

# Training in ade4 in R - Module II: Advanced methods

## Spatial Multivariate Analysis

Stéphane Dray

2022-02-28



# Introduction

```
library(ade4)
library(adegraphics)
adegpar(background.col = "lightgrey")
```

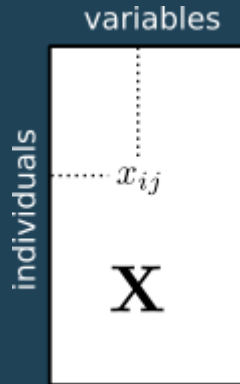
```
## list()
```

```
data(mafragh)
names(mafragh)
```

```
## [1] "xy"           "flo"          "neig"         "env"
## [5] "partition"    "area"         "tre"          "traits"
## [9] "nb"           "Spatial"      "spenames"     "Spatial.contou
```

- Species table (**flo**): abundance indexes of 56 plant species recorded for 97 sites in the Mafragh plain (Algeria)
- Spatial coordinates for the sites (**xy**)

# Introduction



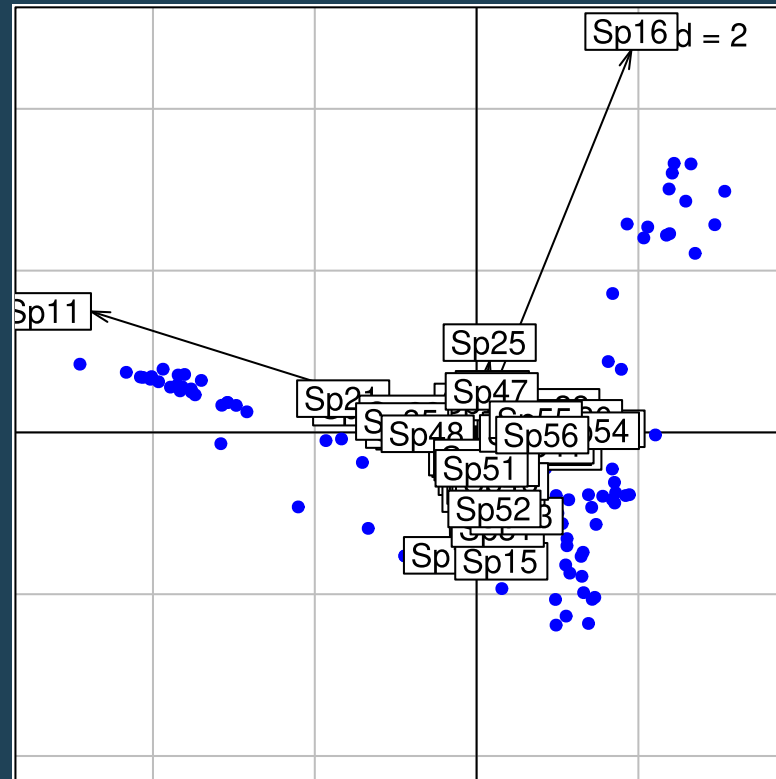
Describe the information contained in the table:

- Identify which variables are the most linked
- Identify the differences/similarities between individuals

One table with  $p$  variables measured on  $n$  individuals

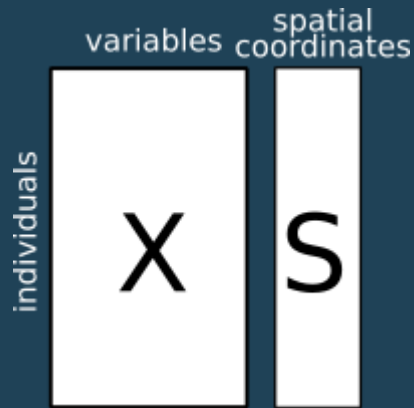
- quantitative (`dudi.pca`)
- categorical (`dudi.coa` or `dudi.acm`)
- both (`dudi.mix` or `dudi.hillsmith`)

```
pca_veg <- dudi.pca(mafragh$flo, scale = FALSE, scannf = FALSE)
biplot(pca_veg, ppoints.col = "blue", row.label.cex = 0,
       posieig = "none")
```



Is community composition spatially structured?

