Analysis of Multivariate Ecological Data

School on Recent Advances in Analysis of Multivariate Ecological Data

24-28 October 2016

Prof. Pierre Legendre
Dr. Daniel Borcard
Département de sciences biologiques
Université de Montréal
C.P. 6128, succursale Centre Ville
Montréal QC
H3C 3J7  Canada
pierre.legendre@umontreal.ca
daniel.borcard@umontreal.ca
Day 5
Multi-scale modelling of the spatial structure of ecological communities
Multi-scale modelling of the spatial structure of ecological communities

Distance-based Moran's eigenvector maps
Main references


Statement of problem

Ecologists want to understand and model spatial/temporal community structures through the analysis of species assemblages.

- Species assemblages are the best response variables available to estimate the impact of [anthropogenic] changes in ecosystems.

- Species assemblages form multivariate data tables (sites $\times$ species).

Spatial structures in communities indicate that some process has been at work that created them.
Two families of mechanisms can generate spatial (or temporal) structures in communities:

- **Autocorrelation** in the species assemblage (response variables).
- **Forcing (explanatory) variables**: environmental or biotic control of the assemblages, or historical dynamics \(\Rightarrow\) **induced spatial correlation**.

To understand the mechanisms that generate these structures, we need to explicitly incorporate the spatial (or temporal) community structures into the statistical model.
Statement of problem

The **scale** issue is important in this context:

- Some processes act at a **global** scale, others are **local**.
- Therefore, not all response variables (species) are structured at the same scale.
- One single response variable can also display structures at more than one spatial or temporal scale.

We need statistical methods to model spatial or temporal structures at **all scales**.
Distance-based Moran's eigenvector maps (dbMEM)

Formerly called principal coordinates of neighbour matrices (PCNM)
Consider the following data:

- species data: $n$ sites, $p$ species
- environmental data: $n$ sites, $m$ environmental variables
- spatial data: $n$ sites, X (and Y) coordinates

What are our goals?
What are our goals?

1. Model the spatial structure of the species data at all scales

2. Identify the scales where structures are present in the response data.

3. Decompose the spatial model into submodels representing these scales

4. Interpret the submodels: reveal the species-environment relationships at the relevant scales
The descriptors of spatial relationships (dbMEM base functions) are obtained by principal coordinate analysis of a truncated matrix of Euclidean (geographic) distances among the sampling sites.
Principal coordinate analysis

Eigenvectors

Euclidean distances

$\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & \ldots & n-1 \\
1 & 2 & 3 & 4 & 5 & \ldots \\
1 & 2 & 3 & 4 & 5 & \ldots \\
1 & 2 & 3 & 4 & 5 & \ldots \\
1 & 2 & 3 & 4 & \ldots \\
1 & 2 & \ldots
\end{bmatrix}$

Multiple regression or canonical analysis

Truncated matrix of Euclidean distances = neighbor matrix

$\begin{bmatrix}
\text{Max x 4} \\
1 & \ldots & \text{max} \\
1 & \ldots & \text{max} \\
1 & \ldots & \text{max} \\
1 & \ldots & \text{max} \\
1 & \ldots & \text{max} \\
1
\end{bmatrix}$

Observed variable $y$

$x$
(spatial coordinates)

$\begin{bmatrix}
Y \\
X(+)
\end{bmatrix}$

Eigenvectors with positive eigenvalues
$= \text{dbMEM variables}$

Principal coordinate analysis

Eigenvectors
Principal coordinate analysis

Eigenvectors

Euclidean distances

n–1
1 2 3 4 5 ...
1 2 3 4 5 ...
1 2 3 4 5 ...
1 2 3 4 5 ...
1 2 3 4 5 ...
1 2 3 4 5 ...
1 2 3 4 ...
1 2 3 ...
1 2

Data

Observed variable y

(x spatial coordinates)

Multiple regression or canonical analysis

Y

X (+)

Eigenvectors with positive eigenvalues

= dbMEM variables

Truncated matrix of Euclidean distances = neighbor matrix

Max x 4

Principal coordinate analysis

Eigenvectors

+ 0 −
Principal coordinate analysis

Eigenvectors

Euclidean distances

Truncated matrix of Euclidean distances = neighbor matrix

Multiple regression or canonical analysis

Eigenvectors with positive eigenvalues = dbMEM variables

Principal coordinate analysis

Observed variable

Data

x (spatial coordinates)

