Spatial eigenfunction modelling: recent developments

Pierre Legendre
Département de sciences biologiques
Université de Montréal

© Pierre Legendre 2018
Outline of the presentation

1. Spatial eigenfunction analysis
2. Distance-based Moran’s eigenvector maps (dbMEM)
3. Generalized MEM analysis
4. Asymmetric eigenvector maps (AEM)
5. AEM analysis of time series
6. Other methods that use spatial eigenfunctions
7. R software
8. References
1. Spatial eigenfunction analysis

This expression was proposed by Griffith & Peres-Neto\(^1\) as the name of a family of methods of statistical analysis …

… based on \textit{eigenvectors} describing the spatial relationships among the study sites.

1. Let us examine first Griffith’s spatial filtering method.

Consider a graph of interactions among neighbouring sites on a map and translate that graph into a binary connectivity matrix:

Graph of connexions among sites

C = connectivity (or contiguity) matrix

Compute the eigenvectors of a centred binary connectivity matrix, and use them in regression models to control for spatial correlation.

• **C** is a binary [0, 1] geographic connectivity matrix among sites.

• Centre **C** so that rows and columns sum to 0 (Gower centring):

\[
G = \left( I - \frac{11'}{n} \right) C \left( I - \frac{11'}{n} \right)
\]

• Compute eigenvectors of **G**.

• Use these spatial eigenvectors as covariables in multiple regression (\( y \sim X \)) or other linear modelling methods.

• The spatial eigenvectors effectively control for spatial correlation when testing the significance of the relationships between \( y \) and \( X \).
2. *Spatial eigenfunction analysis* is a more general method –
   • Eigenvectors of spatial configuration matrices are computed
   • and used as predictors in linear models, including the full range of
     general and generalized linear models. (Other models are possible.)

   The expression covers:
   • the recent methods that take into account the distances among
     localities, described in this talk,
   • and the early methods developed by geographers to analyse binary
     spatial connection matrices (Garrison & Marble, 1964; Gould, 1967;
     Tinkler, 1972; Griffith, 1996); these methods are described as *binary
     MEM analysis* in section 3 of this talk, entitled *Generalized MEM
     analysis*.

   Extension to *multivariate time series* analysis is straightforward\(^1\).

---

Distance-based Moran’s eigenvector maps (dbMEM)

alias Principal Coordinates of Neighbor Matrices (PCNM)

- Borcard & Legendre 2002 [1229 citations]
- Borcard, Legendre et al. 2004 [695 citations]
- Dray, Legendre & Peres-Neto 2006 [1008 citations]
Borcard, Legendre and coauthors’ spatial eigenfunction analysis\textsuperscript{1}

We were primarily interested in estimating the spatial variation, mapping it, and analyzing its relationship with the environmental variables at different spatial scales.

- Griffith’s spatial filters based on a connectivity matrix among sites can be used for these purposes, as will be seen later.

- **Distance-based MEM (dbMEM)** can be used to control for spatial correlation in tests of $y \sim X$ relationships, e.g. species-environment relationships in ecology, just like Griffith’s method.\textsuperscript{2}

---

\textsuperscript{1} dbMEM eigenfunction analysis is also described in Legendre & Legendre (2012, Chap. 14) and Borcard et al. (2018, Section 7.4).

Graphs of 10 of the 49 dbMEM eigenfunctions that represent the spatial variation along a transect with 50 equally-spaced observation points.
Compute dbMEM – Example along a transect

- Truncate the matrix of geographic distances among the sites,
- then decompose $\mathbf{D}$ by principal coordinate analysis (PCoA) = classical MDS (cMDS).

$\Rightarrow$ PCoA: Centre $\mathbf{D}$, then compute eigenvalues and eigenvectors.
dbMEM eigenfunctions display spatial autocorrelation

A spatial correlation coefficient (Moran’s I) can be computed for each dbMEM, indicating the sign of the spatial correlation (+ or −). The eigenvalues are proportional to $I$. Ex.: 50-point transect, 49 dbMEM.
How to find the truncation distance? Example: a 2-dimensional map

<table>
<thead>
<tr>
<th>Point labels</th>
<th>Coordinates</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7.5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
• Spatial data: compute a minimum spanning tree

![Graph showing spatial data with a minimum spanning tree](image)

Truncation distance $\geq$ length of the longest edge.