# One table and spatial information



- Identify spatial patterns
- Identify which variables are spatially structured
- Identify at which scales these spatial patterns occur

One table with  $p$  variables measured on  $n$  individuals

One table with spatial coordinates for the same  $n$  individuals

**But how?**

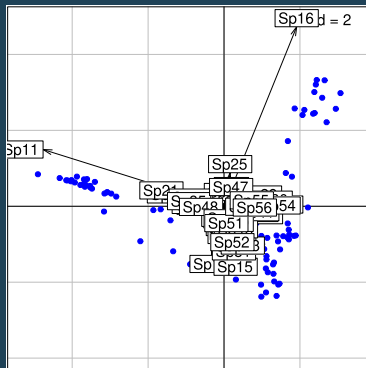
# Indirect approaches



and detect spatial structures using simple tools

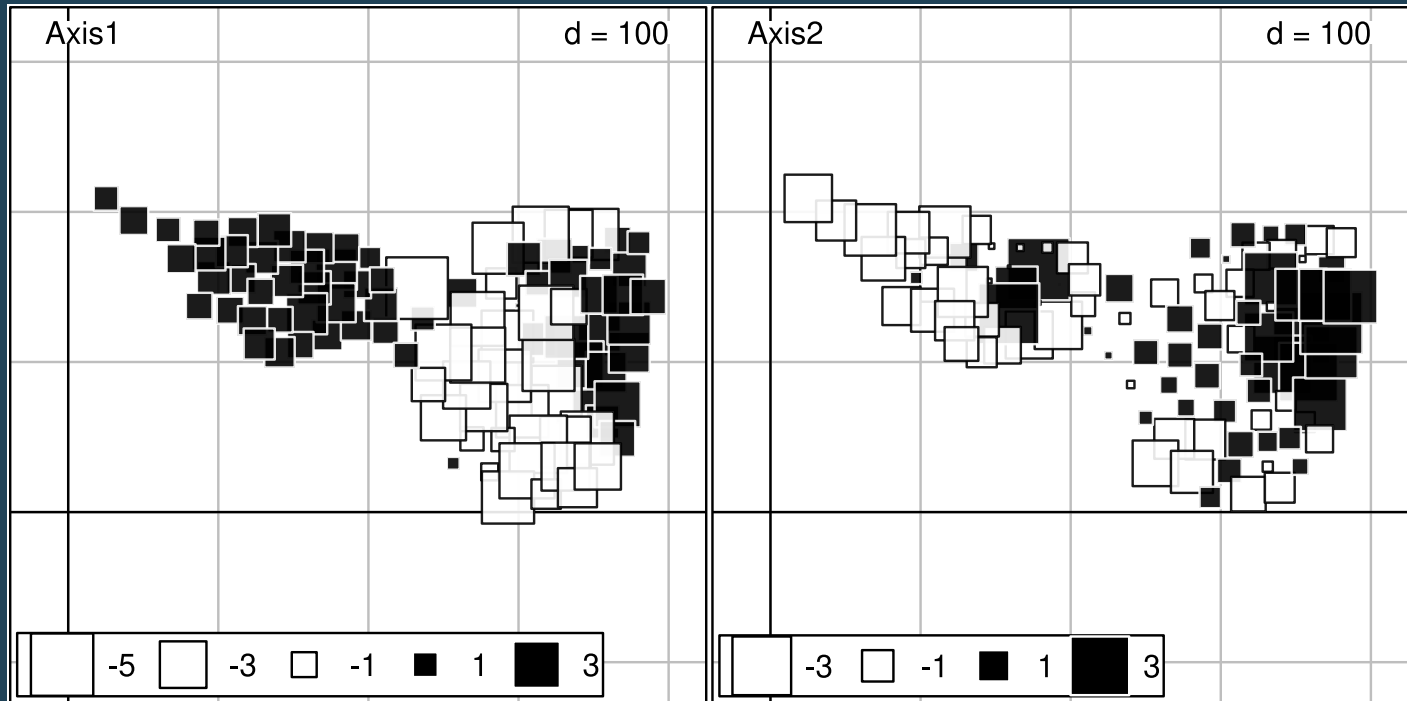
- Mapping
- Regression on spatial predictors (polynomials, MEM)
- Spatial autocorrelation

