

生态学多元数据排序分析软件Canoco 5介绍

赖江山*

(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: 基于样方单元的生物群落调查多元数据是生物多样性研究中最基本的数据类型之一。排序(ordination)作为多元统计最常用的方法之一, 目的是在可视化的低维空间展示多维数据的结构。Canoco是数据排序分析最流行的软件之一。Canoco 4.5自从2002年发布以来, 凭借简单的操作界面和功能齐全的绘图工具, 得到广泛的应用。但随着计算机技术的不断发展和新的排序方法不断出现, Canoco 4.5已经无法满足生态学研究人员对于多元数据深入分析的需求。作为Canoco 4.5的升级版, Canoco 5于2012年10月发布。Canoco 5在Canoco 4.5基础上做了很多改进, 主要体现在简化数据输入、提供更完善的帮助系统和绘图工具、简化方差分解和显著性检验的步骤, 并增加了一些新的分析方法(例如PCNM、NMDS、功能性状关联分析等)。本文概述了Canoco 5所做的这些改进, 并对有些重要操作步骤进行提示, 供同行参考。

关键词: 方差分解, 邻体矩阵主坐标分析, 非度量多维尺度分析, 谱系, 功能属性

Canoco 5: a new version of an ecological multivariate data ordination program

Jiangshan Lai*

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

Abstract: Ordination of multidimensional data on community composition is one of the most important multivariate statistical methods used in biodiversity research. The aim of ordination is to visualize multidimensional data structure at a low-dimensional ordination space. Canoco is one of the most popular programs for ordination analysis and Canoco 4.5 was widely used for such analysis after its release in 2002, because of its simple user interface and powerful graphic tools. A new version of Canoco, Canoco 5 was released in October 2012. This new version simplifies data entry, provides a better help system and graphics tools, simplifies steps of variation partitioning and significance tests, adds some new methods (e.g. PCNM, NMDS, association analysis of functional traits, etc.). This paper provides an overview of the major improvements to Canoco 5, and addresses important steps required for particular analyses.

Key words: variation partitioning, PCNM, NMDS, functional traits, phylogenetic

基于样方单元的生物群落调查数据是生物多样性研究中最基本的数据类型之一。每个样方内通常包含很多物种(或环境因子)的数量信息。这样的原始数据常用多元数据矩阵来表示, 一般是一行代表一个样方, 一列代表一个物种(或环境因子)。样方、物种和环境因子数据结构特征以及它们之间的关系, 是群落生态学研究的主要内容。以研究样方

之间的关系为例, 如果每个物种作为一个维度, 有多少个物种就代表多少维度, 那么这些样方可以被当作多维空间的点的集合。如果所有样方只有2个物种存在, 可以直接用二维平面上的散点图来描述所有样方之间的关系。但样方内物种数通常超过2个, 如果绘制所有物种组合下的平面散点图, 是非常庞大的工作量。比如, 一个包括10个物种的矩阵,

收稿日期: 2013-05-31; 接受日期: 2013-08-22

基金项目: 国家自然科学基金(31200403)

* 通讯作者 Author for correspondence. E-mail: lai@ibcas.ac.cn

可以用45个(组合 C_{10}^2)二维散点图来展示样方之间的关系。但同时观测45张平面散点图,既看不出数据的主要结构,也看不清楚样方之间的关系。排序(ordination)的过程就是在一个可视化的低维空间(通常是二维)重新排列这些样方,使得样方之间的距离最大程度地反映出平面散点图内样方之间的关系信息。此时,低维空间的排序轴不再是代表具体某个物种,而是虚拟的排序轴,反映一定的生态环境因子梯度(张金屯, 2004)。排序运算过程实际上就是降维的过程,降维过程中不可避免地会丢失信息,好的排序方法就是使降维过程中信息损失尽可能减少。

自20世纪50年代开始,生态学家开始用排序的方法研究植被群落的连续分布。经过半个世纪的发展,现在已经开发出种类繁多的排序技术(von Wehrden *et al.*, 2009)。根据不同的数据对象和研究目标,现代的排序方法主要可归为两类:只使用物种组成数据的排序称作非限制性排序(unconstrained ordination),同时使用物种和环境因子组成数据的排序叫作限制性排序(constrained ordination)。常用的非限制性排序方法包括主成分分析(principal components analysis)、对应分析(correspondence analysis, CA)、主坐标分析(principal coordinate analysis, PCoA)、非度量多维尺度分析(non-metric multi-dimensional scaling, NMDS)等。常用的限制性排序方法包括冗余分析(redundancy analysis, RDA)、典范对应分析(canonical correspondence analysis, CCA)等。目前,排序方法不仅仅用于生物群落分析,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。

Canoco是生态学及相关领域多元数据排序分析最流行的软件之一,其名称取自于canonical community ordination (ter Braak & Smilauer, 2002),由美国Microcomputer Power计算机公司开发。从1985年Canoco 1.0发布以来,至今已经发布了6个版本:Canoco 1.0(1985年), Canoco 2.0(1985年), Canoco 3.0(1990年), Canoco 4.0(1998年), Canoco 4.5(2002年), Canoco 5(2012年)。

Canoco 4.5自从2002年发布以来(ter Braak & Smilauer, 2002),凭借简单的操作界面和功能齐全的绘图工具,得到了广泛的应用,也是目前研究人

员比较熟悉的版本。ISI Web of Knowledge的最新引证数据表明,在过去的10年中(2003–2013年),Canoco 4.5使用手册被引用达4,000多次。Canoco 4.5虽然包括大部分常用的排序方法,但随着计算机技术的不断发展和新的排序方法不断出现,已经无法满足生态学研究人员对于多元数据深入分析的需求。作为Canoco 4.5的升级版本,Canoco 5于2012年10月发布(ter Braak & Smilauer, 2012)。Canoco 5在Canoco 4.5基础上作了很多改进,主要体现在简化数据输入、提供更完善帮助系统和绘图工具、简化方差分解和显著性检验的步骤,并增加了一些新的分析方法,例如邻体矩阵主坐标分析(PCNM)、非度量多维尺度分析(NMDS)、功能性状关联分析等。笔者概述了Canoco 5所做的这些改进,并对有些重要操作步骤进行提示,供同行参考。