Longest edge here: $D(7, 8) = 3.0414$

• Irregular time series: truncation distance $\geq$ length of longest edge.
**Technical notes on dbMEM eigenfunctions**

dbMEM variables represent a *spectral decomposition of the spatial/temporal relationships* among the study sites. They can be computed for regular or irregular sets of points in space or time.

dbMEM eigenfunctions are *orthogonal*. If the sampling design is regular, they look like sine waves; this is a property of the eigen-decomposition of the centred form of a distance matrix. If the design is irregular, the sine waves are distorted.
Simulation study

Type I error study
Simulations showed that the procedure is honest. It does not generate more significant results that it should for a given significance level $\alpha$.

Power study
Simulations showed that dbMEM analysis is capable of detecting spatial structures of many kinds:
• random autocorrelated data,
• bumps and sine waves of various sizes, without or with random noise, representing deterministic structures, as long as the structures are larger than the truncation value used to create the dbMEM (PCNM) eigenfunctions.

Detailed results are found in the Borcard & Legendre 2002 paper.
A difficult test case
A difficult test case

a) Linear trend (12.1%)

b) Normal patch (6.3%)

c) 4 waves (13.9%)

d) 17 waves (8.0%)

e) Random autocorrelated deviates (8.9%)

f) Random normal deviates (50.8%)

g) Data (100.0%)

---

Transect coordinates

Dependent variable
A difficult test case

Simulated data

Step 1: detrending

Data (100%)

Dependent variable

Detrended data

Spatial model with 8 PCNM bases (R² = 0.433)

PCNM used is so different from detrended data

A difficult test case
A difficult test case

- g) Data (100%)
  - Simulated data

- h) Detrended data
  - Step 1: detrending

- i) Spatial model with 8 PCNM base functions ($R^2 = 0.433$)
  - PCNM analysis of detrended data

Dependent variable
A difficult test case

Simulated data

Step 1: detrending

PCNM analysis of detrended data

Dependent variable

g) Data (100%)

h) Detrended data

i) Spatial model with 8 PCNM base functions
   \( (R^2 = 0.433) \)

j) Broad-scale submodel \( (R^2 = 0.058) \)
   PCNM #2

k) Intermediate-scale submodel \( (R^2 = 0.246) \)
   PCNM #6, 8, 14

l) Fine-scale submodel \( (R^2 = 0.128) \)
   PCNM #28, 33, 35, 41

20
**dbMEM on a regular grid**

\[ n = 8 \times 12 = 96 \text{ sites.} \]

Maps showing ten of the 48 dbMEM eigenfunctions that display positive spatial correlation.

Shades of grey: values in each eigenvector, from white (largest negative value) to black (largest positive value).
These eigenfunctions are then used as explanatory variables –
• in regression modelling (if there is a single response variable $y$)
• or in canonical analysis (RDA, for multivariate response data $Y$, like community composition or genetic data).

=> Regression analysis estimates weights for the dbMEM variables. We don’t obtain these weights in RDA.

=> Selection of explanatory variables in regression or RDA can be used to obtain a parsimonious spatial model.
Example 1

Regular one-dimensional transect in upper Amazonia

Data: abundance of the fern *Adiantum tomentosum* in quadrats.

Sampling design: 260 adjacent, square (5 m × 5 m) subplots forming a transect in the region of Nauta, Peru.

Questions

• At what spatial scales is the abundance of this species structured?
• Are these scales related to those of the environmental variables?

Pre-treatment

• The abundances were square-root transformed
• and detrended (significant linear trend: $R^2 = 0.102$, $p = 0.001$)

Forward selection
50 dbMEM eigenfunctions (PCNM) were selected (permutation test, 999 permutations).

The dbMEM were arbitrarily divided into 4 submodels. The submodels are orthogonal to one another.

Significant wavelengths (periodogram analysis):
V-br-scale: 250, 355-440 m
Broad-scale: 180 m
Medium-scale: 90 m
Fine-scale: 50, 65 m
**Interpretation**: regression on the environmental variables

<table>
<thead>
<tr>
<th></th>
<th>Adiantum tomentosum</th>
<th>Very broad</th>
<th>Broad</th>
<th>Medium</th>
<th>Fine</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$ of PCNM submodel on <em>A. tomentosum</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ of envir. on submodel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ of envir. on <em>A. tomentosum</em></td>
<td>0.436</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thickness of soil organic horizon (cm)</td>
<td>0.0004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterlogging</td>
<td>0.0393</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy height (m)</td>
<td>0.0011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy coverage (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shrub coverage (%)</td>
<td>0.0621</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Herb coverage (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trees 3 - 7.5 cm DBH</td>
<td>0.0083</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lianas 3 - 7.5 cm diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lianas 8 - 15 cm diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Interpretation**: regression on the environmental variables

