Canoco 5 – a short introduction

Software for multivariable data analysis and visualization

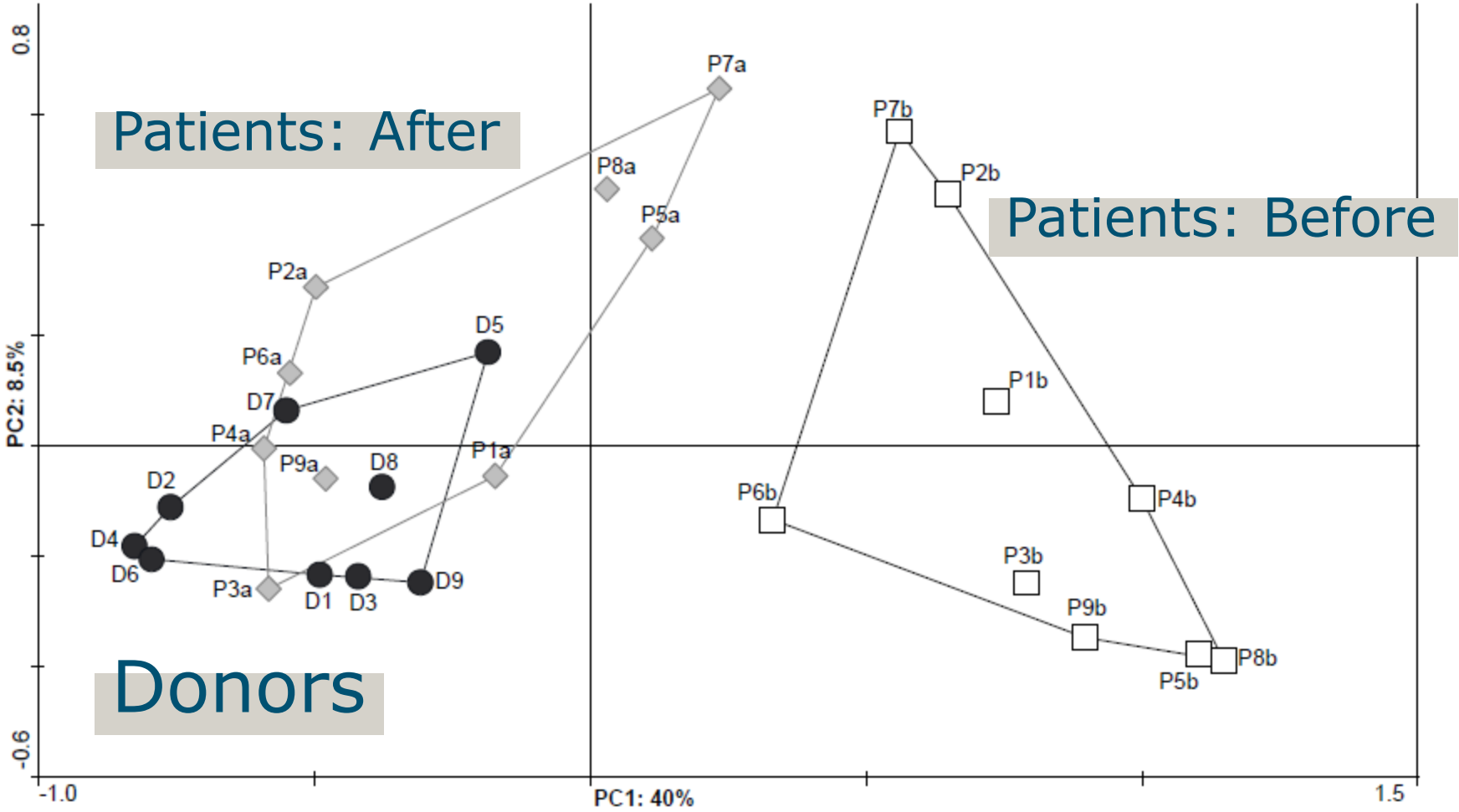
Canoco shows it!

Cajo J.F. ter Braak and Petr Šmilauer

WAGENINGEN UR
For quality of life

Ex1: Comparison of microbiota among three groups

Unconstrained (PCA)

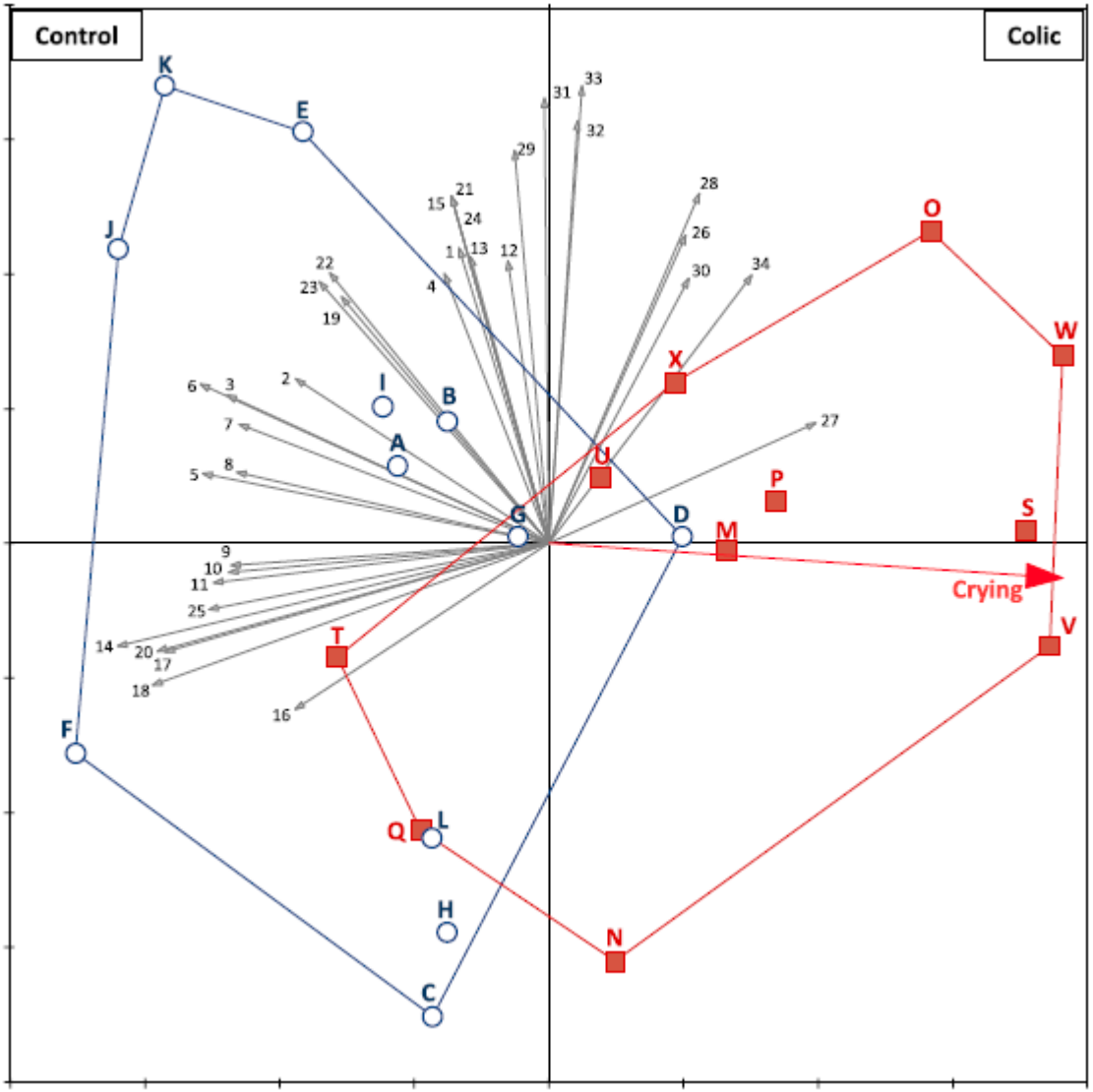


Feces transplant study: van Nood et al. 2013 NEJM
Data: **microbiota taxa** (Susana Fuentes, W. de Vos)

WAGENINGEN UR
For quality of life

Ex2: Comparison among two groups (1)

Constrained (RDA)



- Extension of t-test
- Horizontal (constrained) axis = difference of Control and Colic
- Vertical (unconstrained) axis = main residual pattern
- Correlation with Crying of babies

WAGENINGEN UR
For quality of life

De Weerth et al 2012, Pediatrics
Microbiota (Susana Fuentes, W. de Vos)

Ex2: Comparison among two groups (2)

We see three types of data in this example

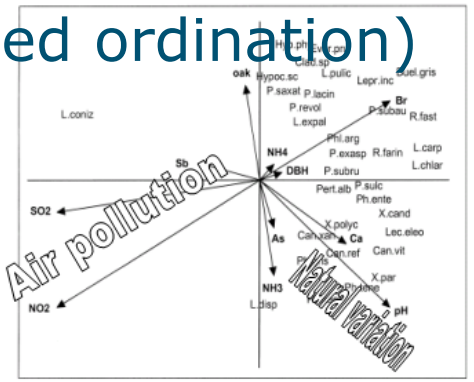
- Response data (the main/focal data :
  - Amounts of 33 microbiota taxa
- Explanatory data:
  - Treatment, a factor with 2 levels (Control and Colic)
- Supplementary data:
  - Crying

Roles of data tables

- Response data (main data table)
  - to be visualized, perhaps in combination with others
- Supplementary data
  - to interpret the response data
- Explanatory data
  - to explain the response data
- Covariate data (for advanced users)
  - to account or adjust for.
  - to enable detection of structure in response *after* accounting for the variation explained by these covariates

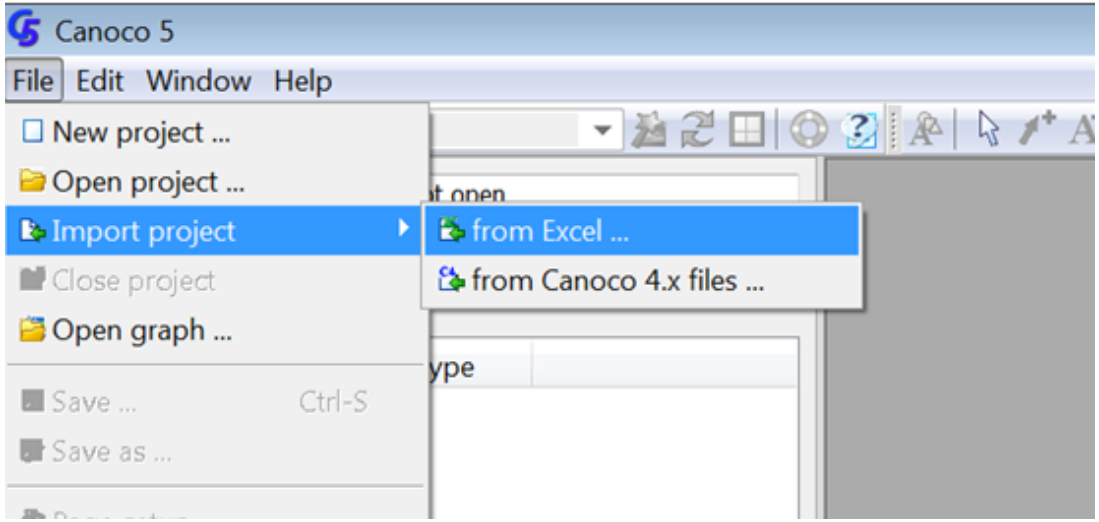
Research questions and methods in Canoco

