

Lab 9 - Spatial analysis: Trend surface analysis and PCNM

by Benoit Gendreau-Berthiaume

Spatial structure in ecological communities is increasingly recognized as being important for the understanding of the processes driving communities. In ecological communities, spatial patterns are either driven by environmental factors or biotic processes. Thus to adequately understand ecological communities it is important to identify the spatial structures and then relate these patterns to underlying processes. Different ecological processes can occur at different spatial scale (fine vs broad scales). It is thus important to detect and quantify spatial patterns over a wide range of scales.

Previously, geographic coordinates have been used directly as explanatory variables in constrained ordinations (RDA, CCA) by using polynomial terms (x , x^2 , x^3). This method is called trend-surface analysis. The problem with this method is that it was conceived to model broad scale patterns and thus finer spatial structure cannot be detected. Nowadays, trend-surface analyses are used to detrend data. This is usually done by regressing all variables on the X-Y coordinates and retaining the residuals.

A recent method has been developed to detect and quantify spatial patterns over a wide range of spatial scales: Principle coordinates of neighbour matrices PCNM (Brocard and Legendre 2002). The PCNM method has been found to be best applied to detrended data (Borcard et al. 2004). This is because if a general trend is present in the data, about half the PCNM variables would be used to describe this general trend and thus their role in modeling finer spatial patterns would go unnoticed.

For this lab we will be using mites species data (mite.csv), environmental variables data (env.csv) and the geographic coordinates of the sites sampled (xy.csv) to explore the PCNM method.

9.1. Install required packages and functions for this lab

You can skip this step and move to the front row of the lab, which has everything installed already. If you want to do it yourself, here are the instructions:

PCNM only works on the latest version of R. You can download it here:

<http://cran.r-project.org/bin/windows/base/>

Then, you need to install the packages "packfor", "AEM", and "PCNM" available here:

https://r-forge.r-project.org/R/?group_id=195

You need another number of packages, so easiest to execute this directly out of R with this code:

```
install.packages("ape")
install.packages("spdep")
install.packages("ade4")
install.packages("vegan")
install.packages("packfor", repos="http://R-Forge.R-project.org")
install.packages("AEM", repos="http://R-Forge.R-project.org")
install.packages("PCNM", repos="http://R-Forge.R-project.org")
```

Now double-click the desktop icon to start R, set your working directory, and add a special function, which is included with the data package that you can download from the course website:

```
setwd("C:/Documents and Settings/866-W24/Desktop/PCNM_Lab")
source("sr.value.R")      #Function must be in working directory
```

Now can start importing the data, which is in a different format this time:

```
mite=read.table("mite.txt")
mite.env=read.table("mite_env.txt")
mite.xy=read.table("mite_xy.txt")
```

Because RDA analyses (used in this lab) have been found to be inappropriate for raw species abundance data involving many null abundances (which is the case with the mite data), a Hellinger transformation of the raw mite data is made prior to the RDA analysis (Legendre and Gallagher 2001).

```
#Hellinger data transformation
mite.h=decostand(mite, "hellinger")
```

9.2. Trend surface analysis

First let's visualize different spatial pattern modelled by trend-surface analysis using a theoretical example.

```
#First simple models on a square, regularly sampled surface
#construct a plot a 10 x 10 grid

xygrid=expand.grid(1:10, 1:10)
plot(xygrid)

xygrid.c=scale(xygrid, scale=FALSE) #centring
plot(xygrid.c)
X=xygrid.c[,1] # give the x coordinate of each sample point
Y=xygrid.c[,2] # give the y coordinate of each sample point

#Plot some first, second and third-degree functions of x and y
par(mfrow=c(3,3))
s.value(xygrid, (X))
s.value(xygrid, (Y))
s.value(xygrid, (X^2))
s.value(xygrid, (Y^2))
s.value(xygrid, (X*Y))
s.value(xygrid, (X^2*Y))
s.value(xygrid, (Y^2*X))
s.value(xygrid, (X^3))
s.value(xygrid, (Y^3))
```

You can play around with the different functions and see what spatial pattern they model

