

4. CANOCO 4.5 软件的使用

Using the Canoco for Windows 4.5 package

4.1 概述 (Overview of the package)

Window 版 CANOCO 4.5 软件由几个相关程序组合成。在这一节我们将简单介绍一下各个程序的功能。然后再接着介绍这些程序的具体使用方法。这一章不能代替 CANOCO 软件的自带说明书，但对于初学者使用 CANOCO 还是很有帮助的。

CANOCO for Window 4.5

这个程序是 CANOCO 的核心。通过使用它，你可以指定要分析的数据和排序模型，选择分析的类型。你也可以选定一部分的被解释的变量和环境因子进入分析，也可以任意改变样方或物种的权重。所以这些选择都被集中在 CANOCO 程序的对话框中。

CANOCO for Window 带有比较全面的排序方法。核心部分是基于线性模型的 PCA 和 RDA，基于单峰的模型的 CA, DCA 和 CCA。在这些基本排序方法基础上，CANOCO 也可以做 CVA (典范变量分析), PCoA (主坐标分析)。但 NMDS 并没有被包括，因为的确用得很少。

WCanoImp

这个程序的功能在 1.7 节已经讨论。但这个程序用法受到了 window 的剪贴板和电子表格文件的限制。比如在 Excel2003 以前的版本，列数仅有 256 列，这就意味着样方数或物种数不能两个同时超过 256 个，否则需要分割。当然行数宽松点，但不能超过 65536 行。如果你的数据超过这个限制，你可以将你的数据分割为几个部分，经过 WCanoImp 转化后，再用 CanoMerge 程序拼接起来。(Excel2007 的行列已经没有限制，这个问题已经不复存在)

CanoMerge

这个程序有如下三个用途，

- 1) 首要的任务是连接两个或多个包含同样样方但不同变量的数据文件。这个功能对于行数或列数两者均超过 255 个变量很有用，能够拼接起来。如果你的数据列数超过了 255 列，就可以分割为多个文件，注意，最好每个文件行数一样，而且行数最好是代表样方。分别将这些分割开的表格用 1.7 节介绍的方法，转化一个个单独文件，然后利用 CanoMerge (如图 4-1) 将几个文件合成一个新的文件。“Add file”是添加文件的按钮，把需要合成的文件加进来，然后按 Merge 按钮，给新的文件名字，就可以了。
- 2) 可以用来将您的数据文件输出带制表符分隔的 ASCII (文本) 文件。这个功能是非常有用的，特别是如果您的数据因为太大而不适合存在 Excel 表格里。而制表符格式文件可以被常有统计软件包直接应用。你直接将需要转为文本文件的数据文件用 add files 按钮选到对话框，然后选中对话框最底下的选项“Write merged table in TAB-separated values format” (图 4-1)，按下“Merge”按钮，即可生成文本文件。
- 3) 最后一个功能是可以过滤掉一些低频率的物种。在“Exclude variables with less than X positive values”中的框中填上一个数字 (图 4-1)，这个数字代表如果一个物种所存在的样方少于 X 个，将被剔除出去。这个功能是相当有用的，因为在 Canoco 程序里面，尽管你可以任意剔除任一变量，也可以用“downweighting of rare species”来减少稀少种的权重，但都是很盲目的，无量化的标准。而 CanoMerge 程序中这个选项可以让你有目标地剔除一些低频率的物种。

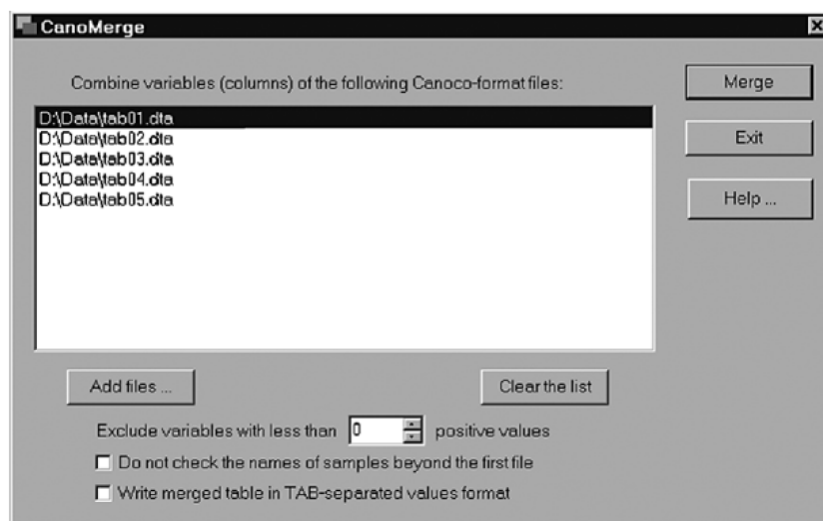


Figure 4-1. CanoMerge program user interface.