Y

X(+)

+ 0 −
Principal coordinate analysis

Eigenvectors

Euclidean distances

Truncated matrix of Euclidean distances = neighbor matrix

Multiple regression or canonical analysis

Eigenvectors with positive eigenvalues = dbMEM variables

Observed variable $y$

(x: spatial coordinates)

Data

Principal coordinate analysis

Eigenvectors

$1 \ 2 \ 3 \ 4 \ 5 \ ... \ n-1$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$

$1 \ 2 \ 3 \ 4 \ 5 \ ...$
Principal coordinate analysis

Eigenvectors

Euclidean distances

Truncated matrix of Euclidean distances = neighbor matrix

Multiple regression or canonical analysis

Data

Eigenvectors with positive eigenvalues = dbMEM variables

Observed variable $y$

(spatial coordinates)

$x$

$1\ 2\ 3\ 4\ 5\ \ldots\ \ldots\ n-1$

$1\ 2\ 3\ 4\ 5\ \ldots$

$1\ 2\ 3\ 4\ 5\ \ldots$

$1\ 2\ 3\ 4\ 5\ \ldots$

$1\ 2\ 3\ 4\ 5\ \ldots$

$1\ 2\ 3$

$1\ 2$

$1$

Max x 4

Principal coordinate analysis

$Y$

$X(+)$

$+$

$0$

$-$
Principal coordinate analysis

Eigenvectors

Euclidean distances

- 0

n–1

1

2

3

4

5

...

Data

Observed variable y

x

(spatial coordinates)

Y

X(+)

Multiple regression or canonical analysis

Eigenvectors with positive eigenvalues = dbMEM variables

Truncated matrix of Euclidean distances = neighbor matrix

Principal coordinate analysis

+ 0 -
Principal coordinate analysis

Eigenvectors

Euclidean distances

Multiple regression or canonical analysis

Truncated matrix of Euclidean distances = neighbor matrix

Data

Observed variable \( y \)

\( x \)
(spatial coordinates)

\[ X(+) \]

\[ Y \]

Eigenvectors with positive eigenvalues
= dbMEM variables

Principal coordinate analysis

Eigenvectors

Truncated matrix of Euclidean distances = neighbor matrix

1...max

Max x 4
Principal coordinate analysis

Eigenvectors

Euclidean distances

\[ \begin{align*}
1 & 2 & 3 & 4 & 5 & \ldots & n-1 \\
1 & 2 & 3 & 4 & 5 & \ldots & \end{align*} \]

...n

Multiple regression
or canonical analysis

Eigenvectors with positive eigenvalues
= dbMEM variables

Truncated matrix of Euclidean distances = neighbor matrix

Max x 4

Data

(x)

(spatial coordinates)

Observed variable y
The difference between the original PCNM and the dbMEM functions is technical. It concerns the values on the diagonal of the truncated distance matrix and the way to compute its principal coordinates. The consequence is that the eigenvalues of the dbMEM are directly proportional to the Moran $I$ coefficient of autocorrelation of the corresponding eigenvectors. Therefore, it is easier to separate the eigenvectors that model positive spatial correlation from those that model negative spatial correlation. In the case of a regular sampling design along a transect, there are approximately $n/2$ dbMEM with positive spatial correlation.
Notes on dbMEM base functions

dbMEM variables represent a spectral decomposition of the spatial relationships among the study sites.

They can be computed for regular or irregular sets of points in space or time.

dbMEM base functions are orthogonal. If the sampling design is regular, they look like sine waves.

If the sampling design is irregular, the dbMEM base functions have irregular shapes as well, but they can still be roughly ordered from broad-scale to fine-scale.
dbMEM variables can be split into groups representing submodels of various spatial scales.

