

基于 Vegan 软件包的生态学数据排序分析

赖江山 米湘成

(中国科学院植物研究所植被与环境变化国家重点实验室, 北京 100093)

摘要: 群落学数据一般是多维数据, 例如物种属性或环境因子的属性。多元统计分析是群落生态学常用的分析方法, 排序 (ordination) 是多元统计最常用的方法之一。CANOCO 是广泛使用的排序软件, 但缺点是商业软件价格不菲, 版本更新速度也很慢。近年来, R 语言以其灵活、开放、易于掌握、免费等诸多优点, 在生态学和生物多样性研究领域迅速赢得广大研究人员的青睐。R 语言中的外置软件包 “Vegan” 是专门用于群落生态学分析的工具。Vegan 能够提供所有基本的排序方法, 同时具有生成精美排序图的功能, 版本更新很快。我们认为 Vegan 包完全可以取代 CANOCO, 成为今后排序分析的首选统计工具。本文首先简述排序的原理和类型, 然后介绍 Vegan 的基本信息和下载安装过程, 最后以古田山 24 公顷样地内随机抽取 40 个 20m × 20m 的样方为例, 展示 Vegan 包内各种常用排序方法 (PCA, RDA, CA 和 CCA) 和排序图生成过程, 希望能为 R 的初学者尽快熟悉并利用 Vegan 包进行排序分析提供参考。

关键词: R 语言, 软件包, 群落分析, 多元统计, PCA, RDA, CA, CCA

Ordination Analysis of Ecological Data Using Vegan Package in R

Jiangshan Lai, Xiangcheng Mi

State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093

Abstract: The multidimensional data on community composition, properties of individual populations, or properties of environment are fundamental materials for vegetation science. They need to be analyzed with taking their multidimensionality into account. Multivariate statistical methods are the perfect approaches for statistical analysis of such data sets. Ordination is fundamental in multivariate statistical methods. The CANOCO is popular program for ordination analysis, however, it is commercial software with high price. R language becomes one of the most popular statistical software in the world, because of open source and free. The Vegan package in R is the statistical analysis tools for community data. It has most basic functions of ordination analysis. Most of its multivariate tools can be used for other data types as well. The

paper firstly briefly describes the basic information of Vegan package and its download and installation process, then display several ordination methods: PCA, RDA, CA, and CCA using Vegan, based a data set form 24 ha Gutianshan forest plot. we hope the paper can help R beginners to familiarize Vegan and use it for their data.

Kew words: R language, package, community analysis, Multivariate statistics, PCA, RDA, CA,CCA

植被生态学研究中最基础的工作是基于样方单元的群落调查。每个样方内通常包含很多物种的数量信息，原始数据常用矩阵来表示，**一般是一行代表一个样方，一列代表一个物种（也可以是其他非生物环境因子）**。根据物种的数量信息，可以通过数量分类方法（如聚类分析）将所调查的样方进行分组归类，可以分出不同的群落类型。分类是分析植被间断性的有效方法，但不能用于描述群落的连续分布。因此，排序方法最初是用于分析群落之间的连续分布关系（张金屯 2001），后来经过不断发展，不仅可以排列样方，也可以排列植物种及环境因素，用于研究群落之间、群落与其环境之间的复杂关系。只使用物种组成数据的排序称作间接排序（indirect ordination），同时使用物种和环境因子组成数据的排序叫做直接排序（direct ordination）。**现在排序不仅仅用于植物群落分析，也适用于大部分以样点为基础的观察数据或实验数据的分析。**

排序的一个重要目的是生成可视化的排序图，这就决定排序过程实际上是将样方或植物种尽可能排列在可视化的低维空间，也就是**让最前面几个排序轴尽可能包含大量的生态信息**。从数学上讲，排序可以理解为一个几何问题：假设要分析含有 P 个物种的 N 个样方之间相互关系，可以把 N 个样方作为 N 个点在 P 维空间进行排列，排列结果虽能准确地反映样方间的相互关系，但是如果 $P > 3$ 维（即物种数大于3），就无法可视化排列结果，只能凭空想象。以最常用的二维排序图为例，排序的过程就是在 P 维空间内找到一个能使 N 个样方在 P 维空间内的点都投影到上面的平面，使得在这个平面内投影后的点与点之间的相对位置能够最大程度反映原来样方间的相对位置。这个**投影过程就是排序运算过程**。好的排序方法是投影过程信息损失最少。

自上世纪五十年代开始,生态学家开始用排序的方法分析生态学数据,经过半个世纪的发展,现在已经创制出种类繁多排序技术(张金屯 2001)。在现代的排序方法中,最常用的有两类:一类是以基于线性模型的主分量分析(Principal components analysis, PCA)及其衍生出来的冗余分析(Redundancy analysis, RDA);另一类基于非线性模型的对对应分析(Correspondence analysis, CA)及其直接梯度分析版本“典范对应分析”(Canonical correspondence analysis, CCA)(TerBraak 1986)。在CA家族中有除趋势对应分析版本(Detrended correspondence analysis, DCA),目的是消除CA排序过程产生的“弓形效应”(Arch effect)(Hill & Gauch 1980)。