CanoDraw for Windows

CanoDraw for Windows 是用来做排图的程序。CanoDraw 的作图基础数据主要来自 Canoco 分析的结果，包括 .con 文件和 .sol 文件。CanoDraw 的输出文件是 .cdw 文件，这个文件独立于 .con 文件和 .sol 文件。CanoDraw 不仅可以生成排序图，也可以生成很多种等值线图 and 回归模型图。详见第 10 章。

4.2 CANOCO 典型分析流程图 (Typical flow-chart of data analysis with Canoco for Windows)

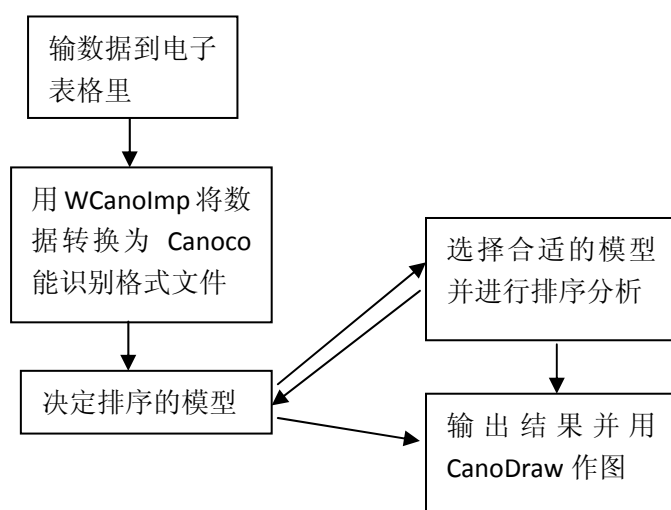


图 4-2 用 CANOCO 作排序分析的简单流程图。

图 4-2 展示的是用 CANOCO 分析多元数据一般的流程。首先你必须打开你的电子表格的数据，然后用 WCanolmp 程序将数据转化为 CACOCO 能兼容的文件。在 Canoco for

Windows 的窗口中，你可以重新创建一个新的“new project”，也可以打开原来已经存储过的分析文件（打开文件后，file>save as, 即可另存一个新的文件。另存的文件自然是保持了原来文件所有的背景选项设置，必要的时候，我们仅仅修改一些选择即可，这样更方便）。

在图 4-2 中关于排序模型的选择，以及模型对应的数据标准化，物种数据和环境因子的取舍、组合等等，可以参考后面将要讲的内容及每个窗口的“help”的命令。

当所有选择完成后，按下“analyze”按钮后，程序即可运行排序分析，分析完后，界面上会保持两个窗口，一个窗口为“Project View”，这个窗口为 project 的一些属性特征（比如，排序的类型、数据的维数、数据文件的名称等等），还有一些相关按钮。另外一个窗口是“Log View”，记录的是分析过程与分析的结果。但仅仅是一部分的统计结果（四个轴的统计总结、轴与环境因子的相关性、蒙特卡罗检验结果）被存储在这个窗口中，其他一些分析结果，比如物种、环境因子及样方的得分（坐标）均存在另外一个与.sol 作为后缀的文件中，此文件一般在当前目录下，以记事本打开即可。

分析完后，直接在“Project View”的窗口选择 CanoDraw 的按钮，便可直接将分析的结果进行做图，无需到桌面或是开始任务栏去打开 CanoDraw 程序。至于如何使用 CanoDraw，可以参考 4.7 节和第 10 章的内容。