- Derive patterns and relationships from data
  - From field or laboratory
  - From designed experiments or surveys
  - **Many noisy variables**, non-linear relationships
- Key methods
  1. Dimension reduction (ordination, factor analysis, multidimensional scaling)
  2. Regression analysis, also non-linear
  3. Combination of 1 and 2 (constrained ordination)
  4. Visualization of results
  5. Statistical testing by permutation



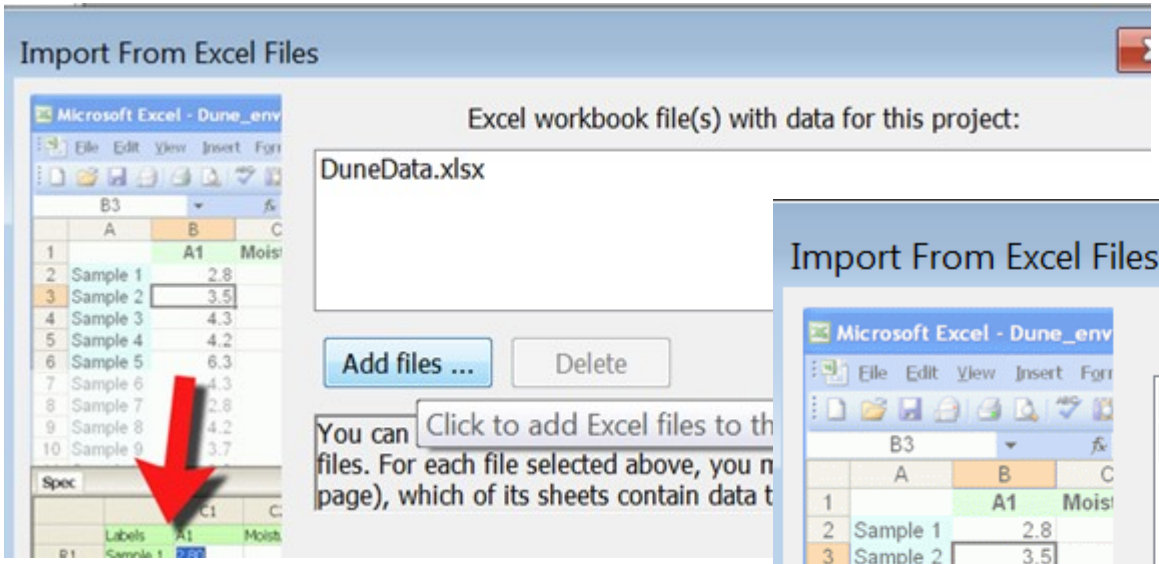
Starting a new Canoco project (1)

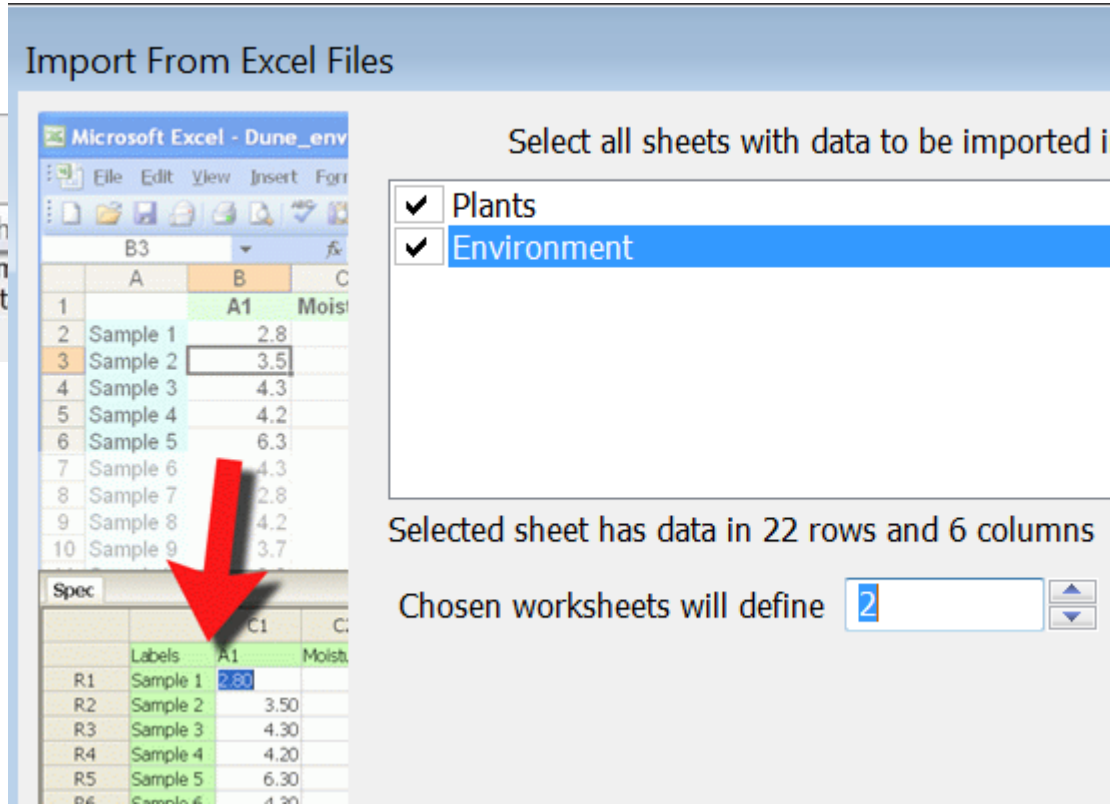
- Canoco 5 focuses on research questions on a set of data
- A Canoco 5 project thus consists of
  - one or more data tables
  - analyses on these data
- Easiest to start a new project with
  File|*Import project*|*from Excel*... (Alt-F-I-Enter)



Starting a new Canoco project (2)

- Select one or more Excel files, here 1
- Select the number of project data tables, here 2





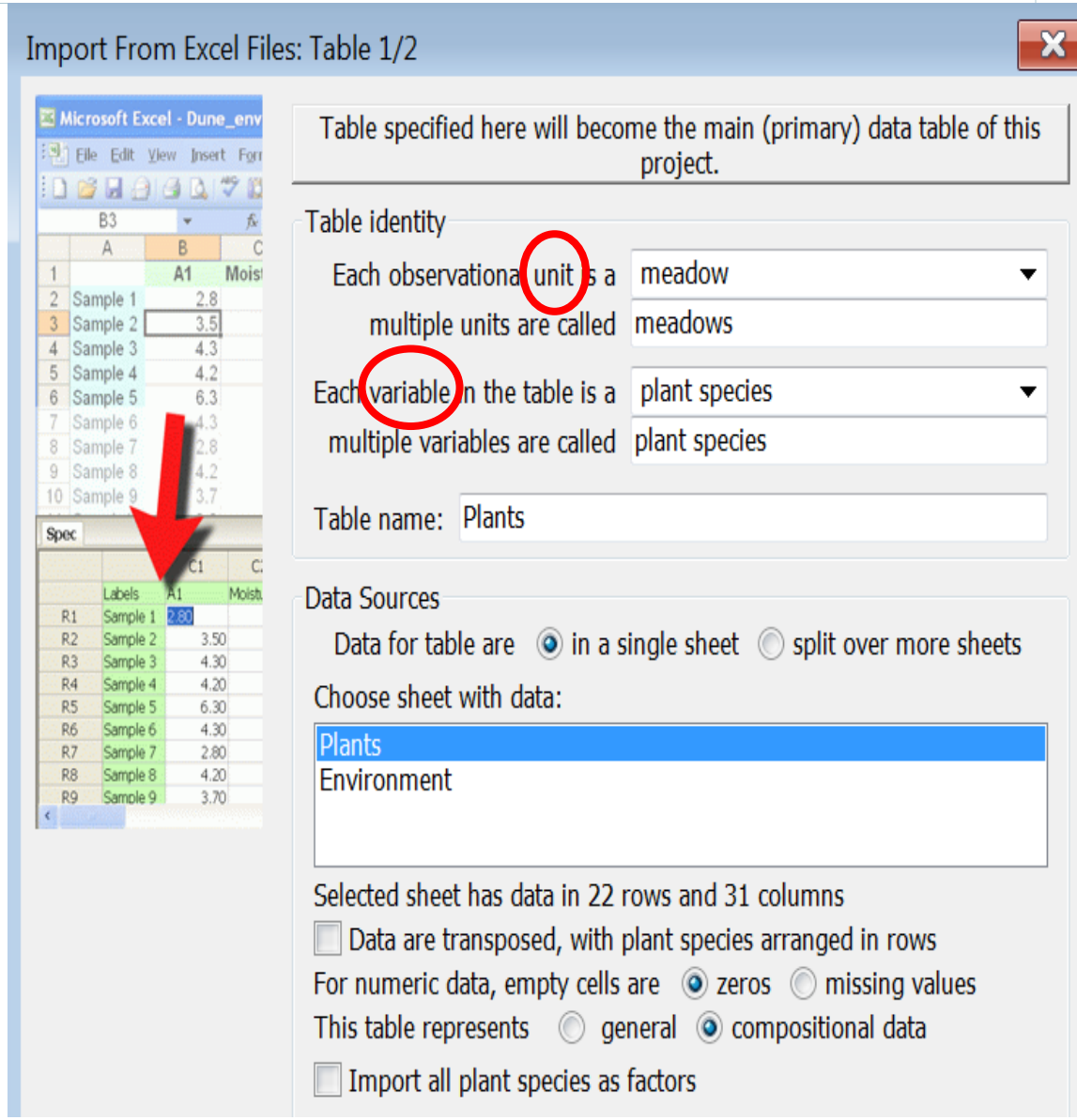
- Excel file can contain more than one sheet
- Each sheet can give ≥1 data tables

WAGENINGEN UR
For quality of life

Starting a new Canoco project (3a)