<table>
<thead>
<tr>
<th></th>
<th>Adiantum tomentosum</th>
<th>Very broad</th>
<th>Broad</th>
<th>Medium</th>
<th>Fine</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$ of PCNM submodel on A. tomentosum</td>
<td>0.333</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ of envir. on submodel</td>
<td>0.347</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ of envir. on A. tomentosum</td>
<td>0.436</td>
<td>0.116</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Very broad</th>
<th>Broad</th>
<th>Medium</th>
<th>Fine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation (m)</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thickness of soil organic horizon (cm)</td>
<td>0.0004</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterlogging</td>
<td>0.0393</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy height (m)</td>
<td>0.0011</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy coverage (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shrub coverage (%)</td>
<td>0.0621</td>
<td>0.0341</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Herb coverage (%)</td>
<td></td>
<td></td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>Trees 3 - 7.5 cm DBH</td>
<td>0.0083</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lianas 3 - 7.5 cm diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lianas 8 - 15 cm diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Interpretation: regression on the environmental variables

<table>
<thead>
<tr>
<th></th>
<th>Adiantum tomentosum</th>
<th>Very broad</th>
<th>Broad</th>
<th>Medium</th>
<th>Fine</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$ of PCNM submodel on A. tomentosum</td>
<td>0.333</td>
<td>0.239</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ of envir. on submodel</td>
<td>0.347</td>
<td>0.334</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ of envir. on A. tomentosum</td>
<td>0.436</td>
<td>0.116</td>
<td>0.080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thickness of soil organic horizon (cm)</td>
<td>0.0004</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterlogging</td>
<td>0.0393</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy height (m)</td>
<td>0.0011</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy coverage (%)</td>
<td>0.0621</td>
<td>0.0341</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shrub coverage (%)</td>
<td>0.0621</td>
<td>0.0341</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Herb coverage (%)</td>
<td>0.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trees 3 - 7.5 cm DBH</td>
<td>0.0083</td>
<td>0.0009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lianas 3 - 7.5 cm diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lianas 8 - 15 cm diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Interpretation:** regression on the environmental variables

<table>
<thead>
<tr>
<th></th>
<th>Adiantum tomentosum</th>
<th>Very broad</th>
<th>Broad</th>
<th>Medium</th>
<th>Fine</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$ of PCNM submodel on A. tomentosum</td>
<td>0.333</td>
<td>0.239</td>
<td>0.126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ of envir. on submodel</td>
<td>0.347</td>
<td>0.334</td>
<td>0.157</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ of envir. on A. tomentosum</td>
<td>0.436</td>
<td>0.116</td>
<td>0.080</td>
<td>0.020</td>
<td></td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thickness of soil organic horizon (cm)</td>
<td>0.0004</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterlogging</td>
<td>0.0393</td>
<td></td>
<td></td>
<td>0.0007</td>
<td></td>
</tr>
<tr>
<td>Canopy height (m)</td>
<td>0.0011</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy coverage (%)</td>
<td>0.0621</td>
<td>0.0341</td>
<td></td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Shrub coverage (%)</td>
<td></td>
<td></td>
<td></td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>Herb coverage (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trees 3 - 7.5 cm DBH</td>
<td>0.0083</td>
<td></td>
<td></td>
<td>0.0009</td>
<td></td>
</tr>
<tr>
<td>Lianas 3 - 7.5 cm diameter</td>
<td></td>
<td></td>
<td></td>
<td>0.0413</td>
<td></td>
</tr>
<tr>
<td>Lianas 8 - 15 cm diameter</td>
<td></td>
<td></td>
<td></td>
<td>0.0183</td>
<td></td>
</tr>
</tbody>
</table>

Spatial eigenfunction modelling
### Interpretation: regression on the environmental variables

<table>
<thead>
<tr>
<th></th>
<th>Adiantum tomentosum</th>
<th>Very broad</th>
<th>Broad</th>
<th>Medium</th>
<th>Fine</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$ of PCNM submodel on A. tomentosum</td>
<td>0.333</td>
<td>0.239</td>
<td>0.126</td>
<td>0.117</td>
<td></td>
</tr>
<tr>
<td>$R^2$ of envir. on submodel</td>
<td>0.347</td>
<td>0.334</td>
<td>0.157</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ of envir. on A. tomentosum</td>
<td>0.436</td>
<td>0.116</td>
<td>0.080</td>
<td>0.020</td>
<td></td>
</tr>
</tbody>
</table>