能做排序的软件有很多,其中使用最广泛的是荷兰生物统计学家TerBraak和捷克植物生态学教授Smilauer编写的CANOCO软件(TerBraak & Smilauer 2002)。CANOCO简单易学,一直以来大部分以排序为分析手段的论文都是引用这个软件。但CANOCO最大的缺点是已经商业化,价格昂贵,版本更新速度慢。近年来,R语言以其灵活、开放、免费等诸多优点,在生态学和生物多样性研究领域迅速赢得研究者的青睐(R Development Core Team 2010)。R语言里的外在软件包“Vegan”是专用于群落生态学分析的工具(Oksanen et al. 2010)。Vegan能够提供所有基本的排序方法,同时能生成精美的排序图,版本更新也很快。我们认为Vegan包完全可以取代CANOCO,成为今后排序分析的首选统计工具。本文首先简介Vegan的基本信息和下载安装过程,然后以古田山24公顷样地内随机抽取的40个20m×20m样方为例,展示Vegan包内各种常用排序方法(PCA, RDA, CA和CCA)和排序图生成过程,同时选择部分参数与CANOCO软件进行对比,希望能为熟悉CANOCO但不熟悉R的读者尽快熟悉并利用Vegan包进行排序分析提供参考。

1. Vegan 软件包简介及安装

Vegan是Vegetation analysis的缩写,是专门用于群落生态学数据分析的R语言外在软件包(Package),由芬兰Oulu大学生物系Oksanen等9位数量生态学者编写并贡献给R使用者无偿使用(Oksanen et al. 2010)。Vegan包提供各种群落生态学分析工具,包括常用的排序方法(PCA, CA, DCA, RDA, CCA 和 NMDS),同时也包括这些方法偏分析版本(Partial methods)。现在升级了的Vegan包几乎包罗了所有群落和植被分析中的一般应用方法,其中大部分多元分析工具也可以用于其他领域的数据分析。有关Vegan包的使用手册

(Reference manual) 可以从 <http://cran.r-project.org/web/packages/vegan/index.html> 下载。Oksanen 还编写了一个 Vegan 包案例分析指南 “Multivariate Analysis of Ecological Communities in R: vegan tutorial”，可以从 Oksanen 个人主页下载：<http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf>

Vegan是R统计的一个软件包，必须在R的环境下使用。安装R的源程序可以按照如下步骤先进行(以Windows版本的R为例)：在网址 <http://cran.r-project.org> 下找到R的最新版安装程序和源代码，选择任一镜像网站点击进入：点击Windows，再选择base，下载SetupR.exe，即R FOR WINDOWS的安装程序，约30多兆。双击SetupR.exe，按照提示一步步安装即可。安装完成后，程序会创建R程序组并在桌面上创建R主程序的快捷方式。通过快捷方式运行R，便可调出R的界面控制台。

Vegan软件包安装到R的源程序可以有两种方式：1.如果电脑已经连接到互联网，在R界面控制台中输入 `install.packages("vegan")`，选择任一镜像网点后，程序将自动下载并安装程序包；2.安装本地zip包：首先从<http://cran.r-project.org/web/packages/vegan/index.html> 下载zip软件包存储在本地磁盘。注意下载的zip包不必解压缩，从R界面上选择路径：Packages>install packages from local files，选择本地磁盘上存储的zip包文件就可以安装Vegan软件包。安装Vegan包之后，在控制台中输入软件包导入命令：`library(vegan)`，便可以使用vegan包内的函数。由于Vegan不属于R的标准包，每次重新打开R后，如果需要用到Vegan包，都必须重新运行一次`library(vegan)`，否则无法调用Vegan包中的函数。

2.案例分析

以古田山常绿阔叶林 24 公顷森林样地内随机抽取的 40 个 20m×20m 样方为例，展示各种常用排序方法（PCA,CA,RDA 和 CCA）在 R 环境下的分析过程。数据有两文件：`gtsdata.csv` 包含 40 个样方内 22 个物种的多度数据（一行代表一个样方，一列代表一个物种，物种名是拉丁名缩写代码）；`gtsenv.csv` 包含 40 个样方内 8 个环境因子的数据（4 个地形因子：海拔 `elevation`、坡度 `slope`、坡向 `aspect` 和凹凸度 `convexity`，4 个土壤因子：N、P、K 和 pH 值）。利用这些数据进行以下分析：1. 通过间接排序分析群落结构特征（物种之间、样方之间、物种与样方之间的关系）；2. 通过直接排序分析物种分布与环境因子之间的关系（8 种环境因子对物种分布的总解释量，以及地形因子与土壤因子单独的解释

