Suivi spatio-temporel des écosystèmes : tester l'interaction espace-temps pour identifier les impacts sur les communautés

Community surveys through space and time: testing the space–time interaction

Pierre Legendre

Département de sciences biologiques, Université de Montréal

Pierre.Legendre@umontreal.ca

Conférence Ifremer

Institut français de recherche pour l'exploitation de la mer (IFREMER)

Plouzané, Bretagne, 20 septembre 2012

1. The general problem

- Ecologists sample portions of the environment repeatedly across time in order to test hypotheses about changes in the environment induced by man, including climatic change.

- Problem: space-time ecological studies are usually done without replication.

1. The general problem

• This talk will describe a method for testing a space–time interaction in repeated ecological survey data, when there is no replication at the level of individual sampling units (sites).

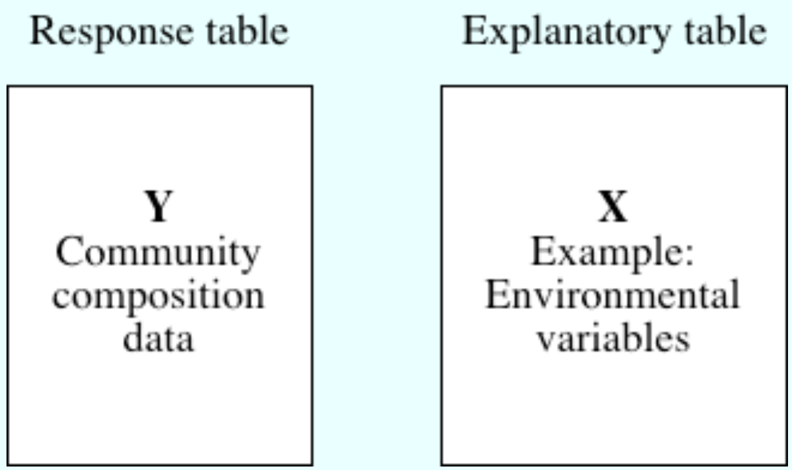
• This methodological development is important for the analysis of long-term monitoring data, including systems under anthropogenic influence. In these systems, an interaction may indicate that the spatial structure of the community composition has changed in the course of time or that the temporal evolution is not the same at all sites.

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

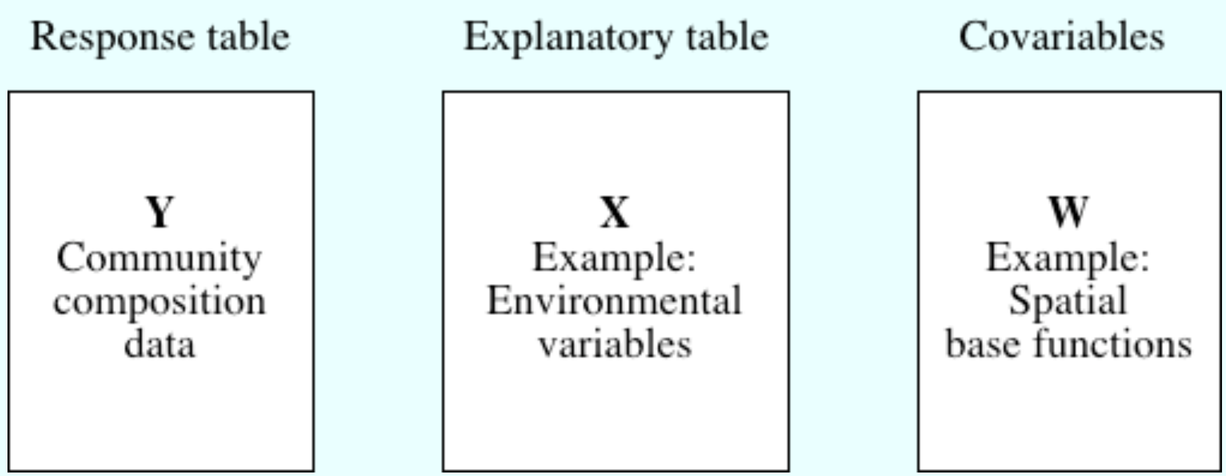
Canonical redundancy analysis (RDA)

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

- Simple RDA



- Partial RDA



RDA as multivariate anova

RDA can be used to test the relationship

• between a response data matrix **Y** containing, for example, community composition data,

• and one or several experimental factors (crossed balanced designs) in matrix **X**. 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)

1	0	0
1	0	0
1	0	0
1	0	0
0	1	0
0	1	0
0	1	0
0	1	0
0	0	1
0	0	1
0	0	1
0	0	1

The anova factors can be coded in two different ways:

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

Matrix X

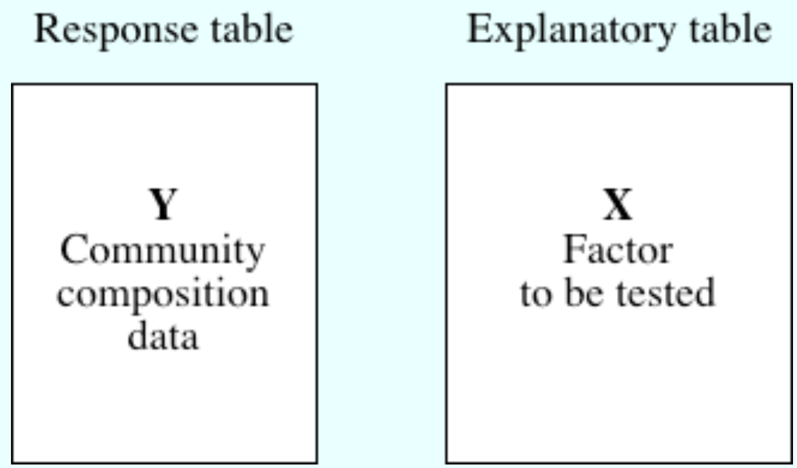
Factor *A* (3 levels)