1 集成了不同功能的程序模块

Canoco 4.5包含5个分离的、不同功能的程序模块,分别是Programs of Canoco for Windows(核心程序)、CanoDraw 4.0 for Windows(绘图程序)、PrCoord(主坐标分析程序)、WcanoImp(数据转化程序)和CanoMerge(数据合并程序)。Canoco 5不再有分离的模块,已经是集数据管理、分析和绘图为一体的单一程序。附图I显示Canoco 5的工作界面。

2 可以直接导入Excel数据表格或Canoco 4.x生成的数据文件

数据输入对于初学者常常是最大的障碍。在Canoco的旧版本中,对数据格式要求严格,必须按照它自带的格式输入才能识别。在Canoco 4.x版本中,Excel数据表格进入分析之前需要由WcanoImp程序转化为Canoco识别的文件。Canoco 5的数据导入功能增强,可以直接导入Excel表格或从Canoco 4.x生成的数据文件,也可以用自带的电子表格编辑器建立空白表格直接输入数据。附图II显示的是从Excel导入Canoco 5的界面。对于变量的名称,不再限于8个字符,可以自己设定其缩略形式,也可以自动从全名中抽取一部分作为变量名称。Canoco 5中,因子变量(等级变量)可直接用字符型变量,而不必预先转为二元变量(0,1数据),运算时程序会自动将因子变量转为二元变量。

3 提供更智能的帮助系统

Canoco 5除了常规的帮助系统, 还以Canoco 顾问(adviser)的形式主动提供智能帮助。Canoco 顾问是真正的专家系统, 能够根据数据类型建议使用什么分析方法、如何运行这些程序, 以及如何获得最佳的可视化排序结果(附图III)。


4 提供完善的绘图工具


Canoco 5提供CanoDraw 4.0的所有绘图工具, 有些工具还得到了优化, 所生成的排序图可以保存为JPEG、TIFF和PDF等格式。附图IV分别显示以多边形(左图)或椭圆形(右图)来区分排序空间内的不同类群。设置多边形和椭圆形的步骤如下:

(1)进行排序分析后, 先对物种(或样方)进行分类: 选择Project-Classifications(注意一定要对所有的物种或样方进行归类, 否则, 做图的时候将不能显示未归类的对象), 然后选择Use selected classification in graphs。

(2) 选择 Analysis-plot creation options—General-plot envelopes for (sample or species)

(3)生成排序图: Graph-Scatterplots-(sample or species)。

(4)选择Graph-Show Attribute Editor 或是直接点击可以对多边形进行编辑, 包括填充颜色、线条粗细等。

(5)如果需要将多边形换成椭圆形: 选择 Edit-Setting-Graphing options—Draw envelopes as ellipses, 然后点击Graph-Recreate graph, 原来排序图上的多边形将被椭圆形替代。一般地, 初始的椭圆形所设定的范围不准确, 可以对其尺寸作进一步修正, 步骤如下: 鼠标左键选定要修正的椭圆形, 点击右键, 选择change ellipses size, 可以根据实际情况设定椭圆形合适的尺寸(因为椭圆的方向和中心点不能改变, 故椭圆形效果一般比多边形差)。选择 Graph-Show Attribute Editor 或是直接点击也可以对椭圆形进行编辑。

5 完善了基于距离矩阵的排序分析

在Canoco 4.5中, 主坐标分析(PCoA)和基于距离的冗余分析(db-RDA)通常预先计算距离矩阵(Conoco 4.5 自带的距离度量模式非常有限), 再输

入PrCoord模块及主程序进行联合运算分析。Canoco 5包含更多的距离度量方式(包括Bray-Curtis、Gower 距离、Jaccard 系数等11种典型距离计算), 可以直接导入原始数据后进行PCoA和db-RDA分析。附图V显示PCoA分析设置向导的界面。另外, Canoco 5增加了NMDS排序。

6 简化方差分解计算步骤

在Canoco 4.5中, 两组或三组解释变量对响应变量总方差的解释率分解需要多次偏分析才能获得最终结果, 而Canoco 5可以直接给出每组变量的单独解释部分、共同解释部分以及总的解释量(附图VI)。在分析方法的选择界面, Canoco 5里将方差分解的方法归为一类: 即Variation Partitioning Analyses, 也见附图IX。

7 增加了单个解释变量的显著性检验功能

在限制性排序分析中, 如果在蒙特卡罗检验选择界面中选中“Summarize effects of expl. variables”(附图VII上图), 分析结果可以显示每个环境因子解释率和显著性检验的结果(附图VII下图)。下图中Simple Effects部分是每个环境因子单独的解释量(等同只导入单个环境因子的限制性排序分析的解释量), Conditional Effects部分是按照解释量大小顺序不断加入环境因子后所增加的解释量。

8 可以显示每个轴的显著性检验结果

在Canoco 4.5中, 限制性排序的显著性检验只能显示前四轴与第一轴的检验结果。如果想检验其他轴的显著性(比如第二轴), 只能首先进行限制性排序分析, 使用样方在第一轴的坐标作为协变量(从承载分析结果的solution(.sol)文件获得)进行偏限制性分析, 这样新检验出的第一轴的显著性便是原来第二轴的显著性(Lepx & Smilauer, 2003)。而Canoco 5可以直接给出各个限制性轴单独的检验结果(附图VIII)。

