

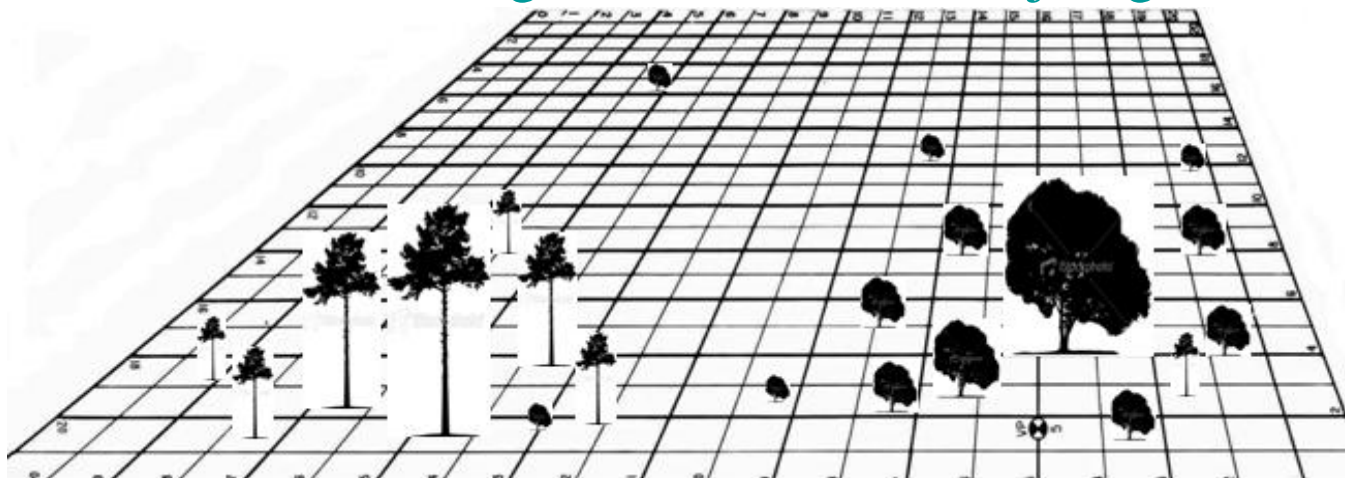
# 生态学数据空间分析 与R实现

赖江山

[lai@ibcas.ac.cn](mailto:lai@ibcas.ac.cn)

[www.planta.cn](http://www.planta.cn)

[blog.sciencenet.cn/u/laijiangshan](http://blog.sciencenet.cn/u/laijiangshan)



» 数据分析与模拟丛书

Daniel Borcard François Gillet Pierre Legendre 著  
赖江山 译

Numerical Ecology with R

# 数量生态学

——R 语言的应用

高等教育出版社



Daniel Borcard 博士

加拿大蒙特利尔大学生物科学系高级研究员。长期从事数量生态学、群落生态学和生物统计学方面的科研和教学工作，曾获蒙特利尔大学优秀教学奖。



François Gillet 博士

法国弗朗什孔泰大学 (Université de Franche-Comté) 科学与技术学院教授。长期从事群落生态学、植被生态学、数量生态学和生态模型方面的科研和教学工作。



Pierre Legendre 博士

加拿大蒙特利尔大学生物科学系教授，加拿大皇家学会会员。在群落生态学、数量生态学、统计生态学 and 系统发育中的数量方法等领域具有很高的造诣，是国际数量生态学界的权威，著有影响力极大的 *Numerical Ecology* 一书，发表论文 200 余篇。

# 主要内容

- 空间相关（Moran相关系数和Mantel空间相关系数）
- 趋势面分析
- PCNM与变差分解（variation partitioning）



# 1.空间相关

空间相关（**spatial correlation**）用来度量这样一种事实：空间接近的两个点之间的值比随机抽取的两个点之间的值更相似（正相关）或更相异（负相关）。这种由真正的自相关或空间依赖引起的空间结构对统计检验有不利的影响。对于具有空间相关的响应变量，如果已知样方的位置和生物过程，至少在某种程度上可以通过其他样方的值预测目标样方的值。这种相关关系意味着样方值之间彼此不是随机独立，这种情况违背了样本独立性的统计假设。换句话说，每个新的样本并未带来一个新的自由度。因此，在分析具有空间相关的数据时，参数检验的自由度往往被高估，导致检验结果偏向“自由”的一边：即导致零假设经常被错误地拒绝。

# 两类空间相关

诱导性空间依赖 (induced spatial dependence)

$$y_j = \mu_y + f(\mathbf{X}_j) + \varepsilon_j$$

空间自相关 (spatial autocorrelation)

$$y_j = \mu_y + \sum f(y_i - \mu_y) + \varepsilon_j$$

# 空间相关系数

Moran's  $I$  (*Moran 1950*) (一元定量变量)