1	0	0
1	0	0
1	0	0
1	0	0
0	1	0
0	1	0
0	1	0
0	1	0
0	0	1
0	0	1
0	0	1
0	0	1

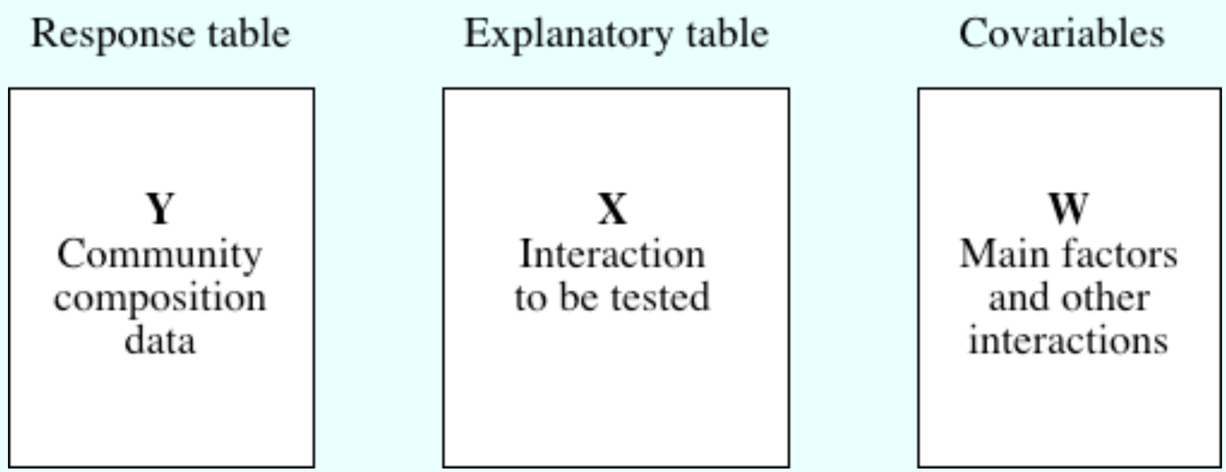
(2) Coding for two crossed factors and their interaction: orthogonal dummy variables, also called Helmert contrasts

Matrix X Interaction *AB*		**Matrix W** (covariables) Factor *A* (3 levels)		Factor *B* (2 levels)
+2	0	+2	0	1
+2	0	+2	0	1
+2	0	+2	0	1
−2	0	+2	0	−1
−2	0	+2	0	−1
−2	0	+2	0	−1
−1	1	−1	1	1
−1	1	−1	1	1
−1	1	−1	1	1
1	−1	−1	1	−1
1	−1	−1	1	−1
1	−1	−1	1	−1
−1	−1	−1	−1	1
−1	−1	−1	−1	1
−1	−1	−1	−1	1
1	1	−1	−1	−1
1	1	−1	−1	−1
1	1	−1	−1	−1

- One factor



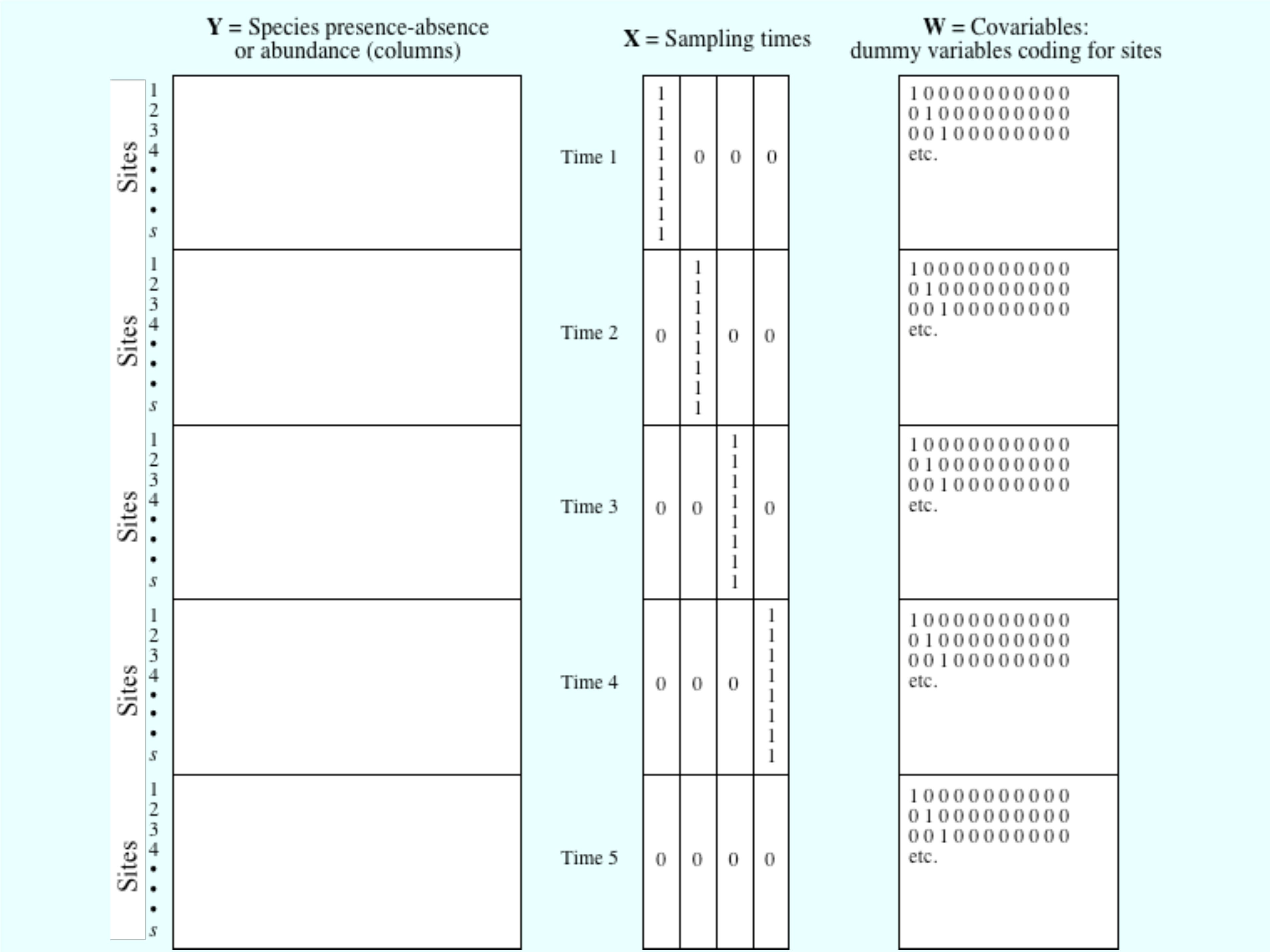
- Two or more crossed factors. Example: test of an interaction



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 is 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.

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 will use Moran's eigenvector maps (MEM)[1].

The type of MEM that we will use in this method, distance-based MEM (dbMEM), were formerly called Principal Coordinates of Neighbour Matrices (PCNM)[1,2].

[1] Dray, S., P. Legendre and P. R. Peres-Neto. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling* 196: 483-493.

[2] 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.