Summarize by a simple ordination method



# Spatial mapping

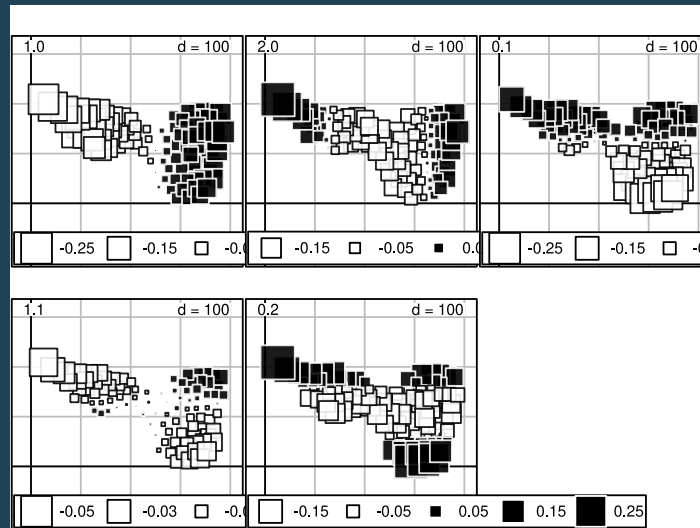
```
s.value(mafragh$xy, pca_veg$li)
```





# Correlation with spatial predictors

```
poly.xy <- poly(as.matrix(mafragh$xy),  
  degree = 2)  
s.value(mafragh$xy, poly.xy)
```



```
cor(pca_veg$li, poly.xy)
```

```
##           1.0           2.0           0.1           1.1           0.2  
## Axis1 -0.2866398  0.30645089  0.40249089 -0.03318448 -0.1706804  
## Axis2  0.5047664  0.01762275 -0.08345049  0.25974753 -0.3269496
```

# Introducing spatial information in multivariate methods

Need of direct and explicit spatial multivariate methods



Spatial information :

- **symmetric** ( $n \times n$ )
  - Distances matrix
  - Spatial weighting matrix
- **raw data** ( $n \times p$ )
  - (Polynomials of) Geographic coordinates
  - Spatial eigenvectors (MEM)

# Space as distances

- Compute spatial distances  $D_S$
- Compute faunistic distances  $D_X$
- Study their link

```
dx <- dist(mafragh$flo)
ds <- dist(mafragh$xy)
plot(dx ~ ds)
```

```
mantel.randtest(dx, ds)
```

```
## Monte-Carlo test
## Call: mantel.randtest(m1 = dx, m2 = 
##
## Observation: 0.2327358
##
## Based on 999 replicates
## Simulated p-value: 0.001
## Alternative hypothesis: greater
##
##           Std.Obs  Expectation      Variance
## 8.1460917683 0.0004854407 0.00081285
```