Now let's run the trend surface analysis of the actual mite data. First we will use a function poly() to construct third-degree orthogonal polynomials of the geographic coordinates. Then we will run a RDA with the polynomials as explanatory variables.

```
mite.poly.ortho=poly(as.matrix(mite.xy), degree=3) #raw=F default
colnames(mite.poly.ortho)=
  c("X", "X2", "X3", "Y", "XY", "X2Y", "Y2", "XY2", "Y3")

#RDA using all 9 polynomial terms
mite.trend.rda.ortho=rda(mite.h~., data=as.data.frame(mite.poly.ortho))
```

```
#Computation of the adjusted R2.
R2adj.poly=RsquareAdj(mite.trend.rda.ortho)$adj.r.squared
```

Adjusted R² measures the unbiased amount of explained variation. This is because the number of explanatory variable in the species matrix influences the R² thus inflating the apparent amount of explained variance because of random correlations.

One could want to reduce to number of explanatory variable to retain only the most significant and important ones. To reduce the risk of incorporating too many variable Blanchet et al. (2008) proposed a forward selection based on two criterions. The forward selection is stop if either the significance level *alpha* is reached (i.e.variable is non-significant) or the global R²_{adj} is exceeded (i.e. a variable brings the R²_{adj} of the current model over the value of the R²_{adj} of the global model).

```
#Forward selection using Blanchet et al.(2008) double stopping criterion
mite.trend.fwd=forward.sel(mite.h, mite.poly.ortho,
adjR2thresh=R2adj.poly)
```

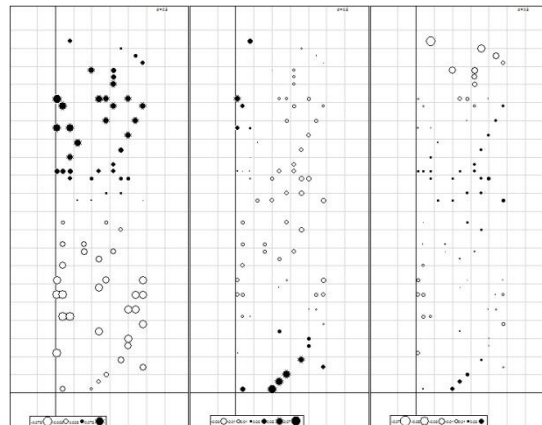
```
#New RDA using the 6 terms retained
mite.trend.rda=rda(mite.h~,
  data=as.data.frame(mite.poly.ortho)[,mite.trend.fwd[,2]])
summary(mite.trend.rda)
```

```
#overall test and test of canonical axes
anova.cca(mite.trend.rda, step=1000)
anova.cca(mite.trend.rda, step=1000, by="axis")
```

Questions: How much of the total variance is explained by the first three axis of the RDA (use summary function)? How much of the variance explained by the constrained ordination is explained by the first axis (See Accumulated constrained eigenvalues)?

Let's Plot the three independent significant spatial structures. For square bubbles type "s.value" instead of "sr.value"

```
mite.trend.fit=scores.cca(mite.trend.rda,
  choices=c(1,2,3),
  display="lc", scaling=1)
#For explanation on scaling see below
par(mfrow=c(1,3))
sr.value(mite.xy, mite.trend.fit[,1])
sr.value(mite.xy, mite.trend.fit[,2])
sr.value(mite.xy, mite.trend.fit[,3])
```



Scaling: There are two types of scaling in ordinations, scaling 1 & 2. In scaling 1 the distance in the biplots are approximations of their Euclidean distances in multidimensional space. In scaling 2 the angles between descriptors in the biplot reflect their correlation. In the present case, scaling 1 is used to display “pure” spatial model i.e. the linear combination of spatial variables, in a projection preserving the Euclidean distances among sites. In this case the descriptors (PCNM variables) are orthogonal thus we know they are independent and we are thus not interested in their correlation with one another.

