmental and spatial components with corresponding *P*-values using (partial) redundancy analysis (Borcard *et al.* 1992; Legendre & Legendre 1998; Cottenie *et al.* 2003). This multivariate extension of linear regression with corresponding $R^2$ measures the amount of variation (computed as the percentage of the total variation in the community matrix) that can be attributed exclusively to one or the other set of explanatory environmental, **E**, or spatial, **S**, variables. The different components are: total explained variation [**E** + **S**], environmental variation [**E**], spatial variation [**S**], environmental variation without a spatial component [**E** | **S**], spatial variation without the environmental component [**S** | **E**]. The significance of these components was evaluated with a Monte Carlo permutation test (1999 new values under the null hypothesis). For the partial RDA analyses [**E** | **S**] and [**S** | **E**], residuals under the 'reduced' model were permuted; for the other RDA analyses ([**E** + **S**], [**E**], and [**S**]), residuals under the 'full' model were permuted (Legendre & Legendre 1998). I computed two other fractions: (i) the unexplained variation (1 − [**E** + **S**]) and (ii) the variation from correlations between environmental and spatial variables ([**E** with **S**] = [**E**] − [**E** | **S**] = [**S**] − [**S** | **E**]). It is important to stress that this approach is essentially correlative, with the different components and associated *P*-values considered as estimates for the observational evidence that a certain process (environmental, dispersal, or a combination of both) is important in determining community structure. As such, I did not adjust the *P*-values for multiple testing (Cottenie *et al.* 2003).

## Construction of initial data sets per study

Community data expressed in absolute abundances were fourth-root transformed to normalize these skewed density data (similar to the logarithmic transformation, without the need to add a constant to zero abundances), while presence/absence data and community data expressed as percentage relative to total density per site were not transformed. The initial set of environmental variables consisted of all environmental variables provided by the original authors. Spatial variables were provided in three basic forms: latitude–longitude coordinates, $X$–$Y$ coordinates with kilometre units based on digitizing map of the different sites with a scale bar, and for sites in a single river or estuary as distance along the river to the most downstream site. The latitude–longitude were converted into projected coordinates. For data sets with projected latitude–longitude or $X$–$Y$ coordinates, the initial set of spatial variables consisted of all third-order polynomials of these two coordinate variables. For data sets with only linear distances, the initial and final spatial data set consisted of this one distance variable.

## Standardization

For the meta-analysis, this (partial) RDA approach has the advantage that it provides a standard and comparable measure, percentage of explained variation, to describe the information present in studies with very different types of variables. However, similar to linear regression, increasing the number of explanatory variables results in an increase of explained variation, while increasing the number of sites will result in a decrease of explained variation. Thus, to remove these potential methodological influences, it was important to standardize across the different data sets by (i) using redundancy analysis for all the analyses, (ii) limiting the two groups of environmental and spatial third degree polynomial variables to the most parsimonious subset of $\leq 4$ significant variables each with a forward selection procedure (and this also eliminated overfitting and problems with co-linear variables), (iii) limiting the number of sites per data set to a maximum of 30. For data sets with information on $> 30$ sites, the parsimonious environmental and spatial variables were determined for all the sites; for 50 random subsets of 30 sites each, I computed all the different components and the associated $P$-values; the final variation components and $P$-values were obtained by averaging these 50 values.

## Metacommunity types

The significance structure of the variation components, based on a $\alpha$ level of 0.05, determined the metacommunity type for each data set (Table 1). In the NM/PD

**Table 1** Decision tree for relationship between significance structure and metacommunity types

| [**E**] | [**S**] | [**E** \| **S**] | [**S** \| **E**] | Metacommunity type |
|---|---|---|---|---|
| (not) sig. | (not) sig. | sig. | not sig. | SS |
| (not) sig. | (not) sig. | sig. | sig. | SS + ME |
| (not) sig. | (not) sig. | not sig. | sig. | NM/PD |
| (not) sig. | (not) sig. | not sig. | not sig. | Undetermined |
| not sig. | not sig. | not sig. | not sig. | No found |

