Using distance-based eigenvector maps (DBEM) in multivariate partitioning

Part 1: PCNM (principal coordinates of neighbor matrices), theory and applications

Daniel Borcard and Pierre Legendre

Département de sciences biologiques

Université de Montréal

Daniel.Borcard@umontreal.ca

Pierre.Legendre@umontreal.ca

Special Session "*Spatial Statistics at Multiple Scales*", ESA-INTECOL 2005 Joint Meeting

Outline of the talk

1. Introduction
2. PCNM: theory
3. Simulation study
4. A difficult simulated test case
5. *An example on real data* ...at the end of the session!

Main references

Borcard, D. and Legendre, P. 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, Avois-Jacquet, C. & Tuomisto, H. 2004. Dissecting the spatial structures of ecologial data at all scales. *Ecology 85(7): 1826-1832.*

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 x species).

Spatial structures in communities indicate that some process has been at work that created them.

Statement of problem

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.

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**.

Principal coordinates of neighbor matrices (PCNM)

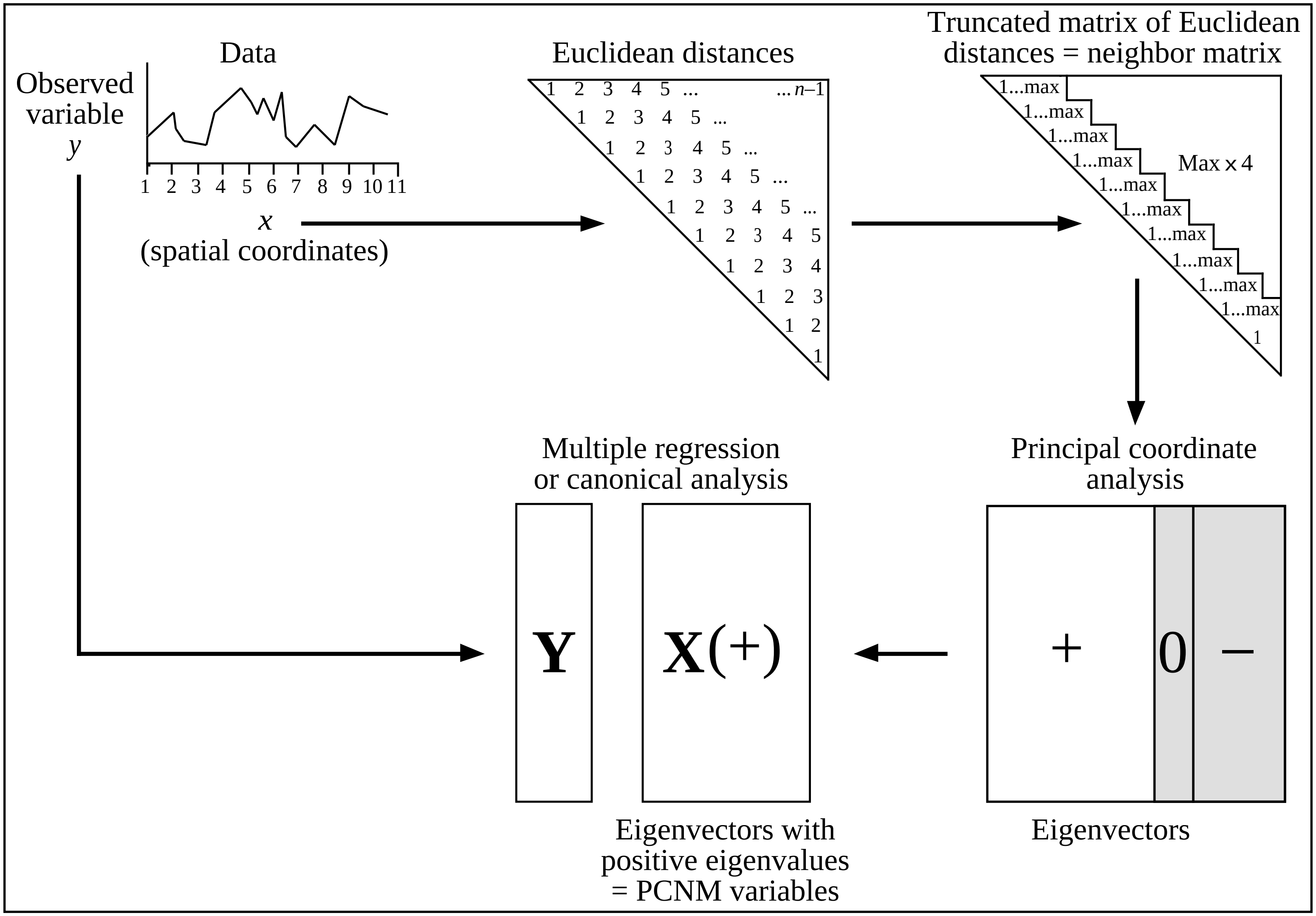
Theory

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?

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 (PCNM base functions) are obtained by principal coordinate analysis of a truncated matrix of Euclidean (geographic) distances among the sampling sites.

Notes on PCNM base functions

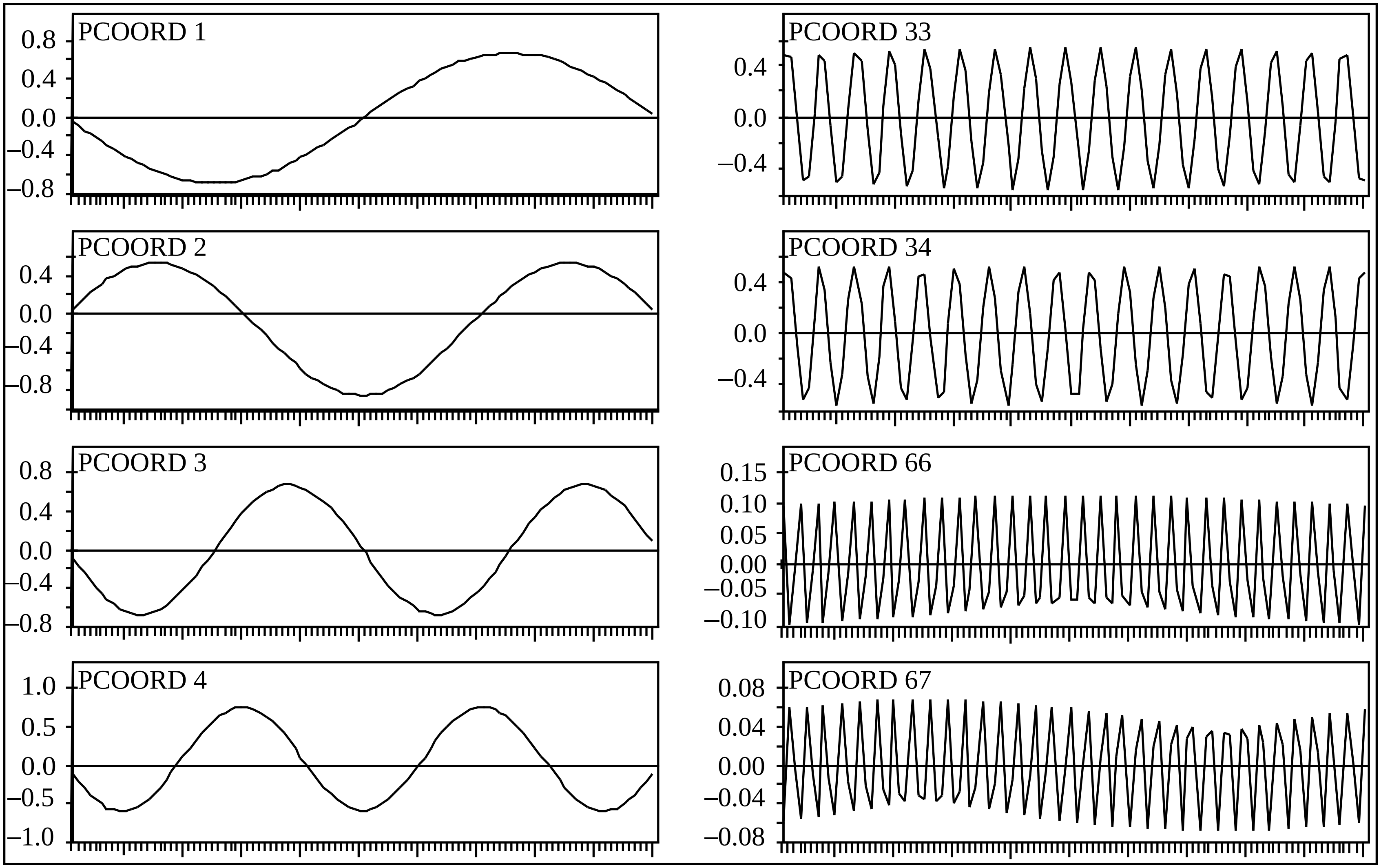
PCNM variables represent a **spectral decomposition** of the spatial relationships among the study sites.

