Training in ade4 in R - Module II: Advanced methods

Spatial Multivariate Analysis

Stéphane Dray

2025-10-19

Introduction

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

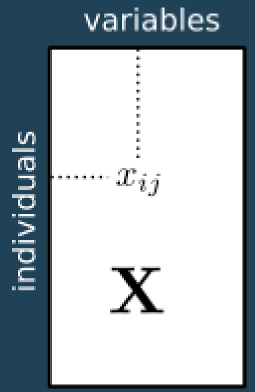
```
## list()
```

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

```
##  [1] "xy"              "flo"             "env"
##  [4] "partition"       "area"            "tre"
##  [7] "traits"          "nb"              "Spatial"
## [10] "spenames"        "Spatial.contour"
```

- 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



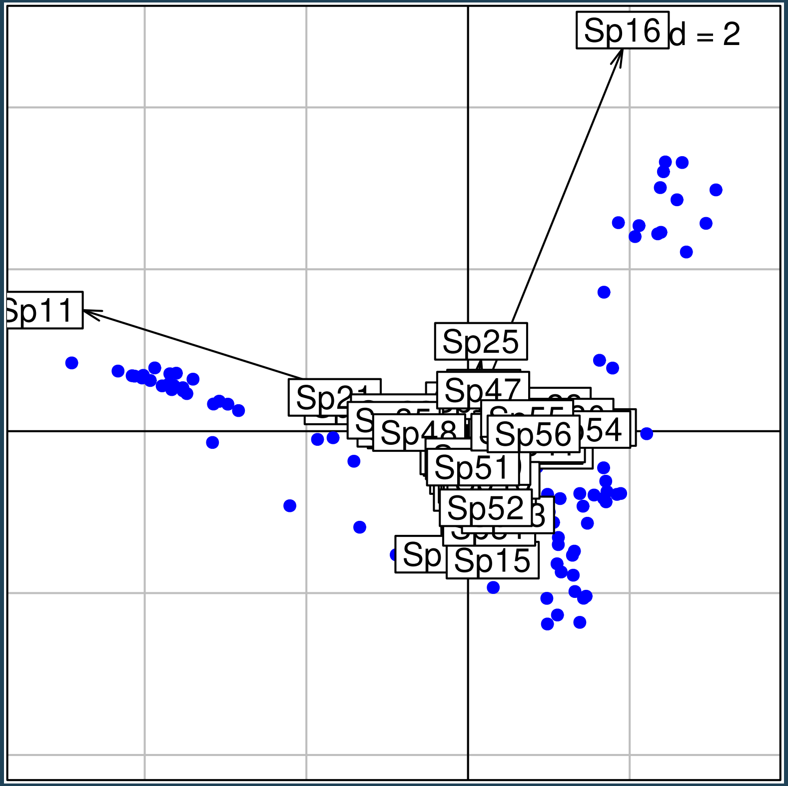
One table with *p* variables measured on *n* individuals

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

Describe the information contained in the table:

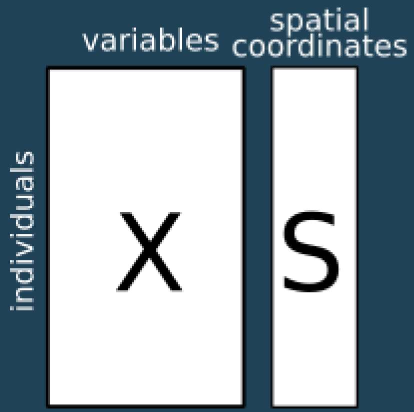
- Identify which variables are the most linked
- Identify the differences/similarities between individuals

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



Is community composition spatially structured?

One table and spatial information



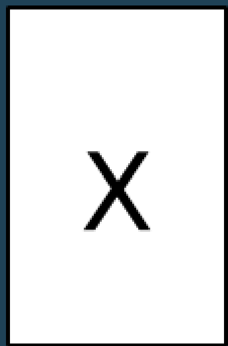
One table with p variables measured on n individuals

One table with spatial coordinates for the same n individuals

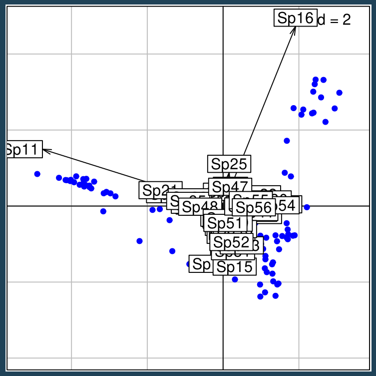
- Identify spatial patterns
- Identify which variables are spatially structred
- Identify at which scales these spatial patterns occur

But how?