9 增加了邻体矩阵主坐标分析(PCNM)方法

邻体矩阵主坐标分析(principal coordinates of neighbor matrices, PCNM)是近期开发的多尺度空间结构建模方法, 现已广泛应用在群落数据空间分析的研究中(Borcard & Legendre, 2002; Legendre *et al.*,

2005, 2008; von Wehrden *et al.*, 2009)。该方法在Canoco 5中被纳入方差分解分析(variation partitioning analyses)范畴。附图IX显示PCNM分析的设置界面。

10 增加了群落内类群功能性状与谱系关联分析

目前,物种的功能性状与谱系分析是生态学研究的热点之一。功能多样性是生物多样性的重要组成部分,一般将其定义为能够影响生态系统功能的物种或有机体性状的价值和范围(Petchey & Gaston, 2006)。目前,功能多样性逐渐被广泛应用于探究陆地、水生生态系统中包括植物、动物、菌类等不同营养级的生物体在群落和生态系统中的功能及其范围。Canoco 5不仅可以管理功能性状数据,并计算群落内和群落间功能多样性相关的指数,计算单个样方内某个性状的平均值等,也可以直接导入谱系树进行相关分析。Canoco 5支持Desdevises方法,允许导入Newick格式系统发育树,并将其转变成为一个Patristic距离矩阵。在Canoco 5中,还可以轻松地做出美观的性状和环境变量的RLQ分析排序图。附图X显示Canoco 5中功能性状与谱系分析的相关界面。

致谢: 感谢Microcomputer Power公司的Richard E. Furnas博士提供Canoco 5试用版本。对Canoco 5试用

版感兴趣的读者请直接向Richard先生索取(FurnasR@microcomputerpower.com)。

参考文献

- Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.
- Legendre P, Borcard D, Peres-Neto PR (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.
- Legendre P, Borcard D, Peres-Neto PR (2008) Analyzing or explaining beta diversity? Comment. *Ecology*, **89**, 3238–3244.
- Lepx J, Smilauer P (2003) *Multivariate Analysis of Ecological Data Using CANOCO*. Cambridge University Press, Cambridge.
- Petchey OL, Gaston KJ (2006) Functional diversity: back to basics and looking forward. *Ecology Letters*, **9**, 741–758.
- ter Braak CJF, 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.
- ter Braak CJF, Smilauer P (2012) *CANOCO Reference Manual and User's Guide: Software for Ordination. Version 5*. Microcomputer Power Ithaca, New York.
- von Wehrden H, Hanspach J, Bruelheide H, Wesche K (2009) Pluralism and diversity: trends in the use and application of ordination methods 1990–2007. *Journal of Vegetation Science*, **20**, 695–705.
- Zhang JT (张金屯) (2004) *Quantitative Ecology (数量生态学)*. Science Press, Beijing. (in Chinese)

(责任编辑: 周玉荣)

附图I Canoco 5的工作界面(<http://www.biodiversity-science.net/fileup/PDF/w2013-133-1.pdf>)

附图II 数据从Excel导入Canoco 5的界面(<http://www.biodiversity-science.net/fileup/PDF/w2013-133-2.pdf>)

附图III 数据导入后Canoco顾问提供的选择排序类型的建议(<http://www.biodiversity-science.net/fileup/PDF/w2013-133-3.pdf>)

附图IV 以多边形或椭圆形区分排序空间内的不同类群(<http://www.biodiversity-science.net/fileup/PDF/w2013-133-4.pdf>)

附图V PCoA分析设置向导界面(<http://www.biodiversity-science.net/fileup/PDF/w2013-133-5.pdf>)

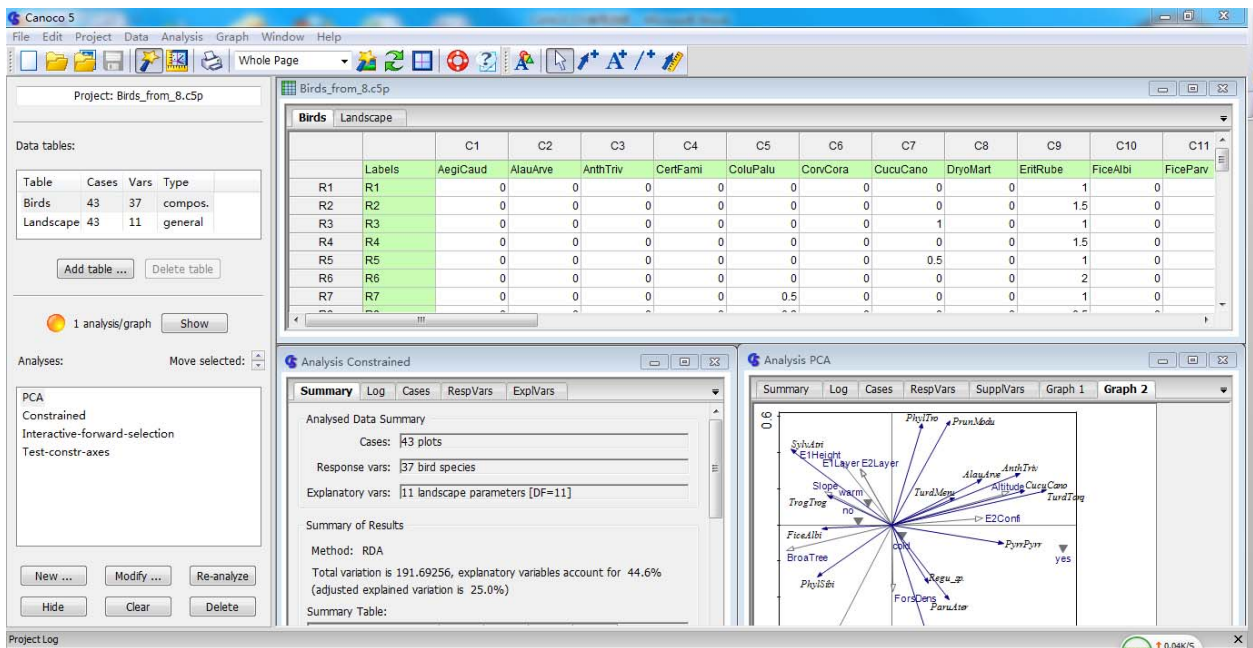
附图VI 方差分解分析的界面(<http://www.biodiversity-science.net/fileup/PDF/w2013-133-6.pdf>)