Relationship between significance structure of the four important variation components and associated metacommunity types. The components are environment [**E**], space [**S**], environment independent of space [**E** \| **S**], and space independent of environment [**S** \| **E**]. sig., variation component explains a significant part of the variation in community structure; not sig., no significant part; (not) sig., either.

perspective, species and habitats are similar such that only dispersal and not environmental processes will relate to community structure. This will result in only spatial patterns in community structure. The PD type is dynamically different from the NM with species experiencing trade-offs between competitive ability and dispersal. However, because of the limitations of my data search with the restriction to time-limited spatial data sets, the used methodology cannot distinguish these two types. SS on the contrary, describes communities where resource gradients or patch types cause sufficiently strong differences in the local demography of species and the outcomes of local species interactions. Thus patch quality and dispersal jointly affect local community structure. Dispersal is important because it allows compositional changes to track changes in local environmental conditions. However, it will not result in spatial patterns independent of the environmental variables as its effect is completely confounded within the environmental processes. ME emphasize the role of spatial dynamics on community structure in a system where species can be rescued from local competitive exclusion in such communities where they are bad competitors by immigration from communities where they are good competitors. This will result in spatial patterns occurring independently of environmental patterns. In the ME context, environmental heterogeneity with corresponding SS dynamics is however present and important in structuring local communities, but its effect is modulated by an independent dispersal effect. While this is implicitly pointed out in Leibold *et al.* (2004), I indicated this explicitly as SS + ME. Theoretically, it is possible that dispersal is so high that it completely overrides SS dynamics, but conceptually such a metacommunity becomes an NM.

## Metacommunity axis

From the above discussion of the relation between variation components and metacommunity types, it is obvious that metacommunity types do not have clear cut boundaries, but form a continuum with transition zones possible between, for example, SS and NM dependent on the amount of environmental heterogeneity for a given amount of dispersal, and between SS and SS + ME dependent on the amount of dispersal for a given amount of environmental heterogeneity. So metacommunity type is probably a continuous characteristic (Leibold *et al.* 2004), while using the significance structure of the different variation components as in Table 1 ignores this important aspect. However, the variation components themselves cannot be directly used as large variation components might not be significant. Therefore, a synthetic metacommunity axis was constructed by combining both the significance structure and the corresponding metacommunity perspective with the different variation components using linear discriminant analysis (Legendre & Legendre 1998). More specifically, it tested whether the four variation components $[E]$, $[S]$, $[E|S]$, and $[S|E]$ can discriminate among the different metacommunity perspectives for data sets that corresponded to the SS, NM/PD, and SS + ME perspective with more than 25 sites ($n = 91$). Moreover, the linear discriminant analysis also estimates the parameters of a predictive linear model of the variation components for discriminating the three metacommunity perspectives. The first linear discriminant axis was termed the 'metacommunity axis', and the position of the different data sets along this metacommunity axis the 'metacommunity scores'.

## Important metacommunity variables

Average metacommunity scores were computed for systems with multiple data sets (number of independent systems is 52) to avoid pseudoreplication. These system metacommunity scores were then related to spatial scale and habitat type of the study site, and main dispersal mechanism and body size of the species. See Supplementary Material (Appendix S1) for the definitions of the variables and input values of the different systems. However, combining the different explanatory variables in one analysis using a linear model is impossible as several combinations of dispersal type are nested in habitat type, and the spatial scale and body size ranges differ in the different groups. Therefore, a regression tree (De'Ath & Fabricius 2000) was used to determine the important variables associated with metacommunity structure. This method can handle both continuous and categorical explanatory variables, is invariant to monotonic

transformations, and can deal with nonlinear relationships and high-order interactions (De'Ath & Fabricius 2000). The regression tree method explained variation in the metacommunity axis by repeatedly splitting the data into more homogeneous groups, using combinations of the selected important explanatory variables.

