Partitioning spatial and temporal dynamics at multiple scales

Pierre Legendre and Daniel Borcard
Département de sciences biologiques
Université de Montréal

Pierre.Legendre@umontreal.ca
Daniel.Borcard@umontreal.ca

Organized Oral Session 53 “*Days to decades and meters to miles*”
ESA-INTECOL 2005 Joint Meeting

1. The general problem

One of the questions addressed by our Working Group was:

Do the spatial and temporal variabilities of ecological processes change in some predictable way with scale?

To answer that question,

• we need a method to compare the variability in community species composition (multivariate data), first between space and time, and then at multiple spatial scales,

• considering the fact that space-time ecological studies are usually done without replication.

This talk will describe a statistical method to achieve that.

2. Two-way anova for space and time crossed factors by canonical analysis (RDA)

Canonical redundancy analysis (RDA)

The most common application of RDA in ecology is to test the relationship between a response **Y** and explanatory variables **X**:

• Simple RDA

Response table	Explanatory table
Y Community composition data	**X** Example: Environmental variables

• Partial RDA

Response table	Explanatory table	Covariables
Y Community composition data	**X** Example: Environmental variables	**W** Example: Spatial base functions

RDA as multivariate anova

RDA can also be used to test the relationship between **Y** and one or several experimental factors (crossed balanced designs). It is then a form of multivariate anova[1].

[1] Legendre, P. and M. J. Anderson. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* 69: 1-24.

The anova factors can be coded in two different ways:

(1) Coding for a single factor: binary dummy variables

Matrix X

Factor A
(3 levels)

$$\left[\begin{array}{ccc}
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
\hline
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
\hline
0 & 0 & 1 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
0 & 0 & 1
\end{array}\right]$$

(2) Coding for two crossed factors and their interaction: orthogonal dummy variables, also called Helmert contrasts

Matrix X

Interaction
AB

$$\left[\begin{array}{cc}
+2 & 0 \\
+2 & 0 \\
+2 & 0 \\
\hline
-2 & 0 \\
-2 & 0 \\
-2 & 0 \\
\hline
-1 & 1 \\
-1 & 1 \\
-1 & 1 \\
\hline
1 & -1 \\
1 & -1 \\
1 & -1 \\
\hline
-1 & -1 \\
-1 & -1 \\
-1 & -1 \\
\hline
1 & 1 \\
1 & 1 \\
1 & 1
\end{array}\right]$$

Matrix W (covariables)

Factor A (3 levels) | Factor B (2 levels)

$$\left[\begin{array}{cc|c}
+2 & 0 & 1 \\
+2 & 0 & 1 \\
+2 & 0 & 1 \\
\cline{3-3}
+2 & 0 & -1 \\
+2 & 0 & -1 \\
+2 & 0 & -1 \\
\hline
-1 & 1 & 1 \\
-1 & 1 & 1 \\
-1 & 1 & 1 \\
\cline{3-3}
-1 & 1 & -1 \\
-1 & 1 & -1 \\
-1 & 1 & -1 \\
\hline
-1 & -1 & 1 \\
-1 & -1 & 1 \\
-1 & -1 & 1 \\
\cline{3-3}
-1 & -1 & -1 \\
-1 & -1 & -1 \\
-1 & -1 & -1
\end{array}\right]$$

- One factor

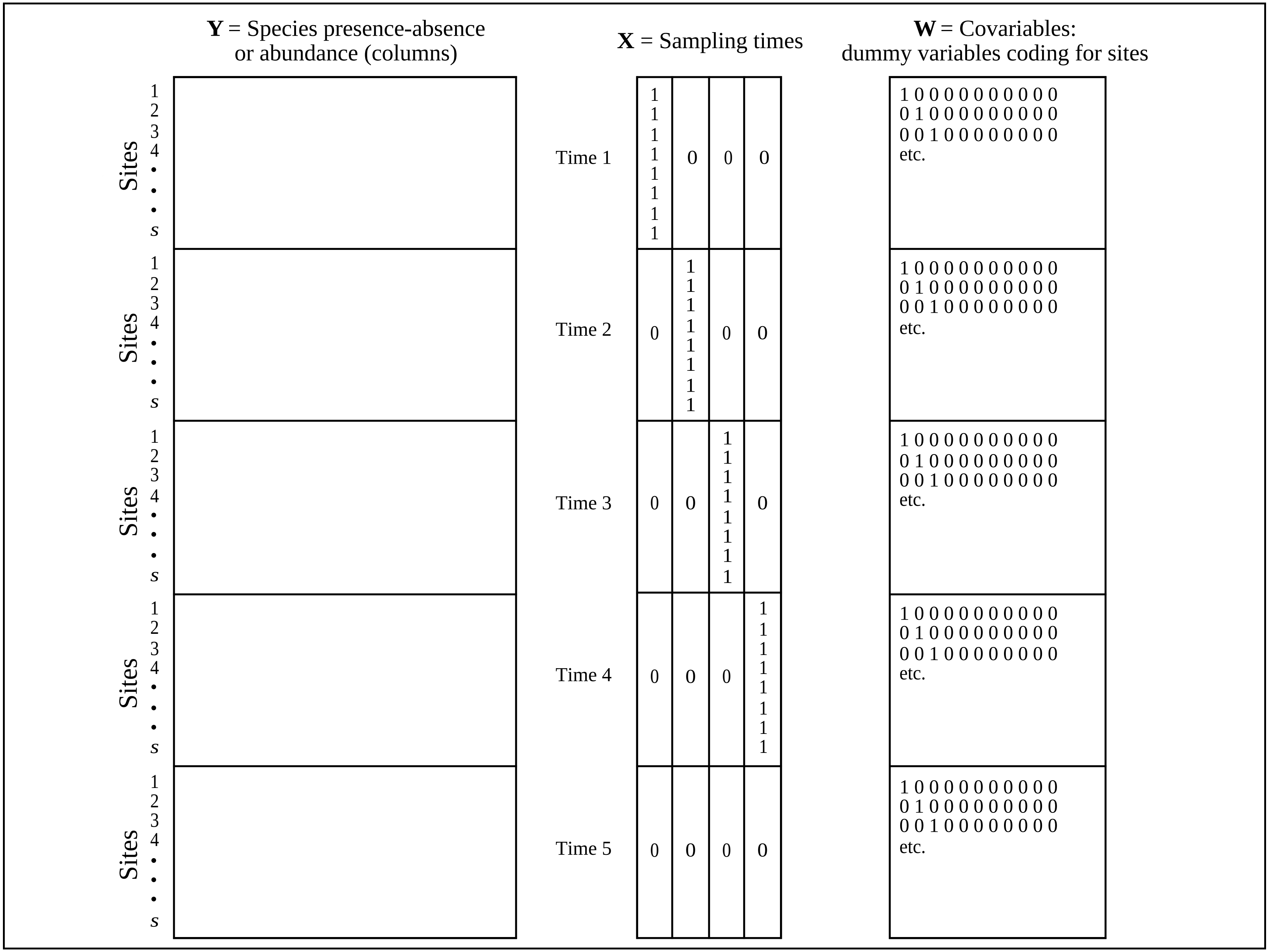
Response table	Explanatory table
Y Community composition data	**X** Factor to be tested