4.3 决定排序的模型：单峰还是线性（Deciding on the ordination method: unimodal or linear?）

这一节提供简单易用的“菜谱”用于判断该选择线性模型还是单峰模型的排序方法，但需要注意的该“菜谱”是难免过于简化，因此它不应该被盲目遵循。

在 Canoco for Windows 的分析过程，有一个关于模型的选择窗口，你必须对排序的模型进行选择。为了选择更合适的模型，你尽可能将各种模型都尝试一下，选择一个你最想要的分析结果及排序图。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。可以先将物种数据进行 DCA 分析，其他的窗口的选项均可以选择默认程序。分析完成后，在“Log View”内，各轴特征根分析那部分，可以找到“Lengths of gradient”这样一栏信息，这个“Lengths of gradient”是展示每个轴的梯度长度。如果这四个轴中梯度最长（最大值）超过 4，选择单峰模型排序（CA、CCA、DCA）更合适。如果是小于 3，选择线性模型（PCA、RDA）比较合理。如果介于 3-4 之间，单峰模型和线性模型都是合适的。当然，这个选择并不是绝对的，最长梯度小于 3，依然可以用单峰模型，因为线性模型可以看作单峰模型一种特例。但是梯度大于 4，最好还是不用线性模型，线性模型可能误差比较大，会丢失比较多的信息。

当你选择单峰还是线形模型的时候，你必须预先认识到，不是所有的数据类型都是适合用单峰模型，比如物种数据是不同量纲的数据（单位不一致）就不适合做单峰模型的排序。还有，如果有个样方的空的（即值都为 0），这样的数据进行单峰排序分析，系统会自动告诉你有空的样方，不能进行分析。如果非要做单峰模型的排序分析，首先要把这些空样方去掉。

4.4 PCA 或 RDA 排序：中心化和标准化（PCA or RDA ordination: centring and standardizing）

CANOCO 里面的线性模型（PCA 和 RDA）过程有个物种中心化和标准化的选择(图

4.3), 这个选择可以很大程度上影响分析的结果。实际上, 你必须根据你的数据情况和所回答的问题来选择物种数据中心化和标准化(将在 14.3 节的案例讨论)。在这里, 我们仅仅简要介绍一下如何正确选择中心化和标准化。

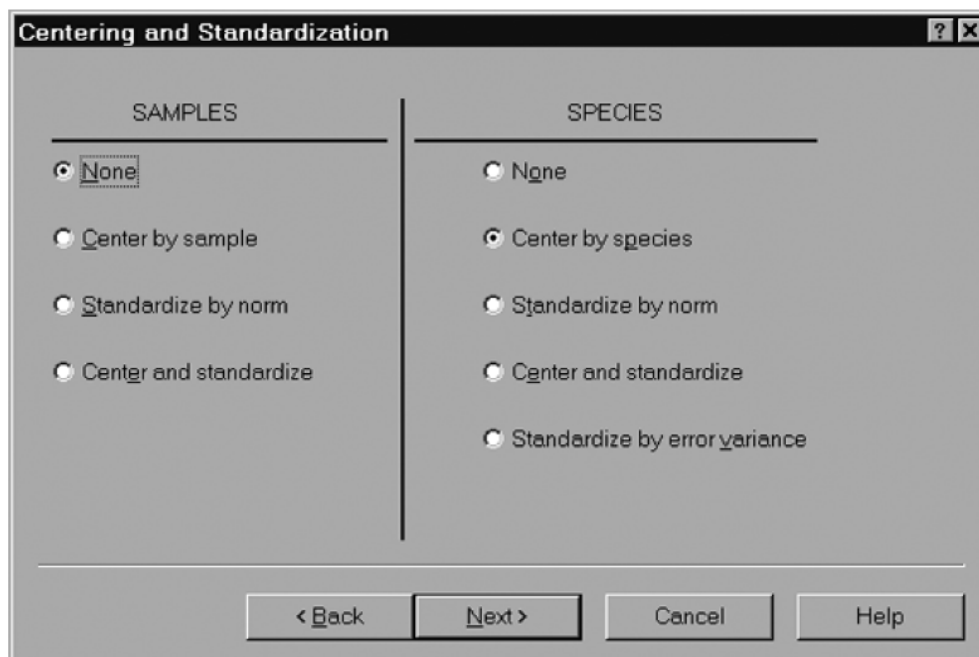


Figure 4-3. Centring and standardization options in the Project setup wizard.