9.3. Detrending the mite data

Because the PCNM method has been found to be best applied to detrended data (see lab introduction) before running the PCNM analysis on the mite data, the data is detrended by regressing all variables on the X-Y coordinates and retaining the residuals. Before detrending we however test if the trend is significant.

```
anova(rda(mite.h, mite.xy)) #result: significant trend

#computation of linearly detrended mite data
mite.h.det=resid(lm(as.matrix(mite.h)~., data=mite.xy))
```

9.4. PCNM visualisation

Before running the PCNM on the mite data let's visualize what PCNM might look like for a two dimensional sampling design. For this purpose we will build an equispaced grid of 20 by 20 points separate 1m apart.

```
# Equispaced 20x20 grid

xygrid2=expand.grid(1:20, 1:20)
plot(xygrid2)
xygrid2.dl=dist(xygrid2)
thresh=1 #Truncation to threshold 1
xygrid2.dl[xygrid2.dl>thresh]=4*thresh #Truncate the distance matrix

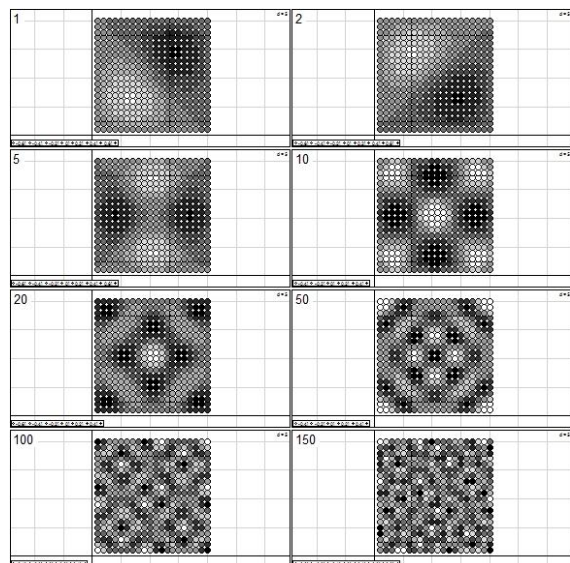
#PCoA of truncated matrix
xygrid2.PCoA=cmdscale(xygrid2.dl, eig=T, k=nrow(xygrid2)-1)
#warning message normal

#Count number of positive eigenvalues
nb.ev2=length(which(xygrid2.PCoA$eig>0.000000001))
#Matrix of PCNM variables
xygrid2.PCNM = xygrid2.PCoA$points[,1:nb.ev2]
```

Now let's plot some PCNM variables modelling positive spatial correlation to better visualise what they represent.

```
par(mfrow=c(4,2))
somePCNM2=c(1,2,5,10,20,50,100,150)
for (i in 1:length(somePCNM2)){
  sr.value(xygrid2,
xygrid2.PCNM[,somePCNM2[i]],
  method="greylevel",
  csize=0.35,sub=somePCNM2[i], csub=2)
}
```

You can explore different PCNM variables by changing the numbers in somePCNM2



9.5. PCNM of the mite data

- Construct the matrix of PCNM variables step by step... (This is an optional section, if you are not interested in doing this manually, skip to the next bullet which does the same automatically with PCNM)

```
xy.dl=dist(mite.xy)
spanning=spantree(xy.dl)
#choosing the threshold as the maximum distance between two points
dmin=max(spanning$dist)
#truncate the distance matrix
xy.dl[xy.dl>dmin]<- 4*dmin

#PCoA of truncated distance matrix
xy.PCoA=cmdscale(xy.dl, k=nrow(mite.xy)-1, eig=T)

#Count the positive eigenvalues
nb.ev=length(which(xy.PCoA$eig>0.0000001))

#construct a data frame with PCNM variables
mite.PCNM=as.data.frame(xy.PCoA$points[1:nrow(mite.xy), 1:nb.ev])
summary(mite.PCNM)
```

The next step would be to calculate Moran's I for each Eigenvectors for the threshold distance (in this case 0 to 1.011187). This is not done here. Instead we use the automatic PCNM function that calculated them for us (next section).