- Two or more crossed factors. Example: test of an interaction

Response table	Explanatory table	Covariables
Y Community composition data	**X** Interaction to be tested	**W** Main factors and other interactions

Analyze Y against space and time without replication

First method: write tables coding for space and time using dummy variables. Example:


Y = Species presence-absence or abundance (columns)
X = Sampling times
W = Covariables: dummy variables coding for sites
Sites
1 2 3 4 • • • s
Time 1
Time 2
Time 3
Time 4
Time 5
1 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0
etc.

Why can't we test the space-time interaction?

	Dummy variables coding	Example dummy var. $s = 30$, $t = 10$
S: s sites	$s - 1$	29
T: t times	$t - 1$	9
S x T interaction	$(s - 1)\,(t - 1)$	261
Total coding variables	$st - 1$	299
Total d.f.	$st - 1$	299
***F*-statistic** for test of interaction		
d.f. numerator = m	$(s - 1)\,(t - 1)$	261
d.f. denominator	0	0

We would still like to test the space-time interaction…

… because a significant interaction would indicate

- that the temporal structures differ from site to site,
- or that the spatial structures differ from time to time.

If the interaction was significant, we should carry out separate analyses of the temporal variance for the different points in space, or separate analyses of the spatial variance for the different times.

The absence of a significant interaction would indicate

- either that the differences among times can be modelled in the same way at all points in space, and conversely;
- or that there were not enough data to obtain a significant result for the test of the interaction (*n* too small, lack of power; type II error).

3. How can we test the space-time interaction in analyses of Y against space and time without replication?

Analyze Y against space and time without replication

Using dummy variables to code for space and time, we did not have enough degrees of freedom, in the no-replication case, to test the S-T interaction.

We can solve that problem by using a more parsimonious way of coding for space and time.

We are proposing to use distance-based eigenvector maps (DBEM), and in particular PCNM[1] base functions which are a type of DBEM[2].

[1] Borcard, D. and P. Legendre. 2005. Using distance-based eigenvector maps (DBEM) in multivariate variation partitioning. Part 1: PCNM (principal coordinates of neighbor matrices), theory and applications. Special Session "*Spatial Statistics at Multiple Scales*", ESA-INTECOL 2005 Joint Meeting, Palais des Congrès, Montréal, August 9, 2005.

[2] Dray, S. 2005. Spatial modeling: a comprehensive framework for distance-based eigenvector maps (DBEM). Special Session "*Spatial Statistics at Multiple Scales*", ESA-INTECOL 2005 Joint Meeting, Palais des Congrès, Montréal, August 9, 2005.

PCNM base functions[1] represent a spectral decomposition of the spatial (or temporal) relationships among sampling sites (or times).