Give names to YOUR units and variables

- choose from list or
- start typing
  - singular, then
  - plural



WAGENINGEN UR
For quality of life

Starting a new Canoco project (3b)

Give names to YOUR units and variables

Empty cells: 0 or mis

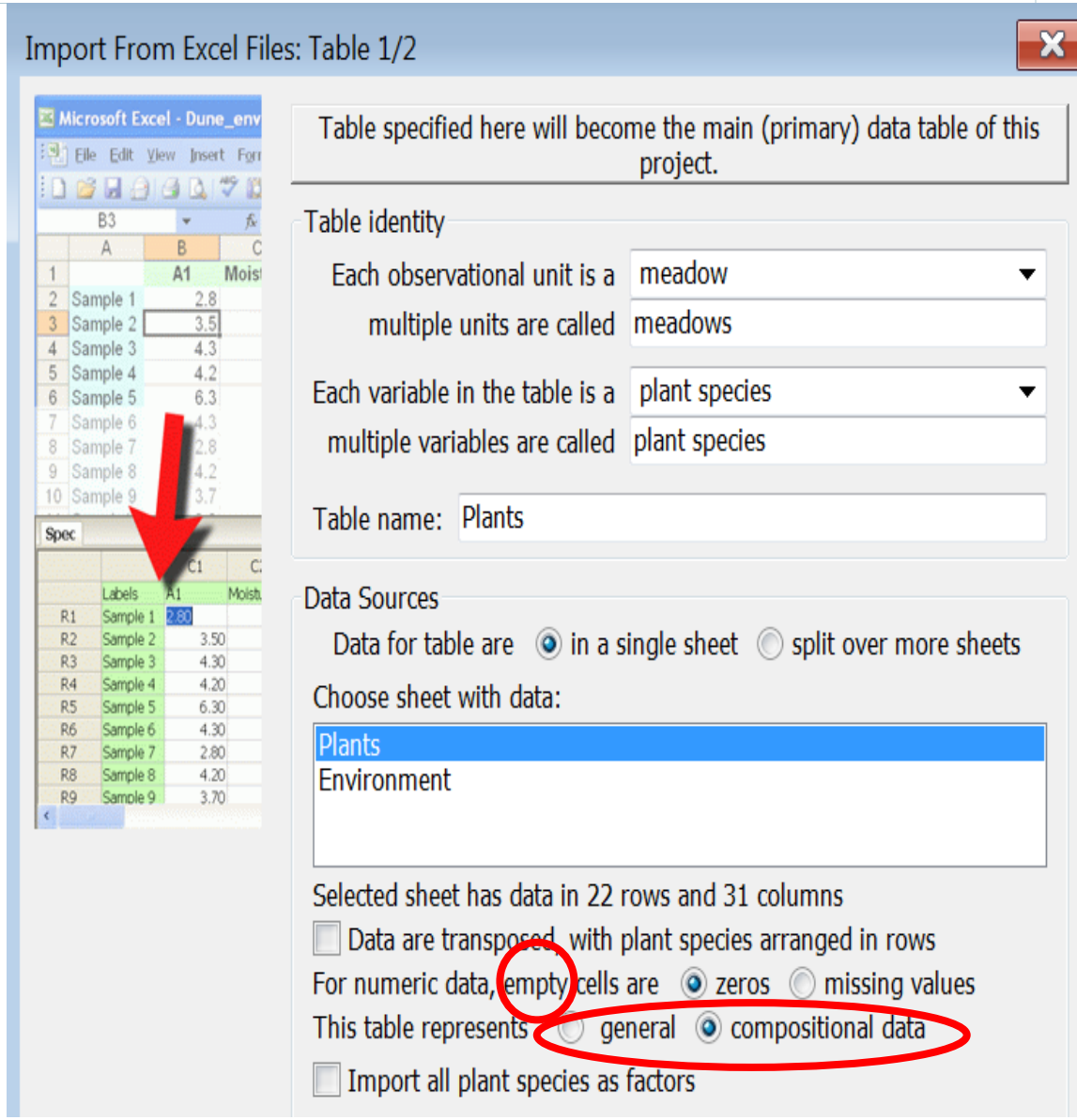
Data kind is

- General or
- Compositional:

-row sum has meaning

-variables measured on the same scale

The right choice helps to select suitable methods



WAGENINGEN UR
For quality of life

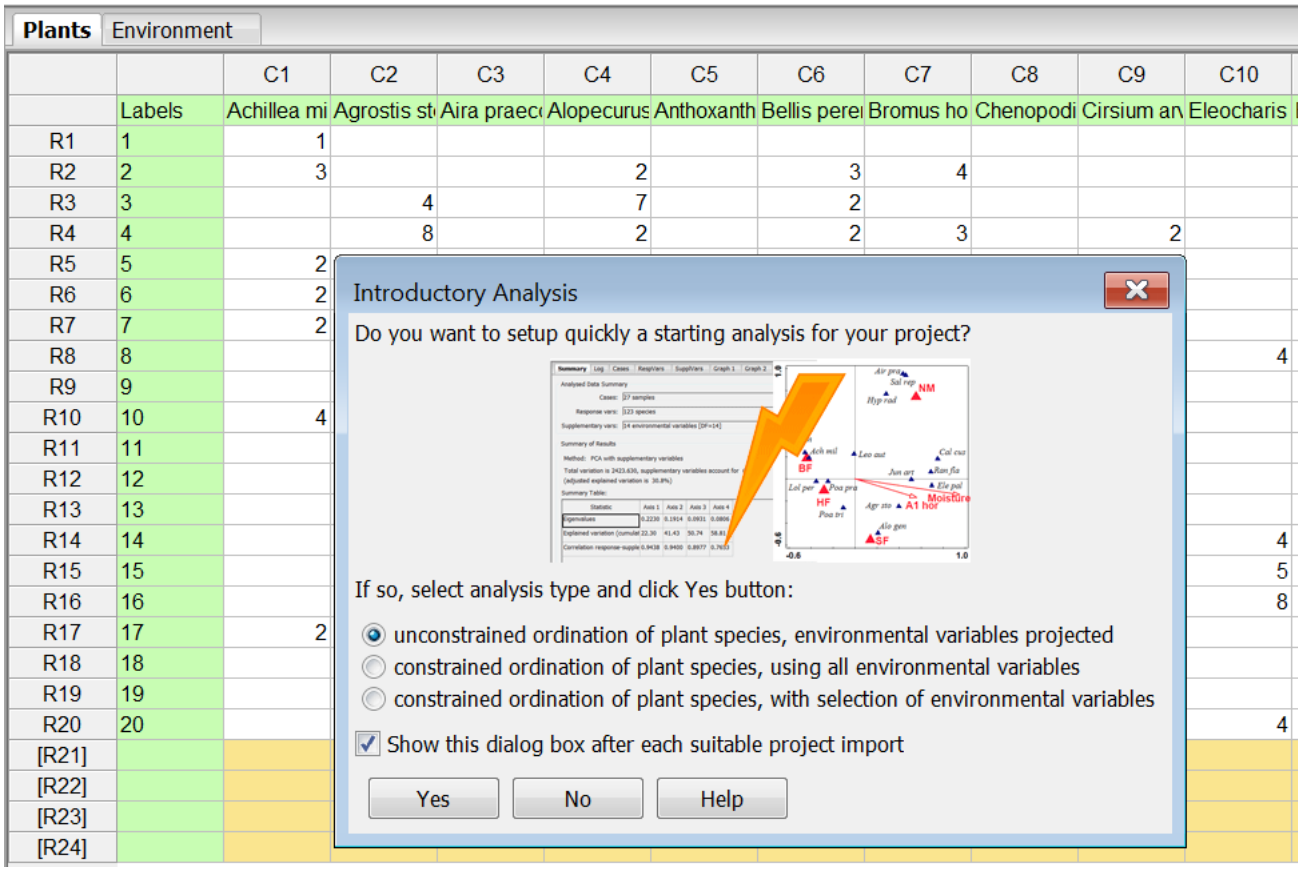
Starting a new Canoco project (5)

Result: two project data tables (Plants and Environment) and offer for starting analysis

Data tables:

you can

- View
- Edit
- Copy
- Export
- Change kind/name etc.



		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	Labels	Achillea mi	Agrostis st	Aira praec	Alopecurus	Anthoxanth	Bellis pere	Bromus ho	Chenopodi	Cirsium ar	Eleocharis
R1	1	1									
R2	2	3			2		3	4			
R3	3		4		7		2				
R4	4		8		2		2	3		2	
R5	5	2									
R6	6	2									
R7	7	2									
R8	8										4
R9	9										
R10	10	4									
R11	11										
R12	12										
R13	13										
R14	14										4
R15	15										5
R16	16										8
R17	17	2									
R18	18										
R19	19										
R20	20										4
[R21]											
[R22]											
[R23]											
[R24]											

WAGENINGEN UR
For quality of life

Starting a new Canoco project (6)