- Elevation (m): $< 0.0001 < 0.0001 < 0.0001$
- Thickness of soil organic horizon (cm): 0.0004 $< 0.0001$
- Waterlogging: 0.0393 0.0007
- Canopy height (m): 0.0011 $< 0.0001$
- Canopy coverage (%): 0.0621 0.0341 0.0001
- Shrub coverage (%): 0.0002
- Herb coverage (%): 0.0002
- Trees 3 - 7.5 cm DBH: 0.0083 0.0009
- Lianas 3 - 7.5 cm diameter: 0.0413
- Lianas 8 - 15 cm diameter: 0.0183

Spatial eigenfunction modelling
Use dbMEM in variation partitioning: *Adiantum tomentosum* at Nauta, Peru ($R^2_a$).
Scalogram of the fern *Adiantum tomentosum* multiscale structure along another transect called Huanta (Peru). Abscissa: the 129 dbMEM eigenfunctions with positive Moran’s $I$. Ordinate: absolute values of the $t$-statistics. The 26 eigenfunctions selected by forward selection ($p \leq 0.05$) are identified by black squares.
Example 2
Gutianshan forest plot in China

• Evergreen forest in Gutianshan Forest Reserve, Zhejiang Province.
• Fully-surveyed 24-ha forest plot in subtropical forest, 29°15'N.
• Plot divided into 600 cells of 20 m × 20 m.
• 159 tree species. Richness: 19 to 54 species per cell.
• Data collection: 2005


1 The Gutianshan forest plot is a member of the *Center for Tropical Forest Science* (CTFS). Details on the plot available at http://www.ctfs.si.edu/site/Gutianshan/.
Questions

• How much of the variation in species composition among sites (beta diversity) is spatially structured?
• Of that, how much is related to the environmental variables?

⇒ Four environmental variables developed in cubic polynomial form:
  Altitude: altitude, altitude\(^2\), altitude\(^3\)
  Convexity: convexity, convexity\(^2\), convexity\(^3\)
  Slope: slope, slope\(^2\), slope\(^3\)
  Aspect (circular variable): sin(aspect), cos(aspect)

(Soil cores collected in 2007. Soil chemistry data not available yet.)

⇒ 599 dbMEM eigenfunctions. 200 model positive spatial correlation.
⇒ Nearly all of them are significant: spatial variation at all scales.
• 63% of the among-cell variation \( (R_a^2) \) of the community composition (159 species) is spatially structured and explained by the 339 PCNM.

• Nearly half of that 63% is also explained by the four environmental variables. Soil chemistry to be added to the model when available.

• Scales of spatial variation: the dominant structure is broad-scaled.

⇒ Balance between neutral processes and environmental control.
**dbMEM eigenfunctions can be used in different ways**

a) We proceeded as follows in example 1 (ferns):

• dbMEM analysis of the response variable $y$;

• Division of the significant dbMEM eigenfunctions into submodels;

• Interpretation of the submodels using explanatory variables.

The objective was to divide the variation of $y$ into submodels and relate those to explanatory environmental variables.

The same type of analysis can be done with a multivariate matrix $Y$.

b) dbMEM eigenfunctions can also be used in the framework of variation partitioning, as in Example #2. The variation of $Y$ can be partitioned with respect to a table of explanatory variables $X$ and (for example) several tables $W_1$, $W_2$, $W_3$, containing dbMEM submodels.
dbMEM eigenfunctions can be used in different ways (cont.)

c) dbMEM can be used to model spatial and temporal variation in the study of spatio-temporal data, and test for the space-time interaction.¹

d) dbMEM can efficiently model spatial structures in data. They can be used to control for spatial autocorrelation in tests of significance of the species-environment relationship (fraction [a]).²


=> See section on Space-time interaction in this course.

3. Generalized MEM analysis

The Moran’s eigenvector maps (MEM) method is a generalization of dbMEM (PCNM) to different types of spatial weights. The result is a set of spatial eigenfunctions, as in dbMEM analysis.

Eigen-decomposition of a spatial/temporal weighting matrix \( W \)

\[
W = B = 0/1 \text{ connectivity matrix among sites} \star \text{ Hadamard product} \star A = \text{edge weighting matrix}
\]

Dray, Legendre & Peres-Neto (2006); Legendre & Legendre (2012); Borcard et al. (2018).
Diagonal values \((D_{ii})\) are 0, indicating that a site is connected to itself.