• They are orthogonal to one another,

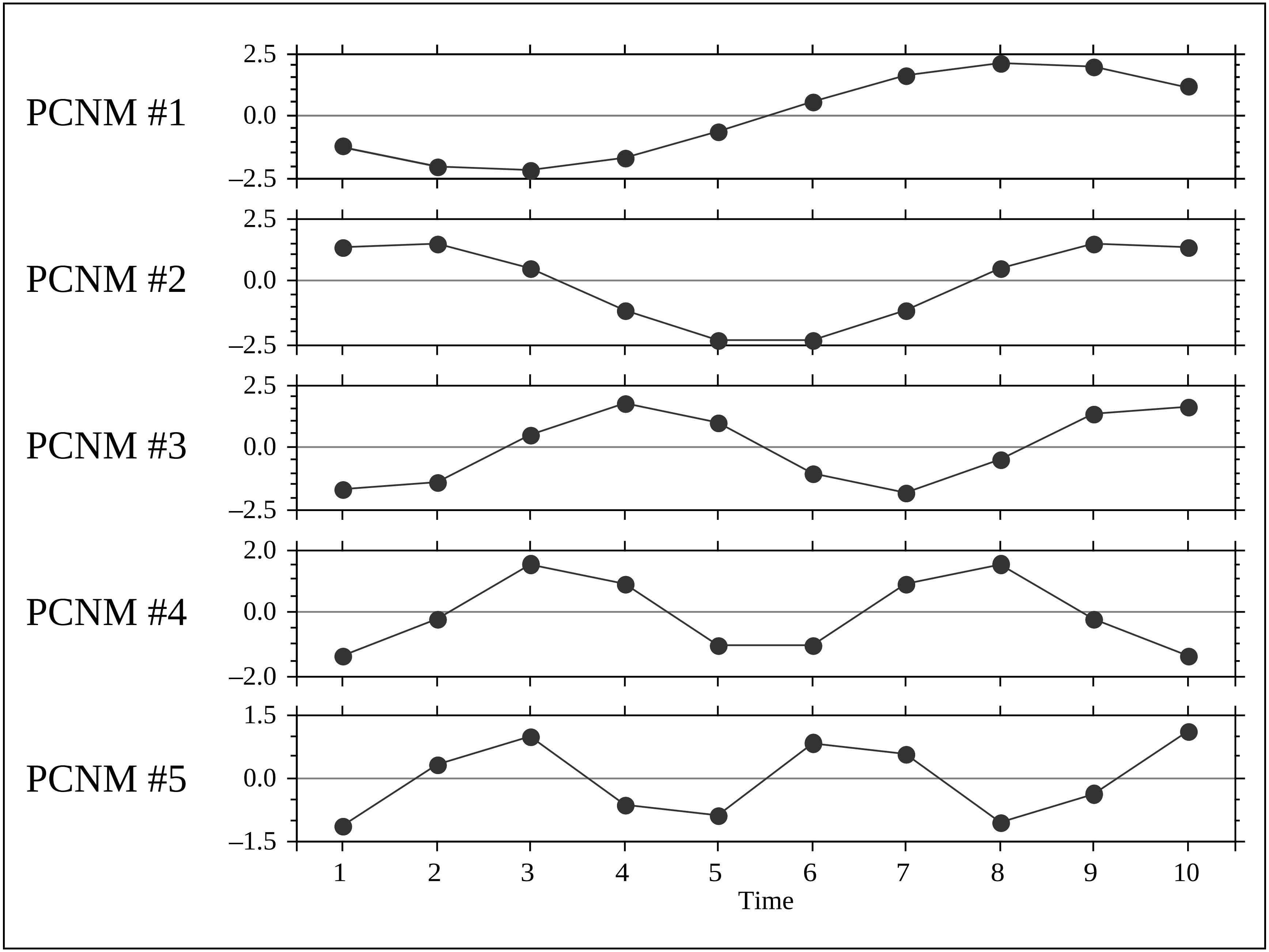
• and fewer in number than dummy variables coding for the same sites (or times).

To model the Space and Time variation, we will use $s/2$ or $t/2$ PCNM functions — actually: round$(s/2 + 0.5)$ and round$(t/2 + 0.5)$.

For example, 10 equispaced sampling times are modelled by the following 5 PCNM functions:

[1] Borcard, D. and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153: 51-68.

Borcard, D., P. Legendre, C. Avois-Jacquet and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85: 1826-1832.


PCNM #1
PCNM #2
PCNM #3
PCNM #4
PCNM #5
2.5
0.0
–2.5
2.0
–2.0
1.5
–1.5
1
2
3
4
5
6
7
8
9
10
Time

Sampling each point in space (S) during each sampling campaign (T) creates an orthogonal design.

For that reason, the PCNM base function, which are orthogonal within each set (S, T), are also orthogonal between sets.

The S-T interaction can be modelled by creating variables that are the products of each S-PCNM by each T-PCNM.

The S-T interaction variables are orthogonal to the S-PCNMs and the T-PCNMs.

Orthogonality of factors
is happiness for statisticians!

Testing interaction in cross-designs without replication

While it takes (s–1) dummy variables to represent s sites, fewer PCNM variables are necessary to analyze the spatial variation; likewise for time.

	Dummy variables coding	Example dummy var. $s = 30$, $t = 10$	PCNM (or DBEM) base functions	Example PCNM $s = 30$, $t = 10$
S: s sites	$s - 1$	29	$u = \text{round}(s/2 + 0.5)$	15
T: t times	$t - 1$	9	$v = \text{round}(t/2 + 0.5)$	5
S x T interaction	$(s - 1)(t - 1)$	261	uv	75
Total coding variables	$st - 1$	299	$u + v + uv$	95
Total d.f.	$st - 1$	299	$st - 1$	299
***F*-statistic** for test of interaction				
d.f. numerator = m	$(s - 1)(t - 1)$	261	uv	75
d.f. denominator	0	0	$(st-1) - (u + v + uv)$	204

Simulated univariate data

1. Data with S-T interaction

• Transect of s = 25 points, t = 5 sampling campaigns.