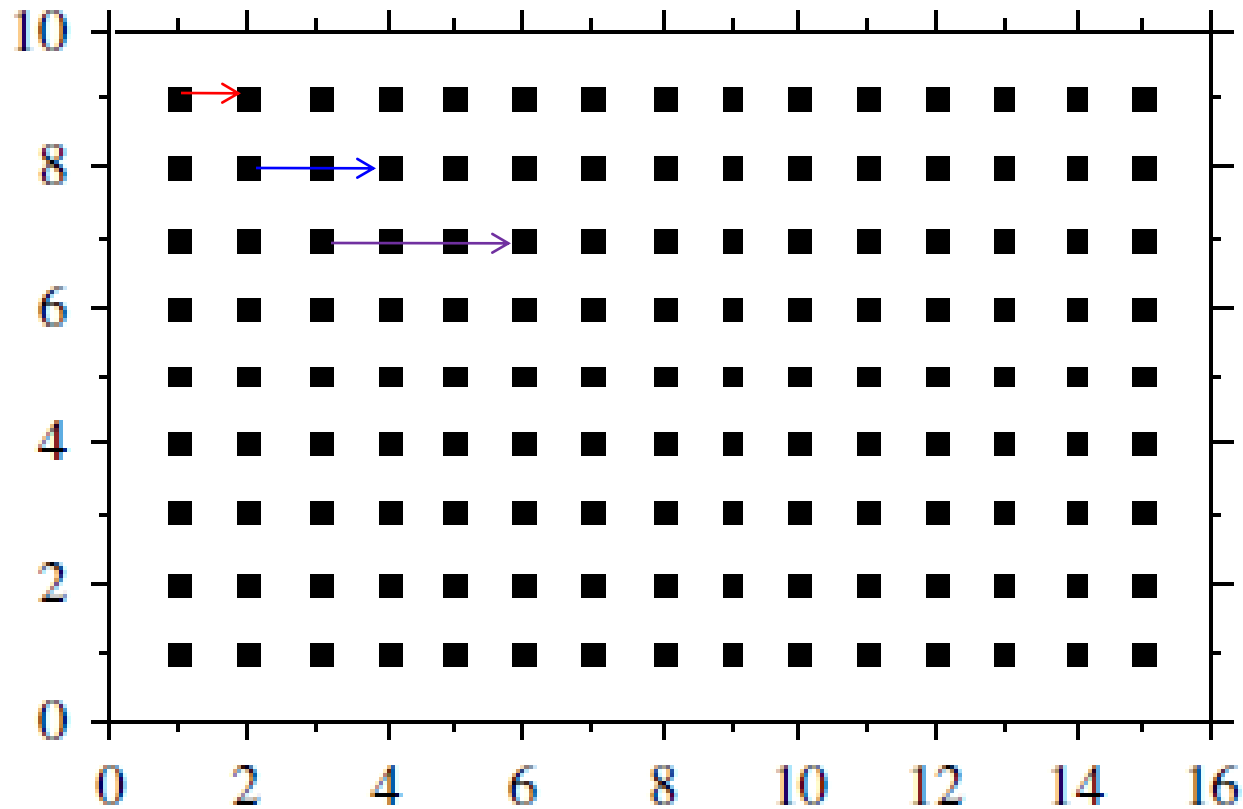
$$I(d) = \frac{\frac{1}{W} \sum_{h=1}^n \sum_{i=1}^n w_{hi} (y_h - \bar{y})(y_i - \bar{y})}{\frac{1}{n} \sum_{i=1}^n w_{hi} (y_i - \bar{y})^2} \quad \text{for } h \neq i \quad (7.3)$$

在没有空间相关时，Moran指数的期望值是：

$$E(I) = \frac{-1}{n-1} \quad (7.4)$$

当Moran指数值小于  $E(I)$  表示负空间相关，大于  $E(I)$  表示正相关。当  $n$  (样本数) 很大时， $E(I)$  接近于0。

# 距离等级



1	2
2	3
6	6
9	3
8	4
9	7
0	4
7	5

经纬度坐标（有时可以简写为Lat/Lon或LatLon）可以通过SoDA程序包内geoXY()函数转换为笛卡尔坐标



# 甲螨数据集



蒙特利尔大学研究站

变量名称	代码	单位
基质密度（干物质）	SubsDens	$\text{g dm}^{-3}$
含水量	WatrCont	$\text{g dm}^{-3}$
土质	Substrate	7个无序的分类
灌丛	Shrub	3个有序的分类
微地形	Topo	平地-山丘

甲螨数据集也由三个文件构成，包含35个形态种多度数据、5个基质和微地形数据和70个取样点笛卡尔坐标（ $x-y$ ）。

```
# 载入本章所用的程序包
```

```
library(ape)
```

```
library(spdep)
```

```
library(vegan)
```

```
library(ade4)
```

```
# 以下几个程序包可以从https://r-forge.r-project.org/R/?group\_id=195下载
```

```
library(packfor)
```

```
library(spacemakeR)
```

```
library(AEM)
```

```
library(PCNM)
```

```
source("plot.links.R") # 脚本函数必须在当前工作文件夹中
```

```
source("sr.value.R") # 脚本函数必须在当前工作文件夹中
```

```
# 导入数据
```

```
mite <- read.csv("mite.csv",header=T, row.names=1)
```

```
mite.env <- read.csv ("mite_env.csv",header=T, row.names=1)
```

```
mite.xy <- read.csv ("mite_xy.csv",header=T, row.names=1)
```

```
mite.h <- decostand (mite, "hellinger")
```

```
mite.xy.c <- scale(mite.xy, center=TRUE, scale=FALSE) #不对坐标进行标准化
```

```
# 空间相关图（基于Moran指数）
```

```
# *****
```

```
# 寻找距离在0.7m范围内的所有样方对，并计算样方对之间的滞后阶数
```

```
plot.links(mite.xy, thresh=0.7)
```

```
nb1 <- dnearneigh(as.matrix(mite.xy), 0, 0.7) #longlat = NULL
```

```
summary(nb1)
```

```
# 基质密度的相关图
```