The grouping of dbMEM variables into submodels of various scales implies arbitrary decisions about the building of the groups.

dbMEM base functions can also be computed for circular sampling designs. An example can be found in Brind'Amour et al. (2005).

dbMEM analysis can be used for temporal analysis, as well as spatio-temporal analysis. A method exists that allows the analysis of spatio-temporal designs without spatial replication while still testing the interaction.
Eight of the 49 orthogonal dbMEM base functions with positive spatial correlation obtained for 100 equally-spaced points along a transect. Truncation after the first neighbour.
None of the 189 orthogonal dbMEM base functions with positive spatial correlation obtained for a grid of 20x20 equispaced points.
Real example

Lac Geai oribatid mites: 70 irregularly positioned sites, 35 species.

dbMEM analyses:

1. Construction of the dbMEM variables: 22 with positive spatial correlation.

2. Global analysis: 8 significant dbMEM.

3. Partition by scale: broad (dbMEM 1, 3, 4), medium (dbMEM 6, 7, 10, 11), fine (dbMEM 20)

4. Separate dbMEM analyses by scale.
70 sites of the Lac Geai data set: the 8 significant dbMEM among the 22 with positive spatial correlation.
dbMEM analysis of the detrended oribatid mite data with the 8 significant dbMEM variables. Maps of the fitted site scores of the first three canonical axes, which are significant at $\alpha = 0.05$. 
dbMEM analysis of the detrended oribatid mite data with the 3 broad-scale dbMEM variables. Maps of the fitted site scores of the first three canonical axes, which are significant at $\alpha = 0.05$. 
dbMEM analysis of the detrended oribatid mite data with the 4 medium-scale dbMEM variables. Maps of the fitted site scores of the first three canonical axes, which are significant at $\alpha = 0.05$. 
Interpretation of the spatial structures revealed by the dbMEM analysis by means of the environmental variables: see Practicals!
Quick and easy dbMEM analysis

Function quickMEM()

Needs two objects: species data table (transformed if necessary) and X or YX coordinates.

After that this is what the function does automatically:

1. Tests if there is a significant linear trend in the species data. If a trend is present, detrends the species data.

2. Computes the dbMEM eigenfunctions with positive Moran's I.

3. Runs and tests an RDA of the species with all dbMEM. If not significant, stops here. If significant, continues to step 4.

4. Runs a forward selection of dbMEM, with the Blanchet et al. (2008) double stopping criteria ($\alpha$ level and $R^2_a$ of complete RDA obtained at step 3).
Quick and easy dbMEM analysis

Function quickMEM()

Needs two objects: species data table (transformed if necessary) and X or YX coordinates.

After that this is what the function does automatically:

5. Runs a new RDA with the dbMEM retained at step 4.
6. Tests the RDA obtained at step 5 globally, and axis by axis.
7. Draws maps of the spatial structures identified on the significant RDA axes.
8. Outputs an object containing the dbMEM, their eigenvalues, the result of the forward selection and the results of the RDA run on the significant dbMEM.

myres <- quickMEM(species,XY)
Day 5

Generalized MEM
Asymmetric eigenvector maps
(AEM)

→ See course notes by Pierre Legendre
Testing interaction between space and time without replication
Space-time interaction without replication

Classical two-way crossed ANOVA design with replication

<table>
<thead>
<tr>
<th>Factor A</th>
<th>Factor B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B1</td>
</tr>
<tr>
<td>A1</td>
<td>0 0 0</td>
</tr>
<tr>
<td>A2</td>
<td>0 0 0</td>
</tr>
<tr>
<td>A3</td>
<td>0 0 0</td>
</tr>
</tbody>
</table>

The replicates provide the within-cell d.f. necessary to test interaction.
### Space-time interaction without replication

#### Space-time survey design without replication

<table>
<thead>
<tr>
<th>Time</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

- **Space:** $s - 1$ d.f.
- **Time:** $t - 1$ d.f.
- **Error:** $(s - 1)(t - 1)$ d.f.

$\Rightarrow 0$ d.f. for the interaction
Space-time interaction without replication

How could we spare some degrees of freedom?

→ by coding space, time or interaction more parsimoniously.

→ code space or time by means of dbMEM, retaining only those with positive spatial or temporal correlation.