附图VII 限制性排序分析中单个解释变量的显著性检验界面(<http://www.biodiversity-science.net/fileup/PDF/w2013-133-7.pdf>)

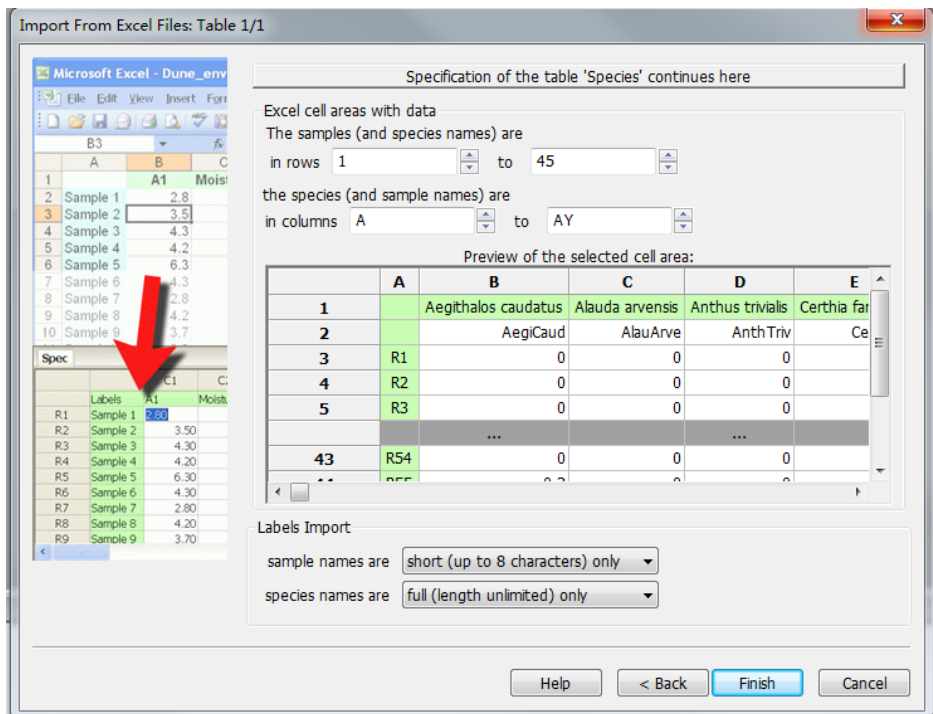
附图VIII 直接显示各个限制性排序轴显著性检验结果(<http://www.biodiversity-science.net/fileup/PDF/w2013-133-8.pdf>)

附图IX PCNM分析设置界面,可以定义样方点的坐标(<http://www.biodiversity-science.net/fileup/PDF/w2013-133-9.pdf>)

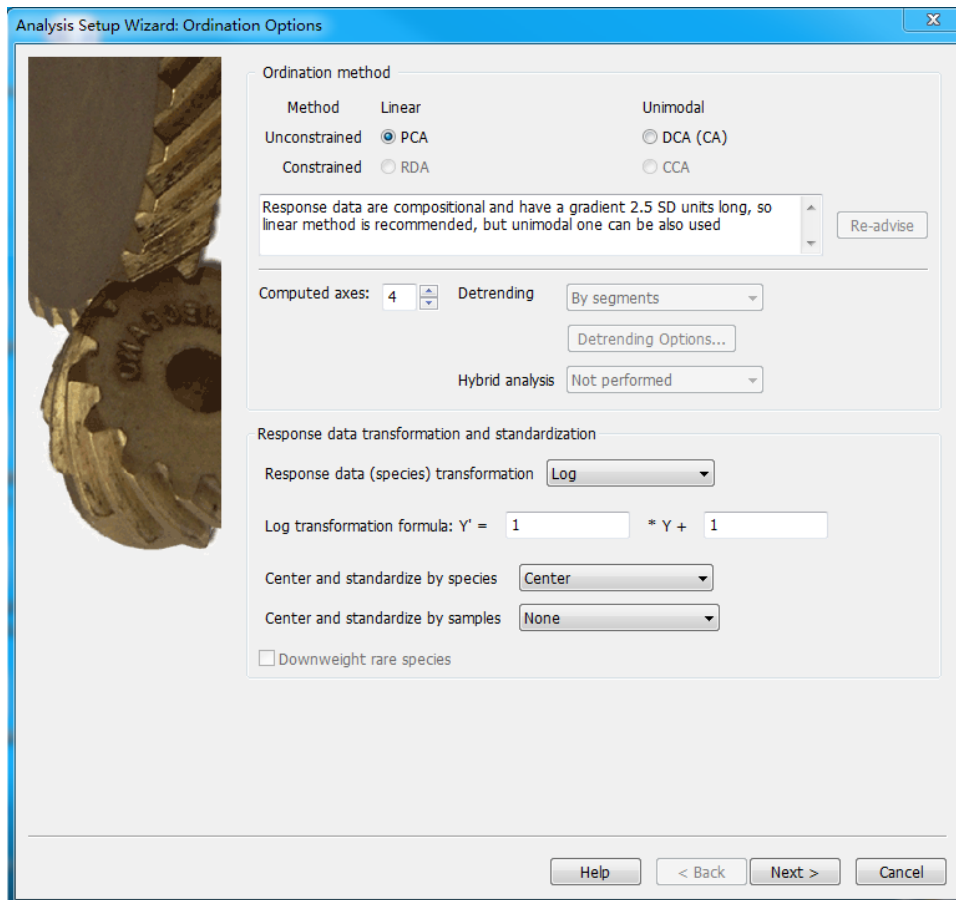
附图X Canoco 5内功能性状与谱系分析相关界面(<http://www.biodiversity-science.net/fileup/PDF/w2013-133-10.pdf>)