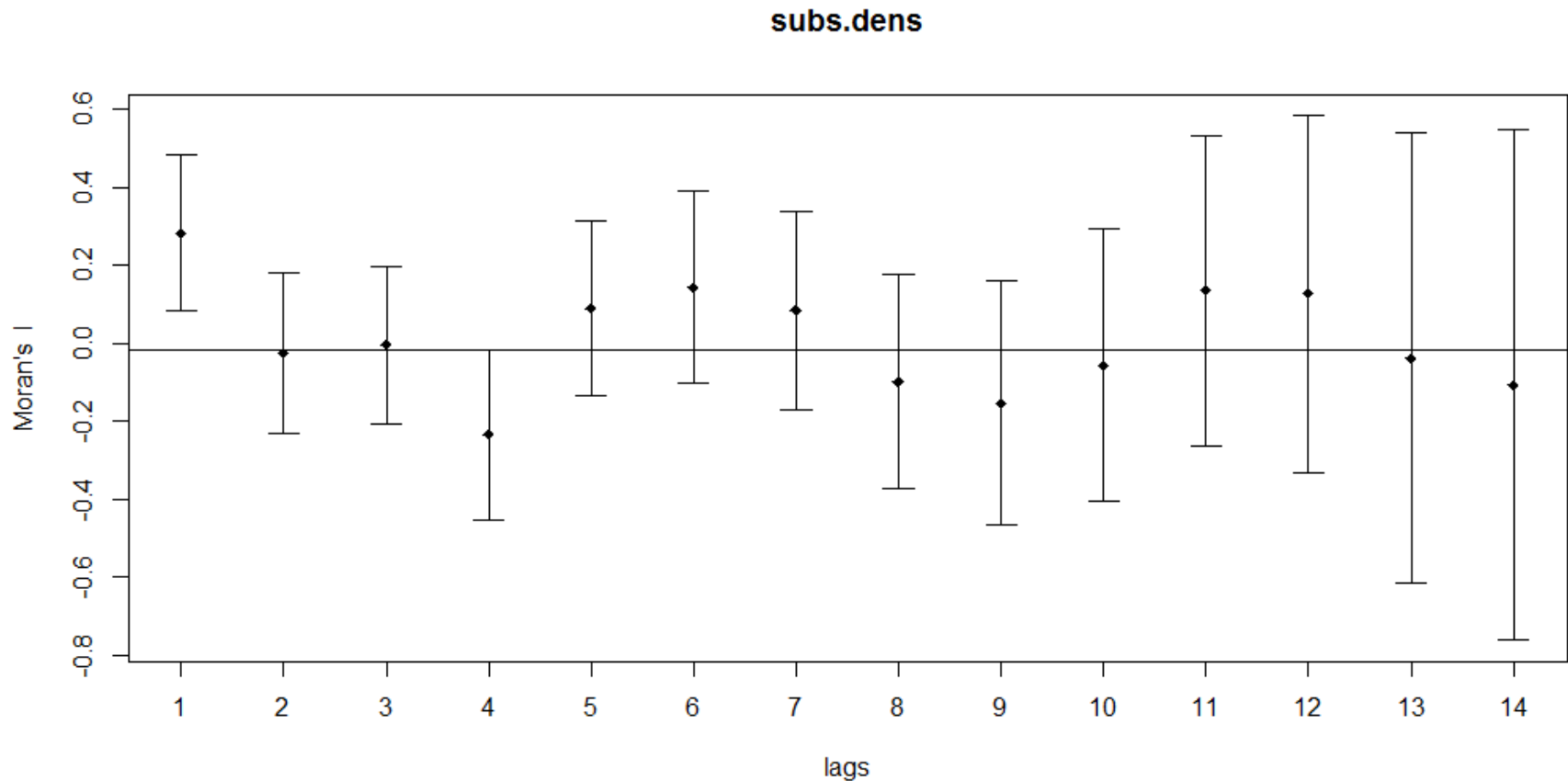
```
subs.dens <- mite.env[,1]
```

```
subs.correlog <- sp.correlogram(nb1, subs.dens, order=14, method="l",
```

```
zero.policy=TRUE)
```

```
print(subs.correlog, p.adj.method="holm")
```

```
plot(subs.correlog)
```



#基质密度的相关图只有一个显著的距离等级：在距离等级为1（即0.0-0.7m）显示正的空间相关。距离等级4似乎显示负的空间相关，但多重检验Holm（1979）校正后的p值表明这个距离等级下Moran指数并不显著。除此之外，没有识别到显著的空间相关，也就是说，当距离超过0.7m或更保守估计超过2.8m（超过第4距离等级）时，样方间的基质密度可以认为空间独立。

# 空间相关系数

## Mantel 相关系数（多元定量变量）

site	Spec01	Spec02	Spec03	Spec04	Spec05	Spec06	Sp
1	17	5	5	3	2	1	
2	2	7	16	0	6	0	
3	4	3	1	1	2	0	
4	23	7	10	2	2	0	
5	5	8	13	9	0	13	
6	19	7	5	9	3	2	
7	17	3	8	2	3	0	
8	5	4	8	2	1	2	
9	3	3	2	2	1	1	
10	22	4	5	3	0	0	
11	36	7	35	9	0	2	
12	28	2	12	13	0	0	
13	3	2	4	12	0	0	
14	41	5	12	0	2	0	
15	6	0	6	0	0	0	
16	7	2	3	2	0	0	
17	0	0	1	0	0	0	

site	x	y
1	0.2	0.1
2	1	0.1
3	1.2	0.3
4	1.4	0.5
5	2.4	0.7
6	1.8	0.9
7	0.05	1.1
8	2	1.3
9	2	1.5
10	1.2	1.7
11	2.4	1.9
12	0.2	2.1
13	0.4	2.1
14	2	2.3
15	2.2	2.3
16	0.05	2.7
17	0.2	2.7

$$r = \frac{1}{(n-1)} \sum_{i=1}^n \sum_{j=1}^n \frac{(x_{ij} - \bar{x})}{s_x} \cdot \frac{(y_{ij} - \bar{y})}{s_y}$$