→ several possibilities to combine dbMEM with the normal coding.

residual sum-of-squares $SS$ would actually be the sum of $SS_{Int1}$ and $SS_{Res1}$ of Model 1. If an interaction truly exists (i.e., if $ab_{ij} \neq 0$ for some $i$ and $j$ combination), then there will be a lack of model fit,

$$SS_{LOF2} = SS_{Int1}$$

where LOF designates lack of fit. The residual mean square will be overestimating $r^2$ and the tests for significant overall spatial and temporal variation will lose power. In other words, using Model 2, the tests of significance for space and time will have correct Type I error only if the interaction is negligible (i.e., if $ab_{ij} = 0$ for all $i$ and $j$). Hence we see that analyzing the interaction is interesting both for testing its possible significant effect and for partialling its effect out in statistical tests of the main effects.

The approach adopted here for testing a space–time interaction considers the analysis of variance as a multiple regression problem using a design matrix, as described for example by Shaw and Mitchell-Olds (1993). Among the various forms of dummy variable coding, we select Helmert contrasts (Chambers and Hastie 1992, Venables and Ripley 2002), also called ''orthogonal dummy variables'' (see Appendix C in Legendre and Anderson 1999). In balanced crossed sampling designs (i.e., when the same number of sites are surveyed during each sampling campaign), the Helmert contrast coding variables are orthogonal to one another (i.e., their scalar products are zero) and each one sums to zero. They are also orthogonal to the column of ''ones'' representing the overall mean effect. The $s$ sites require $s - 1$ Helmert contrast variables for coding (design matrix $X_{s-1}$). Similarly, the $t$ sampling campaigns require $t - 1$ Helmert contrast variables for coding (design matrix $X_{t-1}$). The interaction term is obtained by multiplying each of the $s - 1$ contrast variables coding for sites by each of the $t - 1$ contrast variables coding for times, producing $(s - 1)(t - 1)$ new variables (design matrix $X_{Int}$). These product variables are orthogonal to one another and have zero sums. They are also orthogonal to the main effect Helmert contrasts. As a consequence, the interaction can be tested as a term that is linearly independent of the variables coding for space and for time.

Model 1, re-expressed as a multiple regression model with design matrices, is the following using matrix notation:

$$y = 1 + X_{s-1}C_0^1a + X_{t-1}C_0^1b + X_{Int}(ab) + e$$

For each model $k$,

$$SS_{Res.k} = \Box + \Box$$

---

**ANOVA models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Space</th>
<th>Time</th>
<th>Interaction</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>$SS(X_{s-1})$</td>
<td>$SS(X_{t-1})$</td>
<td>$SS(X_{Int1})$</td>
<td>$SS_{Res1}$</td>
</tr>
<tr>
<td>Model 2</td>
<td>$SS(X_{s-1})$</td>
<td>$SS(X_{t-1})$</td>
<td></td>
<td>$SS_{Res1}$</td>
</tr>
<tr>
<td>Model 3</td>
<td>$SS(X_{u})$</td>
<td>$SS(X_{t-1})$</td>
<td>$SS(X_{Int3})$</td>
<td>$SS_{Res1}$</td>
</tr>
<tr>
<td>Model 4</td>
<td>$SS(X_{u})$</td>
<td>$SS(X_{v})$</td>
<td>$SS(X_{Int4})$</td>
<td>$SS_{Res1}$</td>
</tr>
<tr>
<td>Model 5</td>
<td>$SS(X_{s-1})$</td>
<td>$SS(X_{t-1})$</td>
<td>$SS(X_{Int4})$</td>
<td>$SS_{Res1}$</td>
</tr>
</tbody>
</table>

---

**Model 6a**

<table>
<thead>
<tr>
<th>Space</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SS(X_{u1})$</td>
<td>$SS_{Res}$</td>
</tr>
</tbody>
</table>

**Model 6b**

<table>
<thead>
<tr>
<th>$SS(X_{u1})$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$SS(X_{u2})$</td>
<td>$\ldots$</td>
<td></td>
</tr>
<tr>
<td>$SS(X_{ut})$</td>
<td></td>
<td>$SS_{Res}$</td>
</tr>
</tbody>
</table>
Space-time interaction without replication

Several combinations of Helmert codes and dbMEM

**Model 1:** standard two-way crossed ANOVA design with interaction. \((s - 1), (t - 1)\) and \((s - 1) \times (t - 1)\) Helmert contrasts form the design matrices of space, time and interaction. \(st(r - 1)\) degrees of freedom remain for the error term.
Space-time interaction without replication