MEM eigenfunctions represent a spectral decomposition of the spatial or temporal relationships among the sampling units. [1]

- They are orthogonal to one another,
- and fewer in number than dummy variables coding for the spatial or temporal positions of the sampling units.

[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.

MEM eigenfunctions[1] represent a spectral decomposition of the spatial or temporal relationships among the sampling units.[1]

• They are orthogonal to one another,

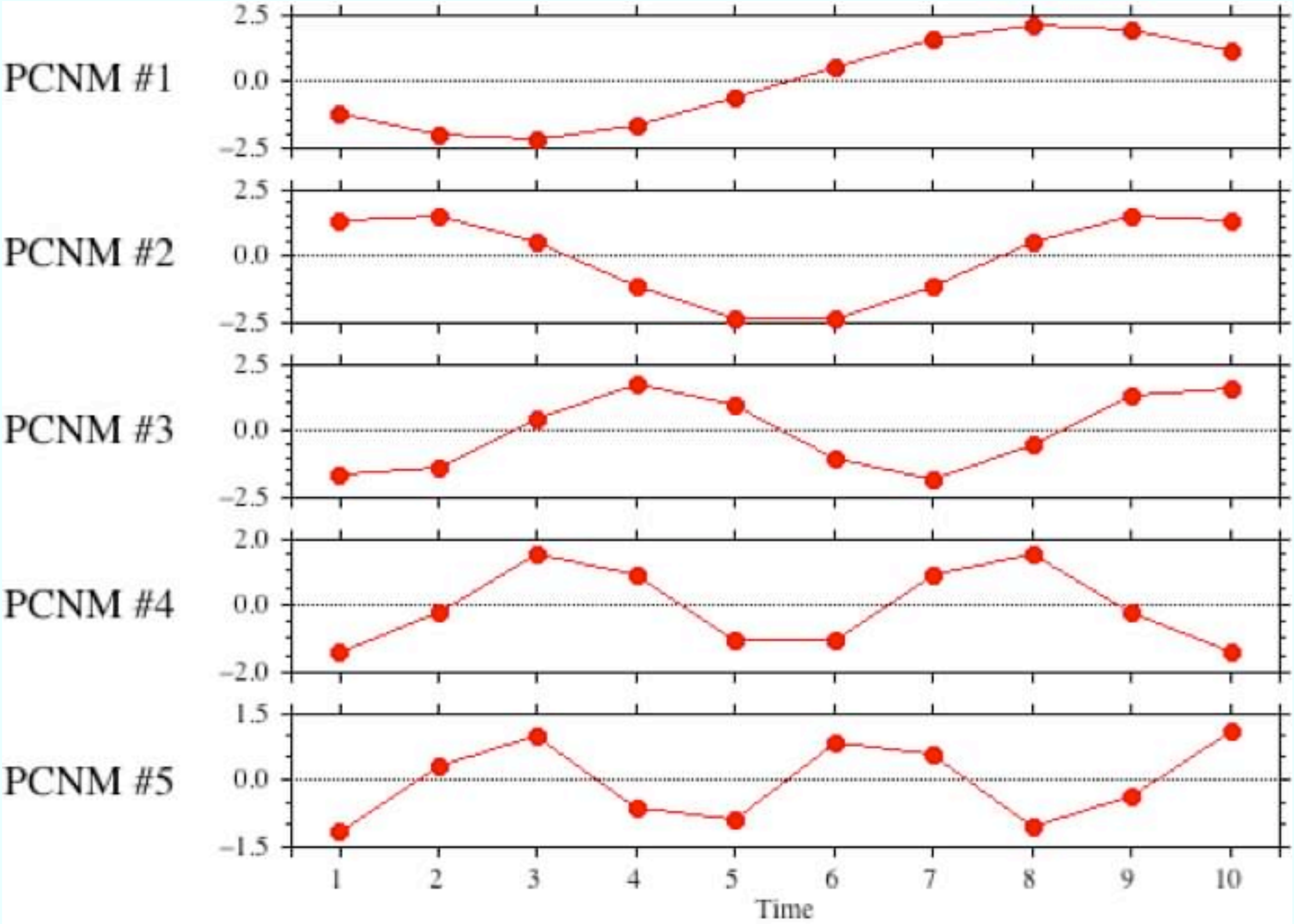
• and fewer in number than dummy variables coding for the spatial or temporal positions of the sampling units.

To model the Space and Time variation, we will use $s/2$ or $t/2$ MEM functions. For a regular design, these MEM are those that are modelling positive spatial or temporal correlation.

For example, 10 equispaced sampling times are modelled by the following 5 MEM 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 MEM eigenfunctions, 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-MEM by each T-MEM.

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

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

For that reason, the MEM eigenfunctions, 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-MEM by each T-MEM.

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

Orthogonality of factors
= happiness for statisticians!

Testing interaction in cross-designs without replication

While it takes (s–1) dummy variables to represent s sites, fewer MEM 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.