Indirect approaches



Summarize by a simple ordination method

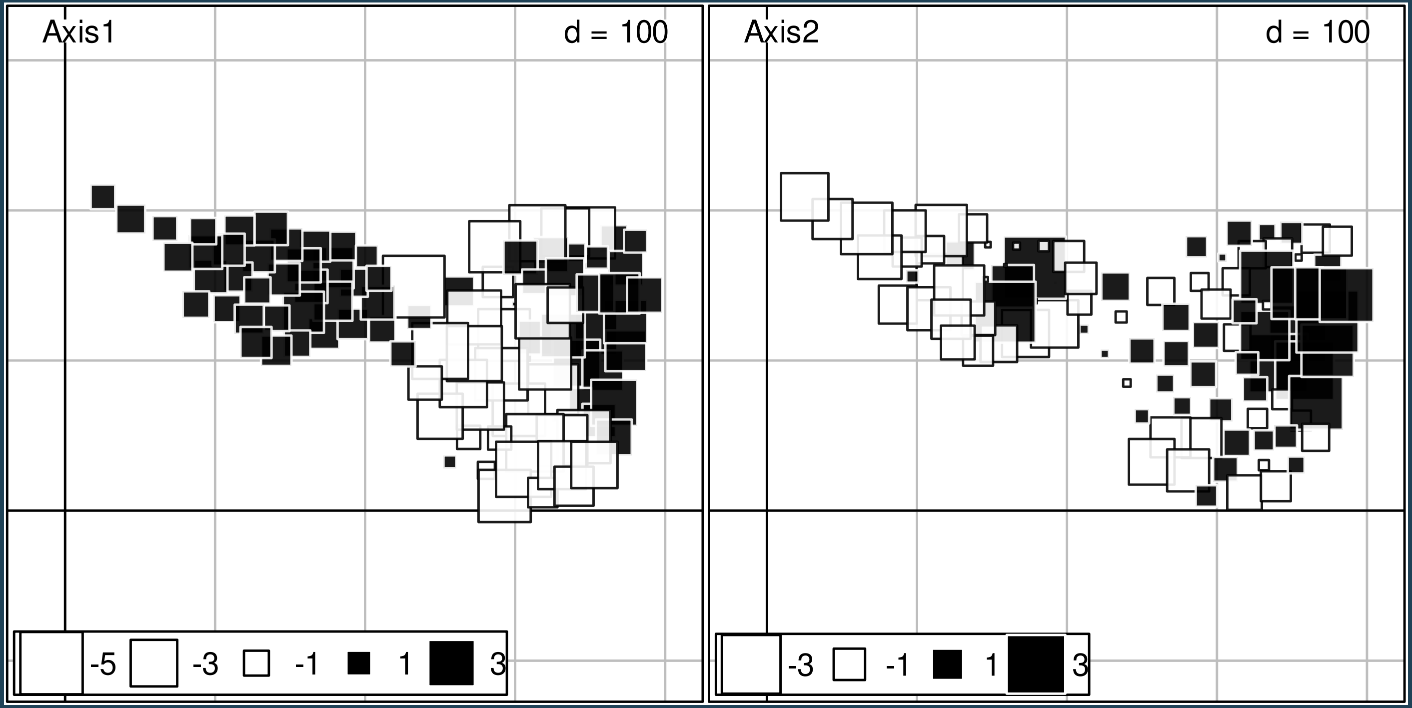


and detect spatial structures using simple tools

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

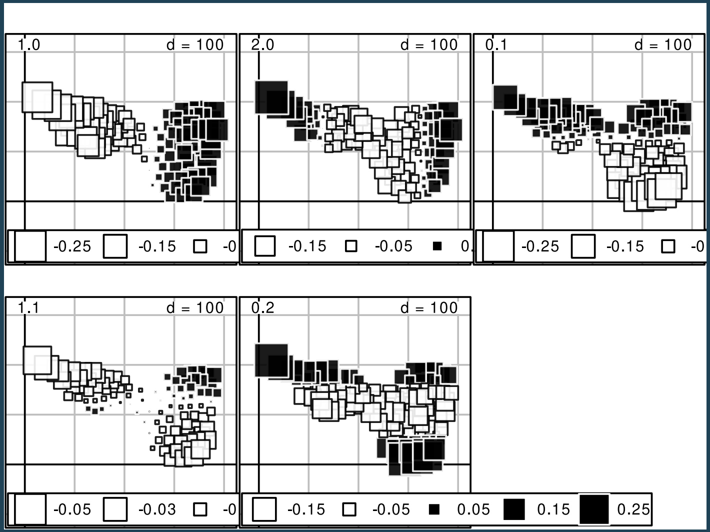
Spatial mapping

```
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)
```