Several combinations of Helmert codes and dbMEM

**Model 2**: standard two-way crossed ANOVA design without replication. \((s - 1)\) and \((t - 1)\) Helmert contrasts form the design matrices of space and time. The remaining \((s - 1) \times (t - 1)\) d.f. are used to estimate the error variance.
Space-time interaction without replication

Several combinations of Helmert codes and dbMEM

**Model 3a:** two-way crossed ANOVA design without replication, but with **space under-fitted:** $u$ dbMEM spatial variables replace the $(s - 1)$ spatial Helmert contrasts; $u < (s - 1)$; an interaction design matrix can be constructed, containing $u \times (t - 1)$ terms. $(s - u - 1) \times t$ d.f. remain to estimate the error variance. A test of the interaction is thus possible.
Space-time interaction without replication

Several combinations of Helmert codes and dbMEM

Model 3b: two-way crossed ANOVA design without replication, but with time under-fitted: $\nu$ dbMEM spatial variables replace the $(t - 1)$ temporal Helmert contrasts; $\nu < (t - 1)$; an interaction design matrix can be constructed, containing $\nu \times (s - 1)$ terms. $(t - \nu - 1) \times s$ d.f. remain to estimate the error variance. A test of the interaction is thus possible.
Space-time interaction without replication

Several combinations of Helmert codes and dbMEM

**Model 4**: two-way crossed ANOVA design without replication, but with **space and time under-fitted**: $u$ dbMEM spatial variables replace the $(s-1)$ spatial Helmert contrasts and $v$ dbMEM variables replace the $(t-1)$ temporal contrasts; $u < (s-1)$ and $v < (t-1)$; an interaction design matrix can be constructed, containing $u \times v$ terms. $[s \times t - (u + v + u \times v) - 1]$ d.f. remain to estimate the error variance. A test of the interaction is thus possible.
Space-time interaction without replication

Several combinations of Helmert codes and dbMEM

**Model 5**: two-way crossed ANOVA design without replication, but with interaction under-fitted: the main factors are coded with their \((s - 1)\) and \((t - 1)\) Helmert contrasts, but the interaction variables are created using the \(u\) and \(v\) dbMEM variables used in Model 4. The d.f. of the interaction term are \(u \times v\) and 
\[\left(\left((s - 1) \times (t - 1) - u \times v\right)\right)\] d.f. remain for the residual error.
Space-time interaction without replication

Several combinations of Helmert codes and dbMEM

**Model 6a:** special case where one wants to test the significance of a spatial effect, or of a temporal effect, in the presence of space-time interaction. Model 6a tests for the presence of spatial structure in turn for each $t$ times (or the reverse), correcting for multiple testing.
Space-time interaction without replication

Several combinations of Helmert codes and dbMEM

**Model 6b**: special case where one wants to test the significance of a spatial effect, or of a temporal effect, in the presence of space-time interaction. Model 6b involves a simultaneous test for spatial structure in all $t$ times, using a staggered matrix of spatial dbMEM variables (or the reverse).
Space-time interaction without replication

Model 5 is recommended to test interaction.

- Permutation test has correct type I error.
- Highest power to detect interaction.

Practicals: functions `stimodels()` and `quicksti()`

These functions will soon be integrated in package `adespatial`. 
Space-time interaction without replication

**Exemple** using function `quicksti()`

**Data:**

Trichoptera (Insecta), emergence traps: 56 species, Hellinger transformation

Space: 22 traps along a stream; outflow of Lac Cromwell, Station de biologie des Laurentides, Université de Montréal.