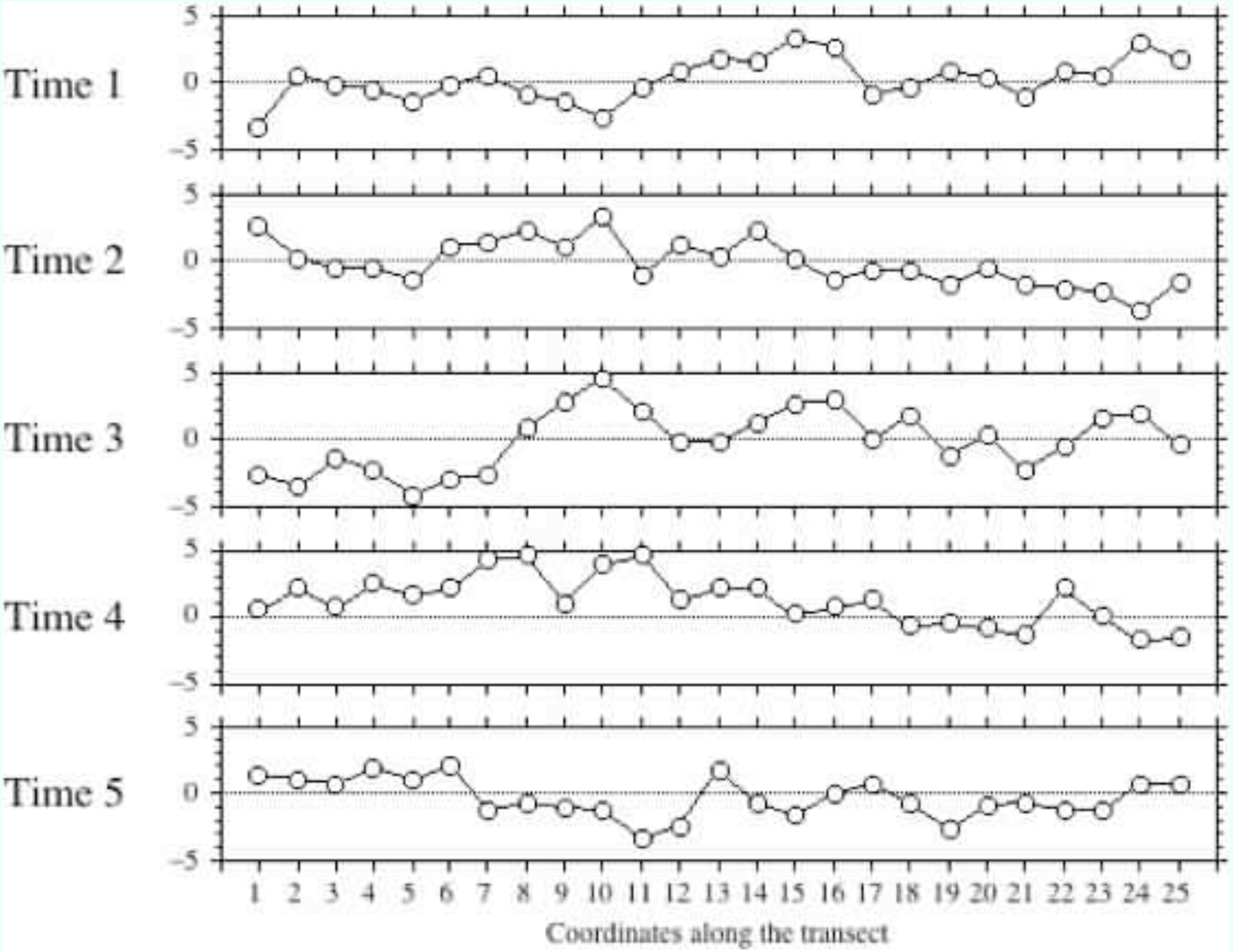
• Spatially autocorrelated data were generated in a 100 x 100 pixel field using the program SimSSD[1]. A transect of 25 equidistant points (spacing = 4 units) was sampled in the middle of the field. A transect variable was the sum of two simulated vectors, with spatial ranges of 10 and 30 units respectively.

• 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. *Ecological Monographs* 75: 435-450.



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

- 13 S-MEM functions were created to model the spatial variation along the 25 points of the transect.
- 3 T-MEM 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-MEM by each T-MEM.

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.)

- 13 S-MEM functions were created to model the spatial variation along the 25 points of the transect.
- 3 T-MEM 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-MEM by each T-MEM.

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.)

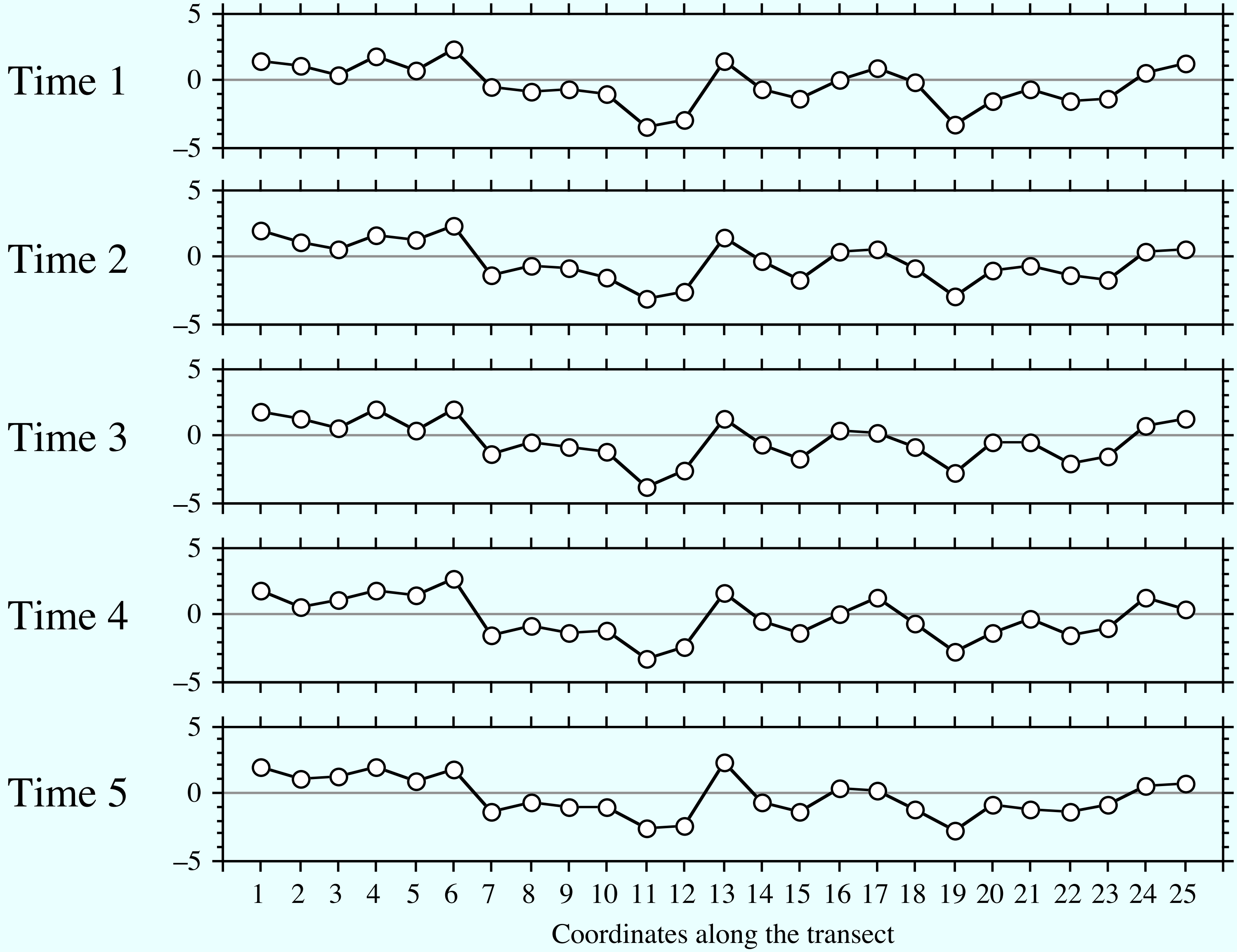
✓ *Correct answer*

The spatial structures differed from time to time. One would have to test the spatial structure of each sampling time separately.