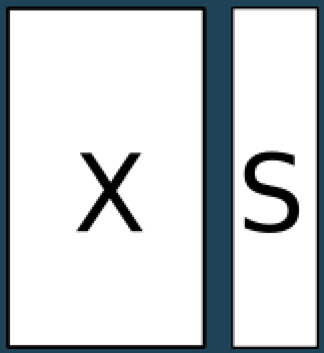



```
cor(pca_veg$li, poly.xy)
```

```
##                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



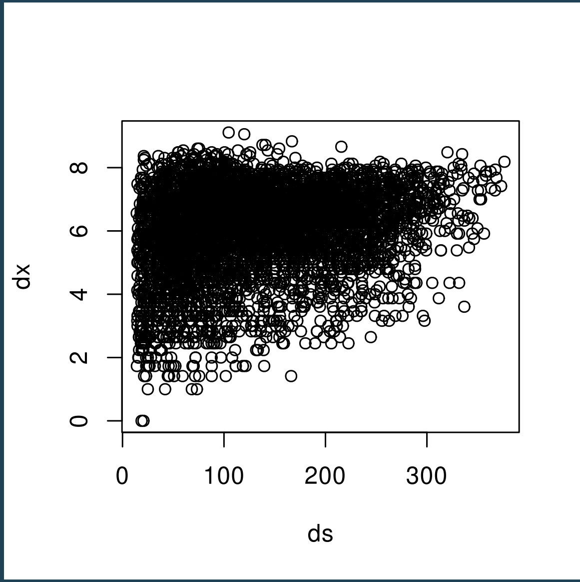
Spatial information :

- **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 $\mathbf{D_S}$
- Compute faunistic distances $\mathbf{D_X}$
- 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 = d
##
## Observation: 0.2327358
##
## Based on 999 replicates
## Simulated p-value: 0.001
## Alternative hypothesis: greater
##
##       Std.Obs   Expectation      Vari
##  8.0283347374 -0.0008646799  0.000846
```

Spatial Weighting Matrix

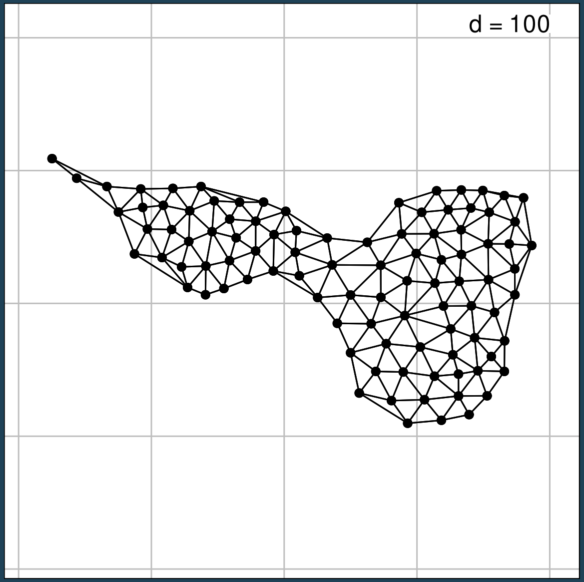
$$\mathbf{W} = [w_{ij}]$$

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

$$w_{ij} = 0 \text{ otherwise}$$

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

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

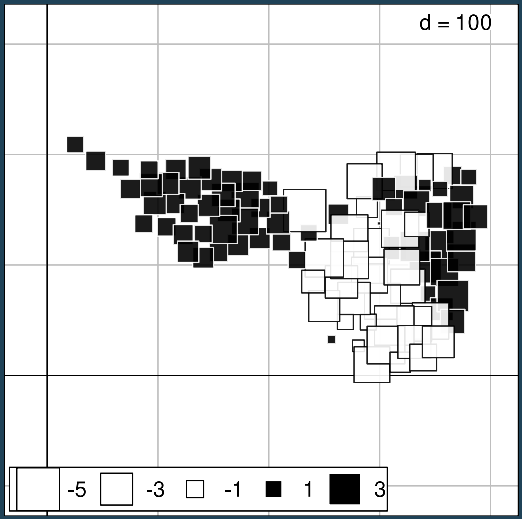
where $\mathbf{z} = \mathbf{x} - \bar{x}$. With row-standardization