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

PCNM base functions are **orthogonal**. If the sampling design is regular, they look like sine waves. This is a property of the eigen-decomposition of the centered form of a distance matrix (Laplacian).

The concept of PCNM has been recently generalized to that of **Distance-Based Eigenvector Maps (DBEM)**; other ways of computing such vectors are now available (Dray et al., submitted)[1].

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



Eight of the 67 orthogonal PCNM base functions obtained for 100 equally-spaced points along a transect. Truncation after the first neighbor.

Principal coordinates of neighbor matrices (PCNM)

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 α.

Power study

Simulations showed that PCNM 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,

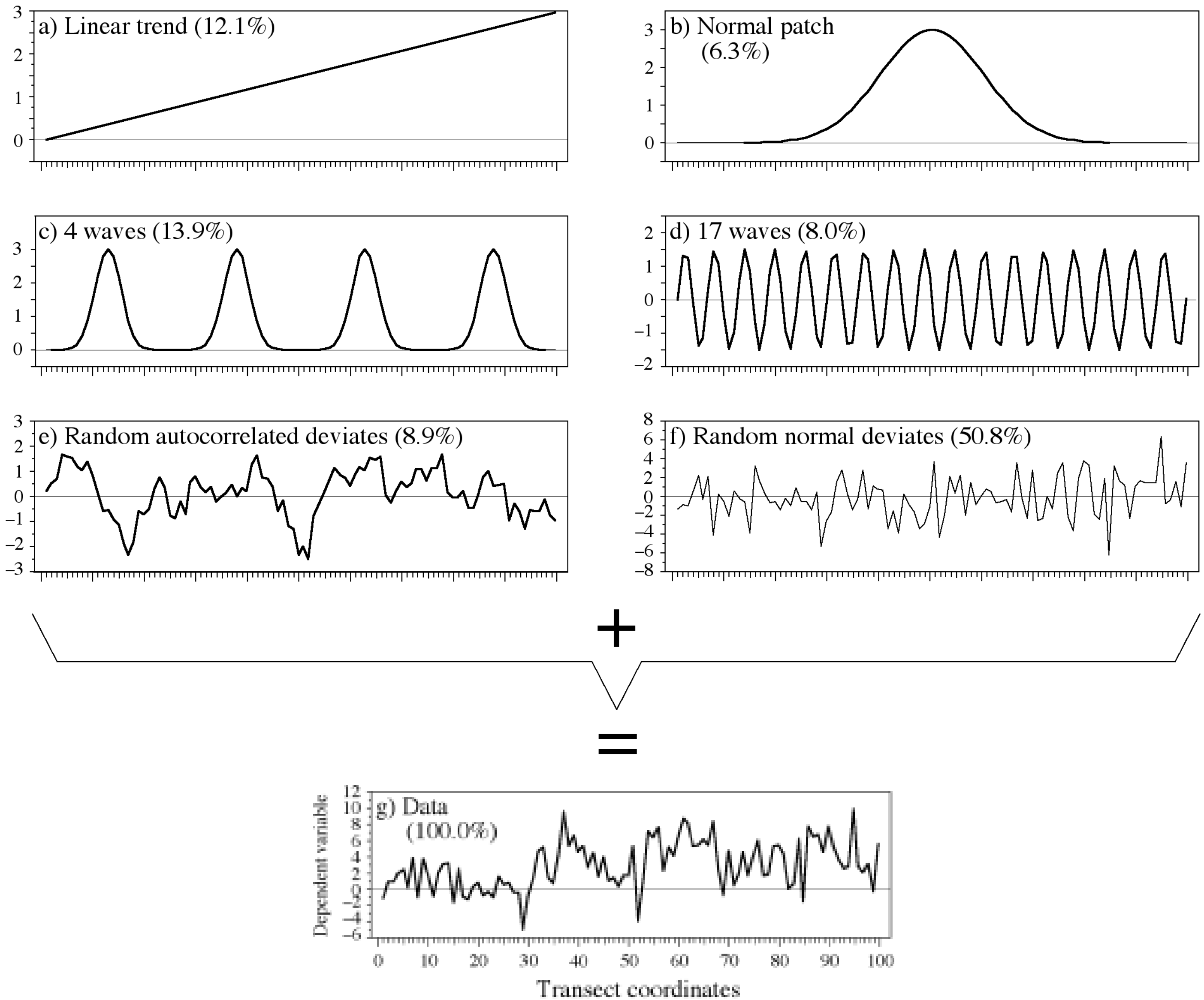
... as long as the structures are larger than the truncation value used to create the PCNM base functions.