# Spatial Weighting Matrix

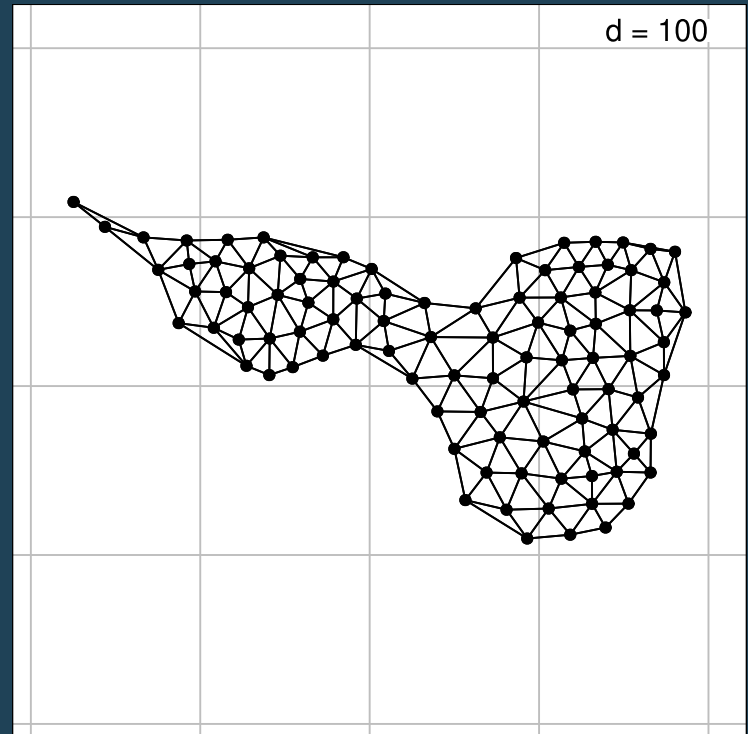
$$\mathbf{W} = [w_{ij}]$$

$w_{ij} = 1$  if sites  $i$  and  $j$  are neighbors

$w_{ij} = 0$  otherwise

Options:

- Non-binary weights
- Standardization



# Spatial autocorrelation

Moran's index

$$MC(\mathbf{x}) = \frac{n \sum_{(2)} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{(2)} w_{ij} \sum_{i=1}^n (x_i - \bar{x})^2} \text{ where } \sum_{(2)} = \sum_{i=1}^n \sum_{j=1}^n \text{ with } i \neq j$$

In matrix notation

$$MC(\mathbf{x}) = \frac{n}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} \frac{\mathbf{z}^\top \mathbf{W} \mathbf{z}}{\mathbf{z}^\top \mathbf{z}}$$

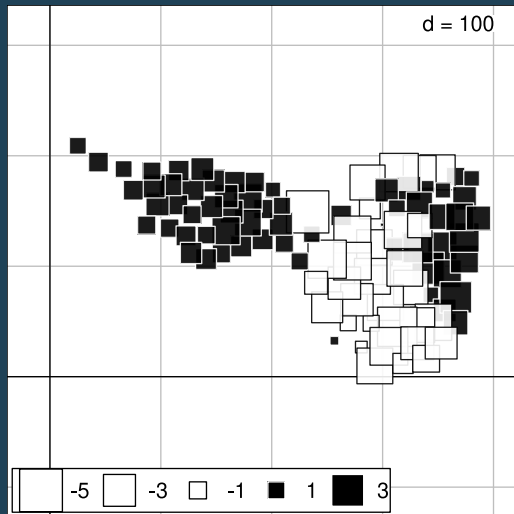
where  $\mathbf{z} = \mathbf{x} - \bar{x}$  . With row-standardization

$$MC(\mathbf{x}) = \frac{\mathbf{z}^\top \tilde{\mathbf{z}}}{\mathbf{z}^\top \mathbf{z}}$$

It is a measure of the link between the original variable (  $\mathbf{z}$  ) and its lagged version (  $\tilde{\mathbf{z}}$  )

# Moran's index and scatterplot

```
library(adespatial)
library(spdep)
s.value(mafragh$xy, pca_veg$li[,
```

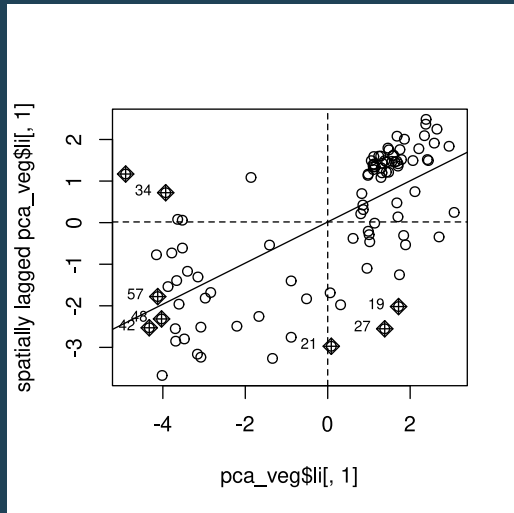


```
moran.randtest(pca_veg$li[, 1], r
```

```
## Monte-Carlo test
## Call: moran.randtest(x = pca_veg$li[
##
## Observation: 0.4947964
##
## Based on 999 replicates
## Simulated p-value: 0.001
## Alternative hypothesis: greater
##
## Std.Obs.statistic      Expectation
##           8.387562399      -0.014619096
```