## Statistical package

All analyses were carried out in R (R Development Core Team 2004) with the following packages: MASS (Venables & Ripley 2002) for the discriminant analysis, mapproj (McIlroy 2004) for converting latitude–longitude coordinates into projected coordinates, vegan (Oksanen 2005) for the variation decomposition, and rpart (Therneau & Atkinson 1997) for the regression tree analysis.

## RESULTS

I collected 158 data sets spanning a broad range of taxa, habitats, spatial scales, body sizes and dispersal mechanisms (Table S1). Most data sets were from northern temperate regions in North America and Europe, and three from Australia and New Zealand (see Figure S2 in Supplementary Material).

Redundancy analysis decomposed the total variation in community structure into variation explained by environmental and spatial variables using a standardized protocol (Fig. 1); 48% of the total variation in community structure was explained by environmental and spatial variables together, 22% by the pure environmental fraction, 16% by the pure spatial component and 10% by the spatially structured environmental fraction.

Moreover, three metacommunity types were distinguishable based on the significance of the environmental and spatial components in the analyses (Table 1). Of these 158 data sets, 69 (44%) best fit the SS type, 46 (29%) a mixed SS + ME type and only 13 (or 8%) the NM/PD type, 19 data sets could not be uniquely associated with these three types, and 11 had no significant components.

As the metacommunity type is probably not a discrete character, metacommunity types of each data set were converted to a continuous score with discriminant analysis. The first linear discriminant axis significantly discriminated the SS, SS + ME and NM/PD types based on variation components. Figure 2 shows the distribution of the different metacommunity types along the axis obtained through linear discriminant analysis, with metacommunities structured by environmental processes having negative scores and metacommunities with an important independent spatial component more positive scores. This metacommunity axis can thus be synthesized as a environment vs. space axis, and the position along this axis indicates the
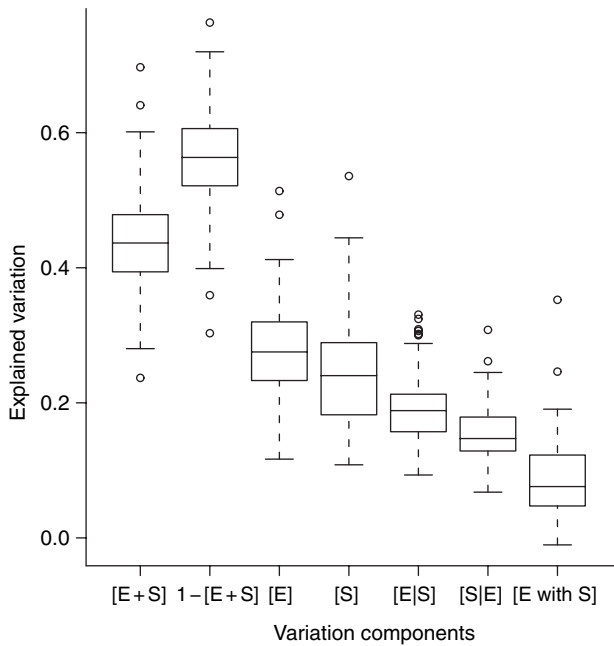
**Figure 1** Variation in community structure explained by environmental and spatial variables. The components are both environment and space [**E** + **S**], unexplained variation $1 - [\mathbf{E} + \mathbf{S}]$, environment [**E**], space [**S**], environment independent of space [**E**|**S**], space independent of environment [**S**|**E**], the spatially structured environmental effect [**E** with **S**]. The boxes represent the 25, 50 and 75 quartiles, the whiskers the minimum and maximum values, the outliers are values outside 1.5 times the interquartile range.

relative importance of local environmental and neutral dispersal processes.