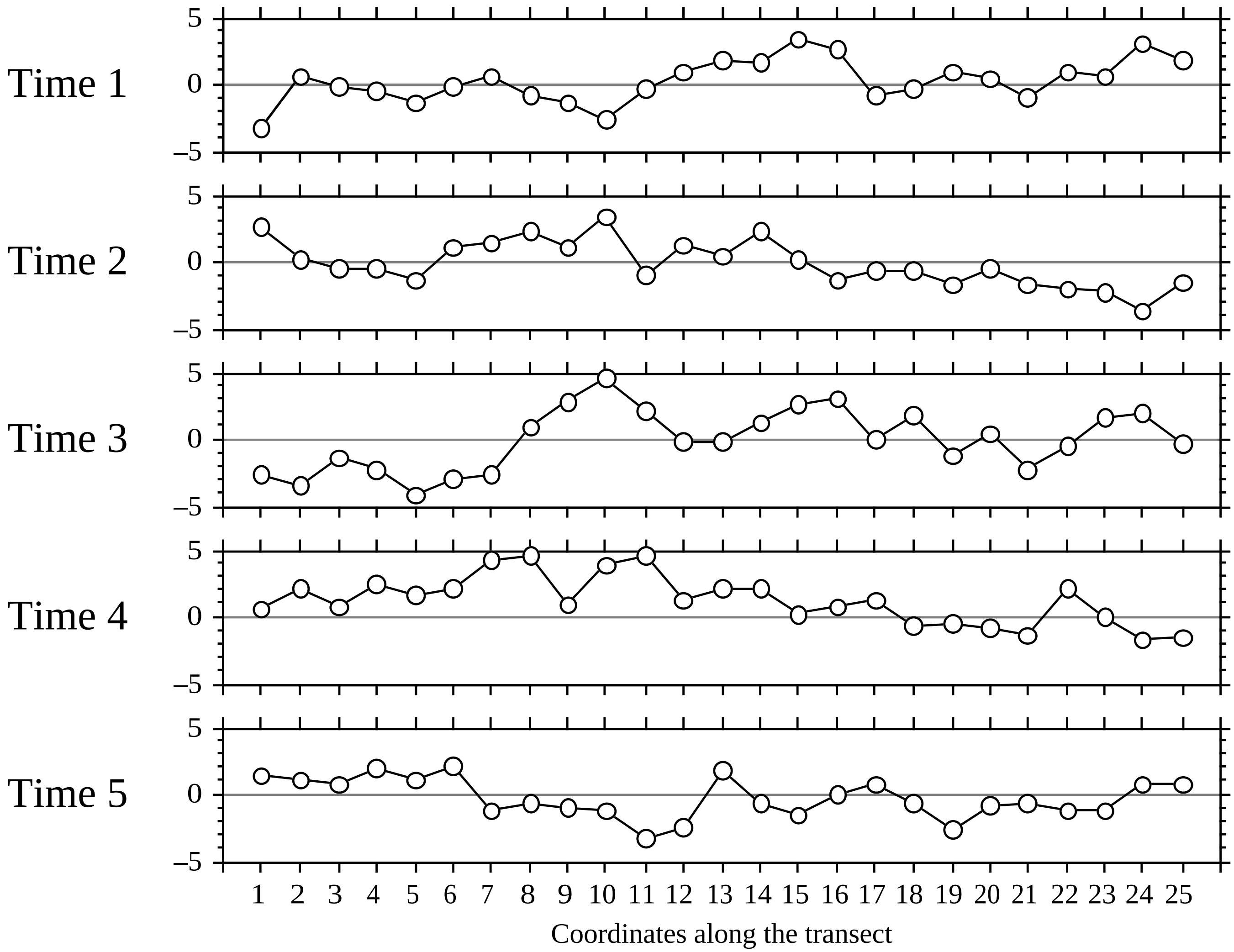
• Spatially autocorrelated data were generated in a 100 x 100 pixel field using the program SimSSD[1]. A transect of 25 equidistant points (spacing = 4 units) was sampled in the middle of the field. A transect variable was the sum of two simulated vectors, with spatial ranges of 10 and 30 units respectively.

• 5 independent "time" realizations of the transect were created.

[1] Legendre, Dale, Fortin, Gurevitch, Hohn and Myers. 2002. *Ecography* 25: 601-615.

Legendre, Dale, Fortin, Casgrain and Gurevitch. 2004. *Ecology* 85: 3202-3214.

Legendre, Borcard and Peres-Neto. 2005. *Ecology* (in press).



There is an S-T interaction because the 5 "time" realizations were created independently of one another.

• 13 S-PCNM functions were created to model the spatial variation along the 25 points of the transect.

• 3 T-PCNM functions were created to model the temporal variation across the 5 sampling times.

• To model the interaction, 39 ST functions were obtained by multiplying each S-PCNM by each T-PCNM.

Canonical RDA was used to test the interaction in the presence of the main factors S and T.

Results

The S-T interaction was significant: $p = 0.0288$ (after 9999 perm.)