## # 甲螨数据Mantel相关图

```
# *****
```

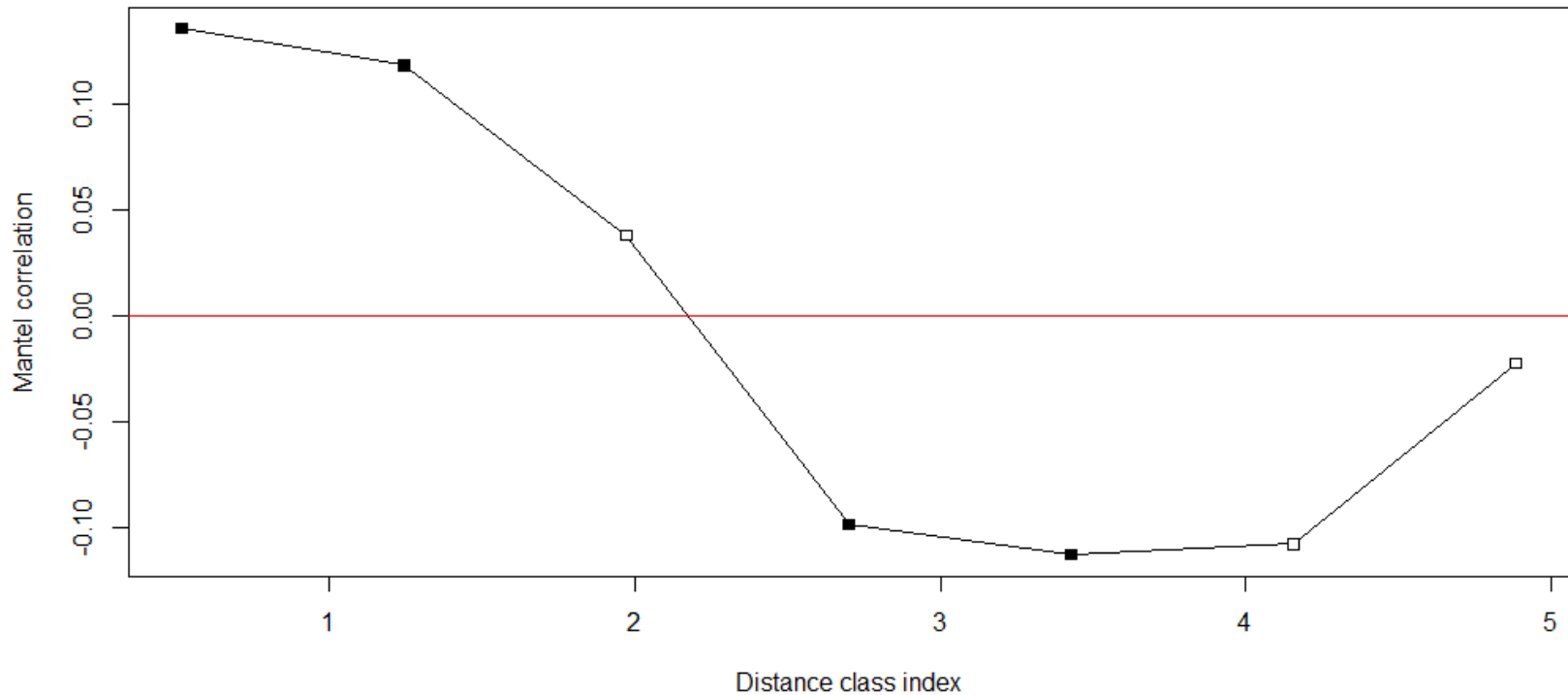
```
mite.h.det <- resid(lm(as.matrix(mite.h) ~ ., data=mite.xy))
```

```
mite.h.D1 <- dist(mite.h.det)
```

```
(mite.correlog <- mantel.correlog(mite.h.D1, XY=mite.xy, nperm=999))
```

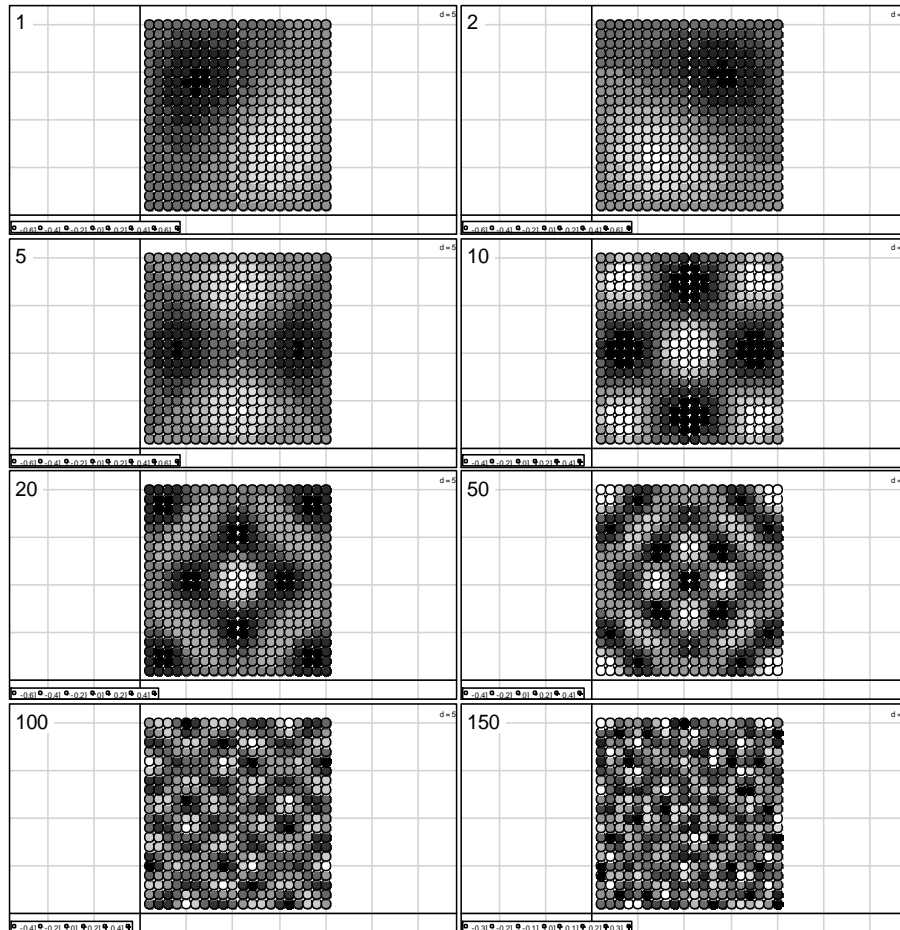
```
summary(mite.correlog)
```

```
plot(mite.correlog)
```



结果表明前两个距离等级有显著的正空间相关（即0.15-1.61m区间），第4至第6个等级（2.34-4.52m区间）显著负相关。查看环境变量可能会提供群落结构空间相关的生态解释。距离接近的样方显示相似的群落组成，原因是土壤条件相似。但是，如果距离超过2.7m（第4距离等级），则可能落入另外一个具有对照意义的土壤斑块，这也解释为什么超过2.7m之后甲螨群落产生差异