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.



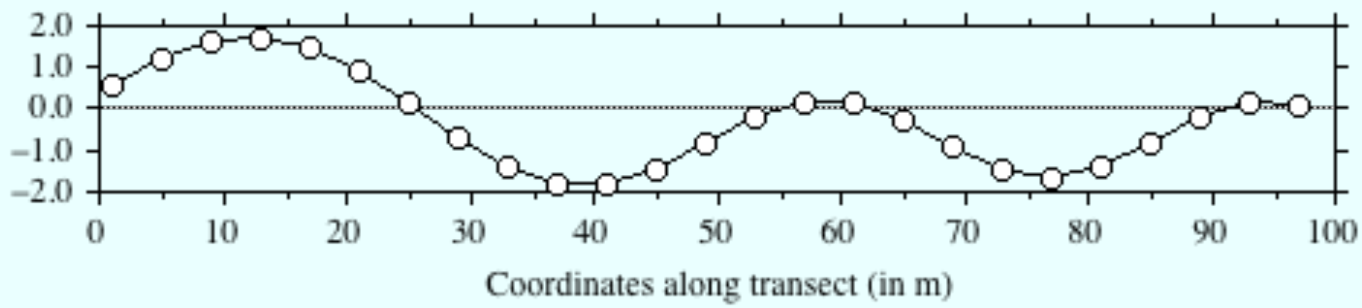
There is no S-T interaction because the 5 “time” realizations were all constructed from the same, common spatial structure.

Results

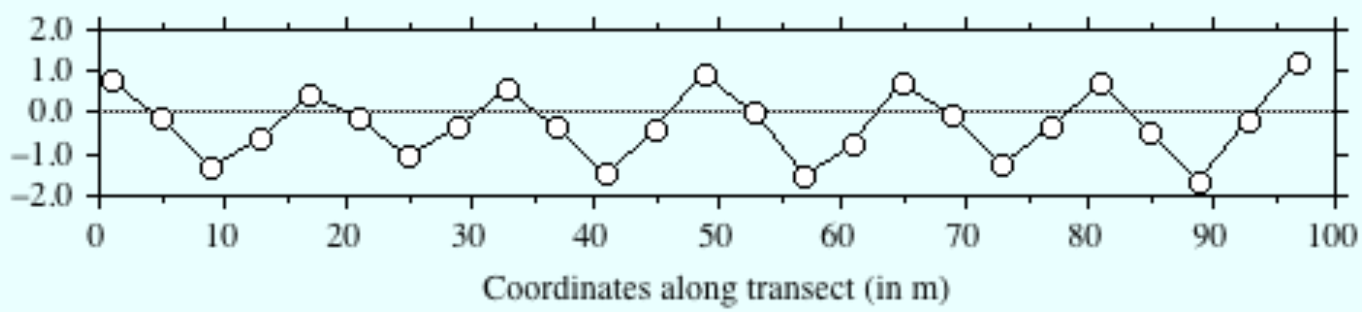
- The S-T interaction was not significant: p = 1.0000 (9999 perm.)
- The main factor Time was not significant: p = 0.9473 (9999 perm.)
- The main factor Space was significant: p = 0.0001 (9999 perm.)

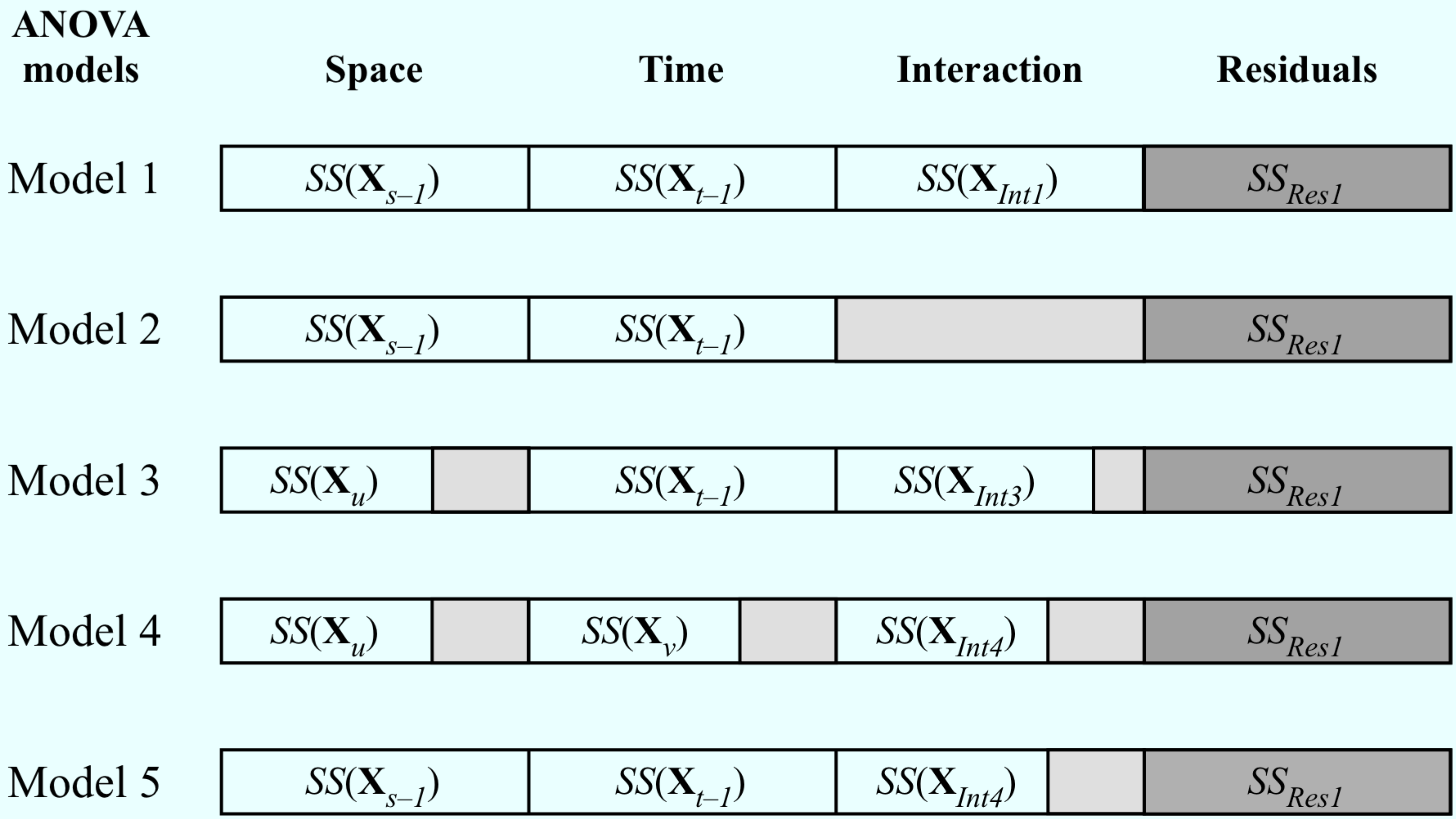
✓ *Correct answer*

Broad-scale model
PCNM #1-7



Fine-scale model
PCNM #8-13





ANOVA models – Sum of squares partitioning for the five models of space-time analysis.