Correct answer

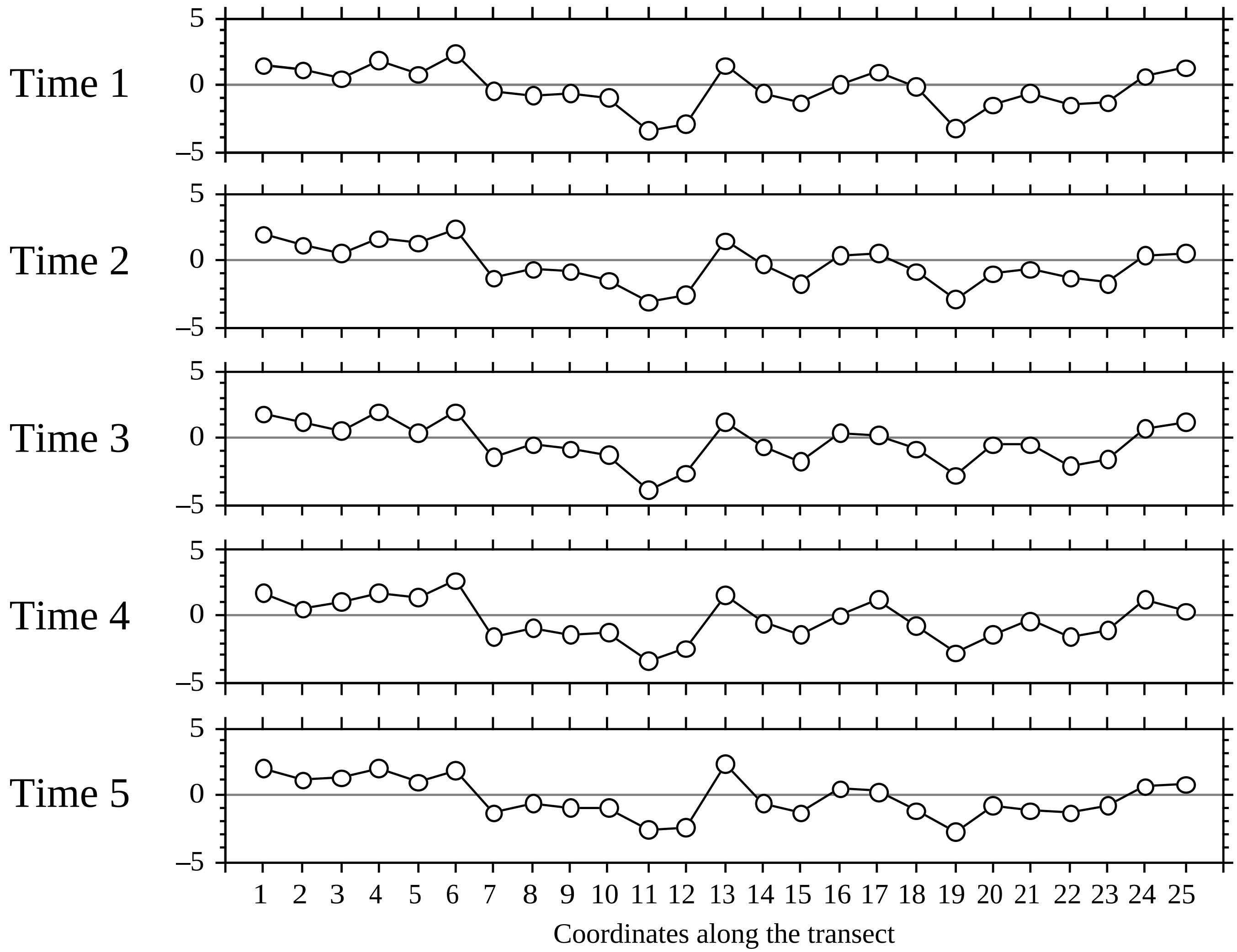
The spatial structures differed from time to time. We could test the spatial structure of each sampling time separately.

Simulated univariate data

2. Data without S-T interaction

• Transect of s = 25 points, t = 5 sampling campaigns.

• We took one of the transects (Time 5) from the previous data set and created 5 new "time" replicates by adding $N(0, 0.3)$ error to all data values.



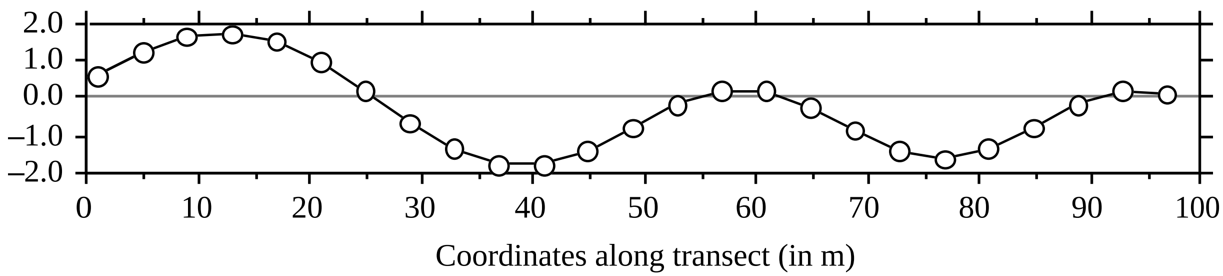
There is no S-T interaction because the 5 “time” realizations were all constructed from the same, common spatial structure.