对话框展示的线性排序 (PCA 和 RDA) 分析前物种数据的处理选择。样方的中心化 (图 4-3 左半部分的对话框) 是让每行 (每个样方) 的平均值为零。同样, 物种的中心化 (图 4-3 右半部分的对话框), 也是让每列 (每个物种) 的和为 0。对于 RDA 及 Partial RDA, 物种的中心化是必须的。

物种 (样方) 的标准化结果是让每列 (行) 范数(norm)等于 1。范数是列 (行) 内各值的平方和的平方根(The norm is the square root of sum of squares of the column (row) values)。如果同时使用中心化和标准化, 中心化必须先进行。物种数据中心化和标准化后, 各列变量平均值为 0, 方差为 1, 这样的话, PCA 的分析是基于物种的相关矩阵。如果没有进行范数标准化, PCA 的分析基于方差-协方差矩阵 (variance-covariance matrix)。需要注意的是, 在物种的平均频率与平均数量差异比较大的时候, 如果进行标准化, 等于的扩大了稀有种的权重。

也要注意的, 如果物种数据的量纲不同的属性数据, 比如不同的物理或化学参数, 物种变量必须标准化 (但样方数据则不必)。

如果是带有环境变量的线性排序 (比如 RDA), 你可以选择误差方差标准化 (standardization by error variance)。在这种情况下, CANOCO 会单独计算每个物种的多少方差可以被环境变量 (协变量) 解释。然后误差方差的倒数被作为物种的权重。这样, 如果一个物种变化被环境因子解释越多, 该物种在最终分析中的权重就越大。

4.5 DCA 排序: 除趋势对应分析 (DCA ordination: detrending)

对应分析(CA)第二轴开始的除趋势的作用在于消除弓形效应(arch effect)的影响, 图 4-4 展示的对对应分析(CA)头两轴的样方排序图。

对应分析中, 样方在第二轴的位置很大程度上依赖于其在第一轴的位置。这个效应源于

方法的局限性。就是 CA 的第二排序轴在许多情况下是第一轴的二次变形，即所谓的“弓形效应”（Arch effect）或者“马蹄形效应”（Horse—shoe effect）。样方在第二轴上的坐标与第一轴的坐标是二次曲线关系，这是正交化的必然结果（Gauch 1982）。

区间除趋势(Detrending by segments)，就是把第一轴分成一系列区间，在每个区间将平均数定为零而对第二轴的坐标值进行调整，从而克服了弓形效应，提高了排序精度。图 4-5 展示的是 CANOCO 里面除趋势的对话框。

对于无环境因子的对应分析（CA），我们推荐使用区间除趋势方法（Detrending by segments）。在一些情况下，多项式的除趋势方法（detrending by polynomials）也是可以推荐的选择。你可以从 CANOCO 使用手册里面得知在什么情况下选择多项式的除趋势方法效果比较好。（Minchin(1987)发现有时区间除趋势方法在有潜在的环境因子没有加入分析的时候，而会把一些变量弄平，使其无法展示出来。）。因此提出用多项式的除趋势方法效果比较。Polynomials 的次数选择（阶数）选择一般可以这样选择：环境因子小于 10 个，用 2nd Polynomials；小于 13 用 3 rd Polynomials；大于 13 用 4th Polynomials (来自 CANOCO 使用手册)。

对于约束排序，除趋势一般是不必要的。比如 CCA 的弓形效应，一般如果出现一些冗余环境变量存在的现象，表示那里有两个或是两个以上的环境变量之间有明显的自相关（正相关或是负相关）。如果在这样一组自相关内的环境变量中只保留一个，弓形效应就会消失了，因此，DCCA 实际上用途并不大。环境变量的选择（剔除）可以通过 CANOCO 里面的环境因子的预选功能来确定(the forward selection of the environmental variables in Canoco for Windows)。