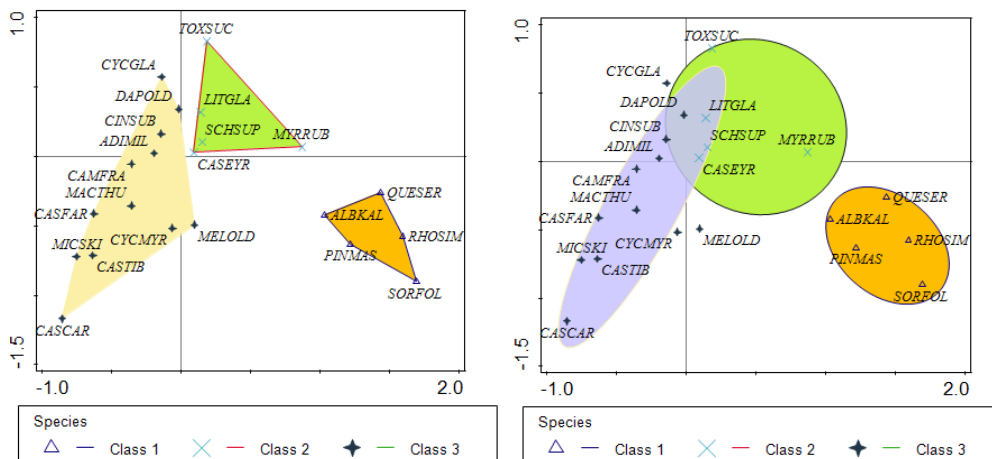
附图I Canoco 5的工作界面。左侧是当前所分析数据的基本信息和分析类型列表，右侧是数据内容、分析结果和排序图。(Canoco 5默认显示简约的排序结果，如果需要显示完整的排序分析结果，在菜单栏Edit-Settings-Canoco5 options内将general-show brief version of notebooks选项的勾去掉)。



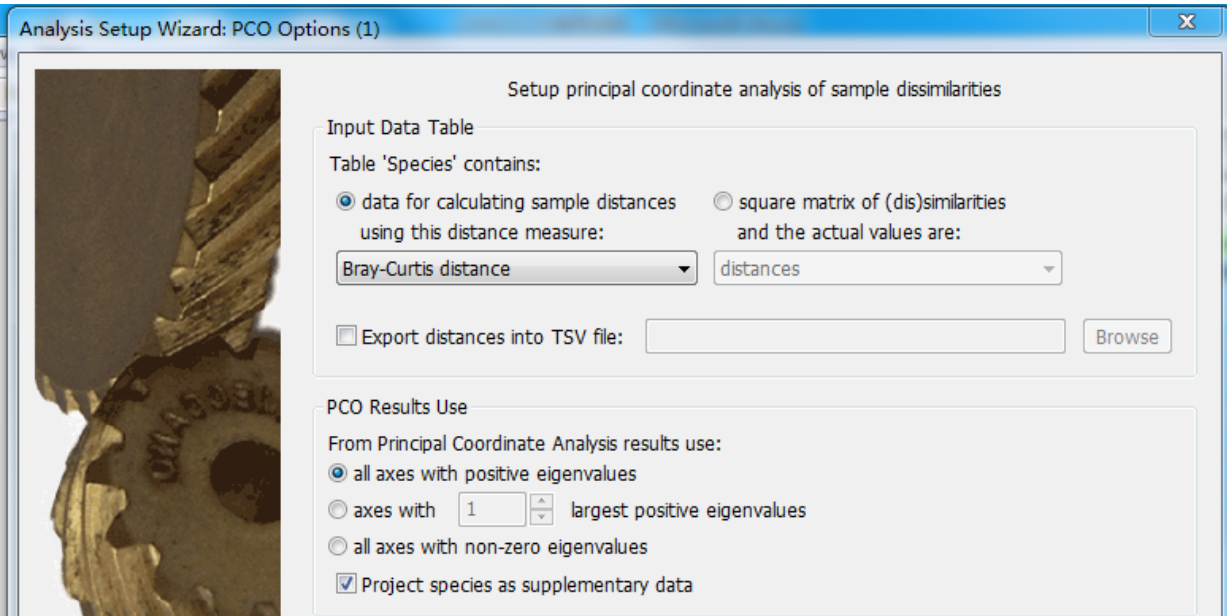
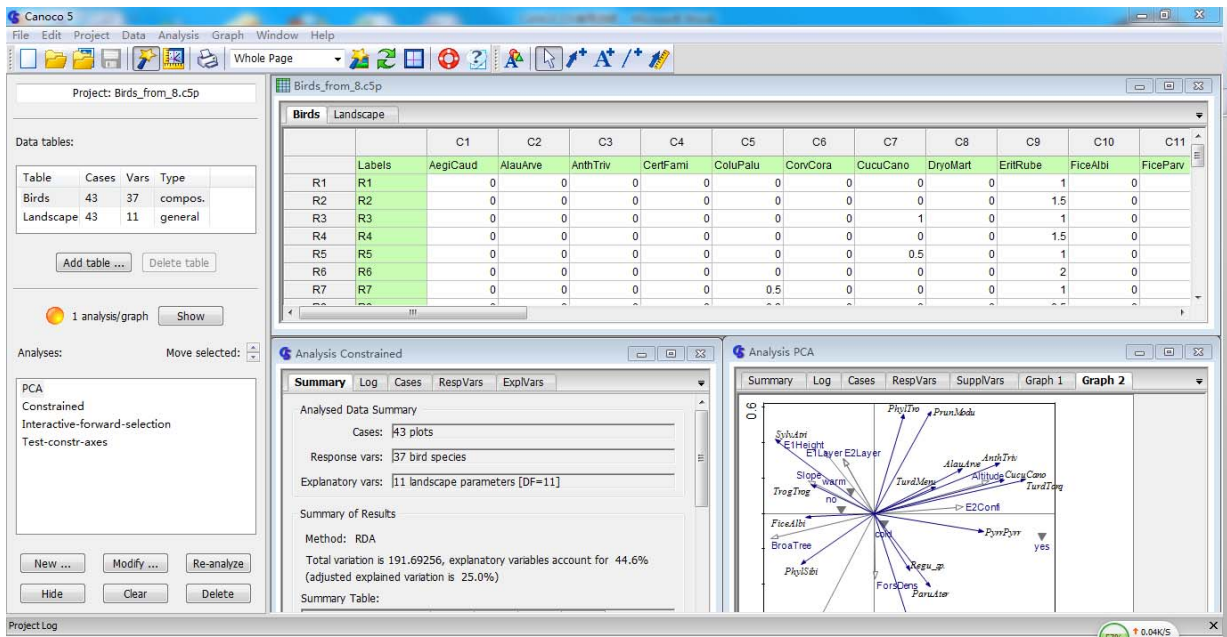
附图II 数据从Excel导入Canoco 5的界面。指定输入数据的范围(Excel cell areas with data)，也可以设定行和列名称模式(Label import)。通过中间的表格，可以浏览数据内容，也可以对数据进行编辑。