Simulation study results (1)

• Model 5 is the model of choice to detect the S-T interaction; its power was always equal to or higher than those of Models 3 and 4.

• Recommendation: perform first a test of the S-T interaction using ANOVA Model 5. Then proceed as follows.

(1) If the hypothesis of no interaction is not rejected:

- In ANOVA for two random factors or a mixed model with a fixed and a random factor, use our Model 2 (ANOVA without replication) to test the effect of space and time.
- In ANOVA for two fixed factors, use the results of ANOVA Model 5 to test for space and time effects.

Simulation study results (2)

(2) If the hypothesis of no interaction is rejected:

- Model the spatial structure of each time period, or the temporal structure of each site, in a separate ANOVA.
- Conduct a single test involving a separate model of the spatial structure for each time period. Details in the paper.
- The temporal structure of each sampling point can also be analyzed in a single test, by interchanging space and time.

=> A test of interaction without replication has, of course, less power than a test conducted using replicated data.

A paper was published in 2010

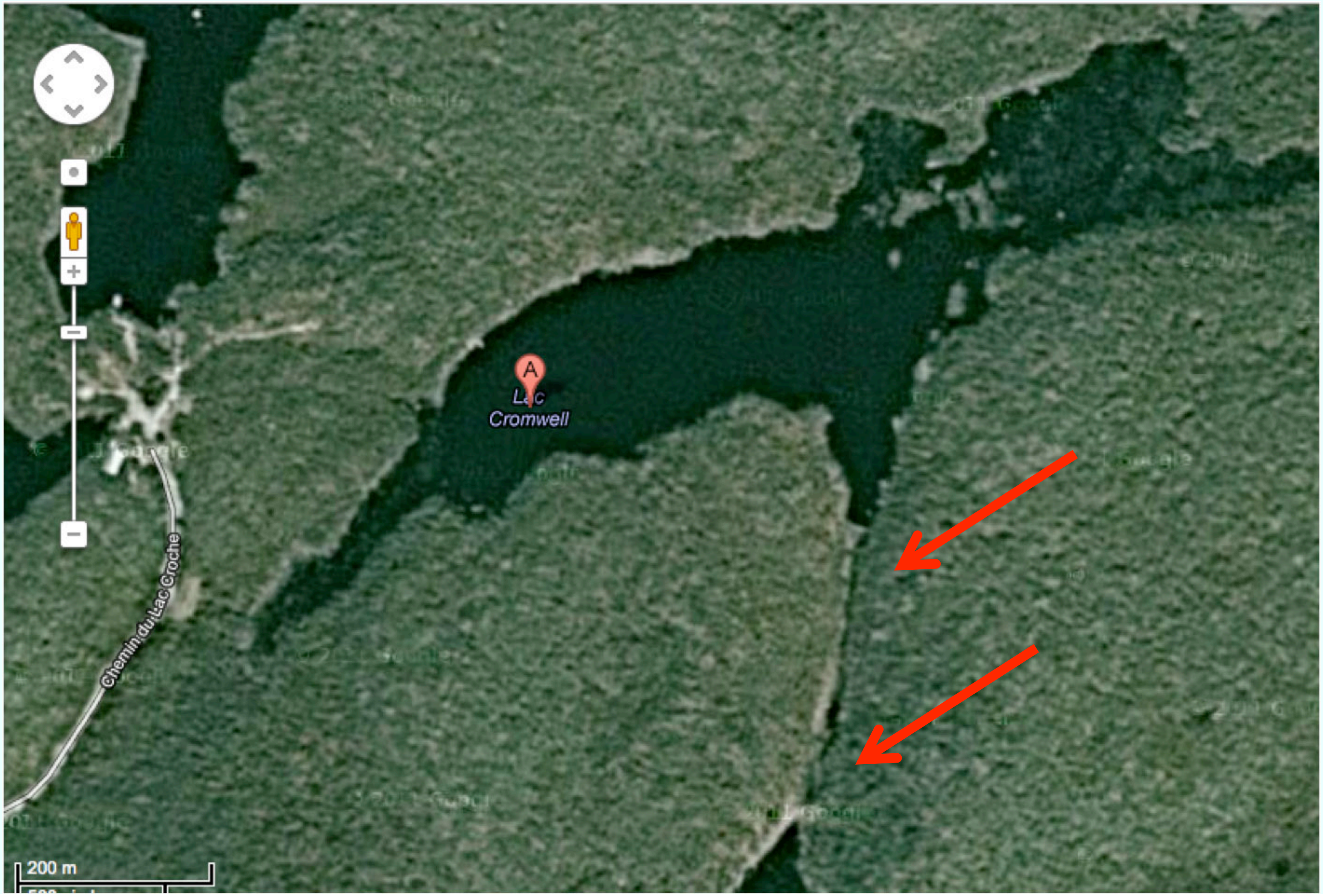
- describing the method,
- reporting the results of an extensive simulation study,
- as well as two examples involving real ecological survey data:

Legendre, P., M. De Cáceres and D. Borcard. 2010. Community surveys through space and time: testing the space-time interaction in the absence of replication. *Ecology* 91: 262-272.

Ecological applications to real data

1. Insect traps

- 22 traps captured emerging insects along a stream during 100 days.



Ecological applications to real data

1. Insect traps

• The captured Trichoptera were pooled into 10 consecutive periods of 10 days each.

• 22 sites × 10 time periods = 220 sampling units.