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

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

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

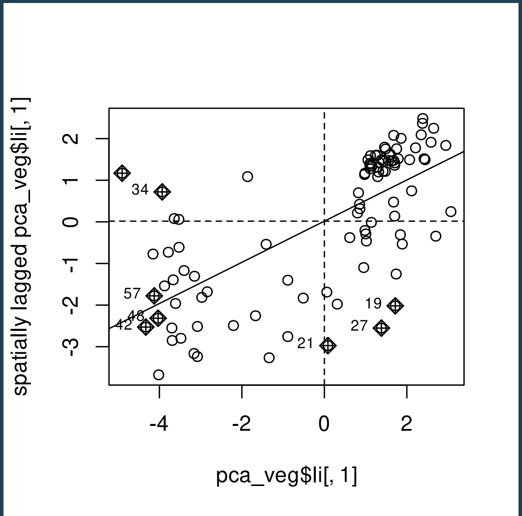



```r
moran.randtest(pca_veg$li[, 1], n
```

```
## 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
##       8.383910406      -0.014727811
```

Moran's index and scatterplot

```r
moran.plot(pca_veg$li[, 1], nb2li
```

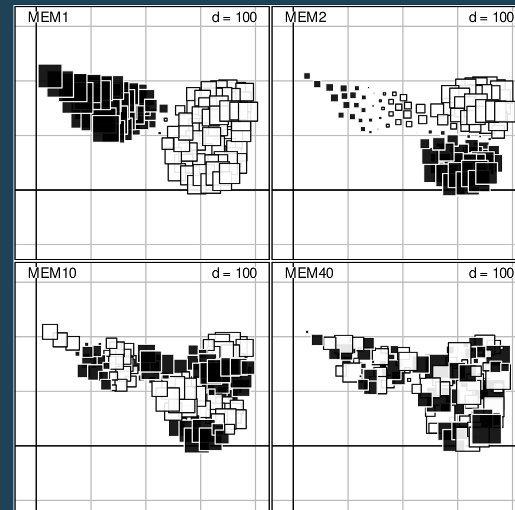


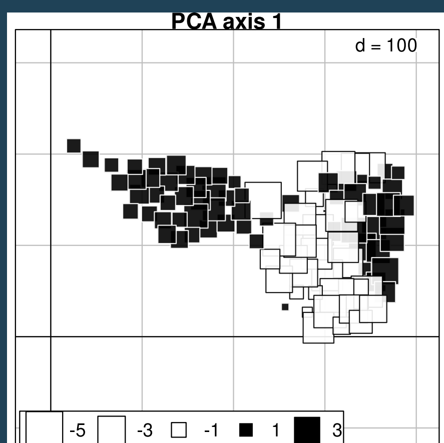

```r
moran.randtest(pca_veg$li[, 1], n
```

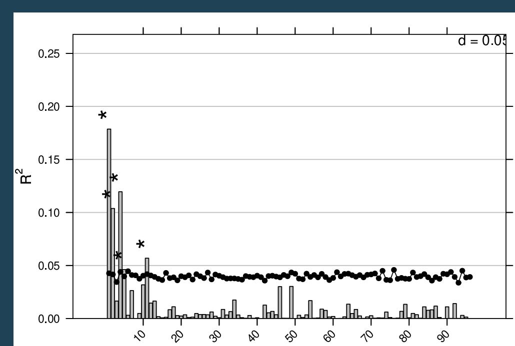
```
## Monte-Carlo test
## Call: moran.randtest(x = pca_veg$li[,
##
## Observation: 0.4947964
##
## Based on 999 replicates
## Simulated p-value: 0.001
## Alternative hypothesis: greater
##
## Std.Obs.statistic       Expectation
##       7.741228586      -0.005461975
```

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)



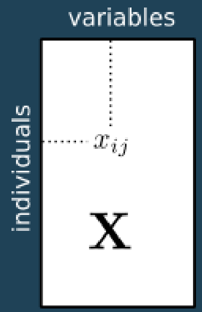


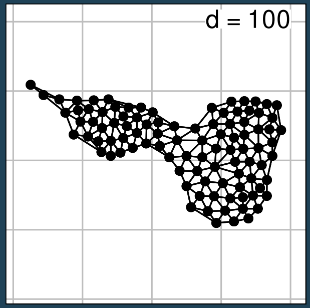


Spatial multivariate methods

- **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

$$MC_{\mathbf{D}}(\mathbf{XQa}) \cdot \|\mathbf{XQa}\|^2_{\mathbf{D}}$$

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

- the criteria optmized in the standard analysis ($\|\mathbf{XQa}\|^2_{\mathbf{D}}$)
- spatial autocorrelation ($MC_{\mathbf{D}}(\mathbf{XQa})$)