I tested the relationship between the metacommunity scores (or the relative importance of local environmental and neutral dispersal processes) and the four potentially determining characters. All the univariate relationships were significant (see Figures S3–S6 in Supplementary Material). To study potential interactions, I constructed a regression tree by repeatedly splitting the scores into more homogeneous groups, using combinations of the selected important explanatory variables. The final model (Fig. 3) explained 40% of the variation in the metacommunity axis. The first split associated predominantly SS dynamics to systems with organisms displaying passive dispersal modes (17 data sets). The second split divided the other dispersal types into lake and marine habitats, with a mixture of SS and SS + ME (nine data sets), from estuary, stream and terrestrial habitats. This last group was then subdivided into metacommunities with a spatial scale smaller than 948 km, with SS + ME (19 data sets), and at spatial scales bigger than 948 with predominantly spatial processes (seven data sets).
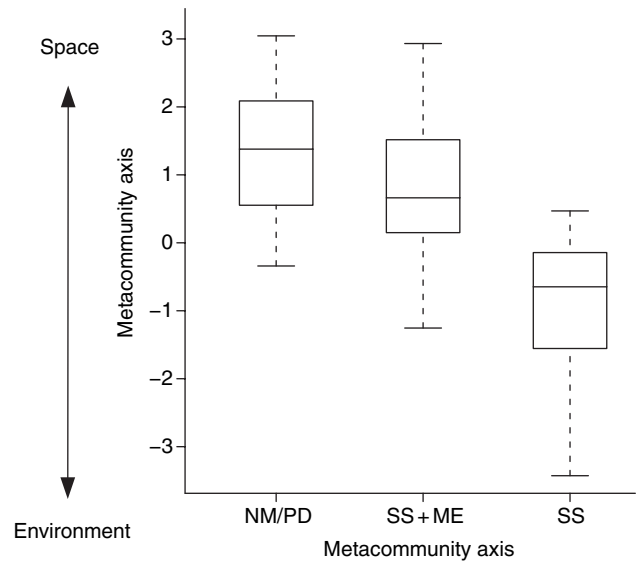


**Figure 2** Distribution of metacommunity types along the first linear discriminant axis. This axis was highly significant (Pillai's trace = 0.67, $F$-value = 10.71, $P$-value < 0.001), and explained 73% of the differences between the three metacommunity perspectives. The boxes represent the 25, 50 and 75 quartiles, the whiskers the minimum and maximum values excluding the outliers, values outside 1.5 times the interquartile range. The SS data sets are the mainly negative values, the NM/PD data sets are the positive values, and the SS + ME are situated around zero.
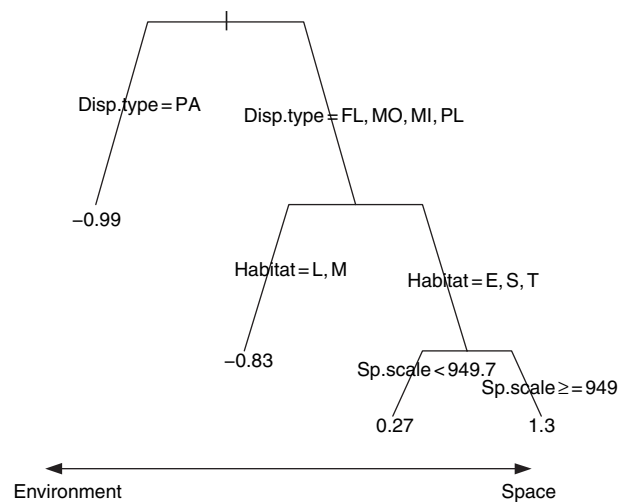


**Figure 3** Regression tree results. Final regression tree results showing the successive splitting and associated splitting criteria, together with the final nodes and the average metacommunity score per final node. To relate those metacommunity scores with metacommunity types, drop scores to the metacommunity axis in Fig. 2. sp.scale, spatial scale; the different habitat codes are: E, estuary; S, stream; T, terrestrial; L, lake; M, marine habitats. disp.type, dispersal type and the codes are: FL, flying; MO, mobility; PL, pelagic larvae; PA, passive; MI, mixed dispersal types.