# 模拟空间结构（建模导向modelling-oriented）





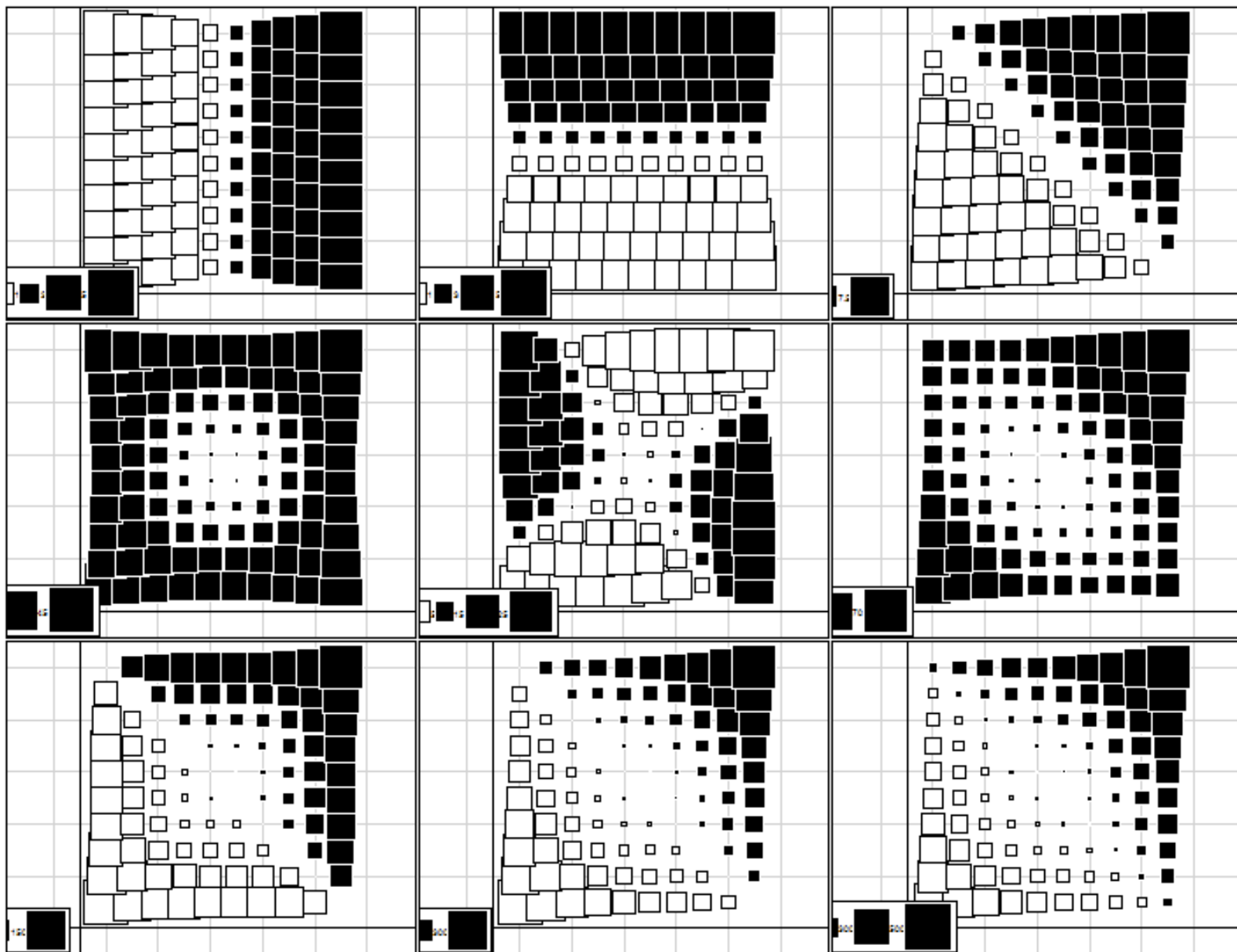
## 2.多元趋势面分析（模拟空间结构）

很多生态学数据取样于地理表面，因此，最原始的空间建模方式是建立响应变量与取样点的空间X-Y坐标之间的回归关系。

$$\hat{z} = f(X, y) = b_0 + b_1X + b_2Y \quad (7.6)$$

$$\hat{z} = f(X, y) = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 \quad (7.7)$$

$$\hat{z} = f(X, y) = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + b_6X^3 + b_7X^2Y + b_8XY^2 + b_9Y^3 \quad (7.8)$$



```
# 趋势面分析
# *****
# 从规则的正方形表面取样的简单模型
# 构建一个10×10的栅格样区
xygrid <- expand.grid(1:10, 1:10)
plot(xygrid)
xygrid.c <- scale(xygrid, scale=FALSE) # 中心化
X <- xygrid.c[,1]
Y <- xygrid.c[,2]
# 绘制X和Y的一阶、二阶和三阶函数
par(mfrow=c(3,3))
s.value(xygrid,(X))
s.value(xygrid,(Y))
s.value(xygrid,(X+Y))
s.value(xygrid,(X^2+Y^2))
s.value(xygrid,(X^2-X*Y-Y^2))
s.value(xygrid,(X+Y+X^2+X*Y+Y^2))
s.value(xygrid,(X^3+Y^3))
s.value(xygrid,(X^3+X^2*Y+X*Y^2+Y^3))
s.value(xygrid,(X+Y+X^2+X*Y+Y^2+X^3+X^2*Y+X*Y^2+Y^3))
```

```
# 甲螨数据趋势面分析
```

```
# 构建中心化甲螨数据X-Y坐标标准（非正交）的三阶多项式函数
```

```
mite.poly <- poly(as.matrix(mite.xy.c), degree=3, raw=TRUE)
```

```
colnames(mite.poly) <- c("X", "X2", "X3", "Y", "XY", "X2Y", "Y2", "XY2", "Y3")
```

```
# 基于9个多项式项的RDA
```

```
mite.trend.rda <- rda(mite.h ~ ., data=as.data.frame(mite.poly))
```

```
# 计算校正 $R^2$ 
```

```
(R2adj.poly <- RsquareAdj(mite.trend.rda)$adj.r.squared)
```

```
# 基于地理坐标正交的三阶项RDA
```

```
mite.poly.ortho <- poly(as.matrix(mite.xy), degree=3)
```

```
colnames(mite.poly.ortho) <- c("X", "X2", "X3", "Y", "XY", "X2Y", "Y2", "XY2", "Y3")
```

```
mite.trend.rda.ortho <- rda(mite.h~., data=as.data.frame(mite.poly.ortho))
```