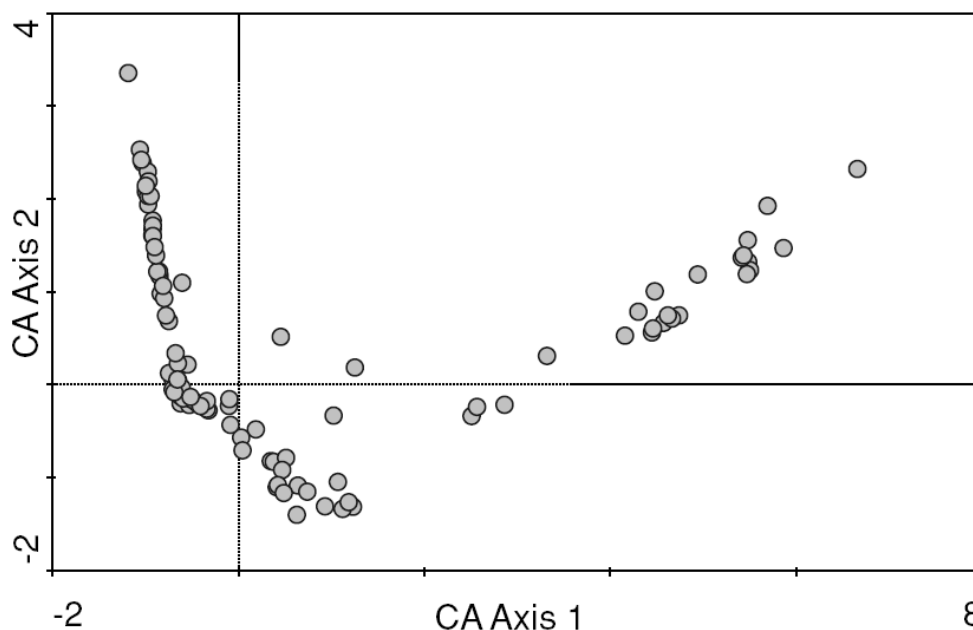


Figure 4-4. Scatter of samples from a correspondence analysis, featuring the so-called arch effect. You can see the ‘arch’ by rotating your book by 180 degrees.

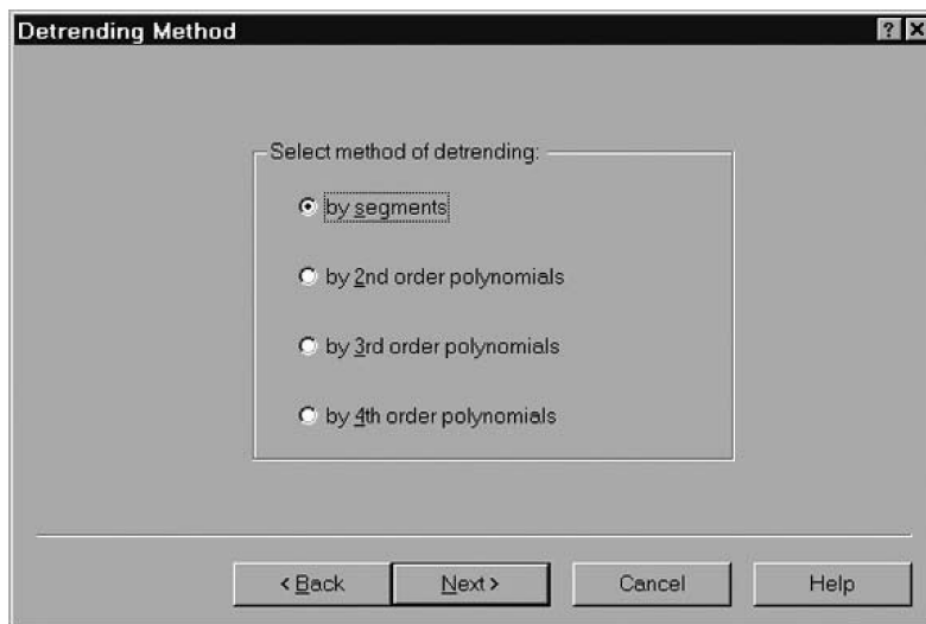


Figure 4-5. Detrending method selection in the Project setup wizard.

4.6 排序得分（坐标）的尺度比例（Scaling of ordination scores）

排序最重要的结果是排序图。通过排序图，理论上讲，不仅可以重建（有一定水平的误差）原始数据结构（物种数据），也可以重建样方之间相似（相异）矩阵和物种之间的相关矩阵。但没有人会想从排序图里面恢复真正的原始被测数据，因为原始数据已经有了。但原始数据对于排序图内容的解释和提炼出感兴趣的假设是很有用的。对于样方的相似度、物种或环境因子的相关性等的解释的准确性，部分取决于在排序轴坐标的相对尺度比例（relative scaling）。Object(样方，物种和环境因子)在排序空间的绝对值没有任何意义。在解释排序图的，我们使用相对距离，相对方向和投影点的相对次序。