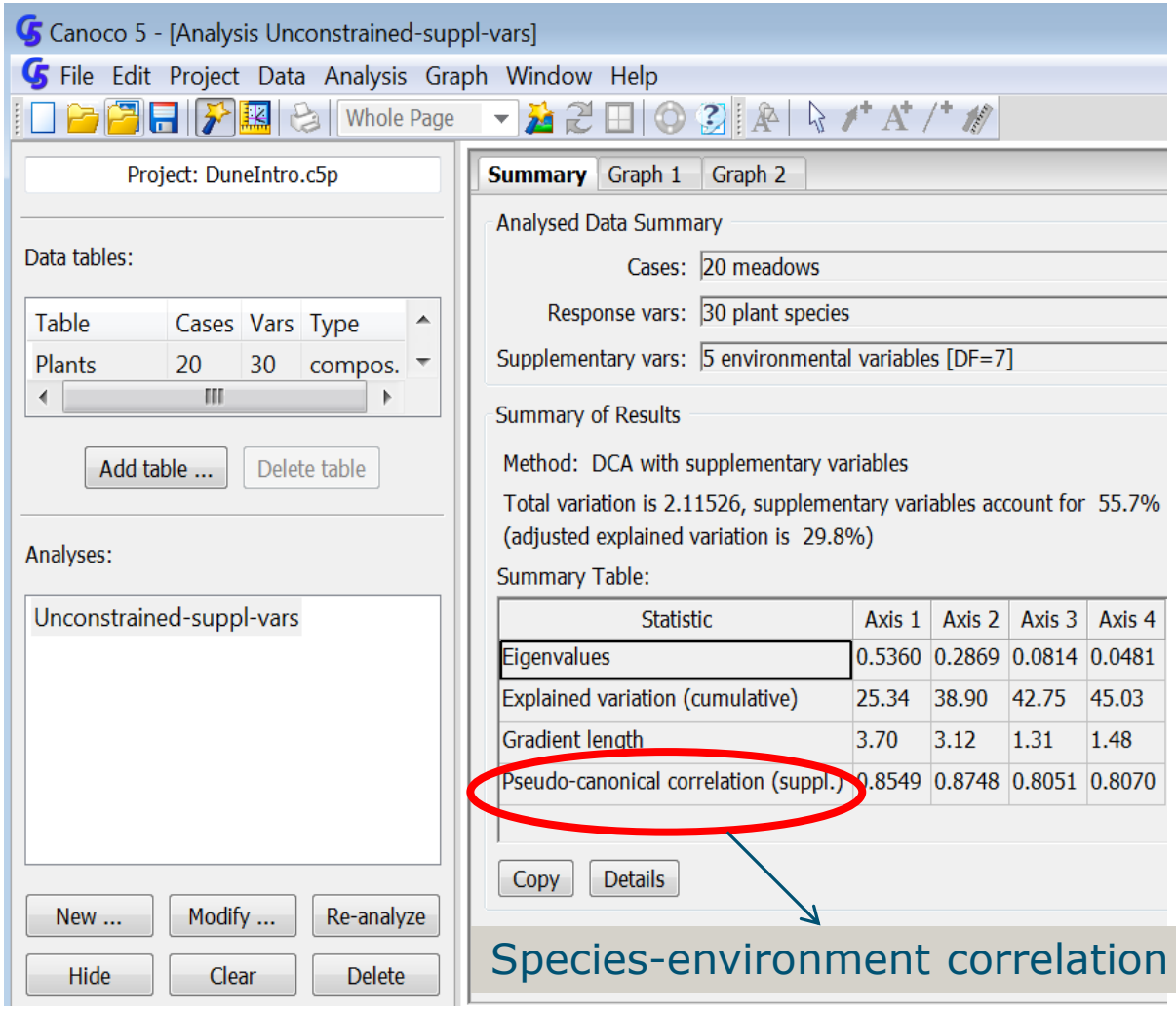
Accepting the offer and all default choices leads to

-Summary of DCA analysis

-Two graphs

Save your project!

- File Save.. or
- Ctrl-S



WAGENINGEN UR
For quality of life

Starting a new Canoco project (6)

Accepting the offer and all default choices leads to

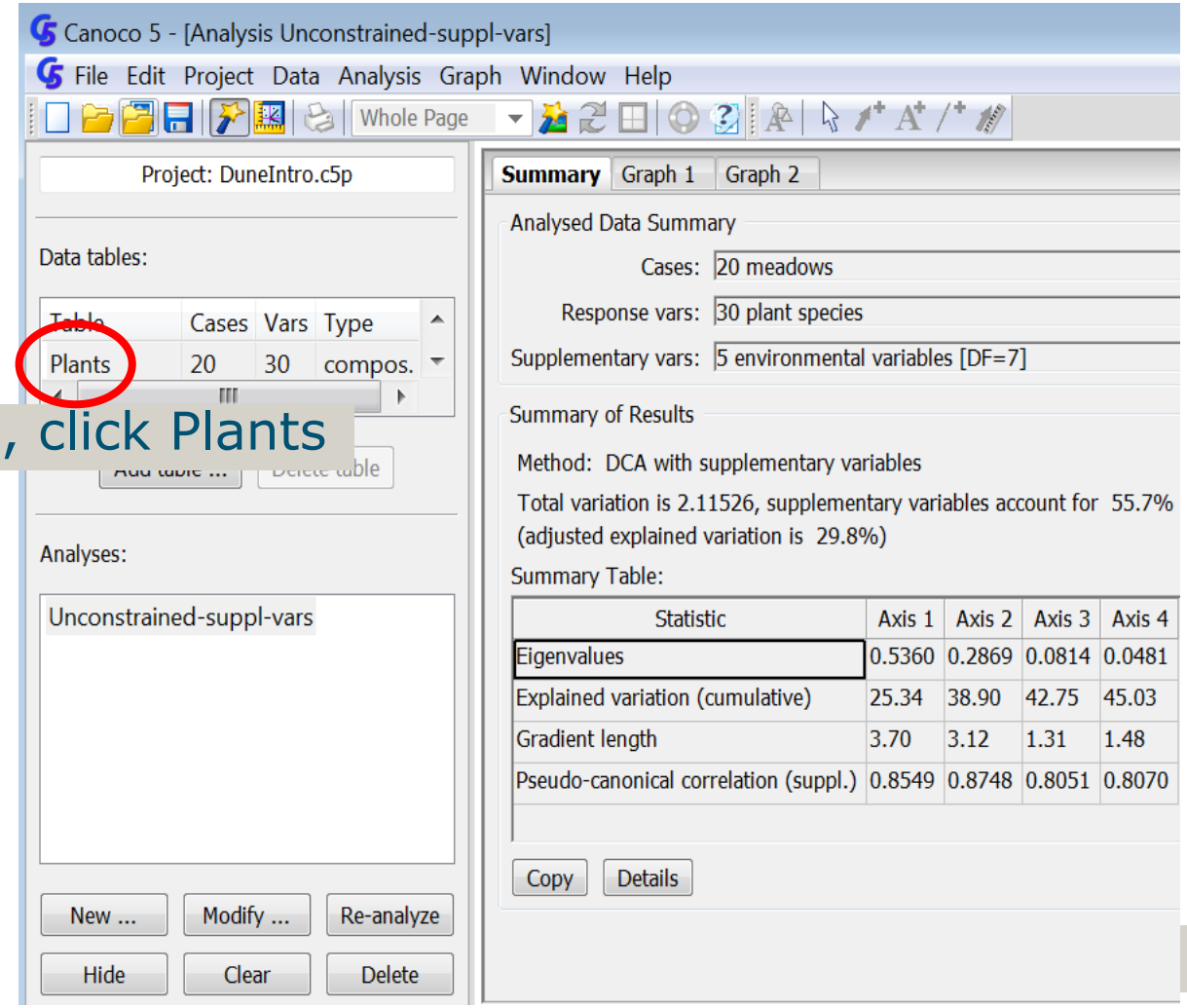
-Summary of DCA analysis

-Two graphs

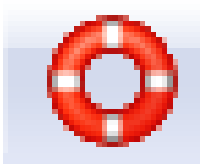
To view the data again, click Plants

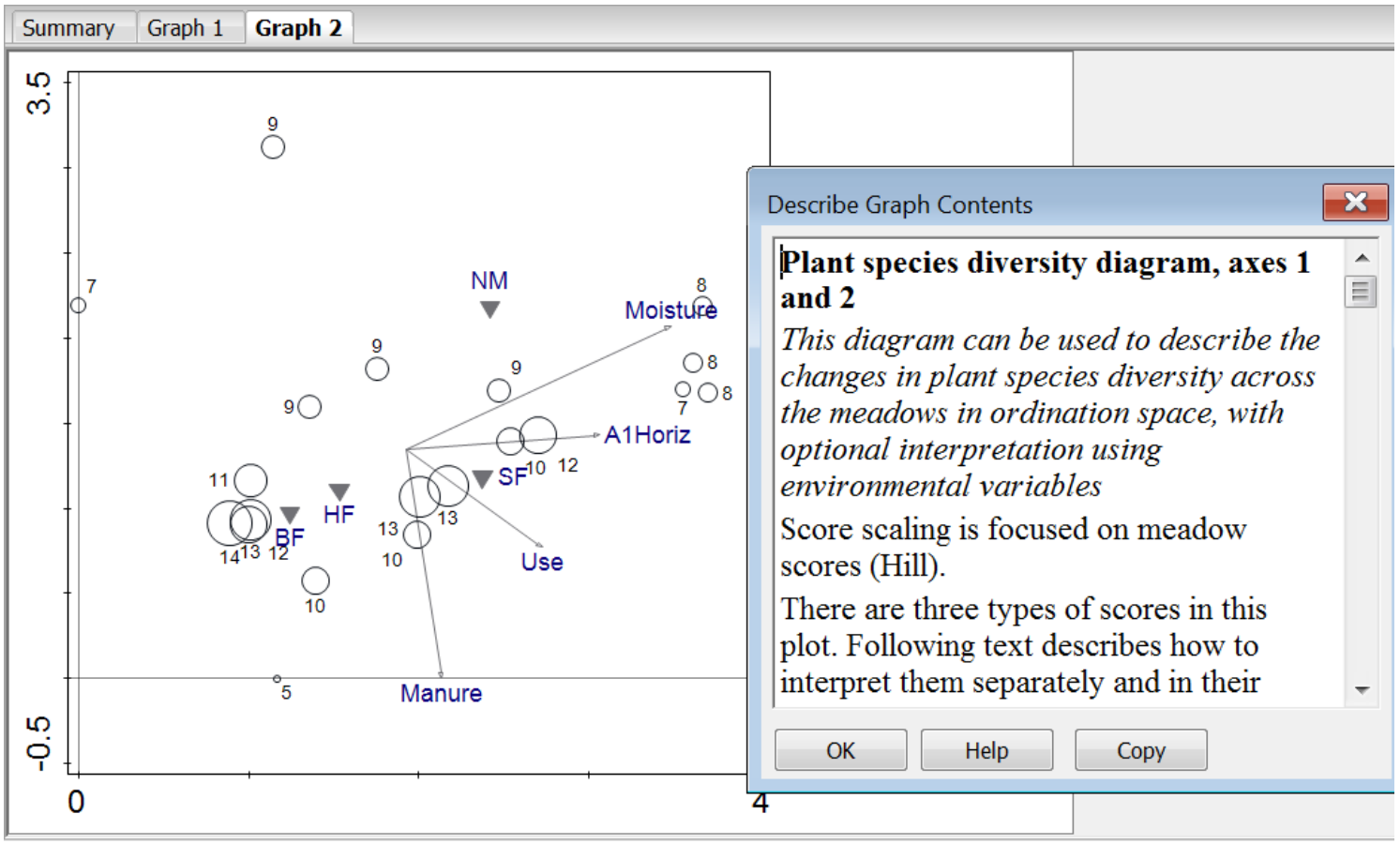
Save your project!