量。)

案例分析的数据可以从赖江山的博客下载 (<http://blog.sciencenet.cn/?267448>)。将下载的数据存储在本地磁盘内,然后将 R 的界面路径改到数据存储的目录下 (File>change direction>选择数据所在地目录)。由于数据是 txt 文本格式的数据 (Excel 里面可以将数据直接存为 txt 格式), 可以用 read.table 函数将数据读入 R 控制台 (以下这些命令可以直接 copy 到控制台运行, #在 R 的运行里表示终止符, #之后的文字是对代码的注解说明):

```
>gtsdata=read.table("gtsdata.txt", header=T) #将物种矩阵数据读入并赋给命名为
“gtsdata”的对象(objects), header=T 表示数据第一行和第一列作为数据表头读
入, 否则系统自动添加表头
```

```
>gtsenv= read.table("gtsenv.txt", header=T) # 将环境因子矩阵数据读入并赋给命名
为 “gtsenv” 的对象
```

2.1 选择排序的模型:线性模型 or 单峰模型:

进行排序分析之前, 首先要判断是选择线性模型 (PCA 和 RDA) 还是单峰模型(CA 和 CCA)的排序方法。一般来说, 如果物种分布变化大, 选择单峰模型效果比较好, 反之, 线性模型也是不错。当然, 为了选择更合适的模型, 你尽可能将各种排序都尝试一下, 选择一个你最想要的分析结果和排序图。当然, 如果你不知道到底是用线性、还好单峰模型好, 有一个比较简单方法可以帮助你来选择。在 CANOCO 里面, 可以先通过 DCA 分析里面 “Lengths of gradient” 来判别式选择线性排序还是单峰排序 (Lepx & Smilauer 2003)。

同样, 在 R-vegan 里面, 可以用 DCA 分析来判别, R 里面 DCA 为函数 decorana():

```
>decorana(gtsdata)
Call:
decorana(veg = gtsdata)
Detrended correspondence analysis with 26 segments.
Rescaling of axes with 4 iterations.
      DCA1  DCA2  DCA3  DCA4
Eigenvalues  0.3939 0.2239 0.09555 0.06226
Decorana values 0.5025 0.1756 0.06712 0.03877
Axis lengths  3.2595 2.5130 1.21445 1.00854
```

这里的 “Axis lengths” 和 CANOCO 里面 DCA 分析结果 “Lengths of gradient” 是一样的。

如果 DCA 排序前 4 个轴中最大值超过 4，选择单峰模型排序更合适。如果是小于 3，则选择线性模型更好(Lepx & Smilauer 2003)。如果介于 3-4 之间，单峰模型和线性模型都可。在本案例中，Axis lengths 最大值为 3.2595，介于 3-4 之间，线性模型（PCA 和 RDA）和单峰模型（CA 和 CCA）都适用。

2.2 PCA 分析

在 R 里面，PCA 和 RDA 都是同一个函数 `rda()` 来实现，如果括号内只用物种矩阵 `rda(X)`，就表示 PCA 分析；如果同时有物种矩阵和环境矩阵即 `rda(X,Y)`，就表示是 RDA 分析：

```
> gts.pca=rda(gtsdata) # 将 PCA 的排序结果直接赋给命名为 “gts.pca”的对象
> gts.pca
Call: rda(X = gtsdata)

Inertia Rank
Total      352.1
Unconstrained 352.1 22
Inertia is variance
Eigenvalues for unconstrained axes:
  PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8
111.779 73.580 54.607 32.959 26.481 18.063 12.763 7.637
(Showned only 8 of all 22 unconstrained eigenvalues)
```

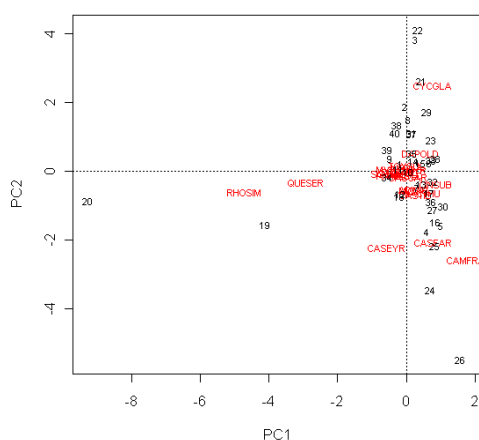
输出结果告诉我们总的特征根（Inertia）为 352.1，这个值也是物种矩阵中各个物种的方差和，可以理解为物种分布的总变化量，可以用 `apply` 函数检验：

```
> sum(apply(gtsdata,2,var)) #apply 是个对数据表或矩阵应用某个函数的函数
[1] 352.0917
```