这个可选的尺度比例让排序图展示不同相对比例，以便更好解释你想要探讨的问题。但是这个选择并不影响整个分析的结果。这里“Focus scaling on”选择在线性和单峰方法中，多少有点相似（图 4-6）。首先，这个选择必须依据分析者在排序图是要关注样方还是物种，如果是关注样方，就选择 inter-sample distances, 如果是关注物种，就选择 inter-species correlations (distances), 当然，如果两个都关心，选择 Symmetric。在在线性模型里面，Species scores 你必须决定箭头的长度是否要反映出物种的多度不同（多度大的物种箭头长），或是物种的多度转化成可以比较的尺度（所能解释的百分比）。如果你选择“divided by the standard deviation”，物种的箭头长度则表示在排序空间内所能展示该物种的变化量的比例（相对值）。如果选择“Do not post-transform”，箭头的长度表示物种变化量的在排序空间内的变化量（绝对值）。在单峰模型中，如果有很长的物种梯度（beta 多样性很高），Scaling types 选择“是 Hill scaling 更好。一般 biplot scaling 就可以的。biplot scaling 让排序图提供一个更量化的解释。（简单讲 biplot scaling 和 Hill scaling 区别，如果有物种比较集中，梯度比较短，可以直接用 biplot 图，如果有些物种有偏离，可以用 Hill scaling）

当然，如果是分离或分散的样方和物种，你选择关注样方或是关注物种，结果都是一样的。因为 CANODRAW 已经自动帮你调节好了。注意，如果好几种实体放在一起，这种自动调节，不能实现。

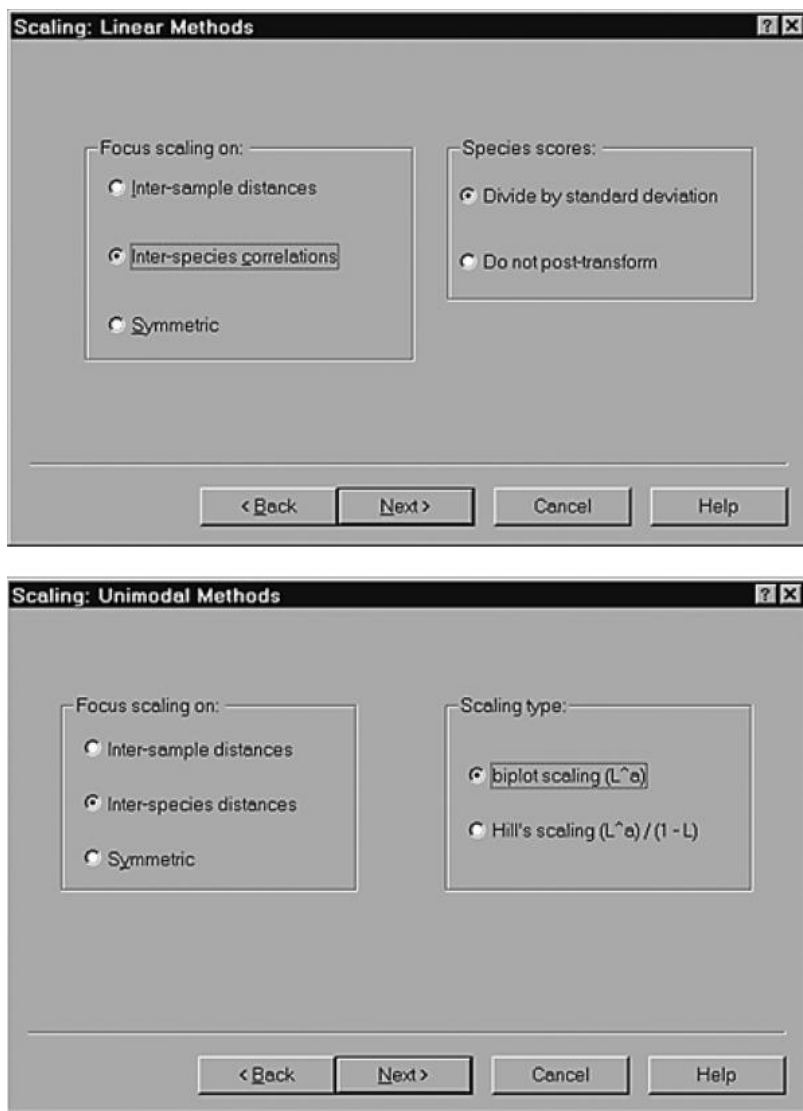


图 4-6 Scaling 选择窗口

4.7 运行出图程序 CanoDraw (Running CanoDraw for Windows 4.0)

排序最重要的结果是排序图。出排序图的程序就是 CanoDraw, 当你用 Canoco for Windows4.5 做完相应的数据分析后,直接在 Project View 的窗口选择 CanoDraw 的按钮,便可直接将分析的结果进行做图,无需到桌面或是开始任务栏去打开 CanoDraw 程序。