## DISCUSSION

This meta-analysis of metacommunities, the first to study metacommunity dynamics with a standard and synthetic method using a broad spectrum of data sets, has several implications. The results form a bench mark study into variation in community composition that is currently explained by environmental, spatial and interactive components, using standardized methods. There is a rich literature on the association between environmental variables and community composition, but comparing different studies was hindered by different statistical methods used, and varying numbers of sites, environmental and spatial variables included. To make meaningful comparisons across different systems, these methodological aspects should be standardized. Moreover, recent theoretical advancements stressed the importance of spatial processes such as dispersal. There is a rich body of literature on the relationship between environmental variables and community structure, but either spatial processes are not incorporated or treated as a 'nuisance' variable that should be eliminated from the analyses (Liebhold & Gurevitch 2002). My result illustrated that in systems with up to 30 sites, the independent variables included in the linear methods explained between 20% and 80% of the variation present in the community composition.

Moreover, investigation of the relative importance of local and regional processes showed the prevalence of three theoretical metacommunity types for real systems in a unified framework (Leibold *et al.* 2004). The important conclusion was that most communities had a significant environmental component, thus habitat heterogeneity and associated SS dynamics were dominant. Importantly, this does not imply that spatial dispersal processes are not present. A necessary underlying assumption for environmental processes to structure communities is that dispersal is sufficient to deliver species to the sites with the correct environmental conditions, that is, dispersal is not limiting (Fuentes 2002).

A second important subset of data sets consisted of metacommunities structured by both environmental and spatial variables, independent of each other, suggesting SS + ME dynamics. This is also an important conclusion as it illustrates the necessity of including spatial processes in community ecology (Levin 1992). Spatial dispersal processes should thus be part of every study in (community) ecology, and should be explicitly modelled in the analyses (Liebhold & Gurevitch 2002). Disregarding neutral dispersal processes in our analyses would have resulted in missing important patterns in 37% of the studied communities: 46 (29%) communities were of mixed SS + ME type and 13 (or 8%) of the NM/PD type.

Only a small fraction of the data sets are structured by only NM/PD dynamics. Despite the considerable controversy surrounding neutral theory in community ecology, it seems to be important in only a limited fraction of real metacommunities in temperate ecosystems. This study did not include tropical data sets, claimed to be an textbook example of NM dynamics (Hubbell 2001). Several tropical rain forest studies that I know of (Duivenvoorden *et al.* 2002; Duque *et al.* 2002; Tuomisto *et al.* 2003; Cannon & Leighton 2004; Vormisto *et al.* 2004), however, performed a similar study. Interpreting their results in the framework of this article, it is also a combination of environmental and spatial dynamics that structure these tropical communities concurrently, and not only neutral dynamics.

Finally, the diversity in metacommunity types (or the relative importance of local environmental vs. neutral dispersal processes) could be partly explained by determinants such as spatial scale, habitat type, organism size and dispersal type. The regression tree results suggested that dispersal type was the most important determinant of metacommunity type. Dispersal type is obviously associated with dispersal rates, and as such it is not surprisingly an important determinant of metacommunity type. Passive dispersal, mainly freshwater plankton and terrestrial plants, results in species composition tracking environmental heterogeneity, even across larger spatial scales. Surprisingly, although marine and lake habitats impose a very different set of dispersal limitations on its constituting species, they resulted in similar metacommunity dynamics. Active dispersers in these two habitats were also strongly related to environmental dynamics, with little evidence for independent spatial dynamics. Active dispersers in estuarine, stream and terrestrial systems finally were split according to spatial scale. However, the results did not support the hypothesized increase of SS dynamics with increased spatial scale. At small spatial scales, SS and NM dynamics were important, while at high spatial scales, an increase of only spatial dynamics becomes apparent. This is probably more the result of limiting dispersal, with suitable sites not colonized by suitable taxa, and thus more neutral dynamics (Fuentes 2002).