Detailed results are found in Borcard & Legendre 2002.

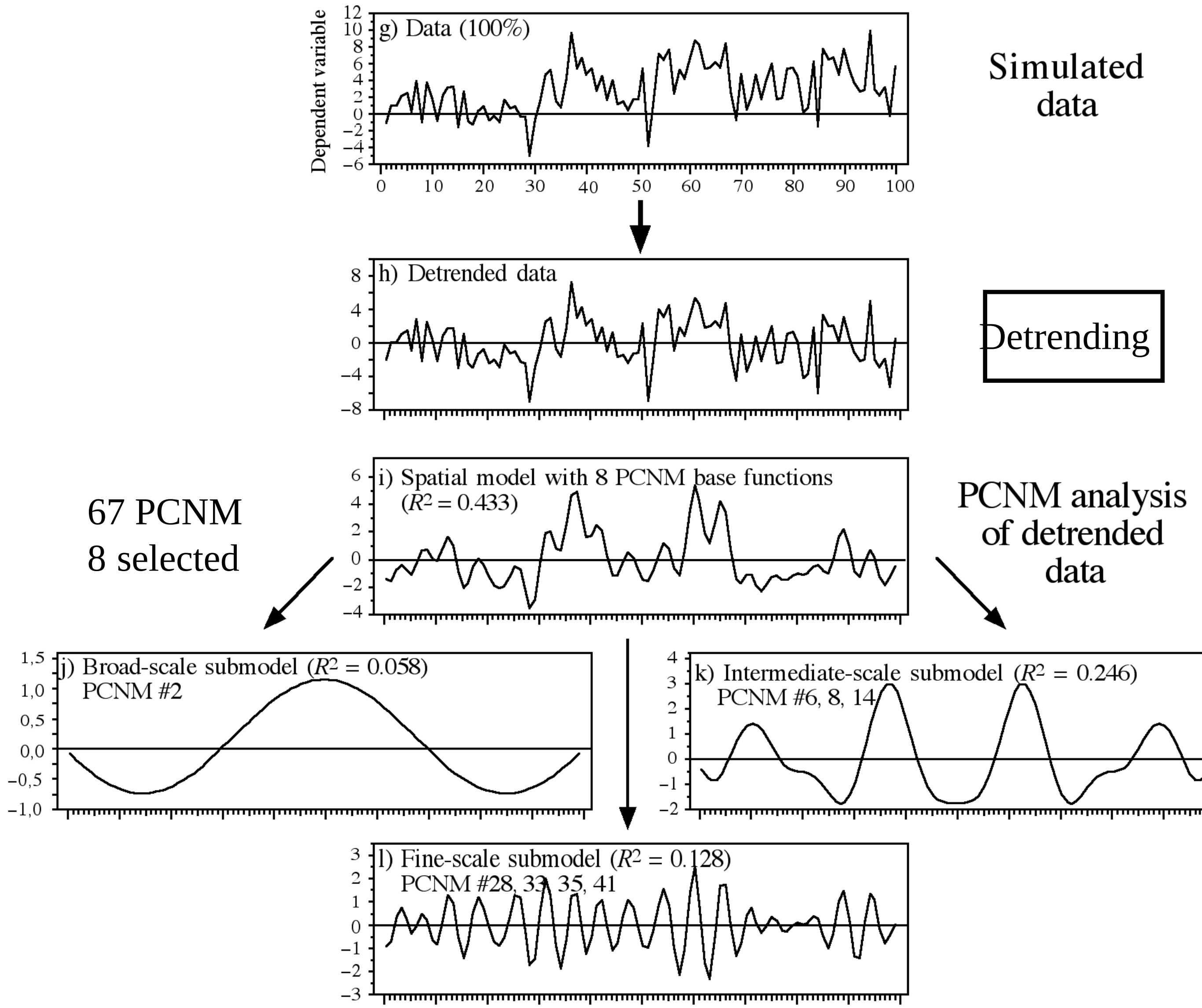
Principal coordinates of neighbor matrices (PCNM)