Results

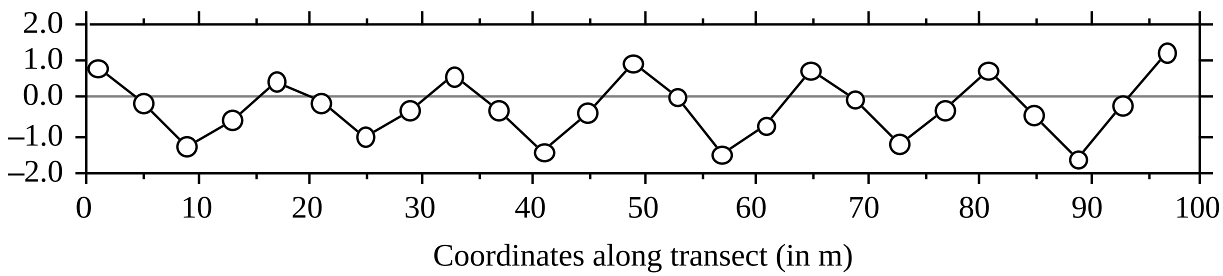
- The S-T interaction was not significant: $p = 1.0000$ (9999 perm.)
- The main factor Time was not significant: $p = 0.9473$ (9999 perm.)
- The main factor Space was significant: $p = 0.0001$ (9999 perm.)

Correct answer

Broad-scale model
PCNM #1-7



Fine-scale model
PCNM #8-13



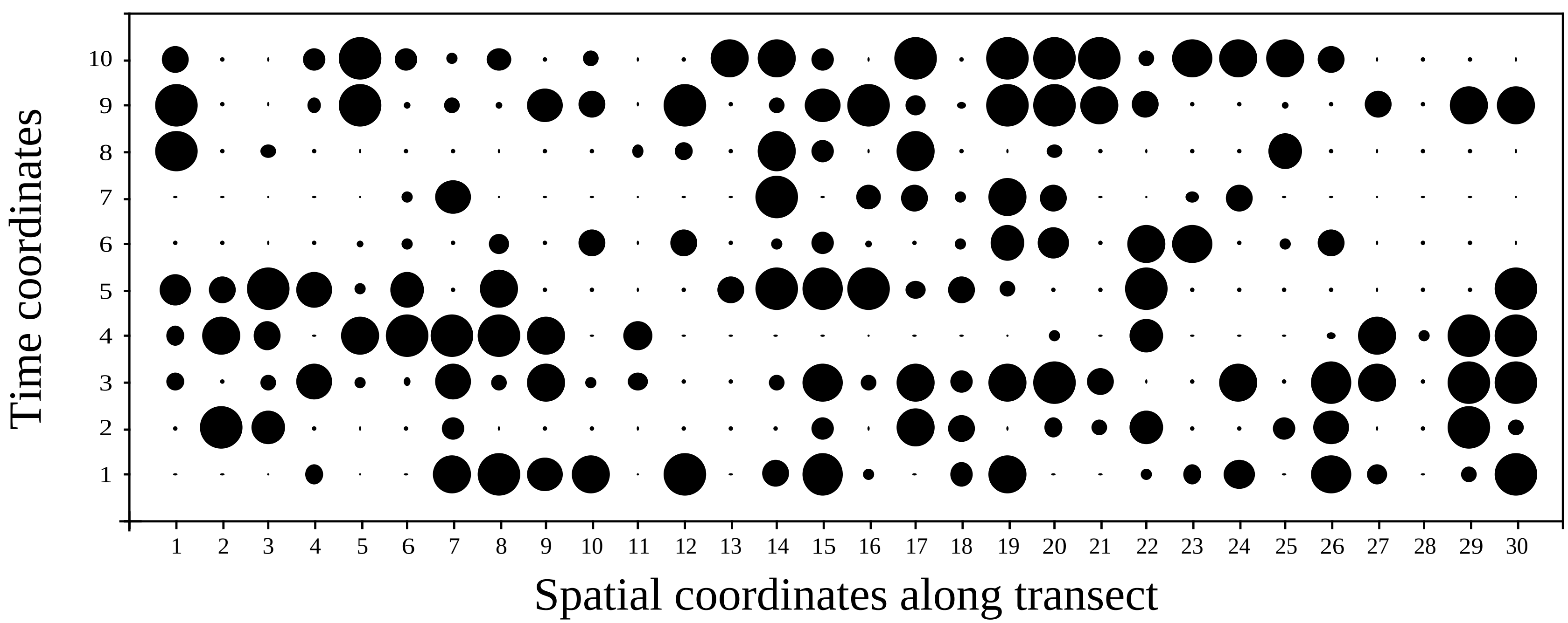
Simulated multivariate data

3. Autocorrelation in S and T

• Transect of s = 30 points, t = 10 sampling campaigns.

• 5 autocorrelated species-like variables with random error were generated in a 10 x 30 pixel field using the program SimSSD[1]. The abscissa of the field represents space, the ordinate represents time.

Example: species #4. Variogram range along abscissa (space): 5. Range along ordinate (time): 2



- 14 S-PCNM functions were created to model the spatial variation along the 30 points of the transect.
- 5 T-PCNM functions were created to model the temporal variation across the 10 sampling times.
- To model the interaction, 75 ST functions were obtained by multiplying each S-PCNM by each T-PCNM.

Canonical RDA was used to test the interaction in the presence of the main factors S and T.

Then, each main factor was tested in the presence of the other factor and the interaction variables.

The presence of autocorrelation in both S and T means that the T structure should differ among positions in S, and conversely. In other words, we expect to find a significant S-T interaction.