- Construct the PCNM variables automatically using PCNM()

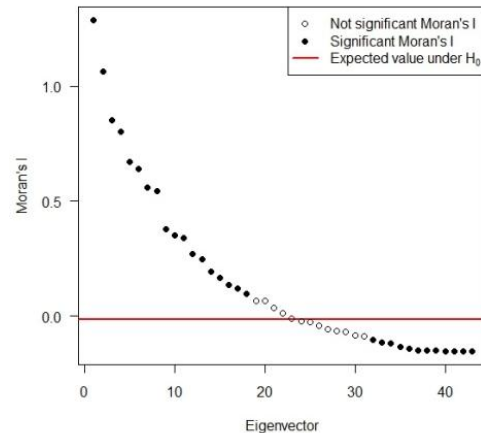
For PCNM analysis it is possible to use the function PCNM() of the package PCNM. This method will automatically construct the PCNM variables and also calculate the Moran's I for each variable. In this function, the default method to determine the threshold distance is by using the minimum spanning tree but one can also decide what threshold distance he wants. The minimum spanning tree method insures that all points are connected which is important for the PCNM analysis.

```
xy.dl=dist(mite.xy)
mite.PCNM.auto=PCNM(xy.dl)
summary(mite.PCNM.auto)
```

```
#plot the minimum spanning tree used to find the truncation distance
plot.spantree(mite.PCNM.auto$spanning, mite.xy)
dmin= mite.PCNM.auto$thresh #truncation distance
nb.ev=length(mite.PCNM.auto$values) #nbr of eigenvalues
```

Moran's I of the PCNM variables (in the first distance class 0 to truncation threshold).

```
#Expected value of I, no spatial correlation
mite.PCNM.auto$expected_Moran
```



```
#Moran's I for each PCNM variables
mite.PCNM.auto$Moran_I

#Eigenfunction with positive spatial correlation
select=which(mite.PCNM.auto$Moran_I$Positive == T)
mite.PCNM.pos=
  as.data.frame(mite.PCNM.auto$vectors)[,select]
```

- Run the global PCNM analysis on the *detrended* mite data

Now that we have selected only the PCNM that model positive spatial correlation we can run the RDA using only these PCNM variables.

```
mite.PCNM.rda=rda(mite.h.det~., mite.PCNM.pos)
anova.cca(mite.PCNM.rda)
```

Since the analysis is significant, compute the adjusted R^2 and run forward selection of the PCNM variables.

```
mite.R2a=RsquareAdj(mite.PCNM.rda)$adj.r.squared
mite.PCNM.fwd= forward.sel(mite.h.det, as.matrix(mite.PCNM.pos),
  adjR2thresh=mite.R2a)
nb.sig.PCNM=nrow(mite.PCNM.fwd)
```

```
#Identity of Significant PCNMs in increasing order
PCNM.sign=sort(mite.PCNM.fwd[,2])
#Write the significant PCNMs to a new object
PCNM.red=mite.PCNM.pos[,c(PCNM.sign)]
```

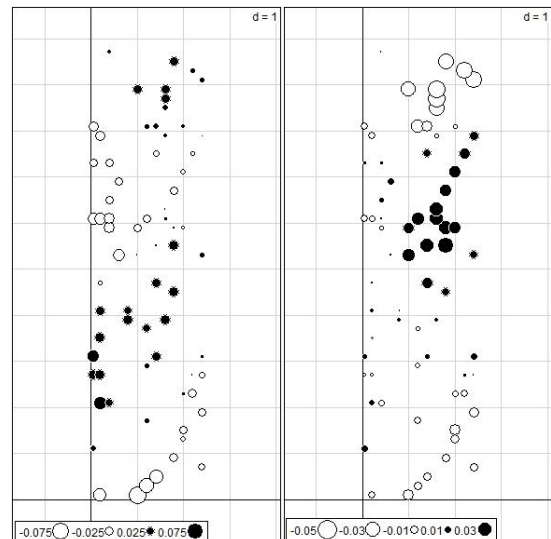
Now that we have selected the best PCNM variables we run a new PCNM analysis with the 10 significant PCNM variables

```
mite.PCNM.rda2=rda(mite.h.det~., data=PCNM.red)
summary(mite.PCNM.rda2)
mite.fwd.R2a=RsquareAdj(mite.PCNM.rda2)$adj.r.squared
anova.cca(mite.PCNM.rda2)
axes.test= anova.cca(mite.PCNM.rda2, by="axis")
nb.ax=length(which(axes.test[,5]<=0.05)) #number of significant axes
```