# Moran's index and scatterplot

```
moran.plot(pca_veg$li[, 1], nb2l)
```

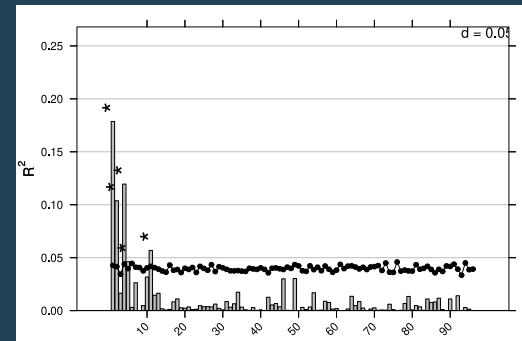
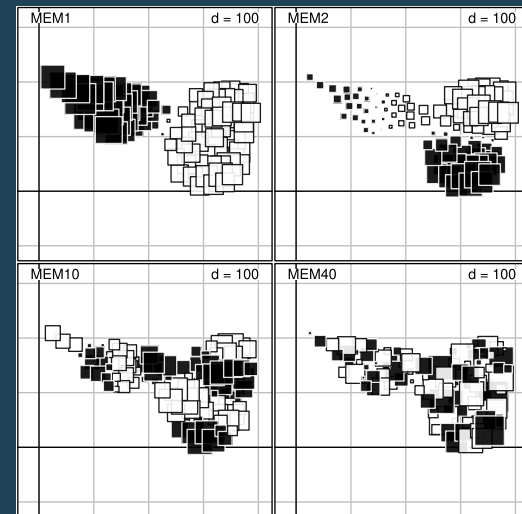


```
moran.randtest(pca_veg$li[, 1], r
```

```
## Monte-Carlo test
## Call: moran.randtest(x = pca_veg$li[, 1], r = nb2l,
##
## Observation: 0.4947964
##
## Based on 999 replicates
## Simulated p-value: 0.001
## Alternative hypothesis: greater
##
## Std.Obs.statistic      Expectation
##      7.748020071      -0.005346483
```

# Moran's Eigenvector Maps

- Eigenvectors of the doubly-centred spatial weighting matrix
- Orthogonal vectors maximizing the spatial autocorrelation
- They can be used as spatial predictors in correlation or regression models (e.g., scalogram)





# Spatial multivariate methods

- **MULTISPATI** integrates a spatial weighting matrix in standard multivariate methods. It identifies spatial structures instead of structures.
- **MEM-based models** introduces spatial predictors in correlation or regression-based approaches (e.g., redundancy analysis, variation partitioning)

# Multispati

This analysis maximizes the product

$$MC_D(\mathbf{XQa}) \cdot \|\mathbf{XQa}\|_D^2$$

and thus finds a linear combination of variables maximizing a compromise between:

- the criteria optimized in the standard analysis (  $\|\mathbf{XQa}\|_D^2$  )
- spatial autocorrelation (  $MC_D(\mathbf{XQa})$  )

It corresponds to a coinertia analysis between the original table  $\mathbf{X}$  and its lagged version  $\tilde{\mathbf{X}} = \mathbf{WX}$ .

# MEM-based models

Spatial eigenvectors can be used as predictors in regression models (e.g., redundancy analysis)

- As there is a high number of predictors, variable selection should be performed prior to the analysis
- When other predictors are considered, variation partitioning can be used. This approach based on combination of several RDA models allows to evaluate the relative contribution of spatial and other predictors to explain the response table