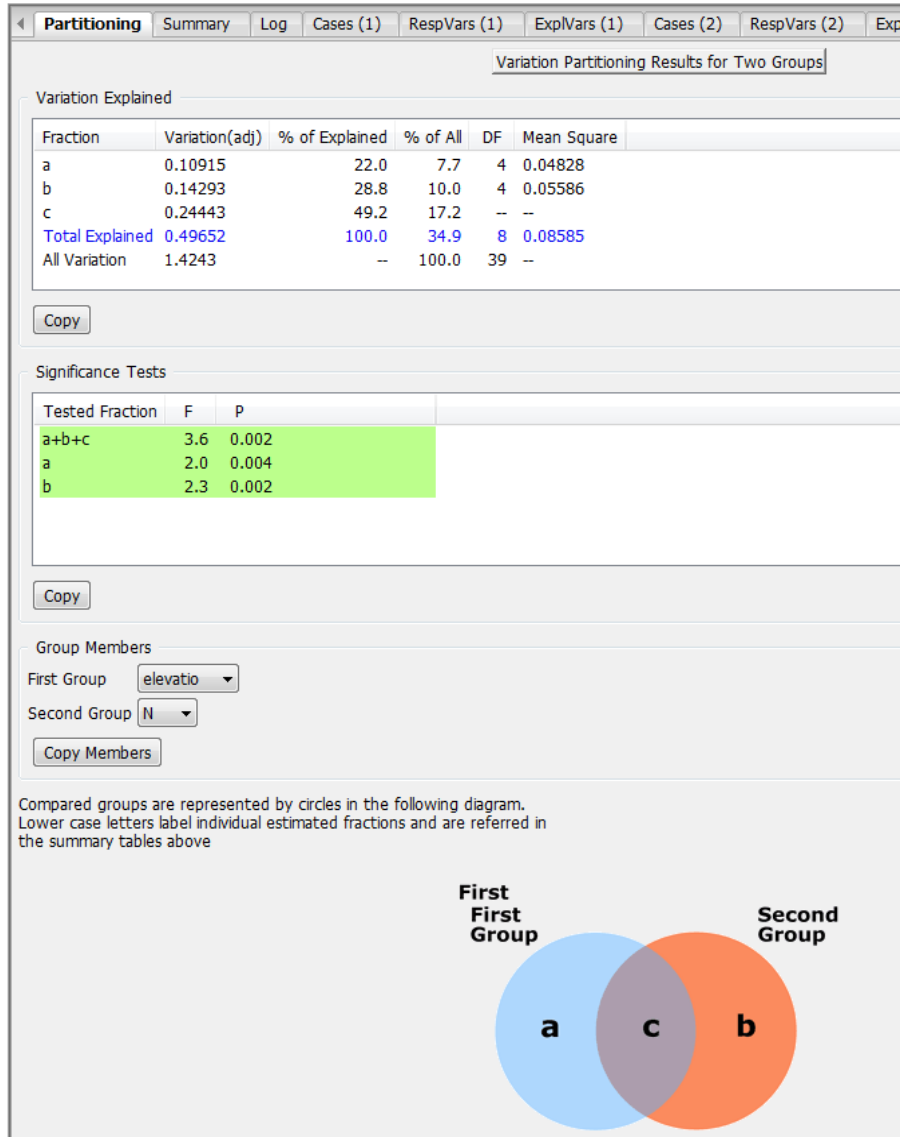
附图III 数据导入后Canoco顾问提供的选择排序类型的建议



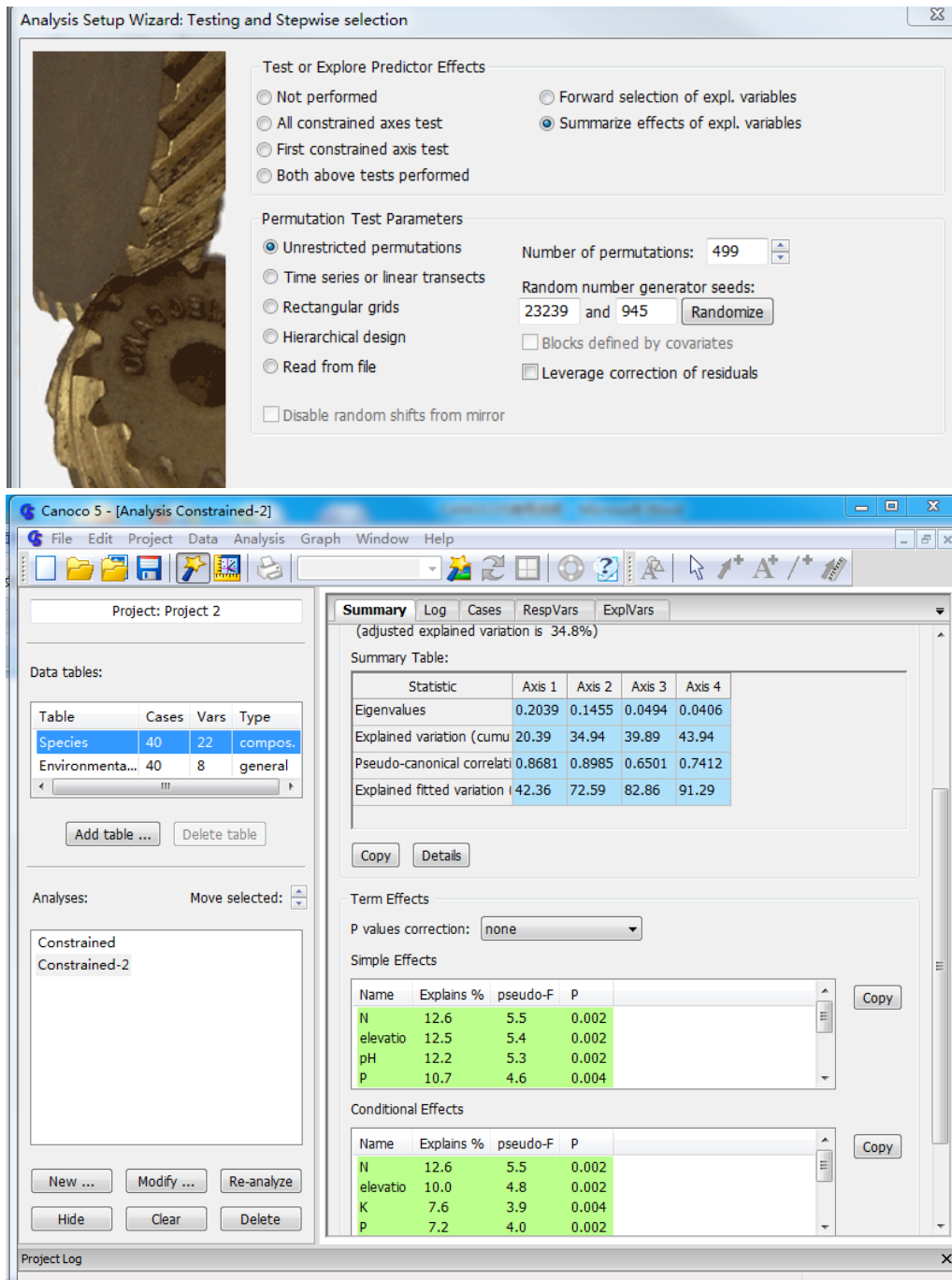
附图IV 以多边形(左图)或椭圆形(右图)区分排序空间内的不同类群



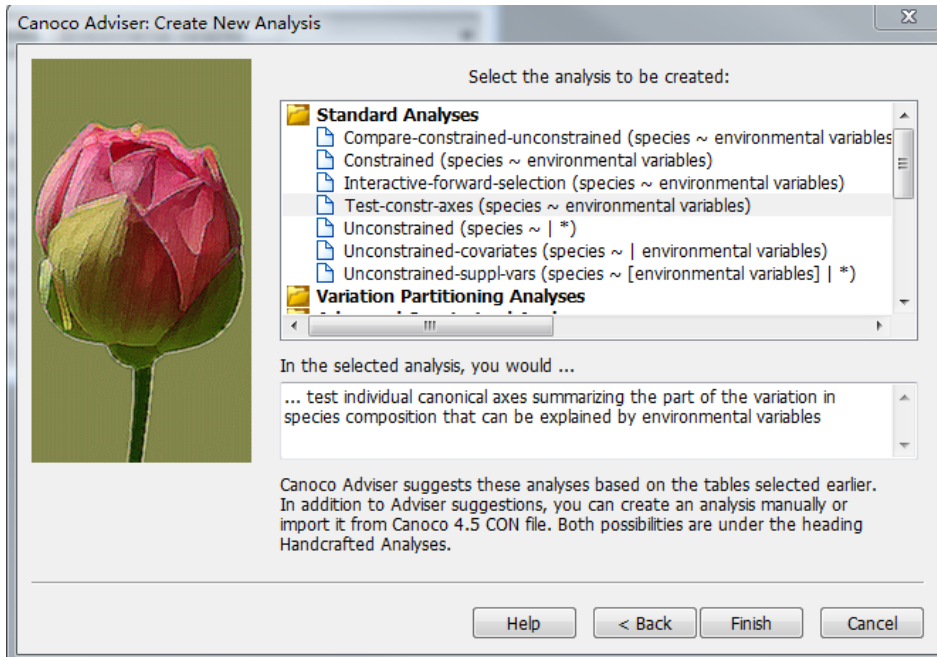
附图V PCoA分析设置向导界面。在此界面中，可以选择计算距离的度量方式，或者使用从外部导入的距离或相似矩阵，并设定显示正特征根范围的排序轴。



附图VI 方差分解分析的界面。上面的表格(Variation Explained)显示响应变量(群落)总方差能够被两组环境因子单独(a和b)及共同解释的部分(c)。中间部分(Significance Tests)显示各部分解释量显著性检验的结果。