Finally lets plot the two significant canonical axes of the RDA with the 10 significant PCNM variables.

```
mite.PCNM.axes=scores.cca(mite.PCNM.rda2,
  choices=c(1,2),
  display="lc", scaling=1)
par(mfrow=c(1,2))
sr.value(mite.xy, mite.PCNM.axes[,1])
sr.value(mite.xy, mite.PCNM.axes[,2])
```

Now that we know that spatial structures are present in the mite data it is possible to determine how these two significant spatial patterns (two significant canonical axes) are related to the environmental variables using linear regression models (using function `lm()`). However before running the regressions we first have to test for normality of residuals of these linear models.



```
#Normality test for first axis
shapiro.test(resid(lm(mite.PCNM.axes[,1]~., data=mite.env)))
hist(resid(lm(mite.PCNM.axes[,1]~., data=mite.env)))
```

It is often more useful to plot histogram of the residuals and assess the normality visually. This is because the Shapiro-Wilk test is overly conservative when you have large sample size and not very meaningful when sample size is small.

```
#Linear model of the first axis
mite.PCNM.axis1.env=lm(mite.PCNM.axes[,1]~., data=mite.env)
summary(mite.PCNM.axis1.env)
```

```
#Normality verification for second axis
hist(resid(lm(mite.PCNM.axes[,2]~., data=mite.env)))
```

```
#Linear model of the second axis
mite.PCNM.axis2.env=lm(mite.PCNM.axes[,2]~., data=mite.env)
summary(mite.PCNM.axis2.env)
```

Question: Are both these spatial pattern related to the same environmental variables? How much variance in these spatial patterns is explained by the environmental variables?

Other question: Would it be possible to evaluate which mite species are most related to the observed spatial patterns? Try: `plot(mite.PCNM.rda2, display="sp", scaling=2)`
Scaling 2 is used because we are now interested in the angle between species vectors.

Questions: What does this graph mean? How do you interpret this result with those regarding the environmental factors?

9.6 Variation partitioning of the Mite – trend – Environment – PCNM variables

Variation partitioning enable us to determine the various unique and combined fractions of variation explained in the mite community data by the environmental data, the broad scale spatial pattern (trend surface analysis) and the finer scale spatial patterns (PCNMs). For this analysis it is better not to detrend the data and instead test for a linear trend and incorporate it explicitly in the partitioning procedure. In the following section we will independently forward select the X-Y coordinates, the environmental variables and the PCNM variables before variation partitioning.

- Test trend. If significant, forward selection of coordinates

```
#Test if trend is significant
mite.XY.rda=rda(mite.h, mite.xy)
anova.cca(mite.XY.rda)

#Forward selection of coordinates
mite.XY.R2a= RsquareAdj(mite.XY.rda)$adj.r.squared
mite.XY.fwd= forward.sel(mite.h, as.matrix(mite.xy),
adjR2thresh=mite.XY.R2a)

#Write the significant coordinates to a new object
XY.sign= sort(mite.XY.fwd$order)
XY.red=mite.xy[,c(XY.sign)]
```

- Test and forward selection of environmental variables

Because the variation partitioning method can only handle numeric variables, we have to recode the environmental variables 3 to 5 into dummy binary variables.

```
substrate= model.matrix(~mite.env[,3])[, -1]
shrubs= model.matrix(~mite.env[,4])[, -1]
topo= model.matrix(~mite.env[,5])[, -1]
mite.env2=cbind(mite.env[,1:2], substrate, shrubs, topo)

#test environmental variables
mite.env.rda=rda(mite.h, mite.env2)
anova.cca(mite.env.rda)

#Forward selection
mite.env.R2a=RsquareAdj(mite.env.rda)$adj.r.squared
mite.env.fwd=forward.sel(mite.h, mite.env2,
  adjR2thresh=mite.env.R2a, nperm=9999)

#Write the significant environmental variables to object
env.sign=sort(mite.env.fwd$order)
env.red=mite.env2[,c(env.sign)]
colnames(env.red) #names of significant environmental variables
```