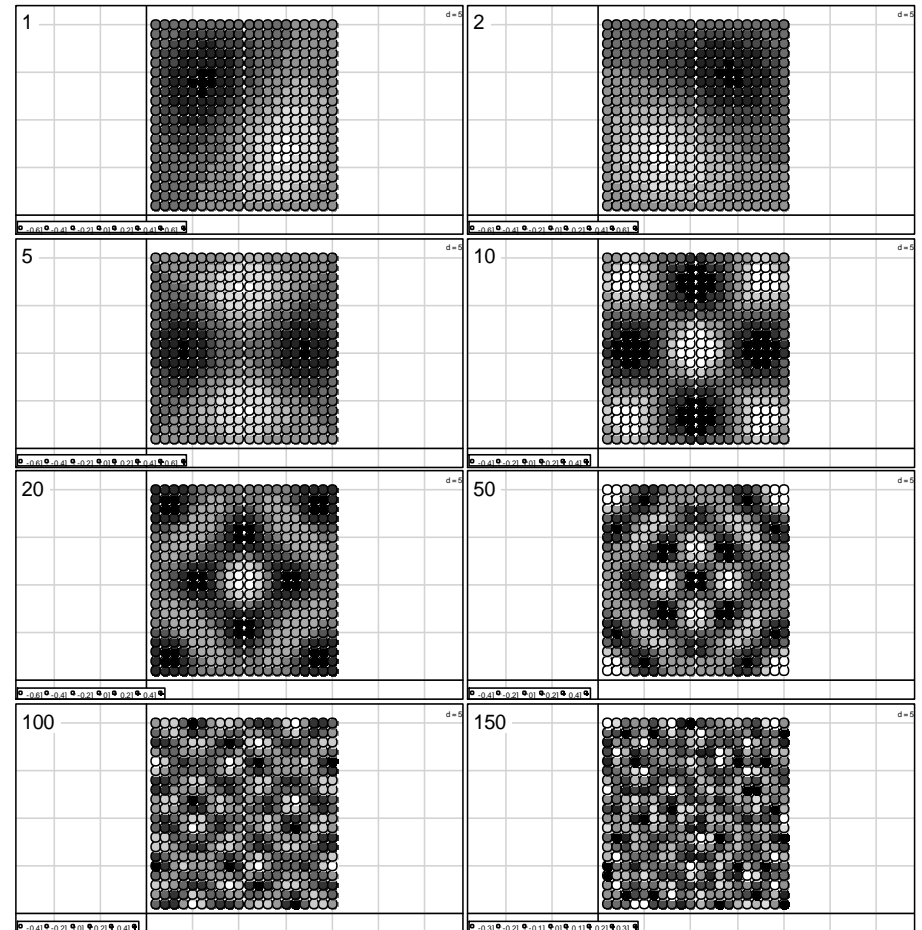
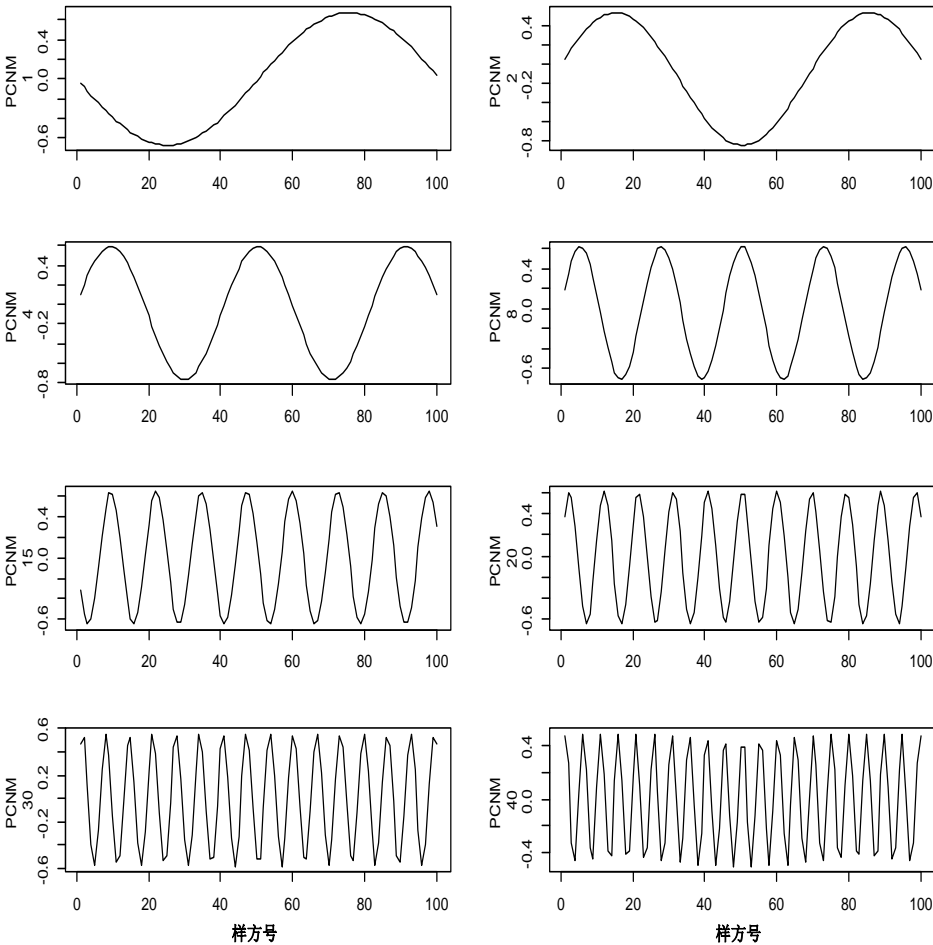
```
(R2adj.poly <- RsquareAdj(mite.trend.rda.ortho)$adj.r.squared)
```

```
anova.cca(mite.trend.rda, step=1000)
```