- File Save.. or
- Cntr-S



Inspecting a graph with Describe Contents





All scores are available too:

Edit | *Settings* | *Canoco5 Options:*

- Click  on the toolbar 

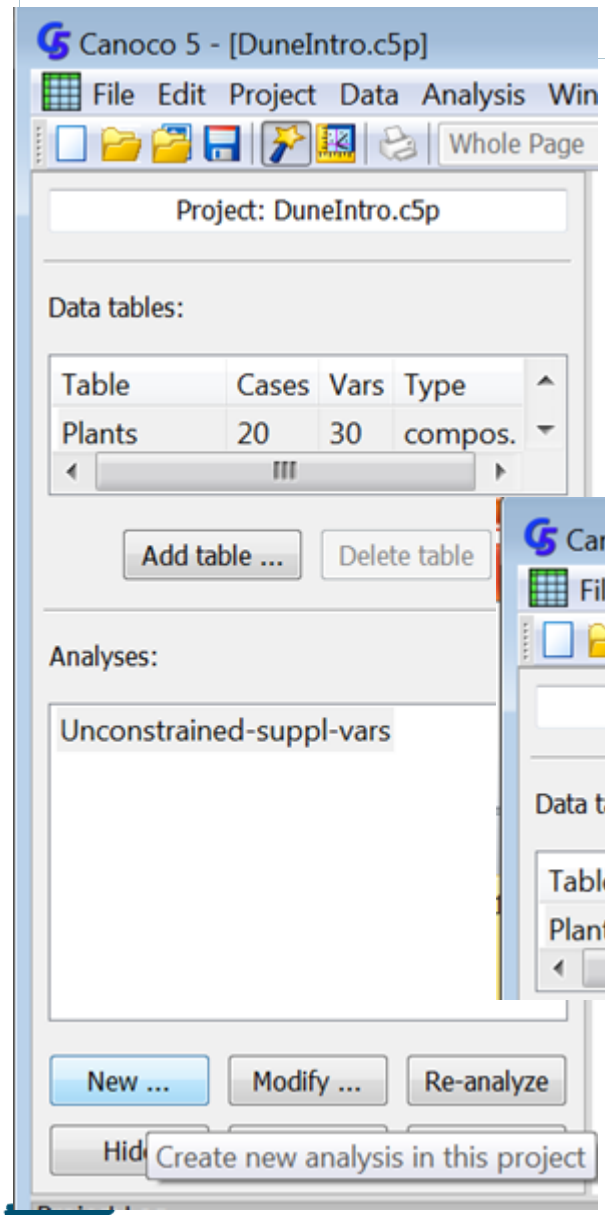
Summary | Log | Cases | **RespVars** | SupplVars | Graph 1 | Graph 2

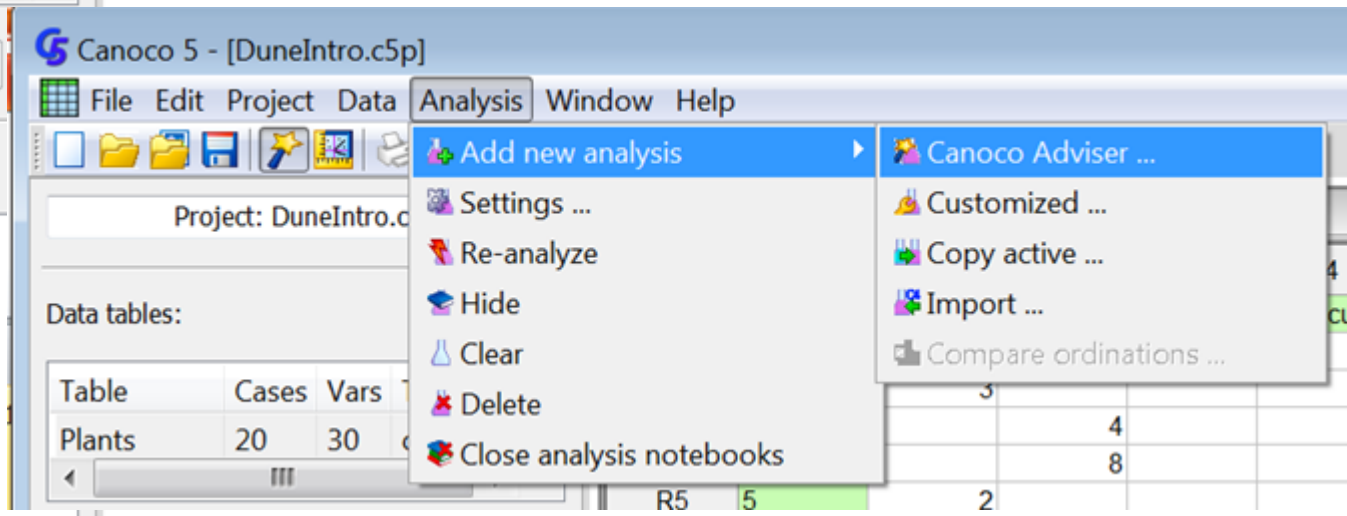
	Resp.1	Resp.2	Resp.3	Resp.4	RespW	RespN2	Tol.1	Tol.2	Tol.3	Tol.4	Tol4
Achillea millefolium	-0.3561	-0.8570	-0.6661	-0.1861	16.0000	6.0952	1.3641	1.9130	1.2816	0.7240	138.6271
Agrostis stolonifera	3.6										4.7858
Aira praecox	-0.7										4.4872
Alopecurus geniculatus	3.0										6.8240
Anthoxanthum odoratum	-0.5796	2.7109	-0.1809	0.0430	21.0000	5.7273	1.4568	1.4038	0.8880	0.5835	114.2609
Bellis perennis	1.0092	0.5036	-0.9428	-0.5007	13.0000	5.8276	0.6525	0.6176	1.4455	1.0899	101.0523
Bromus hordaceus	0.5673	-0.4497	-2.3301	0.2091	15.0000	4.5918	0.8303	1.3136	2.6773	0.3196	155.6025
Chenopodium album	3.0359	1.3324	-2.3487	-1.8633	1.0000	1.0000	0.4963	0.0526	2.8354	1.8633	171.4669
Cirsium arvense	2.1552	2.0161	-4.5330	2.2143	2.0000	1.0000	0.1464	0.9517	4.5330	1.7246	247.2326

Positions of response variables (e.g. species) on ordination axes. In linear methods represent regression coefficients, in unimodal methods represent the "species optima"

- Click again to hide the score tabs

Adding a new analysis to the project (1)





By :

- *New*... (under Analyses) or
- *Analysis* | *Add new analysis* | *Canoco Adviser*... (Alt-a-a-Enter)

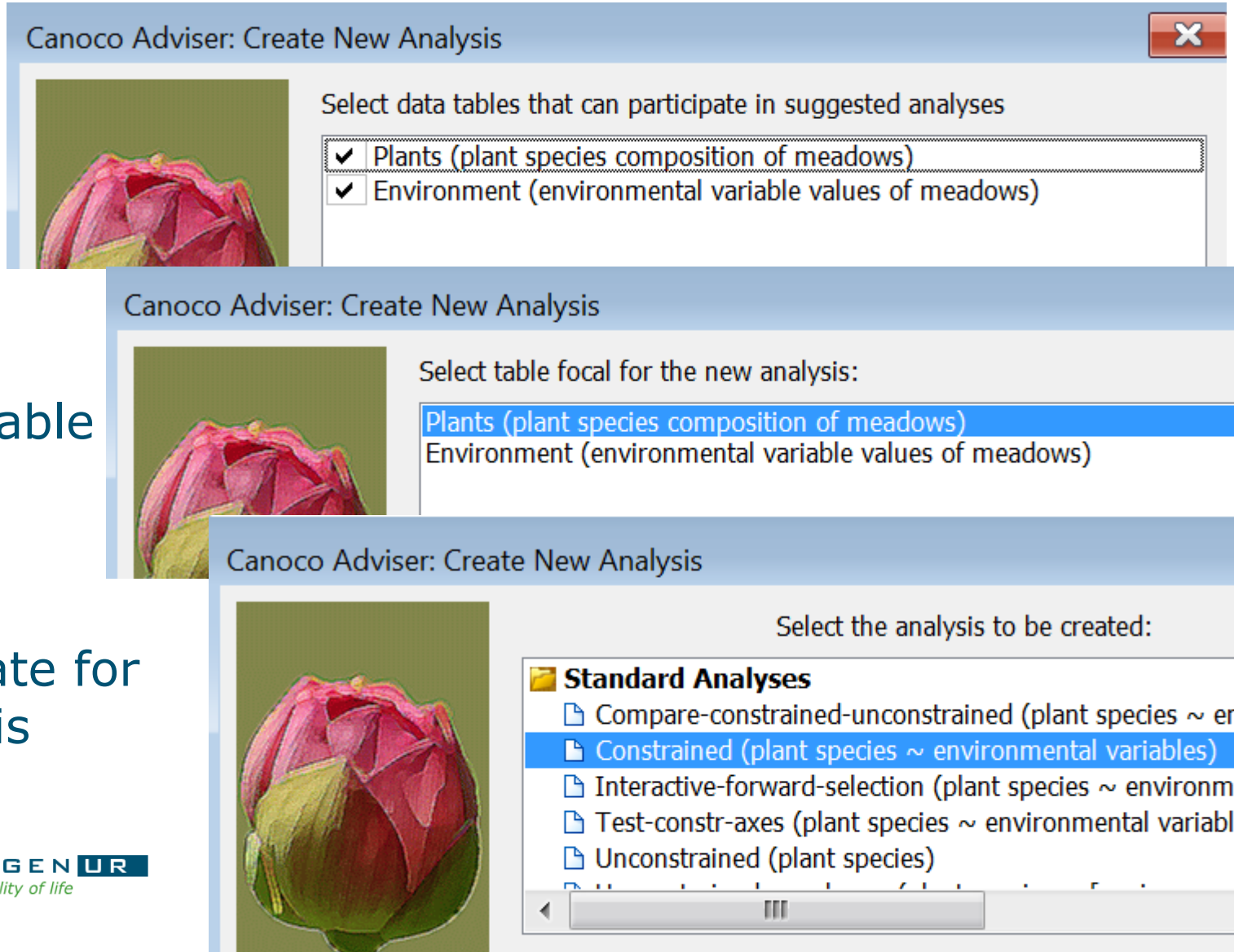
Adding a new analysis to the project (2)

Select:

1. Tables

2. Focal table

3. Template for analysis



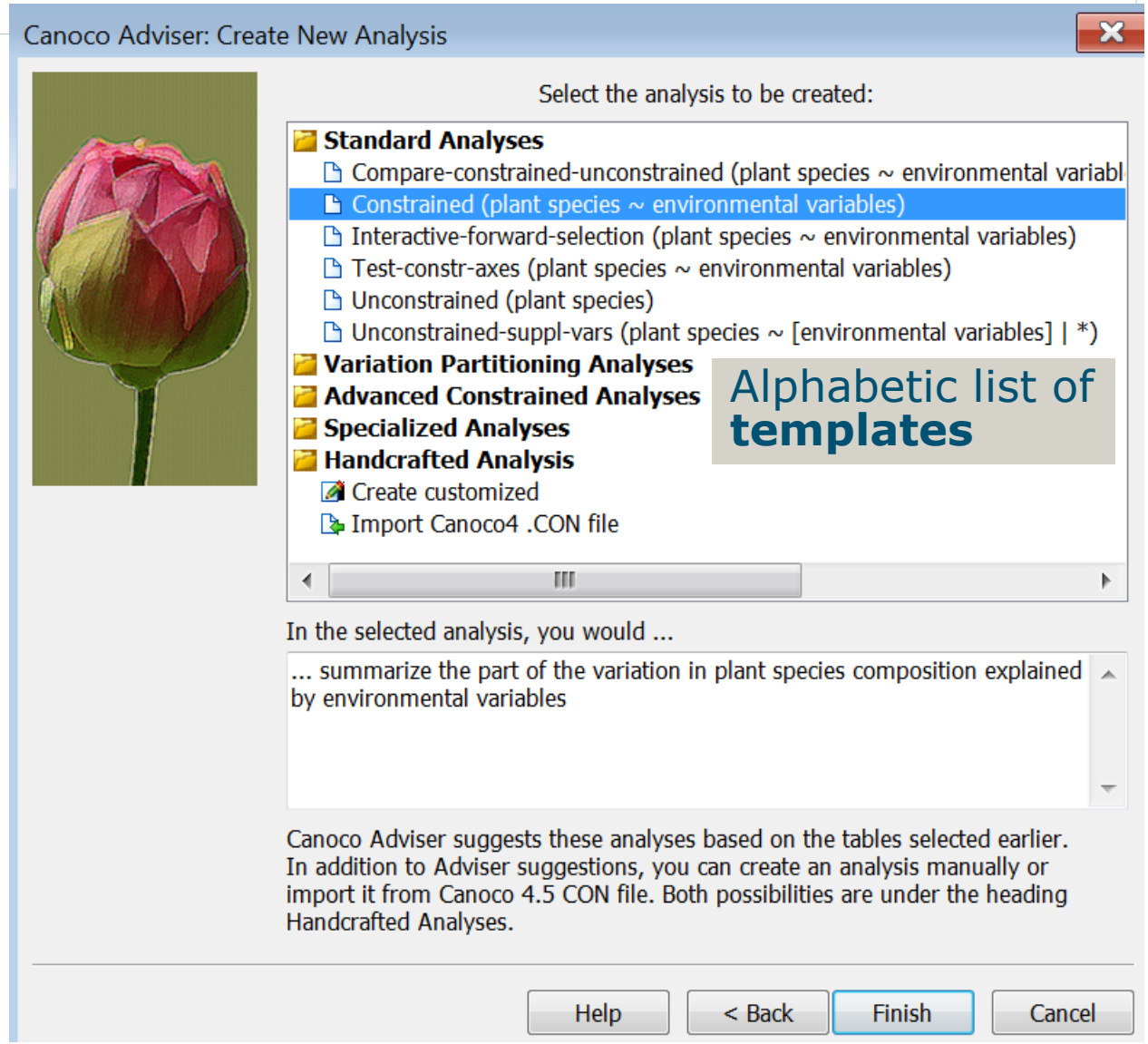
WAGENINGEN UR
For quality of life

Adding a new analysis to the project (3)

3. Select **template**

-double click on bold terms to fold/unfold

(Can enlarge dialog window to see all)



WAGENINGEN UR
For quality of life

Adding a new analysis to the project (4)

Standard analyses:

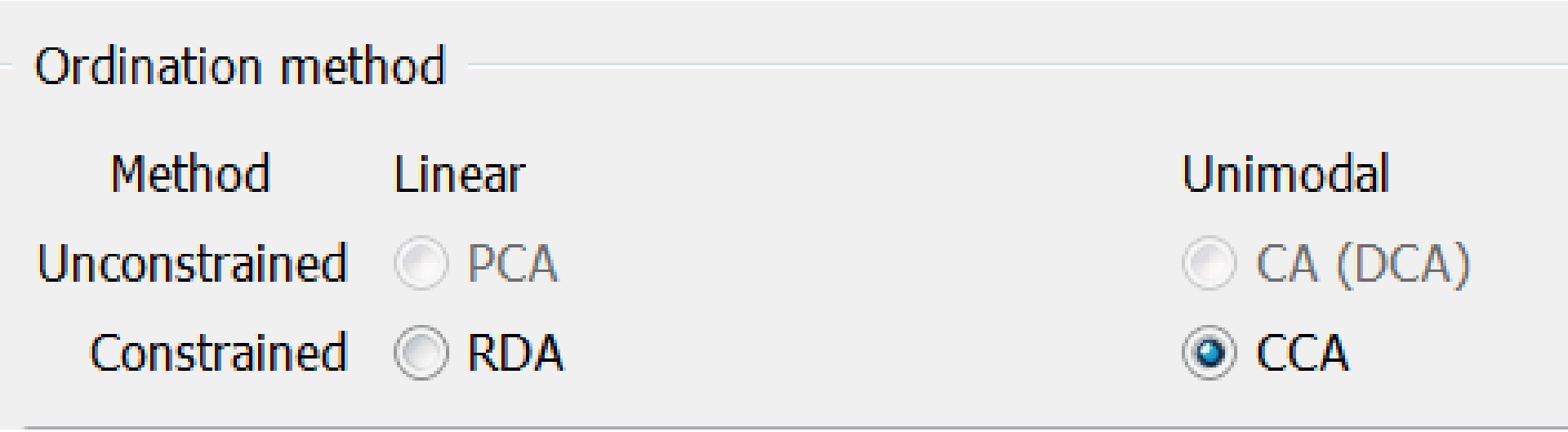
- Constrained: response variables ~ predictors
- Unconstrained: response variables

  response variables ~ [supplementary variables]
- Compare constrained – unconstrained
- Test constrained axes
- Interactive forward selection of predictors
  - See also: Summarize effects of expl variables

See Advanced ... for constrained analysis with covariates

Adding a new analysis to the project (5)

Standard analyses:



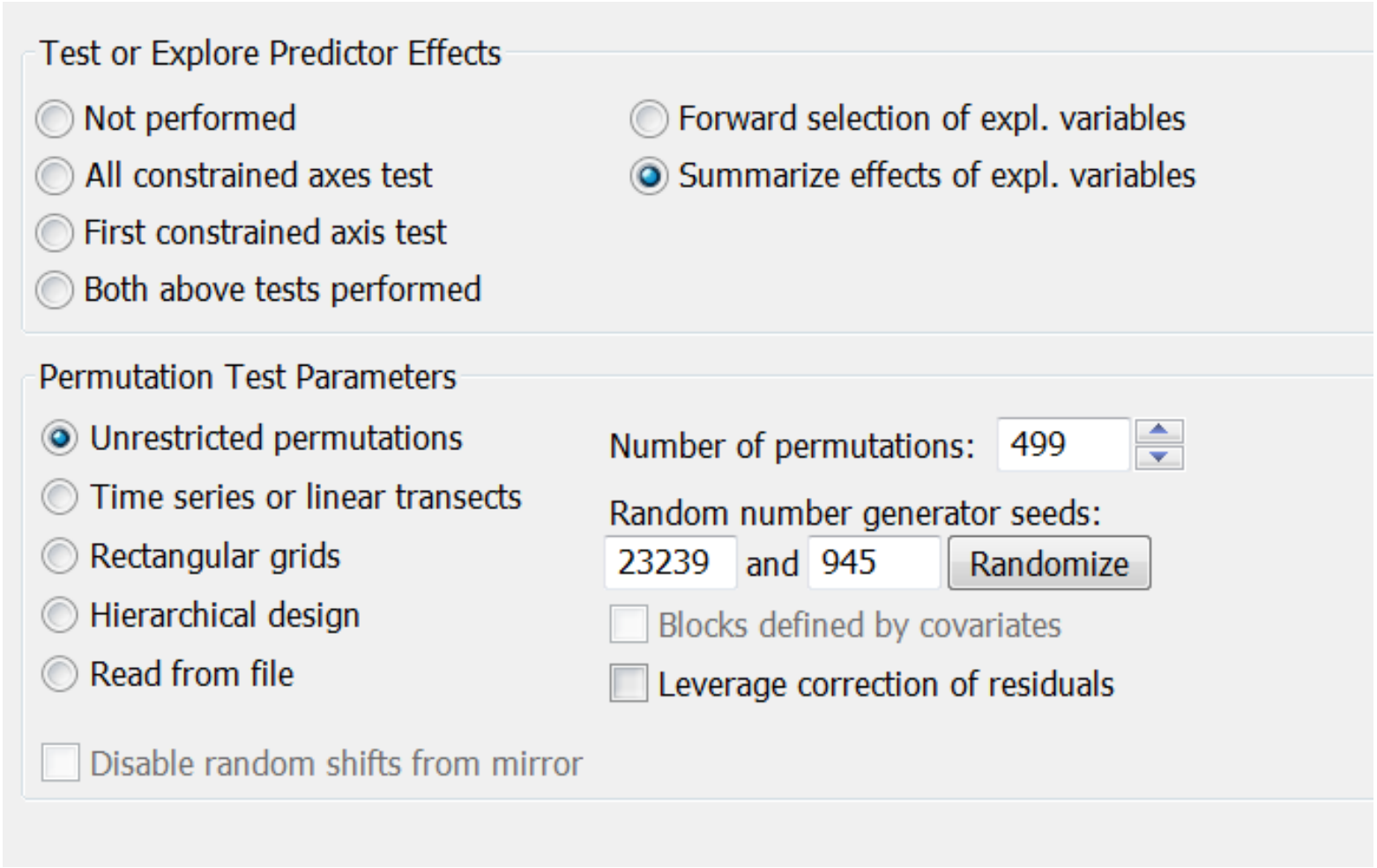
PCA: Principal component analysis

RDA: Redundancy analysis

CA (DCA): Correspondence analysis(Detrended)

CCA: Canonical correspondence analysis

Adding a new analysis to the project (6)



Summarize effects of expl. variables.