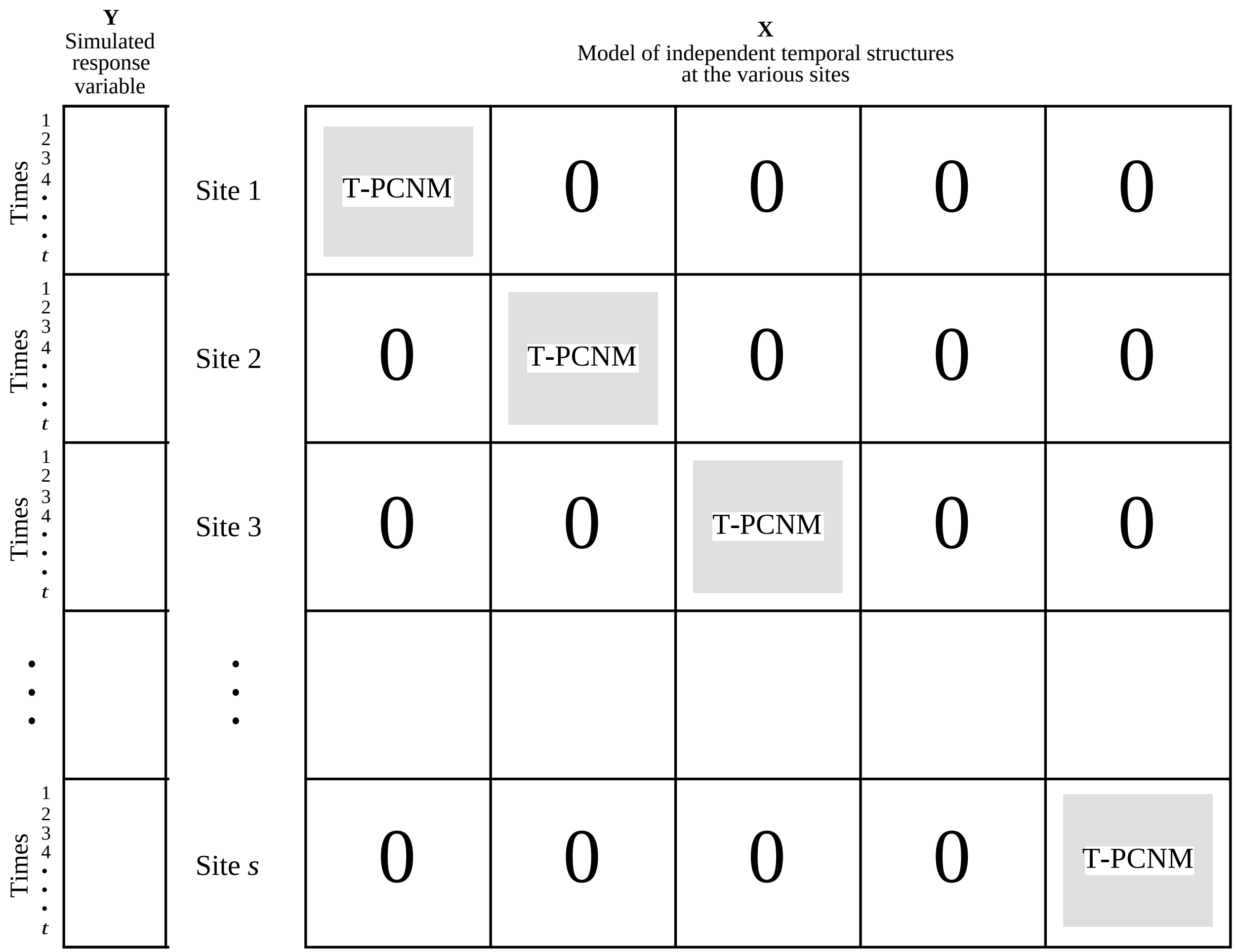
Results

- The S-T interaction was significant: p = 0.0001 (9999 perm.)