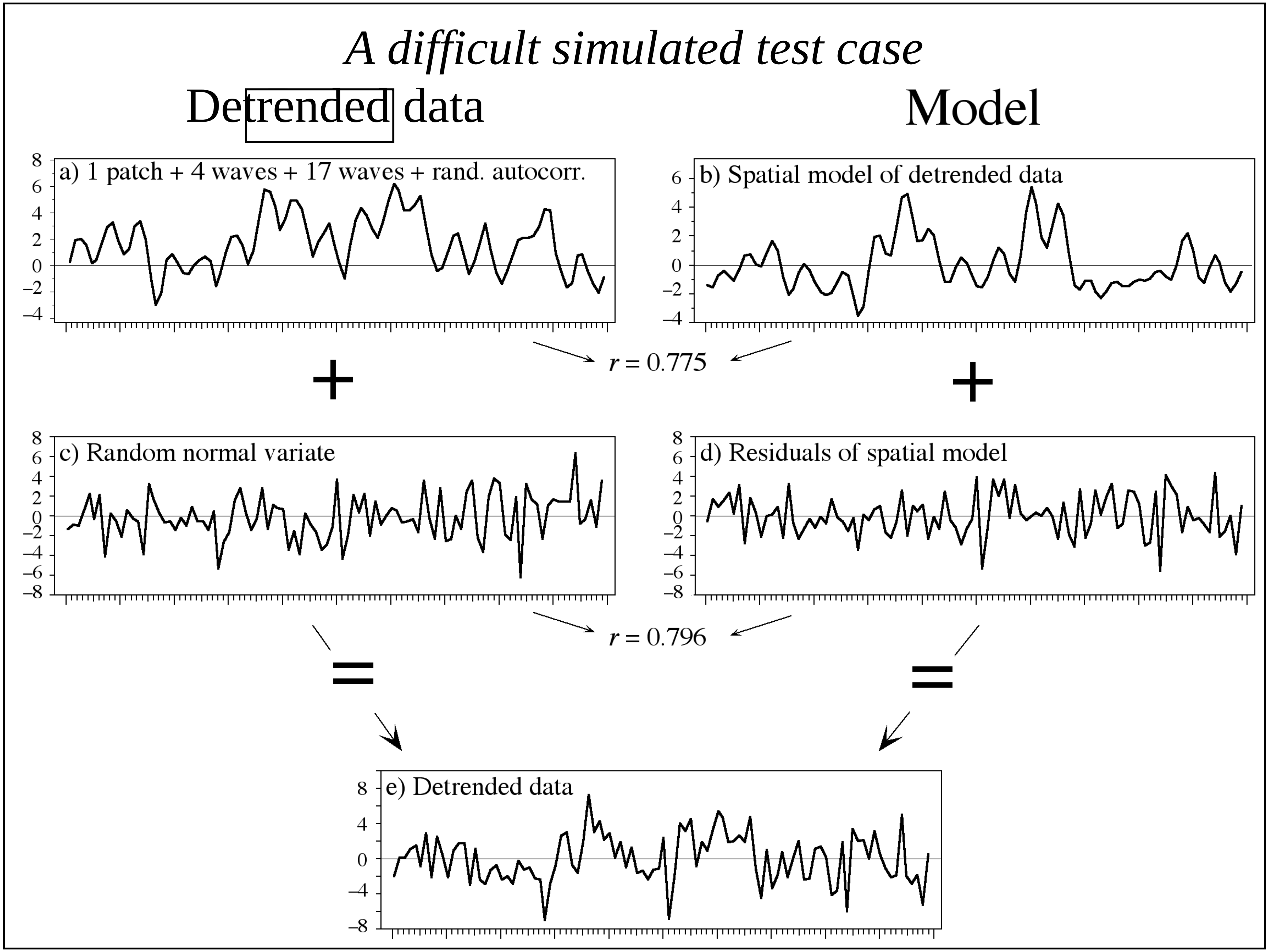
A difficult simulated test case

A difficult simulated test case



A difficult simulated test case



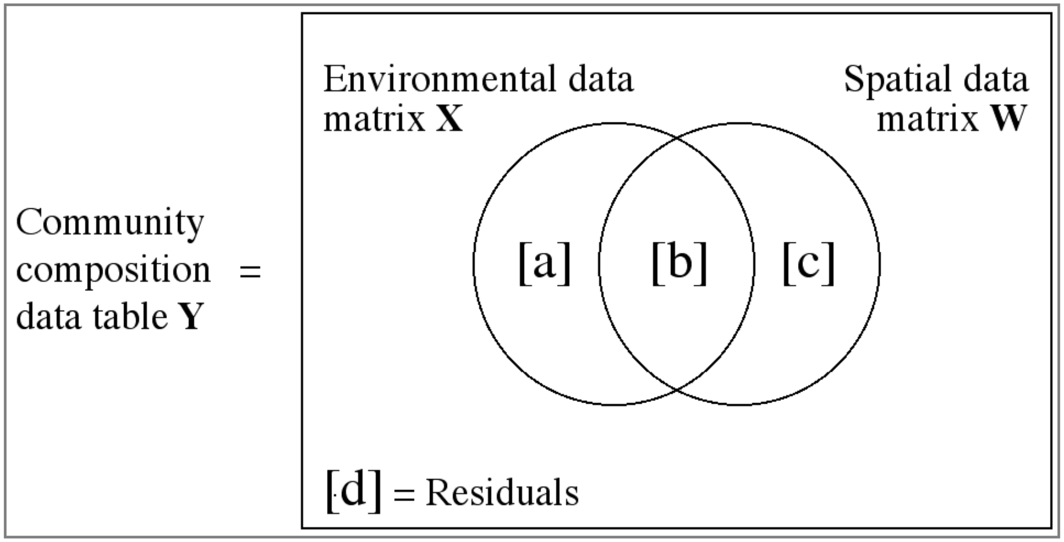


A difficult simulated test case
Detrended data
Model
a) 1 patch + 4 waves + 17 waves + rand. autocorr.
b) Spatial model of detrended data
r = 0.775
+
+
c) Random normal variate
d) Residuals of spatial model
r = 0.796
=
=
e) Detrended data

Putting space and environment together:

variation partitioning

Multivariate variation partitioning (Borcard et al. 1992, Borcard and Legendre 1994) can be applied using PCNM base variables as spatial or temporal descriptors.



Estimation of the R^2 of the various components are best obtained by the three following canonical ordinations:

- Response data | Environment + Space : [a] + [b] + [c]
- Response data | Environment : [a] + [b]
- Response data | Space : [b] + [c]

Individual fractions are obtained by subtractions of the above.