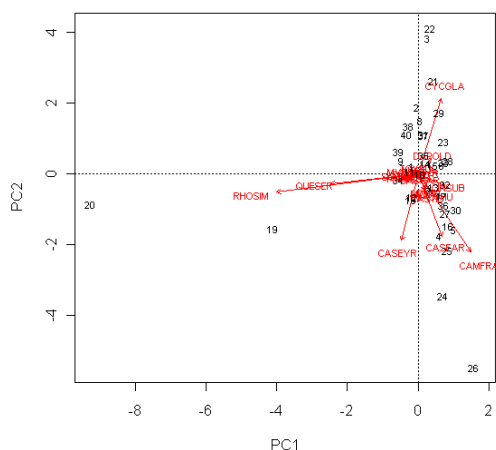
PCA 排序结果中 Eigenvalues for unconstrained axes 表示每个非约束排序轴所负荷的特征根的量，也可以表示每个轴所能解释方差变化的量。例如，对于第一轴而言， $111.779/352.1=31.7\%$ 便是第一轴对物种分布的解释量，其他轴的解释量以此类推，也可以用 `summary(gts.pca)` 直接获得更详细的排序结果分析，`summary(gts.pca)` 所得结果类似于 CANOCO 分析结果里的 log 和 solution 两个文本文件的内容。

排序的重要目的是排序图。标准的 PCA 排序图应该用符号来表示物种 (species) 和样点 (site)，可以直接用做图函数 `plot()` 生成排序图。但有些人更倾向于用箭头来表示物种，这时用函数 `biplot()` 来生成。

```
> plot(gts.pca)
```



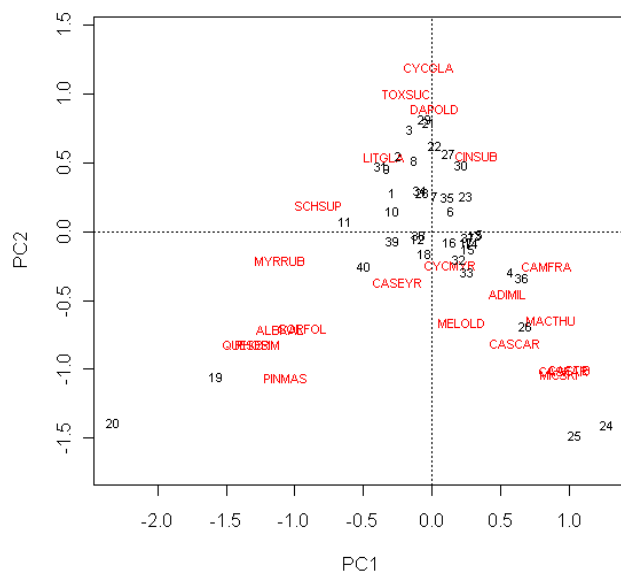
```
> biplot(gts.pca)
```



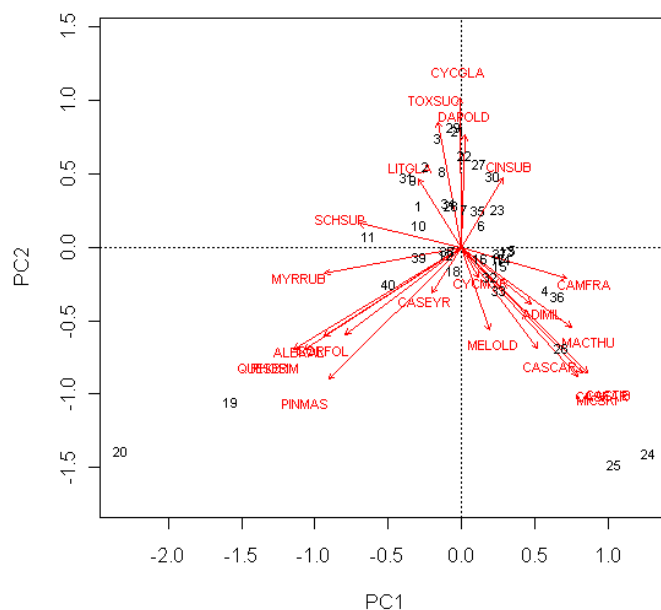
以上排序图可以看出除了少数几个分布在外围物种（比如 RHOSIM、QUESER、CYCGLA 等）能分清之外，很多物种重叠在一起，排序图效果不好。产生重叠的原因是物种分布差异大，某些分布不均匀的物种占据了大部分排序空间。此时可以对物种数据进行单位方差标准化（先中心化再标准化后均值为 0，方差为 1）。单位方差化可以通过参数 `scale` 来实现。重新做图，新的排序图比没有标准化之前效果好。

```
> gts.pca=rda(gtsdata,scale=T) # scale=T 表示对数据进行标准化再进行排序分析
```