Correct answer

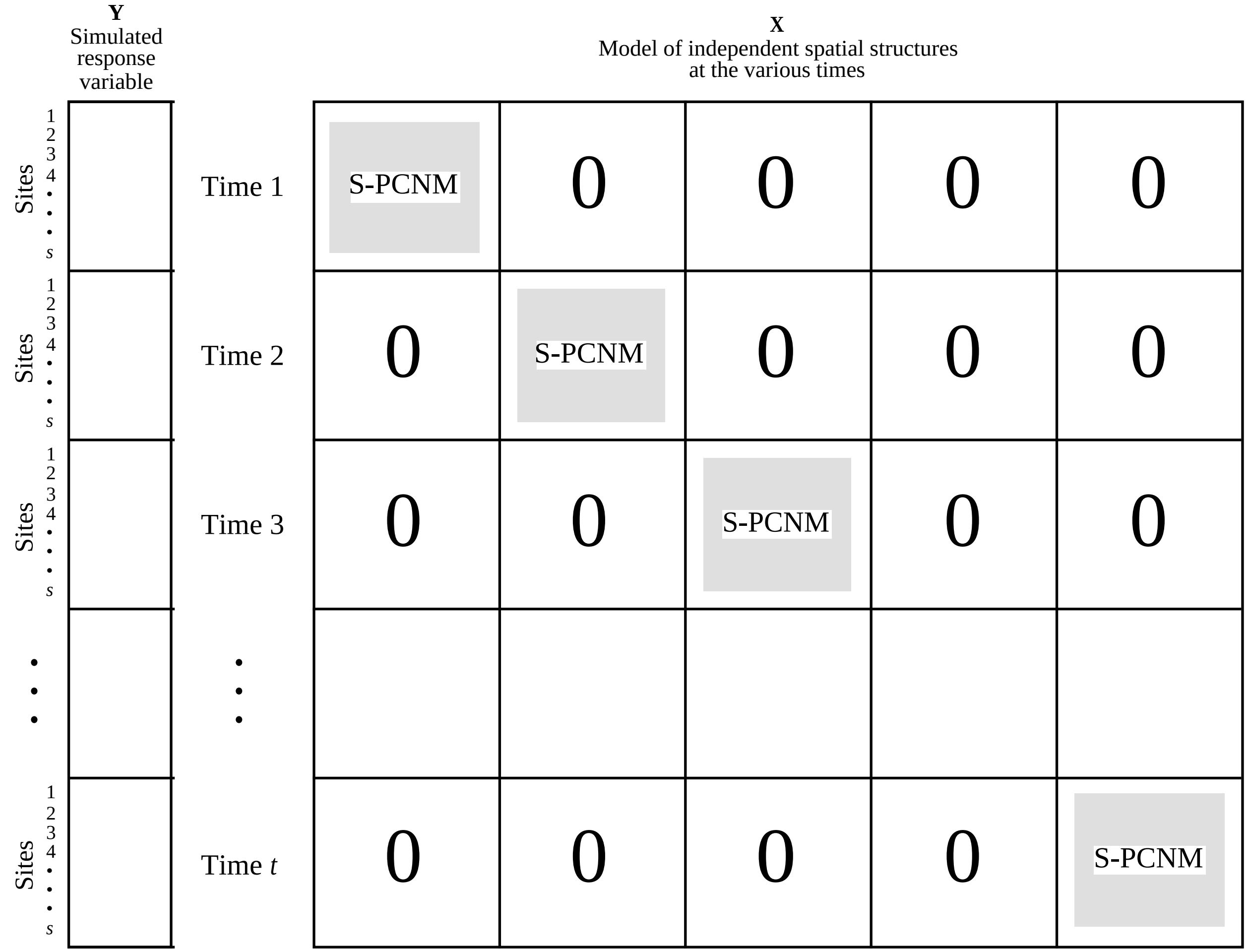
The spatial structures differed from time to time. We could test the spatial structure of each sampling time separately; or, better…

We can test the significance of the temporal structures that were independently generated at the 30 sites along the transect, as follows:



Result: p = 0.0044 (block permutations in Canoco, 9999 perm.)
The independent temporal structures, at the 30 sites, are significant.

Likewise, we can test the significance of the spatial structures that were independently generated at the 10 sampling times, as follows:



Result: p = 0.0001 (block permutations in Canoco, 9999 perm.)

The independent spatial structures, at the 10 times, are significant.

Real multivariate data

4. Servilleta rodent data[1,2]

• 9 sites (irregularly spaced), 15 consecutive sampling years; n = 135.

• Abundance of 23 rodent species.

[1] Serviletta Long-term, Ecological Research site (LTER), New Mexico.

[2] Ernest, S. K. M., J. H. Brown and R. R. Parmenter. 2000. Rodents, plants, and precipitation: spatial and temporal dynamics of consumers and resources. *Oikos* 88: 470-482.

Real multivariate data

4. Servilleta rodent data

• 9 sites (irregularly spaced), 15 consecutive sampling years; n = 135.

• Abundance of 23 rodent species.

• 8 orthogonal dummy variables were created to model the spatial variation among the 9 sites. Contrary to PCNMs, dummy variables (orthogonal or not) do not contain any particular hypothesis of spatial organization of the sites.

• 8 T-PCNM functions were created to model structured temporal variation across the 15 sampling years.

• To model the interaction, 64 ST functions were obtained by multiplying each S-PCNM by each T-PCNM.

That left 54 degrees of freedom in the denominator of the *F*-statistic for the test of the interaction.

Real multivariate data

4. Servilleta rodent data

- 9 sites (irregularly spaced), 15 consecutive sampling years; n = 135.
- Abundance of 23 rodent species.

Canonical RDA was used to test the interaction in the presence of the main factors S and T.

Then, each main factor was tested in the presence of the other factor and the interaction variables.

Test results (999 permutations)

	Raw species abundances	*Hellinger-transformed abundances*
S-T interaction	$p = 1.000^{NS}$	$p = 0.080^{NS}$
Main factor Space	$p = 0.001^{***}$	$p = 0.001^{***}$
Main factor Time	$p = 0.002^{***}$	$p = 0.001^{***}$

The non-significant interaction indicates that:

• the significant spatial variation is common to all years;

• the significant temporal structure is common to all sites.

Based on the results for the Hellinger-transformed species data,

- There is common significant spatial variation among the 9 trapping sites, accounting for 58.7% of the variation in the species table.
- There is common significant temporal variation among the 15 years, accounting for 7.6% of the variation in the species table.

Since that temporal variation is modelled by PCNM functions, we could illustrate it in graphs. We could carry out a partial RDA, retrieve the values along the canonical axes, and plot them as a function of time.

Detailed interpretation

Forward selection of significant variables (dummy variables for space, PCNM for time) should indicate the spatial or temporal scales at which important processes occur.

Discussion

Power is the ability of the method to identify an interaction when there is one in the data.

We need to repeat the simulations a large number of times, with different combinations of $s = \{5, 10, 30, 100\}$ and $t = \{5, 10, 30\}$, different spatial distributions of the sites, different amounts of autocorrelation, and different numbers and types of response variables, to assess the power of this method.

A PDF of this talk is available on

http://biol10.biol.umontreal.ca/ESA_SS/

The End