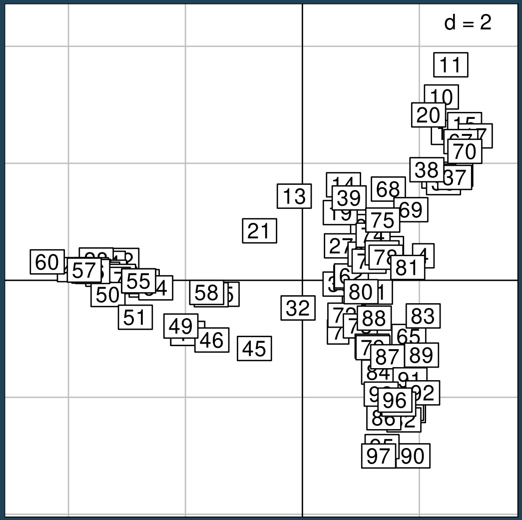
It corresponds to a coinertia analysis between the original table $\mathbf{X}$ and its lagged version $\tilde{\mathbf{X}} = \mathbf{WX}$.

Multispati

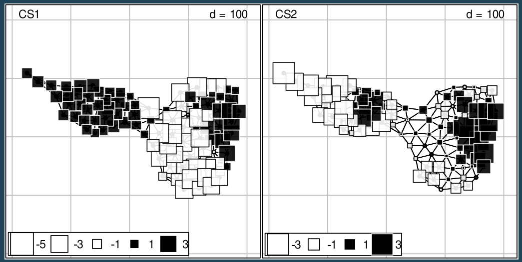
```
ms1 <- multispati(pca_veg, nb2listw(mafragh$nb), scannf = FALSE,
    nfposi = 2, nfnega = 0)
summary(ms1)
```

```
##
## Multivariate Spatial Analysis
## Call: multispati(dudi = pca_veg, listw = nb2listw(mafragh$nb), scannf = FAL
##     nfposi = 2, nfnega = 0)
##
## Scores from the initial duality diagram:
##          var      cum     ratio     moran
## RS1 5.331174 5.331174 0.2834660 0.4947964
## RS2 1.972986 7.304159 0.3883725 0.4435555
##
## Multispati eigenvalues decomposition:
##          eig      var     moran
## CS1 2.992293 4.862003 0.6154445
## CS2 1.164390 1.885904 0.6174172
```

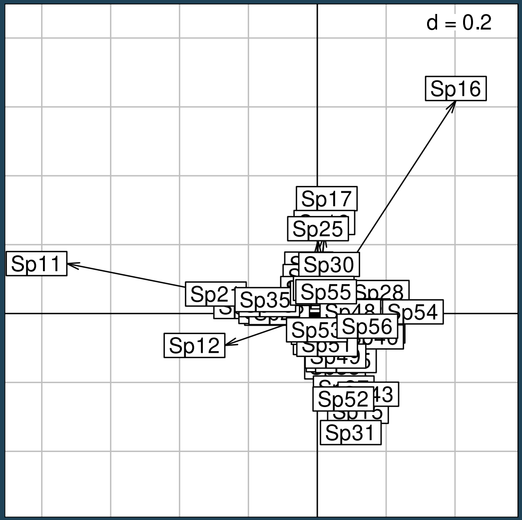
```
s.label(ms1$li)
```



```
s.value(mafragh$xy, ms1$li, nb =
```

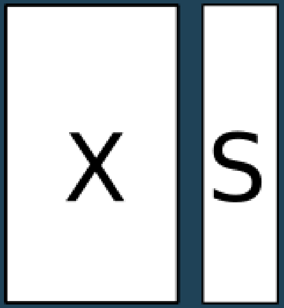


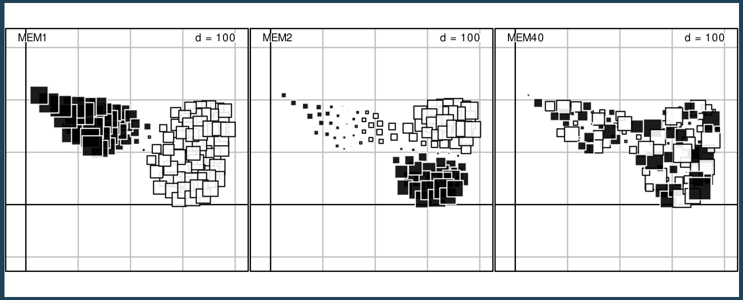
```
s.arrow(ms1$c1)
```



maximization of the product of variance and Moran's I

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 preformed 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 (variation partitioning).
- see adespatial tutorial for more details