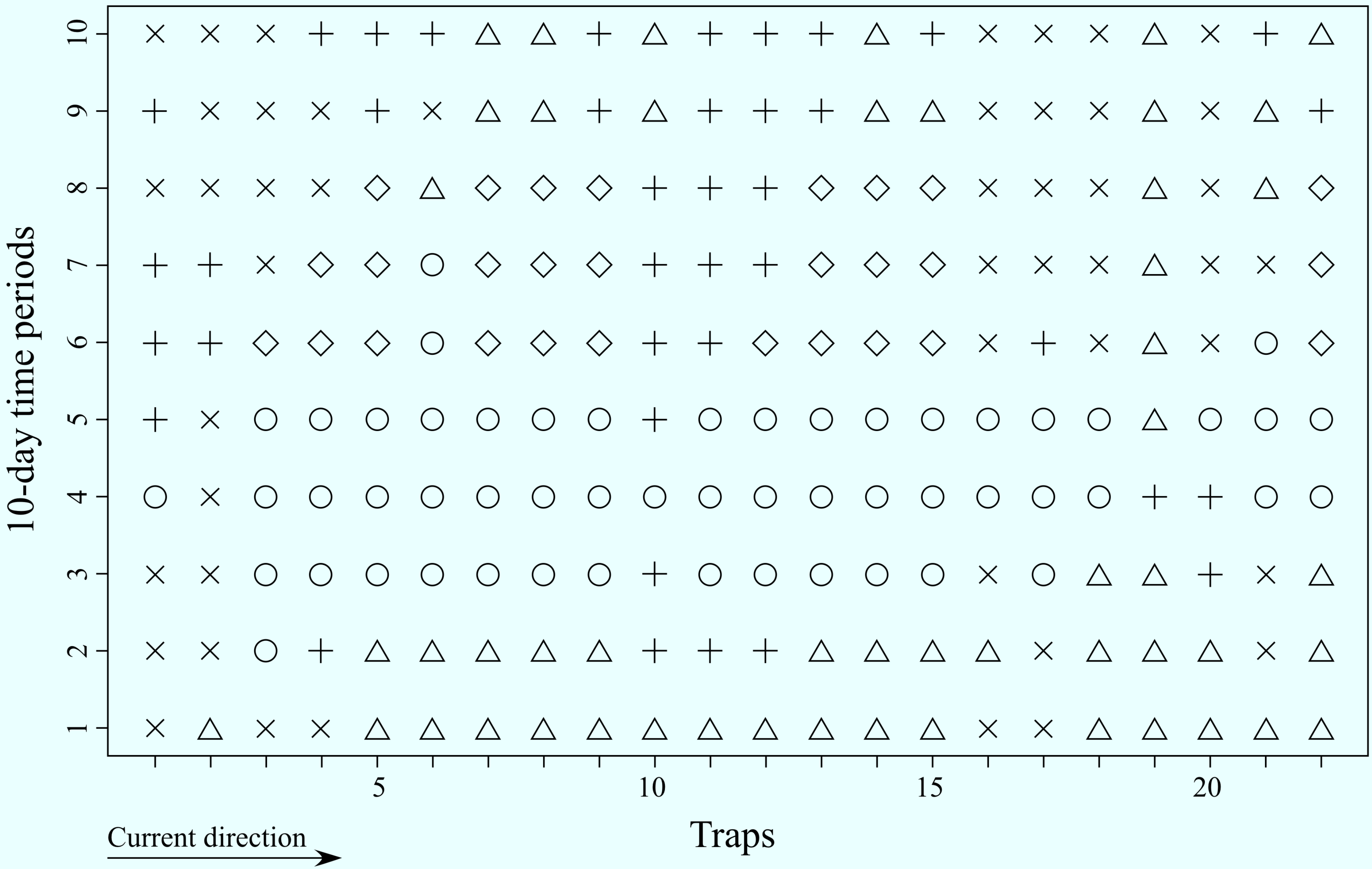
• 56 species.

• The species abundance data were log-transformed: $y' = \log(y + 1)$.

• dbMEM eigenfunctions were used to model space and time.

⇒ The space-time interaction was significant.

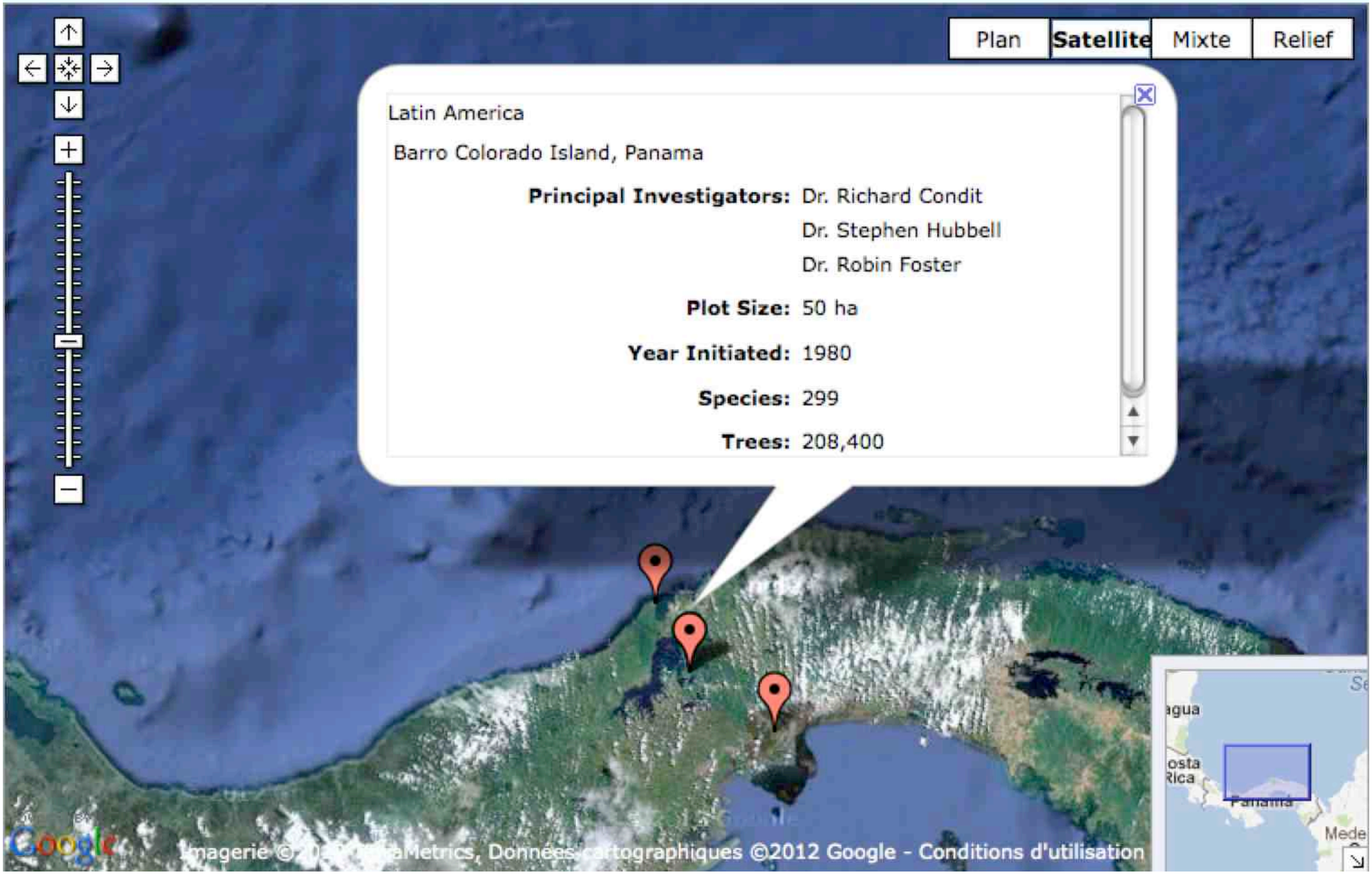
⇒ How can a space-time interaction in multivariate data be represented graphically?