- Test and forward select the PCNM variables

Run the global PCNM analysis on the undetrended mite data.

```
#test the significance of PCNM variables
mite.undet.PCNM.rda=rda(mite.h, mite.PCNM.pos)
anova.cca(mite.undet.PCNM.rda)

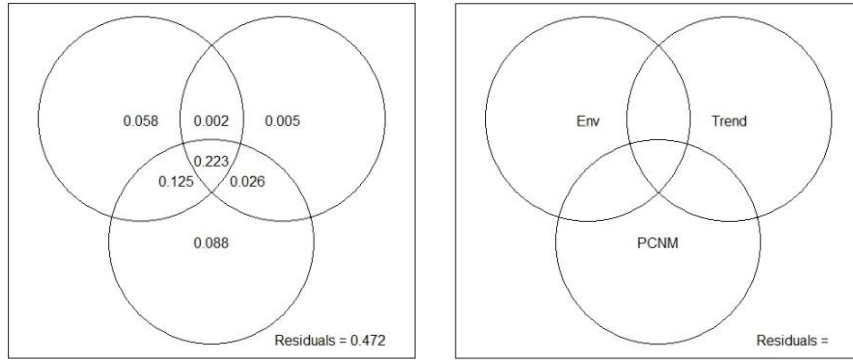
#If significant run forward selection
mite.undet.PCNM.R2a=RsquareAdj(mite.undet.PCNM.rda)$adj.r.squared
mite.undet.PCNM.fwd=forward.sel(mite.h, as.matrix(mite.PCNM.pos),
  adjR2thresh=mite.undet.PCNM.R2a)

#the number of significant PCNM
nb.sig.PCNM = nrow(mite.undet.PCNM.fwd)
# Identity of significant PCNMs in increasing order
PCNM.sign=sort(mite.undet.PCNM.fwd$order)
#Write significant PCNMs to a new object
PCNM.red=mite.PCNM.pos[,c(PCNM.sign)]
```

- Mite - environment - trend - PCNM variation partitioning

```
mite.varpart= varpart(mite.h, env.red, XY.red, PCNM.red)
par(mfrow=c(1,2))
showvarparts(3)
#can also try inserting fraction names using
showvarparts(3, c("Env", "Trend", "PCNM", "", "", "", "", ""))
plot(mite.varpart, digits=2) #see book for other option p 260
```

Question: How much variance is explained by the whole set of environmental variables and spatial variables?



- Tests the significant fractions [a], [b] and [c]

```
#Fraction [a] pure environmental
anova.cca(rda(mite.h, env.red, cbind(XY.red, PCNM.red)))
#Fraction [b] pure trend
anova.cca(rda(mite.h, XY.red, cbind(env.red, PCNM.red)))
#Fraction [c] pure PCNM spatial variation
anova.cca(rda(mite.h, PCNM.red, cbind(env.red, XY.red)))
```

Questions: Are all the pure fractions (environmental, trend and PCNM) significant?
How much variance is explained by the environmental variables? How much of this variance is not spatially structured?

The fraction of variance explained by “pure” (not spatially structured) environmental variables represents species-environment relationships associated with local environment.

Question: How much variance is explained only by the trend and the PCNM variables respectively?

Questions: How much variance is jointly explained by the environmental variables, the Y coordinate of the sampling sites (only significant coordinate in the trend) and the PCNM variation?

“This is a typical case of induced spatial variation, where spatial structure of the environmental factors produces a similar spatial structure in the response data” (Brocard et al. 2011). However one should be careful before inferring causal species-environment relationships i.e. the environmental variables directly influence the species. This is because the correlation between environmental variables and mite community can also be due to unmeasured processes that influence both the mite community and the environmental variables.

References

- Blanchet, F. G., P. Legendre and D. Borcard. 2008. Forward selection of explanatory variables. *Ecology* 89(9): 2623-2632.
- Borcard, D. and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153(1-2): 51-68.
- Borcard, D., P. Legendre, A.-J. Carol and T. Hanna. 2004. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85(7): 1826-1832.
- Borcard, D., F. Gillet and P. Legendre. 2011. *Numerical ecology in R*. Springer, New York, p. 306.
- Legendre, P. and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129(2): 271-280.