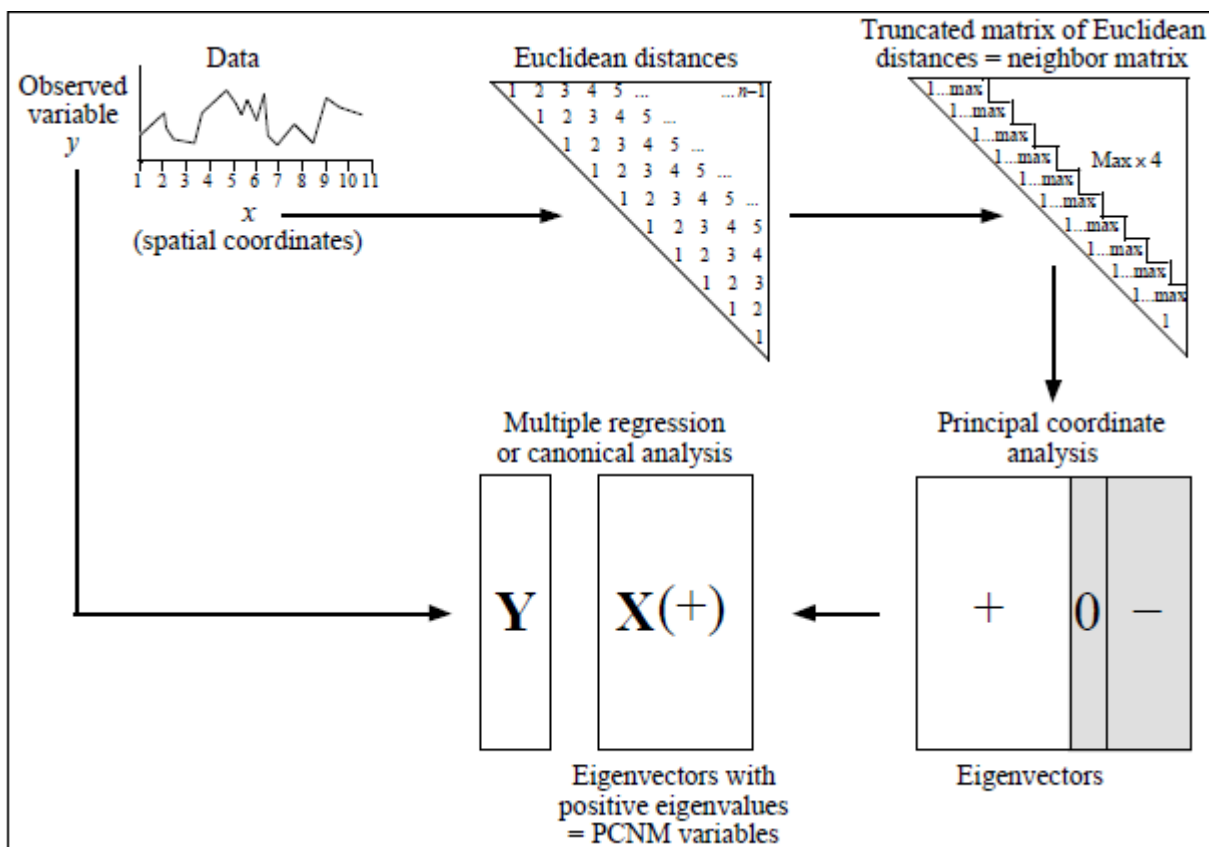
```
anova.cca(mite.trend.rda, step=1000, by="axis")
```

# PCNM (邻体矩阵主坐标分析)

## principal coordinates of neighbour matrices



# PCNM的计算步骤

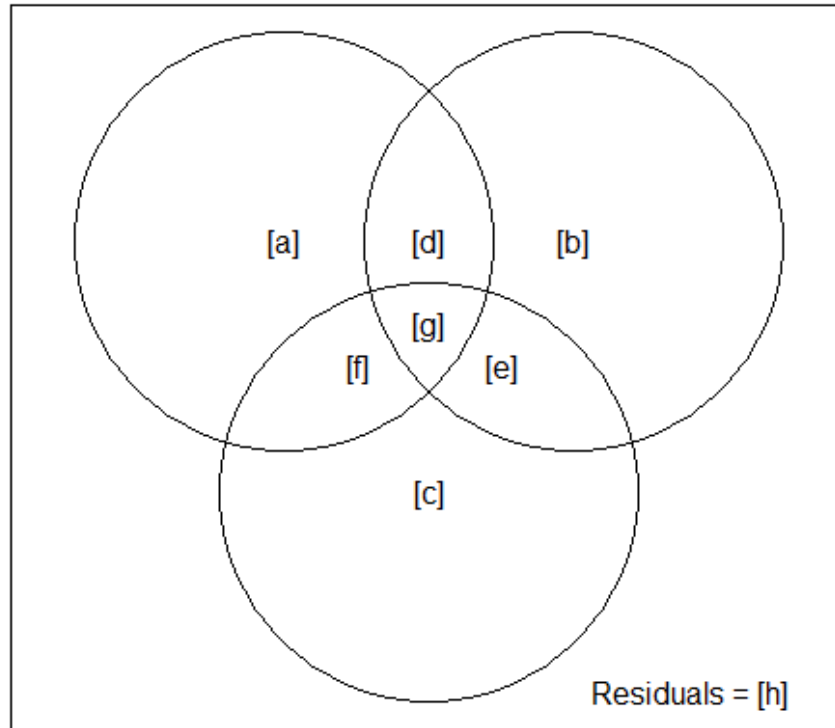


Borcard, D. and Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153: 51-68.

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

# 组合PCNM分析和变差分解

环境变量



X-Y坐标

PCNM变量

```
# 1b. ...或自动构建PCNM变量
# library(PCNM) # 如果还未加载PCNM程序包
xy.d1 <- dist(mite.xy)
mite.PCNM.auto <- PCNM(xy.d1)
summary(mite.PCNM.auto)
# PCNM变量Moran指数（由第一距离等级0到削减阈值）；也见PCNM（）函数
# 产生的图（此处无显示图）
# Moran指数的期望值（代表无空间相关）
mite.PCNM.auto$expected_Moran
mite.PCNM.auto$Moran_I
# 正空间相关的特征函数
(select <- which(mite.PCNM.auto$Moran_I$Positive == TRUE))
length(select) # I > E(I)条件下PCNM变量的数量
mite.PCNM.pos <- as.data.frame(mite.PCNM.auto$vectors)[,select]
```



```
# 甲螨-趋势-环境-PCNM变差分解

# *****

# 1.检验趋势，如果显著，对坐标进行前向选择

# -----

mite.XY.rda <- rda(mite.h, mite.xy)

anova.cca(mite.XY.rda)

(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))

XY.sign <- sort(mite.XY.fwd$order)

# 将显著的坐标变量赋予新的对象

XY.red <- mite.xy[,c(XY.sign)]
```

```
# 2. 环境变量检验和前向选择
```

```
# 将环境变量3-5重新编码成二元变量
```

```
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)
```

```
# 环境变量的前向选择
```

```
mite.env.rda <- rda(mite.h, mite.env2)
```

```
(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)
```

```
env.sign <- sort(mite.env.fwd$order)
```

```
env.red <- mite.env2[,c(env.sign)]
```

```
colnames(env.red)
```