As with any meta-analysis, there are several methodological design decisions that could influence some of the results. The first is the nature of space in the above analyses, at three different levels. I decided to treat space as potentially structuring variable as it can both effect environmental heterogeneity and dispersal limitation. This made it fundamentally different from other methodological variables such as the number of sites or variables included in the variation decomposition that I tried to standardize as much as possible. The second level concerns the interpretation of the pure spatial component ($[\mathbf{E}\,|\,\mathbf{S}]$). As it is highly unlikely that authors measured all the relevant environmental variables, it is probable that this spatial component

captured part of the environmental variation not measured by the environmental variables included in the study. However, this would only influence the results in the direction of the importance of SS (and ME) dynamics, and diminish even more the prevalence of neutral dynamics. The third problem with space is that it is treated as a crude proxy for dispersal limitation between the sites in a metacommunity. While third order polynomials of site coordinates can adequately model broad-scale spatial trends in data (Legendre & Legendre 1998), it will potentially miss to include local dispersal patterns and thus potential neutral dynamics. However, the result showed that a large range of metacommunity dynamics were present, but with a tendency for the importance of environmental heterogeneity and SS, at small spatial scales.

A second methodological problem bears upon the standardization to a maximum of four explanatory environmental and spatial variables. While, in general, forward selection without this standardization would have resulted in the inclusion of more environmental than spatial variables, it would have been difficult to compare among studies, with methodological issues confounding the structuring process variables. This would result in a bias towards a higher prevalence of spatial and neutral dynamics. As the results clearly showed the opposite effect, they are probably robust enough, despite this potential bias.

A third methodological problem stems from potential biases in obtaining the different studies. This is for instance obvious in the unintended restriction to north temperate systems (see above). This limits all generalizations of the results to temperate ecosystems. However, most, if not all, data sets were initially collected for other purposes than determining metacommunity structure in the study system. Thus biases of the results towards one of the metacommunity perspectives within the temperate ecosystems will be limited. One could argue that investigators only incorporate environmental variables if they consider them to be important in structuring communities, and as such biased the results towards SS dynamics. However, as the communities included in this meta-analysis are common and well-studied communities, these results imply that neutral dynamics will only be important in limited (sub-)sets of communities with, for instance, very restricted dispersal capabilities (Hubbell 2005).

The presented results clearly illustrated the necessity to include both local environmental and regional dispersal dynamics in metacommunities. The majority of the data sets included relationships with both environmental variables and spatial variables (either directly on indirectly). Only a small subset experienced only neutral dynamics. The next important step in metacommunity theory is to explain the diversity in metacommunity dynamics. While the current analyses succeeded in establishing some broad scale

patterns, the number of individual data sets per final subgroup was still low (on average 14), and confounding several explanatory variables still an issue. With more data sets per system, and including more relevant information on, for instance, species-specific dispersal characteristics, short-range dispersal, seasonal patterns, our understanding of metacommunity dynamics and ecology in general will greatly improve.

## REFERENCES

Bell, G. (2001). Neutral macroecology. *Science*, 293, 2413–2418.

Borcard, D., Legendre, P. & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73, 1045–1055.

Cannon, C.H. & Leighton, M. (2004). Tree species distributions across five habitats in a bornean rain forest. *J. Veg. Sci.*, 15, 257–266.

Chase, J.M. & Leibold, M.A. (2003). *Ecological Niches*. University of Chicago Press, Chicago, IL.

Chave, J. (2004). Neutral theory and community ecology. *Ecol. Lett.* 7, 241–253.

Cohen, J.E., Jonsson, T. & Carpenter, S.R. (2003). Ecological community description using the food web, species abundance, and body size. *Proc. Natl Acad. Sci. USA*, 100, 1781–1786.

Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B. *et al.* (2002). Beta-diversity in tropical forest trees. *Science*, 295, 666–669.

Cottenie, K., Michels, E., Nuytten, N. & De Meester, L. (2003). Zooplankton metacommunity structure: regional versus local processes in highly interconnected ponds. *Ecology*, 84, 991–1000.

De'Ath, G. & Fabricius, K.E. (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81, 3178–3192.

Duivenvoorden, J.F., Svenning, J.C. & Wright, S.J. (2002). Ecology – beta diversity in tropical forests. *Science*, 295, 636–637.

Duque, A., Sanchez, M., Cavalier, J. & Duivenvoorden, J.F. (2002). Different floristic patterns of woody understory and canopy plants in Colombian Amazonia. *J. Trop. Ecol.*, 18, 499–525.