附图 VII 限制性排序分析中单个解释变量的显著性检验界面



Analysis Test-constr-axes-2

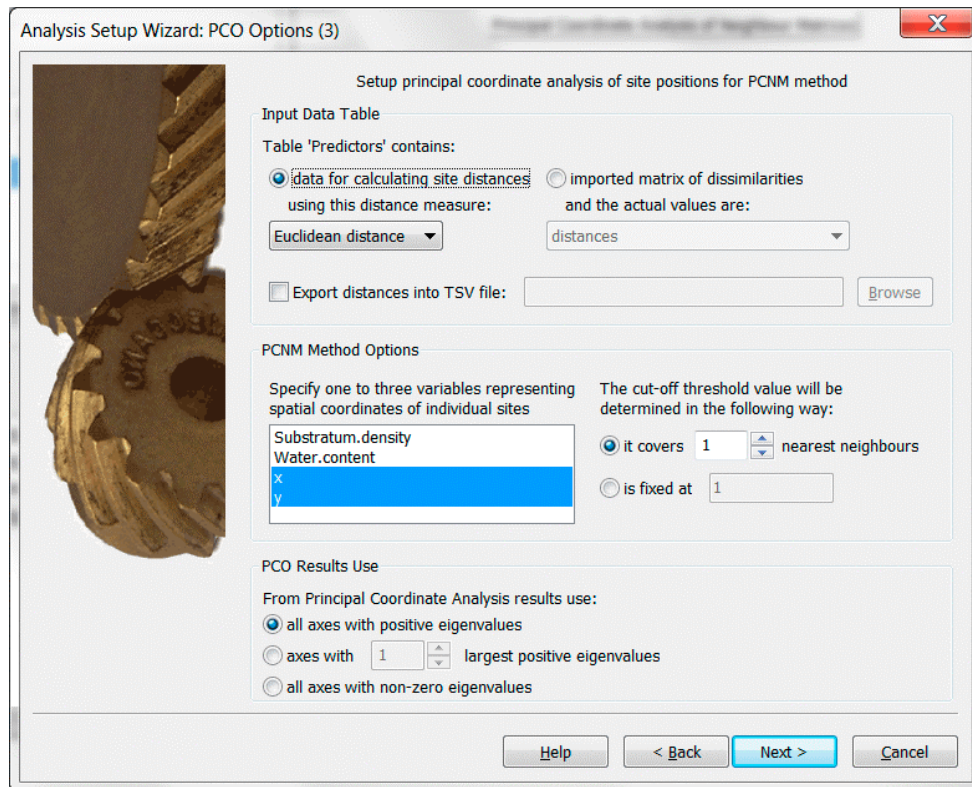
Axes significance Summary Log Cases (1) RespVars (1) ExpIVars (1) Cases (2: Te

Following table summarizes the results of testing individual constrained axes of a constrained analysis. Colouring in the first row (with percentage of total variation explained by particular axis) reflects the P values shown in the third row.

	1	2	3	4	5	6	7	8
Explained by constrained axis [%]	20.39	14.55	4.94	4.06	1.86	1.26	0.63	0.44
Pseudo-F value	7.9	6.9	2.5	2.2	1.1	0.7	0.4	0.3
P value	0.002	0.002	0.188	0.214	0.95	0.994	1.	0.998

Copy Help

附图VIII 直接显示各个限制性排序轴显著性检验结果



附图IX PCNM分析设置界面，可以定义样方点的坐标

Species		Environment	Traits	Trait Averages		Functional Diversity	
			C1	C2	C3	C4	C5
	Labels	Polycarpic	CNratio	seed.mass.log	SLA	height	
R1	ACHIMILL	polycarp	12.94	-0.876	19.63	2	
R2	AGROCAPI	polycarp	17.421	-1.444	29.54	21	
R3	ANTHODOR	polycarp	13.778	-0.678	29.97	14	
R4	BRIZMEDI	polycarp	33.69	-0.578	25.93	20	
R5	BROMHORD	monocarp	16.183	0.305	26.19	12	
R6	CAPSBURS	monocarp				15	
R7	CAREFLAC	polycarp				2	
R8	CAREHIRT	polycarp				15	
R9	CARENIGR	polycarp				15	
R10	CENTJACE	polycarp				15	
R11	CERARVE	polycarp				15	
R12	CERAGLOM	monocarp				15	
R13	CIRSARVE	polycarp				56	
R14	CYNOCRIS	polycarp				1	
R15	DAUCCARO	polycarp				3	
R16	DESCCESP.CE	polycarp				36	
R17	EQUIPALU	polycarp				16	
R18	FESTOVIN	polycarp				16	
R19	FESTPRAT	polycarp				16	
R20	FESTRUBR	polycarp	23.679	-0.235	18.6		
R21	GALIMOLL	polycarp	17.856	-0.372	21.11	4	
R22	GALIULIG	polycarp	27.043	-0.772	29.63		
R23	HIERACIU	polycarp	24.123	-0.377	32.35		
R24	HOLCLANA	polycarp	16.803	-0.333	29.43	23	
R25	HYDRVULG	polycarp	16.045	-0.62	33.77		
R26	LEONAUTU	polycarp	18.093	-0.554	25.18	7	
R27	LOLIPERE	polycarp	14.972	0.132	25.93		
R28	LOTUULIG	polycarp	14.871	-0.311	25.13		
R29	LUZUCAMP	polycarp	15.258	-0.298	27.75	3	
R30	MATRMARI.INC	monocarp	18.446	-0.896	22.2		
R31	MOLICAER	polycarp	30.106	-0.572	18.72		
R32	PHRAAUST	polycarp	16.219	-0.921	12.31		

Import Phylogenetic Tree

Name of imported file:

Default branch length:

Terminal items of the tree correspond to:

Item labels in the tree correspond better to:
 short labels full labels

附图X Canoco 5内功能性状与谱系分析相关界面。物种数据表(species)是植物群落物种组成, 性状数据表(Traits)显示已知的群落内物种的功能性状。