Example 3a - PCIA

Maxime Hervé, Florence Nicolè and Kim-Anh Lê Cao

16/10/2017

Contents

Packages to load

```
library(ade4)          # Functions needed: is.euclid, procuste, randtest
library(vegan)         # Functions needed: dbrda, scores
library(RVAideMemoire) # Functions needed: MVA.synt, MVA.plot
```

Data loading

Geographic data

The first dataset is the distance matrix based on geographic data, composed of 9 rows (9 populations) crossed by 9 columns (the same 9 populations).

```
mat.dist.geography <- read.table("Example 3a - Geographic data.txt",header=TRUE)
```

R-related step: The distance matrix is tranformed in a correct format for R:

```
mat.dist.geography <- as.dist(mat.dist.geography)
```

Chemical data

The second dataset is the distance matrix based on chemical data, composed of the same rows and columns as the geographic distance matrix.

```
mat.dist.chemistry <- read.table("Example 3a - Chemical data.txt",header=TRUE)
```

R-related step: The distance matrix is tranformed in a correct format for R:

```
mat.dist.chemistry <- as.dist(mat.dist.chemistry)
```

Analysis

Step 1: PCoA

Geographic data

We first need to check wether the distance matrix has Euclidian properties:

```
is.euclid(mat.dist.geography)
```

```
[1] FALSE
```

As it is not the case, we perform the PCoA with a special correction:

```
PCoA.geography <- dbrda(mat.dist.geography~1,add=TRUE)
```

How much total variance does each component explain?

```
MVA.synt(PCoA.geography)
```

```
Criterion: total variance (%)
 Axis Proportion Cumulative
    1      92.50      92.50
    2       7.49      99.99
    3       0.00      99.99
    4       0.00     100.00
    5       0.00     100.00
```

Two components are sufficient to explain 100 % of the total variance. We will use only these ones in the PCIA. We extract the coordinates of the populations on these components, which will be used as input variables in the PCIA:

```
scores.geography <- scores(PCoA.geography,choices=1:2,display="sites")
```

Chemical data

We first need to check wether the distance matrix has Euclidian properties:

```
is.euclid(mat.dist.chemistry)
```

```
[1] TRUE
```

As it is the case, we perform the PCoA:

```
PCoA.chemistry <- dbrda(mat.dist.chemistry~1)
```

How much total variance does each component explain?

```
MVA.synt(PCoA.chemistry)
```

```
Criterion: total variance (%)
 Axis Proportion Cumulative
    1      19.57      19.57
    2      17.03      36.60
    3      14.16      50.76
```

```
4      12.34      63.09
5      10.99      74.09
```

We keep components 1 to 5 for the PCIA (74 % of total variance). We extract the coordinates of the populations on these components, which will be used as input variables in the PCIA:

```
scores.chemistry <- scores(PCoA.chemistry,choices=1:5,display="sites")
```

Step 2: PCIA

We perform the PCIA:

```
PCIA <- procuste(scores.geography,scores.chemistry)
```

We test for the concordance between the two distance matrices:

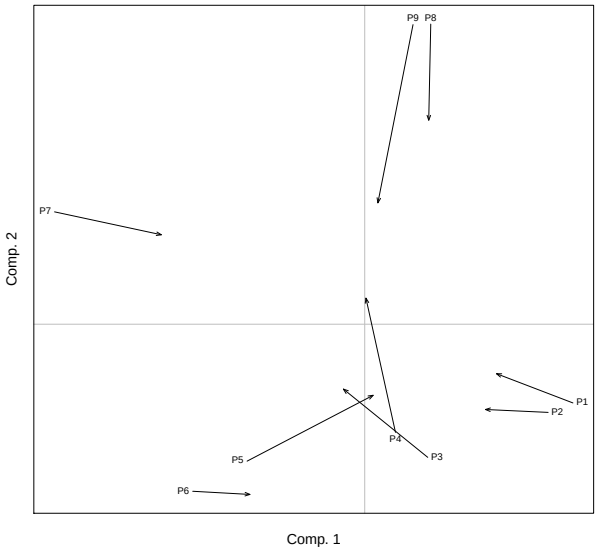
```
randtest(PCIA)
```

```
Monte-Carlo test
Call: randtest.procuste(xtest = PCIA)

Observation: 0.584303

Based on 999 replicates
Simulated p-value: 0.025
Alternative hypothesis: greater

    Std.Obs Expectation    Variance
1.708797109 0.486806561 0.003255339
```

The concordance is significant, indicating that the higher the spatial distance between two populations, the higher the difference in their cuticle's composition.

We draw a score plot where each individual is represented by an arrow. The arrow starts at the position of the individual in the geographic space and ends at the position of the same individual but in the chemical space. Such plot helps us to visualize the concordance between the two datasets:

```
MVA.plot(PCIA,"pairs",drawextaxes=FALSE)
```



Information on the current R session

```
sessionInfo()
```

```
R version 3.4.0 (2017-04-21)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 7 x64 (build 7601) Service Pack 1

Matrix products: default

locale:
[1] LC_COLLATE=French_France.1252  LC_CTYPE=French_France.1252
[3] LC_MONETARY=French_France.1252 LC_NUMERIC=C
[5] LC_TIME=French_France.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] RVAideMemoire_0.9-68 vegan_2.4-4          lattice_0.20-35
[4] permute_0.9-4        ade4_1.7-8           knitr_1.17

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.13       cluster_2.0.6      magrittr_1.5
 [4] splines_3.4.0      MASS_7.3-47        minqa_1.2.4
 [7] stringr_1.2.0      car_2.1-5          tools_3.4.0
[10] pbkrtest_0.4-7     nnet_7.3-12        parallel_3.4.0
[13] grid_3.4.0         nlme_3.1-131       mgcv_1.8-22
[16] quantreg_5.33      MatrixModels_0.4-1 htmltools_0.3.6
[19] lme4_1.1-14        yaml_2.1.14        rprojroot_1.2
[22] digest_0.6.12      Matrix_1.2-11      nloptr_1.0.4
[25] evaluate_0.10.1    rmarkdown_1.6      stringi_1.1.5
```

```
[28] compiler_3.4.0     backports_1.1.1    SparseM_1.77
```