>plot(gtspca,scaling=3) #参数 scaling 的值类似于 CANOCO 里面 scaling 的界面。如果排序图内关注物种之间的关系，可以选择 scaling=1；如果是关注样方之间的关系，scaling=2；如果是关注样方与物种之间的关系，可以选择 scaling=3。



> biplot(gtspca,scaling=3)



上面的排序图都是将样方与物种同时展示，可以通过 display 参数分别生成物种和样方排序图，可以用 choices 参数选择所要展示的轴，轴不一定是连续的，比如

choices=c(1,3) 表示第一、三轴，可以尝试以下命令，了解 display 和 choices 的用途：

```
> plot(gts.pca,display="sp", choices=c(1,3)) #sp 代表物种
> plot(gts.pca,display="si") #si 代表样方
> biplot(gts.pca,display="sp",choices=c(2,3))
> biplot(gts.pca,display="si")
```

2.3 CA 分析

CA 和 CCA 在 R 软件里也是同一个函数 cca() 来实现，如果括号内只用物种矩阵，就表示 CA 分析；如果同时有物种矩阵和环境矩阵，就表示是 CCA 分析：

```
> gts.ca=cca(gtsdata)
> gts.ca
Call: cca(X = gtsdata)

      Inertia Rank
Total      1.424
Unconstrained 1.424 21
Inertia is mean squared contingency coefficient
Eigenvalues for unconstrained axes:
  CA1  CA2  CA3  CA4  CA5  CA6  CA7  CA8
0.50253 0.26564 0.14023 0.10502 0.09127 0.05540 0.05063 0.04204
(Showned only 8 of all 21 unconstrained eigenvalues)
```

CA 中排序结果的 Total inertia 只有 1.424，比 PCA 的 Total inertia 小很多，这是因为 PCA 以欧氏距离(Euclidern distance)作为排序标准，而 CA 是卡方距离(chi-square distance)进行排序。Total inertia 值可以通过卡方检验验证：

```
> chisq.test(gtsdata/sum(gtsdata))

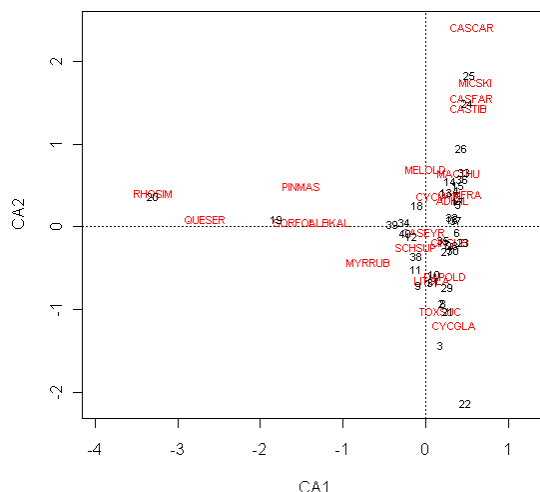
      Pearson's Chi-squared test

data: gtsdata/sum(gtsdata)
X-squared = 1.4243, df = 819, p-value = 1
```

关于每个轴的解释量，可以用与 PCA 相同的方式进行计算，也可以用函数

summary(gts.ca) 直接获得每个轴的解释量和物种与样方的排序坐标。同样，也可以用 plot() 函数生成排序图：

```
> plot(gts.ca, scaling=1)
```



对于 CA 的排序图，可以这样解读：如果一个物种靠近某个样方，表明该物种可能对样方的位置起很大的作用。如上图，20 号样方与短柄枹(QUESER)靠得近，表明短柄枹表征了 20 号样方的特征，19 号样方与 20 号样方距离近，物种组成结构特征也较近。只在少数样方出现的物种通常在排序空间的边缘，表明它们只是偶然发生，或它们只存在于稀有生境，如米楮 (CASCAR)。在排序空间中心的物种，可能在取样区域是该物种最优分布区，如甜楮 (CASERY)。

2.4 RDA 分析

PCA 和 CA 都属于间接分析，只能分析一个矩阵的数据。在分析物种分布与环境因子关系的时需要用直接排序，也就是约束排序 (Constrained ordination)。约束排序的主要类型有 RDA 和 CCA，分别是 PCA 和 CA 在排序过程加入环境因子进行线性回归最终得到的排序结果。所以，约束排序与非约束排序的区别不仅在于约束排序图加入环境因子，也在于非约束排序轴展示的总特征根是所有物种分布的变化量 (方差)。而在约束排序里，只展示能被环境因子所解释的物种分布变化量。从约束排序结果里可以看出，约束排序轴明显比非约束排序轴的解释量小。