Time: 10 periods
Space-time interaction without replication

Data layout

<table>
<thead>
<tr>
<th>Sites</th>
<th>Time 1</th>
<th>Time …</th>
<th>Time 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sp1 \[ \ldots \] Sp56
Space-time interaction without replication

quicksti(trich.hel, S=22, Ti=10)

Number of space points (s) = 22
Number of time points (tt) = 10
Number of observations (n = s*tt) = 220
Number of response variables (p) = 56
Significance level for the interaction test (alpha) = 0.05

Computing dbMEMs to code for space
Truncation level for space dbMEMs = 1
Computing dbMEMs to code for time
Truncation level for time dbMEMs = 1

Number of space coding functions = 10
Number of time coding functions = 4

------------------------------------------

Testing space-time interaction (model 5)
------------------------------------------

Number of space variables = 21
Number of time variables = 9
Number of interaction variables = 40
Number of residual degrees of freedom = 149

Interaction test:  R2 = 0.1835   F = 2.313   P( 999 perm) = 0.001
Space-time interaction without replication

Testing for the existence of separate spatial structures (model 6a)

- Number of space variables = 100
- Number of time variables = 9
- Number of residual degrees of freedom = 110

Space test: \( R^2 = 0.4739 \quad F = 2.0256 \quad P(999 \text{ perm}) = 0.001 \)

Testing for separate temporal structures (model 6b)

- Number of space variables = 21
- Number of time variables = 88
- Number of residual degrees of freedom = 110

Time test: \( R^2 = 0.4981 \quad F = 2.4936 \quad P(999 \text{ perm}) = 0.001 \)
Space-time interaction without replication

Invasive process. However, while *Drypetes standleyi* (Euphorbiaceae) is considered to be invading the plot from the east (Harms et al. 2001), we found that its spatial structure did not change significantly during the study period ($P = 0.498$). Other spatial changes may have originated due to strong perturbations or climatic change. Condit (1998) reported a decline in nearly all BCI tree and shrub species associated with moist microhabitats due to a reduction in rainfall and a lengthening of the dry seasons. We found that among the 31 species associated with slope habitats (Harms et al. 2001), which are moister than the plateaus in BCI, 23 ($74\%$) had significantly changed their spatial distributions. Examples are the crash of dominant moisture specialists like *Poulsenia armata* (Moraceae, $P = 0.001$) and the displacement of the population of *Beilschmiedia pendula* (Lauraceae, $P = 0.001$) towards areas of higher slope (Fig. 4).

We also carried out a space–time interaction test on the multivariate Hellinger-transformed abundance data, which turned out to be highly significant ($P = 0.001$ after 999 permutations). Hence, the changes at the species level were strong enough to allow the detection of changes in the spatial distribution of species composition at the community level.

In another application, Laliberte´ et al. (2009) studied the space and time factors, as well as the space–time interaction, in a temperate forest understory where tree seedling abundances had been monitored during a 9-yr period at 40 permanent plots.

**Fig. 3.** Spatiotemporal map showing the $K$-means partition of the emergence trap observations into five groups. Abscissa: trap line with traps numbered 1 (upstream) to 22 (downstream). Ordinate: 10-day time periods numbered 1 to 10. Symbols for groups: 1, circle; 2, triangle; 3, “plus” sign (+); 4, “times” sign ($\times$); 5, diamond.
10.8). The test can be carried out separately for each sampling time (Appendix C: Model 6a: one-factor ANOVA model). One can also conduct a single test involving a separate model of the spatial structure for each time period (Appendix C: Model 6b: stacked one-factor ANOVA model). The temporal structure of each sampling point can be analyzed in the same way by interchanging space and time.

A test of an interaction without replication has, of course, less power than a test conducted using replicated data. The problem challenged in this paper is, however: What can one do to test an interaction in the absence of replication? This question is of interest to all researchers who conduct spatial ecological surveys repeated across time. Identifying an interaction in the community composition data is a clear signal that the community has reacted to changing environmental conditions (or other causes) by modifying its spatial structure. The new test should prove useful to detect natural or man-made changes in ecological communities and ecosystems.

**ACKNOWLEDGMENTS**

We are grateful to Pierre-Paul Harper, Université de Montreàl, for permission to use the Trichoptera emergence data as an example in this paper. This research was supported by NSERC Grant no. 7738 to P. Legendre.

**LITERATURE CITED**


**Fig. 4.** Changes in numbers of individuals in the Barro Colorado Island (BCI), Panama, permanent forest plot for two species associated with slopes that significantly changed their distributions between the 1982–1983 and 1995 censuses. Light gray squares indicate loss, whereas dark gray squares indicate gain of individuals.