Etienne, R.S. & Olff, H. (2005). Confronting different models of community structure to species-abundance data: a Bayesian model comparison. *Ecol. Lett.*, 8, 493–504.

Finlay, B.J. (2002). Global dispersal of free-living microbial euka-ryote species. *Science*, 296, 1061–1063.

Forbes, A.E. & Chase, J.M. (2002). The role of habitat connectivity and landscape geometry in experimental zooplankton meta-communities. *Oikos* 96, 433–440.

Fuentes, M. (2002). Seed dispersal and tree species diversity. *Trends Ecol. Evol.*, 17, 550.

Gilbert, B. & Lechowicz, M.J. (2004). Neutrality, niches, and dis-persal in a temperate forest understory. *Proc. Natl Acad. Sci. USA*, 101, 7651–7656.

Gonzalez, A., Lawton, J.H., Gilbert, F.S., Blackburn, T.M. & Evans-Freke, I. (1998). Metapopulation dynamics, abundance, and distribution in a microecosystem. *Science*, 281, 2045–2047.

Hastings, A. (1980). Disturbance, coexistence, history, and com-petition for space. *Theor. Popul. Biol.*, 18, 363–373.

Holt, R.D. (1993). Ecology at the mesoscale: the influence of regional processes on local communities. In: *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (eds Ricklefs, R.E. & Schluter, D.). University of Chicago Press, Chicago, IL, pp. 77–88.

Holyoak, M. (2000). Habitat subdivision causes changes in food web structure. *Ecol. Lett.*, 3, 509–515.

Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.

Hubbell, S.P. (2005). Neutral theory in community ecology and the hypothesis of functional equivalence. *Funct. Ecol.*, 19, 166–172.

Legendre, P. & Legendre, L. (1998). *Numerical Ecology*. Elsevier Science B.V., Amsterdam.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F. *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.*, 7, 601–613.

Levin, S.A. (1992). The problem of pattern and scale in ecology. *Ecology*, 73, 1943–1967.

Liebhold, A.M. & Gurevitch, J. (2002). Integrating the statistical analysis of spatial data in ecology. *Ecography*, 25, 553–557.

McGill, B.J. (2003). A test of the unified neutral theory of bio-diversity. *Nature*, 422, 881–885.

McIlroy, D. (2004). Packaged for R by Browning R. and Minka T.P. *Mapproj: Map Projections*. R package version 1.1.7. Available at http://www.r-project.org.

Mouquet, N. & Loreau, M. (2002). Coexistence in metacommu-nities: the regional similarity hypothesis. *Am. Nat.*, 159, 420–426.

Mouquet, N. & Loreau, M. (2003). Community patterns in source-sink metacommunities. *Am. Nat.*, 162, 544–557.

Oksanen, J. (2005). Vegan: R functions for vegetation ecologists. Available at: http://cc.oulu.fi/∼jarioksa/softhelp/vegan.html.

R Development Core Team (2004). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Therneau, T.M. & Atkinson, E.J. (1997). *An Introduction to Recursive Partitioning using the RPART Routines*. Technical Report, Mayo Foundation, Rochester, Minnesota.

Tilman, D. (1994). Competition and biodiversity in spatially structured habitats. *Ecology*, 75, 2–16.

Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, 299, 241–244.

Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S*. Springer-Verlag, New York.

Vormisto, J., Svenning, J.C., Hall, P. & Balsley, H. (2004). Diversity and dominance in palm (Arecaceae) communities in terra firme forests in the western Amazonian basis. *J. Ecol.*, 92, 577–588.

## SUPPLEMENTARY MATERIAL

The following supplementary material is available online for this article from http://www.Blackwell-Synergy.com:

**Table S1** Data set descriptions.
**Figure S2** Data set locations.
**Figure S3** Metacommunity axis vs. spatial scale.
**Figure S4** Metacommunity axis vs. body size.
**Figure S5** Metacommunity axis vs. habitat type.
**Figure S6** Metacommunity axis vs. dispersal type.