- The data were subjected to *K*-means clustering, producing 5 groups.
- The 5 groups (symbols) are shown on a space-time map.

Ecological applications to real data

2. The Barro Colorado Island (BCI) permanent forest plot



2. The Barro Colorado Island (BCI) permanent forest plot

- 50-ha forest plot in Panama divided into 1250 (20 × 20 m) cells
- 315 tree species
- Four censuses analysed paper: 1982-83, 1985, 1990, and 1995.
- The species abundance data were log-transformed: $y' = \log(y + 1)$.

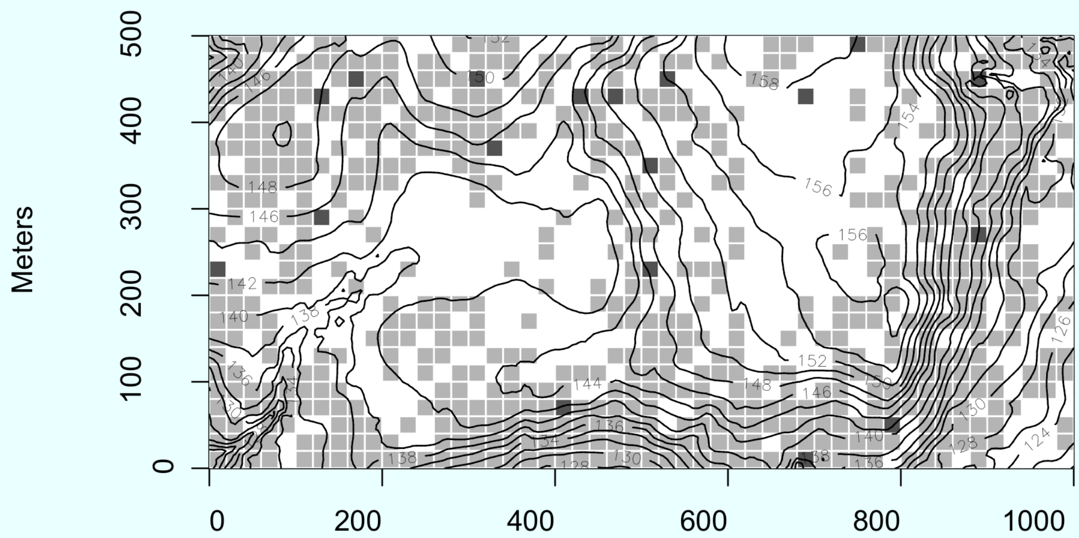
2. The Barro Colorado Island (BCI) permanent forest plot

- dbMEM were used to model space
- Helmert contrasts were used to model the 4 surveys through time
- Space-time interaction on all 315 species: significant
- By species: 43% of the species had significant interactions

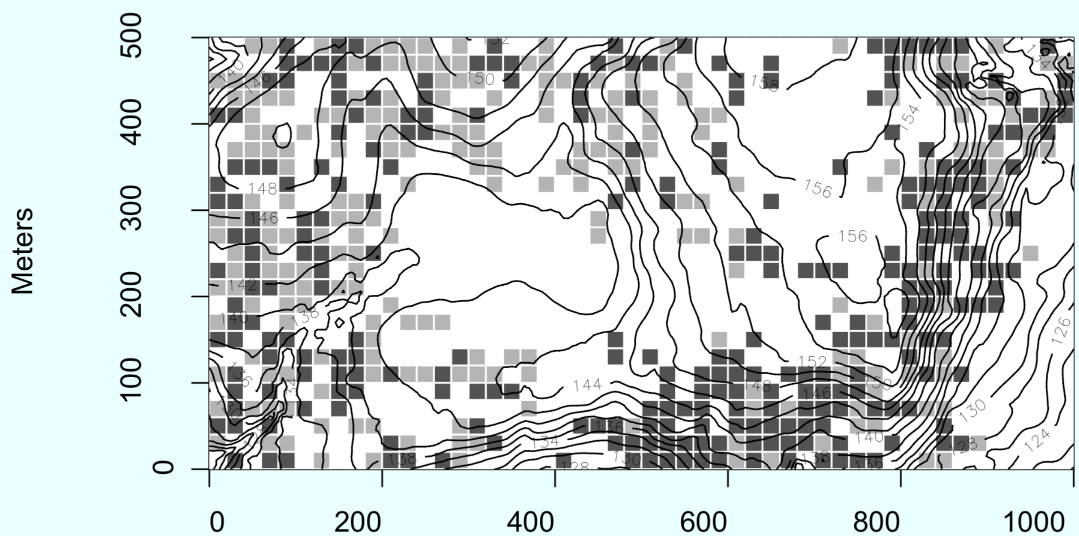
⇒ The spatial structure of these species had significantly changed across the four censuses.

⇒ Possible cause: climatic (change in rainfall, longer dry season; Condit 1998).

a) Poulsenia armata



b) Beilschmiedia pendula



R package

Package STI is available on
Ecological Archives E091-019-S1 of the *Ecological Society of America*
and on the Web page of Miquel De Cáceres:
http://sites.google.com/site/miqueldecaceres/software

References

Legendre, P., M. De Cáceres, and D. Borcard. 2010. Community surveys through space and time: testing the space-time interaction in the absence of replication. *Ecology* 91: 262-272.

Legendre, P. and L. Legendre. 2012. *Numerical ecology, 3rd English edition*. Elsevier Science BV, Amsterdam.

Do you have questions?