Classical PCNM eigenfunctions can be generated with function `pcnm()` of the vegan package.
Difference between classical PCNM and dbMEM

\[ D'_\text{PCNM} = \begin{pmatrix}
0 & d_{12} & 4t & 4t \\
 d_{21} & 0 & & \\
 4t & 0 & & \\
 4t & 0 & & \\
\end{pmatrix} \]

Diagonal values \( (D_{ii}) \) are 0, indicating that a site is connected to itself.

\[ D'_\text{dbMEM} = \begin{pmatrix}
4t & d_{12} & 4t & 4t \\
 d_{21} & 4t & & \\
 4t & 4t & & \\
 4t & 4t & & \\
\end{pmatrix} \]

Diagonal values \( (D_{ii}) = 4 \times \text{threshold} \): a site is not connected to itself.
The eigenvalues become proportional to Moran’s \( I \).

The eigenvalues are different but the eigenvectors (i.e. the spatial eigenfunctions) are the same.
Another improvement in the dbmem() function of adespatial –

In regular PCoA, the eigenvectors are scaled to a norm of $\sqrt{\lambda}$, which is an imaginary number when $\lambda$ is negative.

Principal coordinates with negative eigenvalues thus contain complex numbers.

$=>$ In the PCoA calculation that is part of the construction of MEM eigenfunctions, the eigenvectors computed by function eigen() are not multiplied by the square roots of their eigenvalues, $\sqrt{\lambda}$.

They are kept with the norm=1 that they received from eigen(). [That norm is multiplied by $\sqrt{n}$ in function dbmem().]

Hence they are not genuine principal coordinates.

This does not change their explanatory powers in multiple linear regression or RDA.
Rationale –

Explanatory variables can be scaled by any linear transformation in linear models. It does not change the fitted values they produce.

Indeed, the regression coefficients have dimensions that transform the contributions of the explanatory variables to the dimension of the response variable they model. [See the course on multiple regression].

For that reason, a principal coordinate with a negative eigenvalue, which contains complex numbers, can be post-divided by sqrt(λ), giving it back its norm of 1.

Eigenfunctions with norms of 1 now contain real numbers, even those that have negative eigenvalues.

It is thus easier to use them as explanatory variables in functions \texttt{lm()}, \texttt{rda()}, \texttt{varpart()}, etc. It also makes the calculation of the fitted values easier to compute.
Different forms of Moran’s Eigenvector Maps (generalized MEM) can be constructed (Dray et al. 2006):

• Binary MEM: double-centre matrix $B$, then compute its eigenvalues and eigenvectors. This is Griffith’s spatial filtering method.

• To obtain dbMEM, matrix $A$ contains the distances.

• Replace matrix $A$ by some function of the distances.

• Replace $A$ by some other weights, e.g. resistance of the landscape.

$$W = \begin{array}{c}
B = 0/1 \\
\text{connectivity matrix among sites}
\end{array} \quad \text{Hadamard product} \quad \begin{array}{c}
A = \text{edge weighting matrix}
\end{array}$$
4. Asymmetric eigenvector maps (AEM)

Spatial eigenfunction method developed to model multivariate (e.g. species communities, genetic data) spatial distributions generated by an asymmetric, directional physical process.\(^1\)

AEM can also be applied to time series. Temporal processes are asymmetric.


See also Legendre & Legendre (2012, Section 14.3); Borcard et al. (2018, Section 7.4.5).
Nodes-by-edges matrix $E$

<table>
<thead>
<tr>
<th>Nodes</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
<th>E4</th>
<th>E5</th>
<th>E6</th>
<th>E7</th>
<th>E8</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N2 (Lake 6)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N3 (Lake 3)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N4 (Lake 2)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N6 (Lake 1)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N7 (Lake 4)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>N8 (Lake 5)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

$w =$ vector of weights

PCA, SVD or PCoA

(c) 

Y 

AEM

Response data  Spatial eigenfunctions

Figures 14.11 from Legendre & Legendre 2012
For 8 nodes, 7 AEM eigenfunctions are produced.

Shades of grey represent the values in each eigenvector, from white (largest negative value) to black (largest positive value). Signs are arbitrary; they may be reverted with no consequence for the analysis.