用古田山样地数据来展示 RDA，同时使用物种矩阵和环境因子矩阵，用函数 rda() 来

实现。

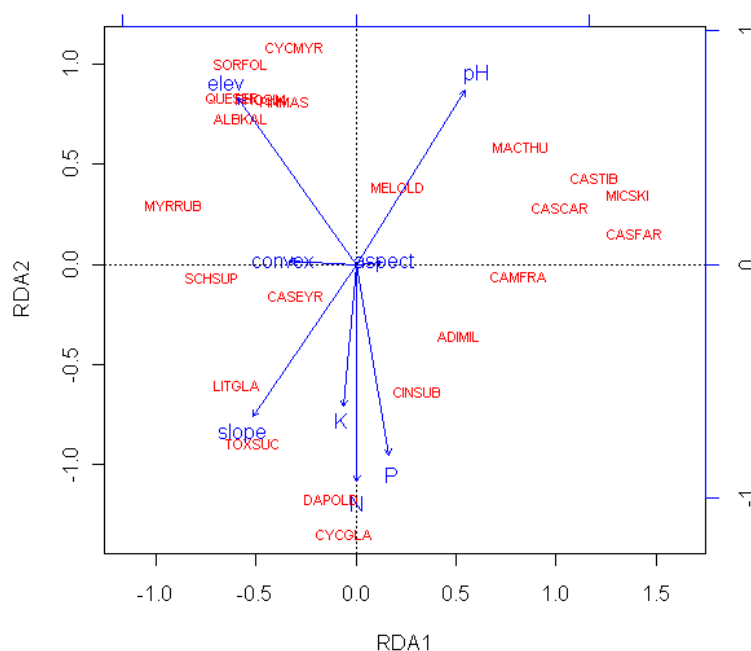
```
> gts.rda=rda(gtsdata,gtsenv)
>gts.rda=rda(gtsdata~elev+convex+slope+aspect+N+K+P+pH,data=gtsenv);#以上两
种表达式结果相同，第二种的优点是可以选择所需要的环境因子进入分析，
data=gtsenv 表示环境因子是数据框 gtsenv 里的列变量。
> gts.rda
Call: rda(X = gtsdata, Y = gtsenv)

          Inertia Rank
Total      352.1
Constrained 137.4   8
Unconstrained 214.7 22
Inertia is variance
Eigenvalues for constrained axes:
  RDA1  RDA2  RDA3  RDA4  RDA5  RDA6  RDA7  RDA8
56.3864 42.7769 17.8270 13.5066 2.5020 2.1217 1.6616 0.6203
Eigenvalues for unconstrained axes:
  PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8
72.287 54.891 26.618 17.959 12.730 9.918 5.659 5.349
(Showed only 8 of all 22 unconstrained eigenvalues)
```

RDA 的 Total Inertia 与 PCA 相同，但 Constrained Inertia 为 137.4，表明 8 个环境因子对物种分布的解释量为 $137.4/352.1=39.02\%$ 。不能解释的物种分布变量为 60.98%。rank 表示排序轴的数据，本例中有 8 个环境因子，就会有 8 个约束排序轴。Eigenvalues for constrained axes 表示各个轴承载的特征根，只显示 8 个轴表示 8 个环境因子。比较 PCA 和 RDA 中 Eigenvalues for unconstrained axes，RDA 中 Eigenvalues for constrained axes 与 Eigenvalues for unconstrained axes 之和等于 PCA 中 Eigenvalues for unconstrained axes 的特征根。RDA 更详细的分析，可以用 summary(gts.rda) 获得。RDA 的排序图，同样可以用 plot() 函数生成

```
> plot(gts.rda,display=c("sp","bp"),scaling=3)
```

display=c("sp","bp") 表示显示物种与环境因子。如果显示样方与环境因子，可以表示为 display=c("si","bp")，如果物种、样方和环境因子同时显示，可以设定 display=c("sp","si","bp")。



在 RDA 排序图内，环境因子一般用箭头表示，箭头连线的长度代表某个环境因子与群落分布和种类分布间相关程度的大小，连线越长，说有相关性越大，反之越小。箭头连线和排序轴的夹角代表某个环境因子与排序轴的相关性大小，夹角越小，相关性越高；反之越低。

约束排序有偏分析法 (Partial methods)，可以分别分析主环境变量与协变量对物种分布的影响。

在本例中，分析地形因子（海拔、坡度、坡向和凹凸度）和土壤因子（N、P、K 和 pH 值）分别对物种分布的解释量及两组环境因子的共同解释部分。Partial RDA 同样也使用 rda() 函数，函数表达式为 $\text{rda}(X,Y,Z)$ ， X 为物种矩阵， Y 为主环境变量， Z 为协变量 (Covariables)。如果分析排除土壤因子后地形因子对物种分布的解释量，则 Y 为地形因子矩阵， Z 为土壤因子的矩阵。

```
> gts.prda1=rda(gtsdata,gtsenv[,1:4], gtsenv[,5:8]) #gtsenv[,1:4]表示环境矩阵只取前
```