### # 3. PCNM变量的前向选择

# 运行未去趋势甲螨数据的全模型PCNM分析

```
mite.undet.PCNM.rda <- rda(mite.h, mite.PCNM.pos)
```

```
anova.cca(mite.undet.PCNM.rda)
```

# 如果分析表明显著，计算校正 $R^2$ 和运行PCNM变量前向选择

```
(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))
```

# 根据R2a准则，如果保留12个PCNM变量，获得的校正R2已经稍大于全模

# 型的校正R2。但这个“稍微超过”也是可行，并不一定很严格。

```
(nb.sig.PCNM <- nrow(mite.undet.PCNM.fwd)) # 显著的PCNM变量的数量
```

# 按顺序排列显著的PCNM变量

```
(PCNM.sign <- sort(mite.undet.PCNM.fwd$order))
```

# 赋予所有显著PCNM变量一个新的对象

```
PCNM.red <- mite.PCNM.pos[,c(PCNM.sign)]
```

## # 5. 甲螨-环境-趋势-PCNM变差分解

```
(mite.varpart <- varpart(mite.h, env.red, XY.red, PCNM.red ))
```

```
par(mfrow=c(1,2))
```

```
showvarparts(3)
```

```
plot(mite.varpart, digits=2)
```

```
# 检验单独解释部分[a], [b], [c] 和 [d]
```

```
# *****
```

```
# [a]部分，环境变量单独解释部分
```

```
anova.cca(rda(mite.h, env.red, cbind(XY.red, PCNM.red )))
```

```
# [b]部分，趋势单独解释部分
```

```
anova.cca(rda(mite.h, XY.red, cbind(env.red, PCNM.red )))
```

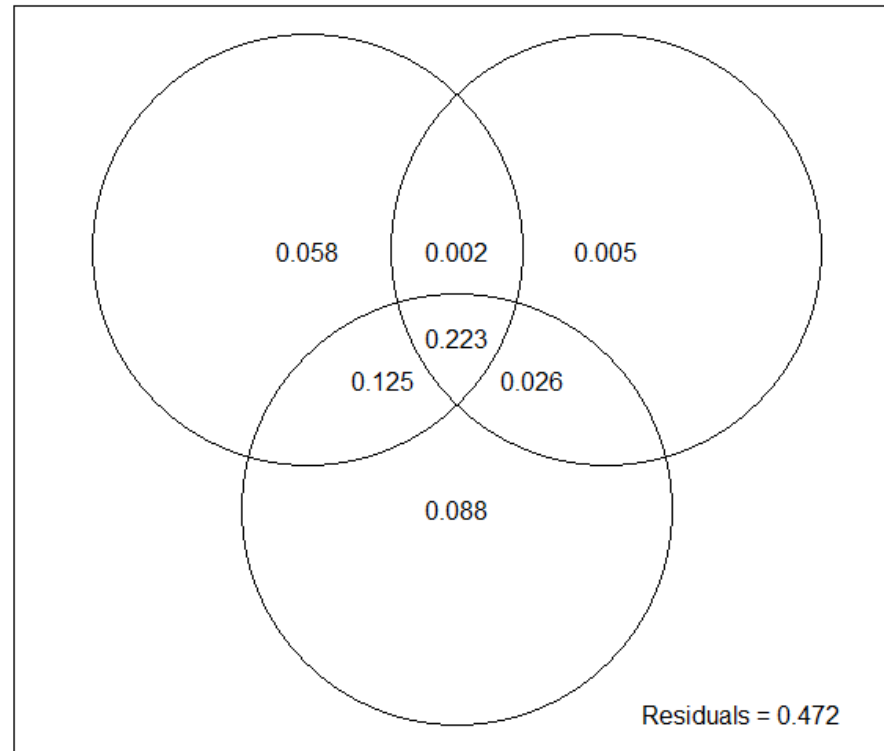
```
# [c]部分，宽尺度空间变量单独解释部分
```

```
anova.cca(rda(mite.h, PCNM.red , cbind(env.red, XY.red)))
```

```
#仅有环境变量和宽尺度空间变量单独解释部分显著。
```

## 组合PCNM分析和变差分解

环境变量



X-Y坐标

PCNM变量

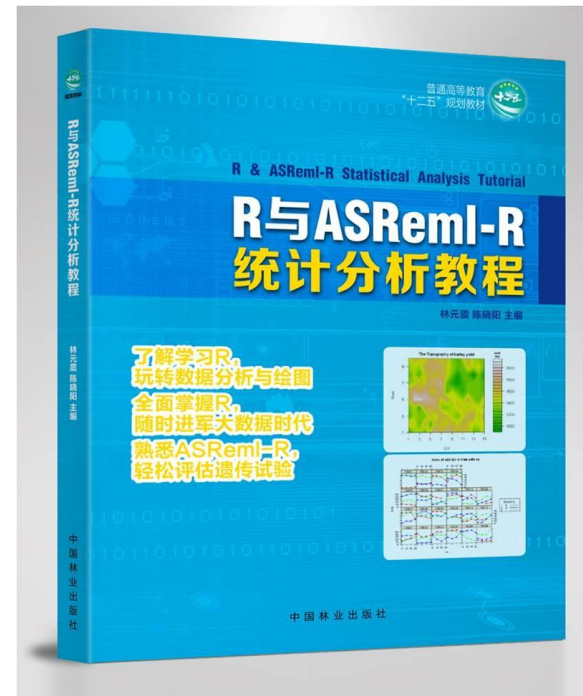
All=52.8%;环境变量（结果输出的X1矩阵）能够解释40.8%变差，但只有5.8%与空间变量无关（[a]部分）。[a]部分也代表与局部环境条件相关的物种-环境关系。[g]（解释12.5%变差）是环境变量和PCNM变量共同解释的部分，[i]是（23.0%）是环境变量、Y坐标和宽尺度PCNM变量共同解释的部分，这两部分比例较大，表明是典型的“诱导性空间变差（induced spatial variation）”，即环境变量空间变化引起响应变量产生类似的空间变化。

# 2014年暑期R语言与数据分析培训班通知

(8月23-27日北京林业大学)

( [moqinsmile@163.com](mailto:moqinsmile@163.com) 手机: 13121612257 )

<http://www.planta.cn/forum/viewtopic.php?t=36839>



# 致谢！



Daniel Borcard



François Gillet



Pierre Legendre

感谢高等教育出版社！