Figure 14.12 from Legendre & Legendre 2012
Example – AEM analysis of data from a river network

42 lakes of the Mastigouche Reserve, Québec, Canada

Diet composition in 20 stomach contents of brook trout (*Salvelinus fontinalis*) in each lake.

Response variables for each lake: percent wet mass for nine functional prey categories (mean across the 20 fish)

- zoobenthos
- amphipods
- zooplankton
- dipteran pupae
- aquatic insects
- terrestrial insects
- prey-fish
- leeches
- other prey
**Question** – Is diet variation related to the genetics of the populations of trout that successively invaded the river network after the last glaciation?

The nodes-by-edges matrix $E$ was constructed with 42 nodes and 65 edges. No weights (e.g. inverse of distances) were placed on the edges. See Blanchet *et al.* (2008, Table 2).

The AEM eigenfunctions, representing the structure of the network, explained $R^2_{\text{adjusted}} = 0.636$ of the variation in trout diet among lakes. Classical dbMEM analysis based on geographic distances among lakes only explained $R^2_{\text{adjusted}} = 0.199$ of that variation.

$\Rightarrow$ The trout population variation among lakes (reflected in differences in diet) is better explained by the AEM eigenfunctions (directional model).

$\Rightarrow$ A portion of the variation was non-directional.
Three other applications of AEM analysis to spatial data –


Here is one of them =>
Example 2 – Dominique Monti, UAG

*Atya innocuous*

94 sites were sampled in Rivière Capesterre
93 AEM variables were constructed based on this connection diagram; 38 measured positive autocorrelation. 12 were selected.
AEM model: $R^2_{adj} = 59.8\%$ using 12 selected AEMs
5. AEM analysis of time series

Example: 10 sampling units along time, equal spacing –

Nodes-by-edges matrix $E$:

<table>
<thead>
<tr>
<th></th>
<th>E0</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
<th>E4</th>
<th>E5</th>
<th>E6</th>
<th>E7</th>
<th>E8</th>
<th>E9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time.1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time.2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time.3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time.7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Time.10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Nine AEM eigenfunctions are produced by PCA of matrix $E$ –

• 4 AEM modelling positive temporal correlation
• 5 AEM modelling negative temporal correlation
Eigenfunction analysis of multivariate time series

Analysis of multivariate community composition data:
A full Practical exercise in R about temporal eigenfunction analysis of the Chesapeake Bay Benthic Monitoring Program (USA) is available as Appendix S2 of a paper by Legendre & Gauthier (2014).

The Appendix is entitled:

Temporal eigenfunction methods – Practicals in R
**AEM and MEM analyses are complementary**

- To emphasize the directional nature of the process influencing the data, **AEM analysis**, which was designed to take trends into account, should be applied to the non-detrended series.

- **MEM analysis** can be applied to data series that were detrended to remove the directional component.

⇒ By applying both methods to spatial data, one can differentiate the directional and non-directional components of variation in a [multivariate] data matrix.

⇒ Processes acting along time series produce a single gradient in data. For time series, MEM and AEM modelling can both be applied to the undetrended data, even when a trend is present.

⇒ Example: **detrended** palaeoecological core data could be studied by MEM analysis, the **undetrended** data by MEM or AEM analysis.
6. Other methods that use spatial eigenfunctions

• Multi-scale ordination (MSO): Are explanatory variables responsible for the spatial correlation observed in multivariate data \( Y \), e.g. community composition data?

• Test the space-time interaction in space-time community surveys: code the space and time factors by MEM and compute the interaction terms, which can be tested.\(^2\)

• Multiscale codependence analysis\(^3,4\): at what scales are two var. / mat correlated? What is their correlation at each spatial/temporal scale?

---


On CRAN page: http://cran.r-project.org

**adespatial package**: functions
- eigenfunctions: mem(), dbmem(), aem(), create.dbMEM.model();
- variable selection: forward.sel();
- scalogram: scalogram();
- space-time interaction: stimodels();
- time series analysis: WRperiodogram(), mfpa(), Cperiodogram;

**vegan package**: function varpart() for multivariate variation partitioning;
- ordistep() and ordiR2step() : forward selection of explanatory variables.
- pcnm(): construction of classical PCNM for RDA and CCA;
- mso(): multiscale ordination;

**codep package**: multiscale codependence analysis.
8. References

Books describing MEM and AEM theory


Papers on MEM and AEM theory available in pdf at http://numericalecology.com/


Papers on MEM + AEM theory available in pdf at http://numericalecology.com/ (continued)


Other references mentioned in this presentation


End of the presentation