四列，即为 4 个地形因子，直接运行 `gtsenv[,1:4]` 看数据结构。

```
> gts.prda1
```

```
Call: rda(X = gtsdata, Y = gtsenv[, 1:4], Z = gtsenv[, 5:8])
```

Inertia Rank

Total 352.09

Conditional 95.03 4

Constrained 42.37 4

Unconstrained 214.69 22

Inertia is variance

Eigenvalues for constrained axes:

RDA1 RDA2 RDA3 RDA4

27.522 9.087 3.442 2.320

Eigenvalues for unconstrained axes:

PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8

72.287 54.891 26.618 17.959 12.730 9.918 5.659 5.349

(Showed only 8 of all 22 unconstrained eigenvalues)

Constrained Inertia 表示 `gtsenv[,1:4]` 4 个环境变量，即地形因子变量单独所能解释的特征根占总特征根为 $42.37/352.09=12.03\%$ 。如果将 Y, Z 的矩阵调换位置，便可以求得土壤因子单独的解释量：

```
>gts.prda2=rda(gtsdata,gtsenv[,5:8], gtsenv[,1:4])
```

```
> gts.prda2
```

```
Call: rda(X = gtsdata, Y = gtsenv[, 5:8], Z = gtsenv[, 1:4])
```

Inertia Rank

Total 352.09

Conditional 82.72 4

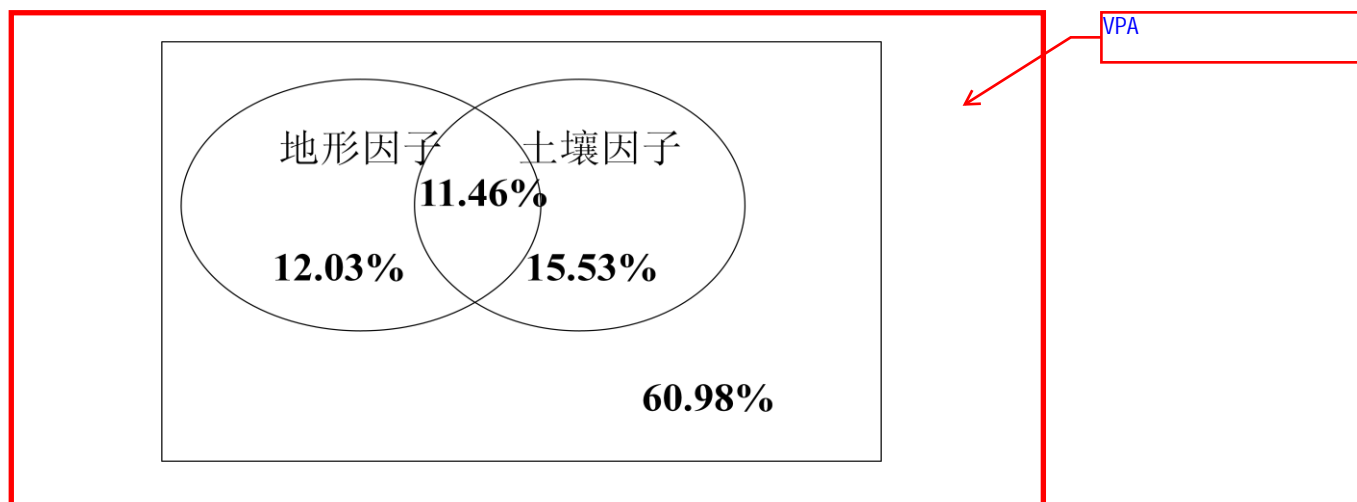
Constrained 54.69 4

Unconstrained 214.69 22

Inertia is variance

土壤因子所能解释的变量为 $54.69/352.09=15.53\%$ 。之前的 RDA 分析可以得知，8 个环境

因子总解释量为 39.02%，所以两组环境变量的共同解释量为 39.02%-15.53%-12.03%=11.46%。用维恩图三部分的解释量和不能的解释量表示如下：



2.5 CCA 分析

在 R 里面，CCA 也是通过 `cca()` 函数来实现，算法和解读与 RDA 类似。

```
> gts.cca=cca(gtsdata,gtsenv)
```

```
> gts.cca
```

```
Call: cca(X = gtsdata, Y = gtsenv)
```

Inertia Rank

```
Total      1.4243
```

```
Constrained 0.6928  8
```

```
Unconstrained 0.7315 21
```

Inertia is mean squared contingency coefficient

Eigenvalues for constrained axes:

```
CCA1  CCA2  CCA3  CCA4  CCA5  CCA6  CCA7  CCA8
0.338081 0.209369 0.066810 0.036824 0.013992 0.011367 0.009057 0.007305
```