当产生一个排序图之前,你必须先想想你自己究竟需要什么图。如果你关注物种与环境因子之间的关系,你就要选择带有物种和环境因子的排序图(如图 4-7 所示)。但在作图之前,你必须考虑下面两点:

首先,你要先明确环境因子里面是否有定性(等级)变量(比如因子)。这些变量在排序图里面是用样方的坐标的质心来表示。从 Project>Nominal variables 子菜单进去,将因子变量预先选出来,然后再生成排序图。

第二,一般来说,你通常要将所有的样方或物种都在排序图展示出来,但是有时你觉得有些许稀有种,在群落内的分量很小,对于排序图的解释没有多少参考价值。你可以将这些物种不纳入排序图。我们可以通过两个标准来保留物种:物种的适合度(species fit):你可以自己设适合度百分之多少以上的物种才入选排序图里面。物种适合度表示每个物种被排序空间(通常是头两轴)所能展示的变化百分比例。物种的权重(species weight):你可以自己设权重多少以上的物种才入选排序图里面(也就是在相对多度多少以上才能够入选)。物种的权重仅仅在单峰排序里面有,它等于物种在所在样方内多度之和比上所有物种多度之和。这两个标准均可以在 CanoDraw 里面 Project>Settings> Inclusion Rules 实现。

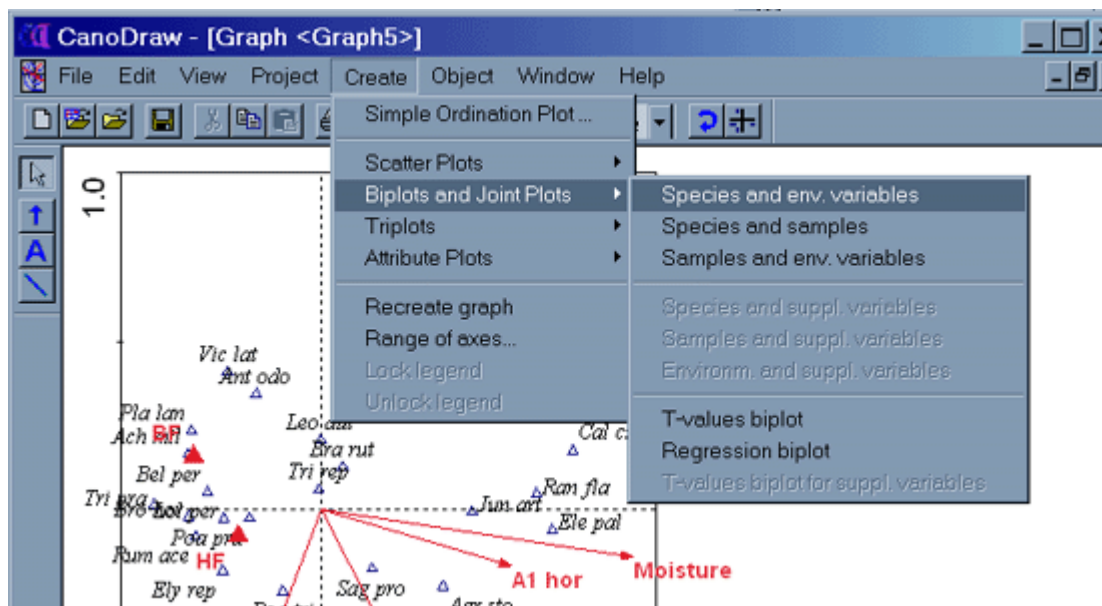


图4-7 CanoDraw 界面

当你生成排序图后,你可以对图上的各种图标进行修改,比如,你选择一个图标,然后按 F5 键,就可以根据你的需要,对图标的形状、颜色、线条等等进行修改。点右键可以对图标进行同类全选、锁定、修改标题等修改。如果你对整个图的内容(比如说是图上物种的数量太多,有点乱),你可以调整图的显示内容,通过命令栏目行中 View 和 Project 的子菜单内选项来调整,做完调整后,在执行 Create> Recreate graph 命令,便可重新出图,出的新图就会覆盖你原来的图。

CanoDraw 作出来的图能够直接打印,也可以复制到 window 的剪贴板,还可以输出各种格式的图像文件供选择(BMP, PNG, Adobe Illustrator, and Enhanced Metafile Format).