Eigenvalues for unconstrained axes:

```
CA1  CA2  CA3  CA4  CA5  CA6  CA7  CA8
0.24676 0.11172 0.06530 0.05198 0.04321 0.03871 0.03436 0.02685
```

(Showed only 8 of all 21 unconstrained eigenvalues)

在 CCA 分析结果中，8 个环境因子的解释量为 $0.6928/1.4243=48.64\%$ ，明显比 RDA 的 39.02% 高。一般来说单峰模型拟合效果比线性模型好，也反映了物种与环境因子之间的复杂关系。同样，CCA 也有偏分析算法，算法和分析与 pRDA 类似。

2.6 检验环境因子相关显著性：

在 CANOCO 软件中，检验环境因子与物种分布相关的显著性，可用蒙特卡罗置换检验实现 (Monte Carlo permutation test)。在 R 软件中，可用函数 `permutest()` 进行实现。以 CCA 的结果 `gts.cca` 进行蒙特卡罗置换检验，分析 8 个环境因子对物种分布的解释量是否具有显著性：

```
> permutest(gts.cca,permu=999) # permu=999 是表示置换循环的次数
Permutation test for cca
Call: cca(X = gtsdata, Y = gtsenv)
Permutation test for all constrained eigenvalues
Pseudo-F:      3.669857
Significance:   0.001
Based on 999 permutations under reduced model.
```

Significance 为 0.001，说明排序的结果可以接受环境因子对物种分布的解释量，但只是 8 个环境因子综合分析的结果，并不代表每个环境因子都对物种分布具有显著影响。如果需要检验每个环境因子的显著性，可以用函数 `envfit()`。`envfit()` 函数类似 CANOCO 中环境变量的预选 (Forward Selection Environmental Variables) 界面。`envfit()` 除了用到 `cca` 的排序结果外，还用到所检验的环境矩阵参数，`permu` 表示置换的模拟次数，一般选择 999。

```
> ef=envfit(gts.cca,gtsenv,permu=999)
> ef
***VECTORS
CCA1    CCA2    r2    Pr(>r)
elev -0.9999873 -0.0050479 0.5005 0.001 ***
convex -0.7746504 0.6323898 0.0844 0.342
```



```

slope 0.1526685 0.9882775 0.5296 0.001 ***
aspect -0.1364818 -0.9906426 0.0169 0.818
N      0.6878942 0.7258110 0.6342 0.001 ***
P      0.8101502 0.5862224 0.5856 0.001 ***
K      0.6004984 0.7996260 0.3364 0.002 **
pH     -0.1929061 -0.9812172 0.6313 0.001 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
P values based on 999 permutations.

```

这里 CCA1 和 CCA2 两列所对应的值是环境因子箭头与排序轴夹角的余弦值，表示环境因子与排序轴的相关性。 r^2 表示环境因子对物种分布的决定系数， r^2 越小，表示该环境因子对物种分布影响越小。Pr 表示相关性的显著性检验。从 Pr 的值看，表明凹凸度 convexity 和坡向 aspect 对物种分布影响不显著。如果要简化模型，可以将这两个因子去掉。从 CCA 的排序图看，凹凸度 convexity 和坡向 aspect 的箭头长度也最短。实际上，envfit 函数所得的结果与 CCA 排序图是一一对应的，检验的结果是将排序图量化。

由于篇幅、能力等客观原因限制，在此只介绍了最基本的排序方法在R软件中的使用。对于熟悉R的读者，这些内容可能过于简单，但是对于R的初学者却有帮助。Vegan 包功能非常强大，几乎包罗了所有群落和植被分析中的一般方法。对Vegan包感兴趣的读者可以进一步参考Vegan使用手册和Oksanen 编写的Vegan使用指南（Multivariate Analysis of Ecological Communities in R: vegan tutorial）。此外，网站 <http://ecology.msu.montana.edu/labds/R/labs/> 有专人维护和培训Vegan包，也是初学者很好的学习材料。

参考文献

- Hill, M.O. & Gauch, H.G. (1980) Detrended Correspondence-Analysis - an Improved Ordination Technique. *Vegetatio*, 42, 47-58.
- Lepx, J. & Smilauer, P. (2003) *Multivariate Analysis of Ecological Data Using CANOCO*. Cambridge University Press, Cambridge, UK.

- Oksanen, J., F. Blanchet, R. Kindt, P. Legendre, and other five authors (2010) *vegan: Community Ecology Package*. R package version 1.17-3. <http://vegan.r-forge.r-project.org/>
- R Development Core Team. (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.*
- TerBraak, C.J.F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67, 1137-1179.
- TerBraak, C.J.F. & Smilauer, P. (2002) *CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination. Version 4.5*. Microcomputer Power, Ithaca, New York, USA.

张金屯 (2004) *数量生态学* 北京: 科学出版社