Dune meadow data

- Plant species ~ Environment (CCA)

Term Effects

P values correction: False discovery rate

Simple Effects

Name	Explains %	pseudo-F	P	P(adj)
Moisture	19.4	4.3	0.002	0.016
Management.NM	15.0	3.2	0.004	0.016
Manure Amount	11.3	2.3	0.016	0.04267

Copy

Conditional Effects

Name	Explains %	pseudo-F	P	P(adj)
Moisture	19.4	4.3	0.002	0.008
Management.NM	12.2	3.0	0.002	0.008
A1 horizont	6.7	1.7	0.06	0.16

Copy

Term Effects

P values correction: False discovery rate

none
False discovery rate
Holm correction
Bonferroni correction

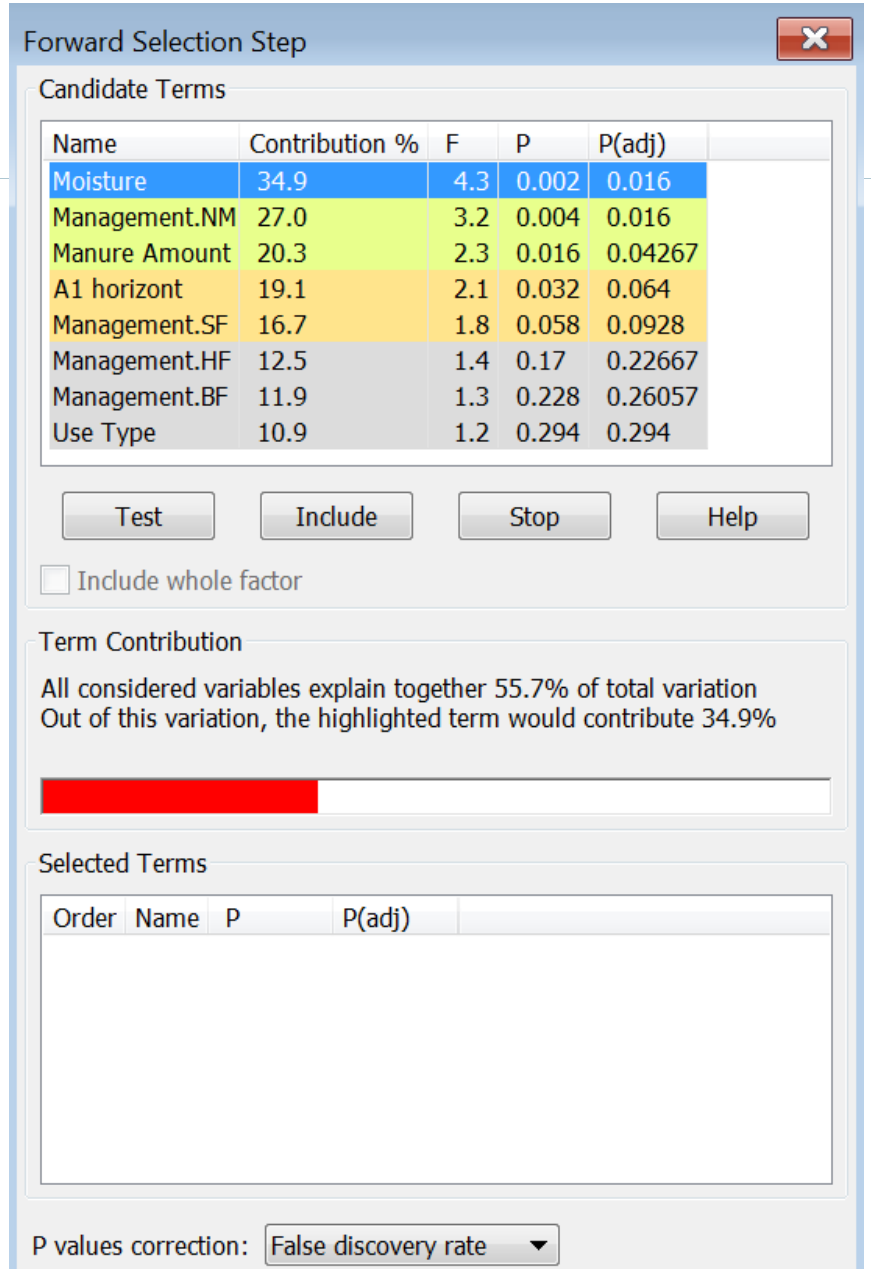
Simple Effects

Name Ex P

Moisture 19.4 4.3 0.002

Forward selection of expl. variables

- Color code for significance
- FDR testing on-line, but only for viewed variables
  - Tip: increase window size to get correct FDR



Canoco Adviser

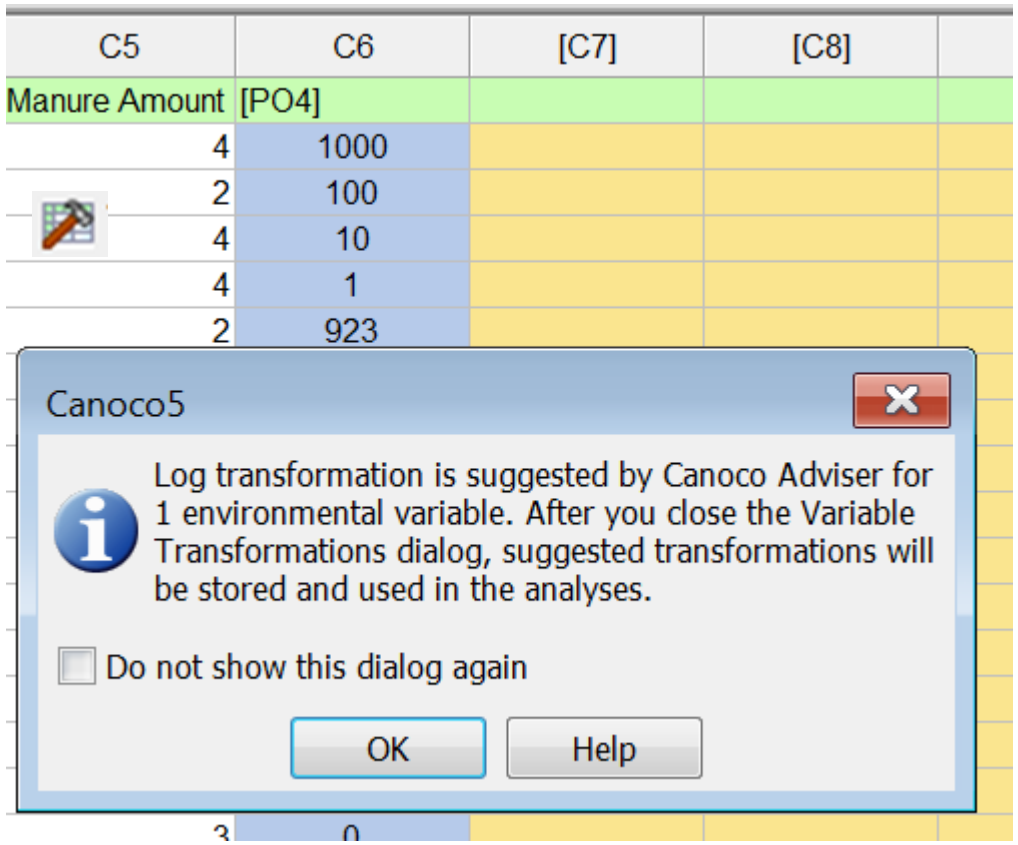
On the basis of the data properties the Adviser suggests

- Transformation and standardization of variables

right-click on top-left cell in data sheet

Or use

Data | Default transformation and ...



Canoco Adviser

On the basis of the data properties the Adviser suggests

-  Transformation and standardization of variables
- Common analyses via templates
- Choice between Linear and Unimodal

Ordination method

Method	Linear	Unimodal
Unconstrained	○ PCA	○ CA (DCA)
Constrained	○ RDA	◉ CCA

Response data are compositional and have a gradient 3.7 SD units long, so unimodal method is suggested, but linear one would be also a good choice

Re-advise

Methods in Canoco 5

- Standard multivariate methods such as [partial] PCA/RDA/CA/DCA/CCA

But also:

- Response curves (regression by GLM/GAMs with one predictor)
- Contour plots (GLM/GAM with two predictors)
- Distance-based methods incl. nonmetric multidimensional scaling (NMDS)
- Variation partitioning
- Principal response curves (PRC)
- Generalized linear models (GLM) with permutation tests
- Double-constrained analyses (dc-CA and dc-PCA)

and more...

Nonlinear response curves via GLM or GAM

Expected number of 12 spider species plotted against a synthetic explanatory variable

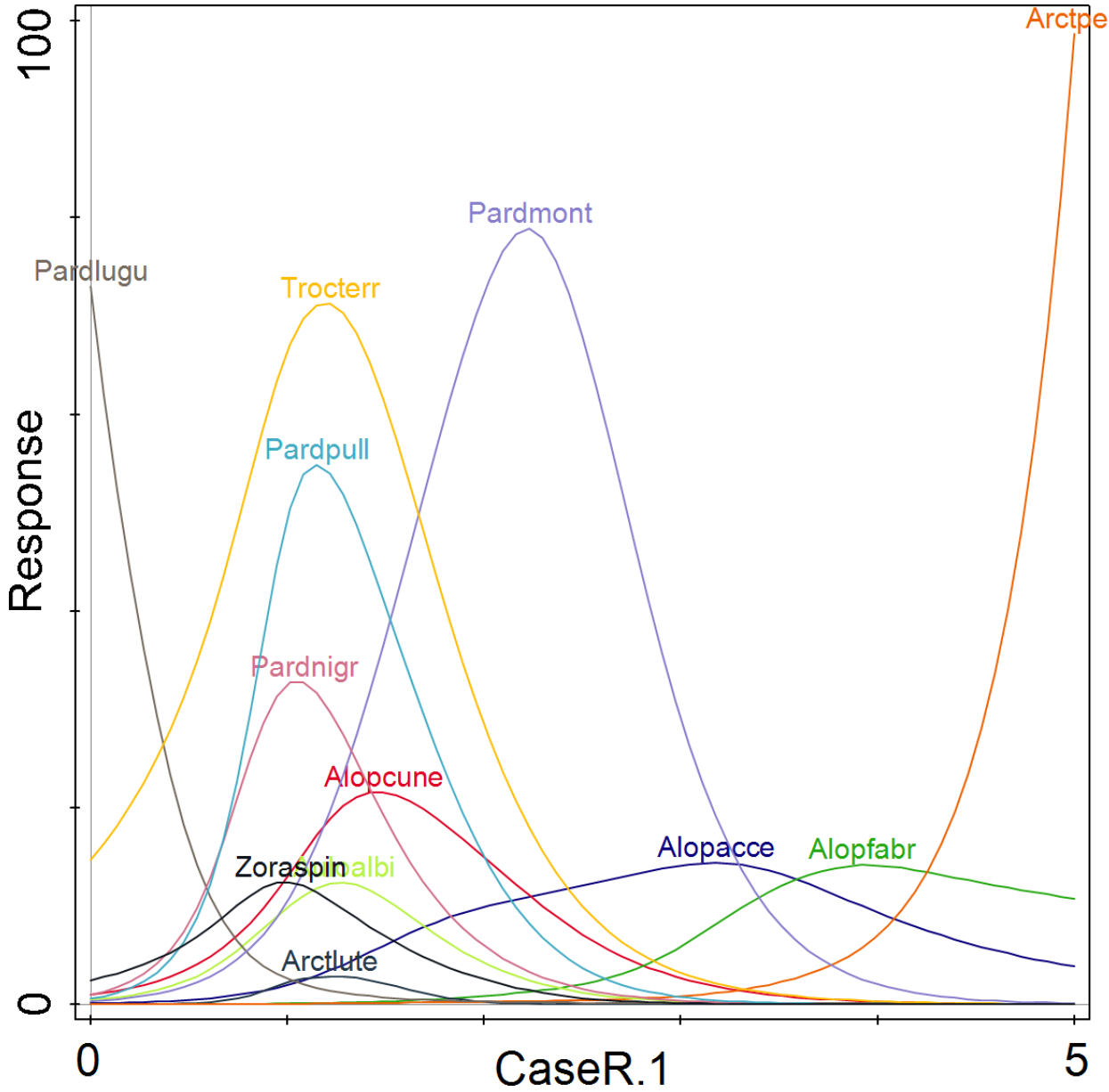
Fitted Generalized additive models

Predictors	CaseR.1
Distribution	quasi-Poisson
Link function	log

GAM fitted for 12 response variables:

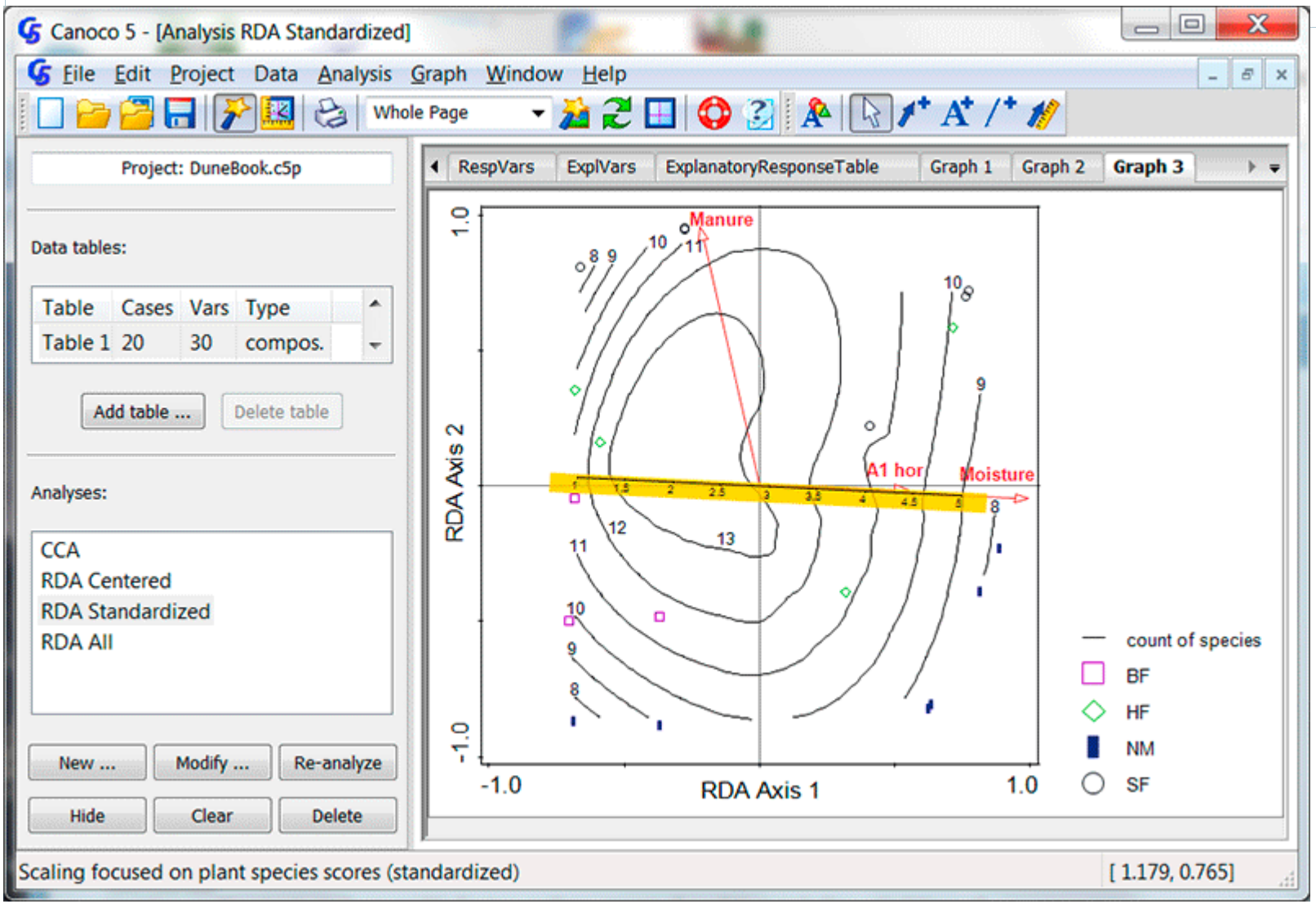
Response	T...	R2[%]	F	p
Alopacce	s3	63.7	53.0	<0.0000
Alopcune	s3	52.2	35.7	<0.0000
Alopfabr	s3	84.2	130.1	<0.0000

OK Help Copy



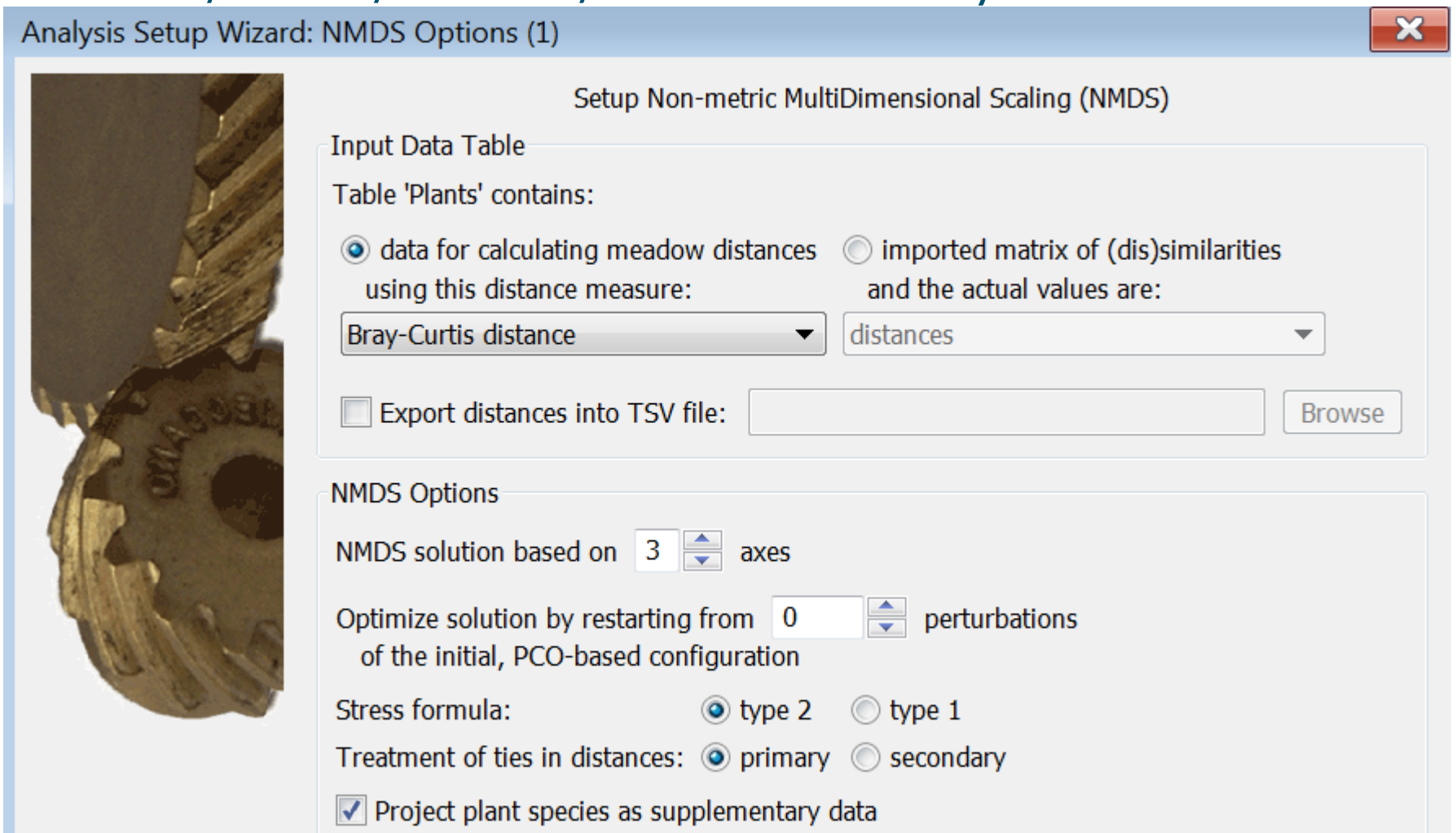
WAGENINGEN UR
For quality of life

GAMs or GLMs with two predictors



Distance-based methods eg. NMDS

- E.g. from intercity train-time to a map of cities
- PCO/NMDS/db-RDA/Procrustes analysis



Variation partitioning

Which part of variation is due to

(a) Environment

and which to

(b) Management

and which part is

(c) shared?

two or three groups of variables

Partitioning

Variation Partitioning Results for Two Groups

Variation Explained

Fraction	Variation(adj)	% of Explained	% of All	DF	Mean Square
a	0.30775	50.2	14.5	2	0.20054
b	0.26405	43.1	12.5	4	0.13812
c	0.041408	6.8	2.0	--	--
Total Explained	0.61321	100.0	29.0	6	0.18126
All Variation	2.1153	--	100.0	19	--

Copy

Significance Tests

Tested Fraction	F	P
a+b+c	2.3	0.001
a	2.5	0.004
b	1.7	0.002

Copy

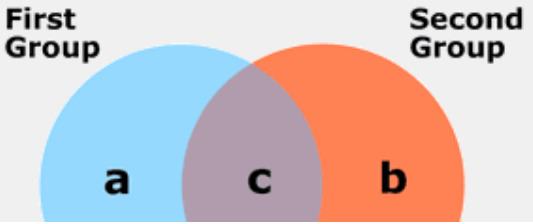
Group Members

First Group: A1 horizont

Second Group: Management

Copy Members

Compared groups are represented by circles in the following diagram. Lower case letters label individual estimated fractions and are referred in the summary tables above

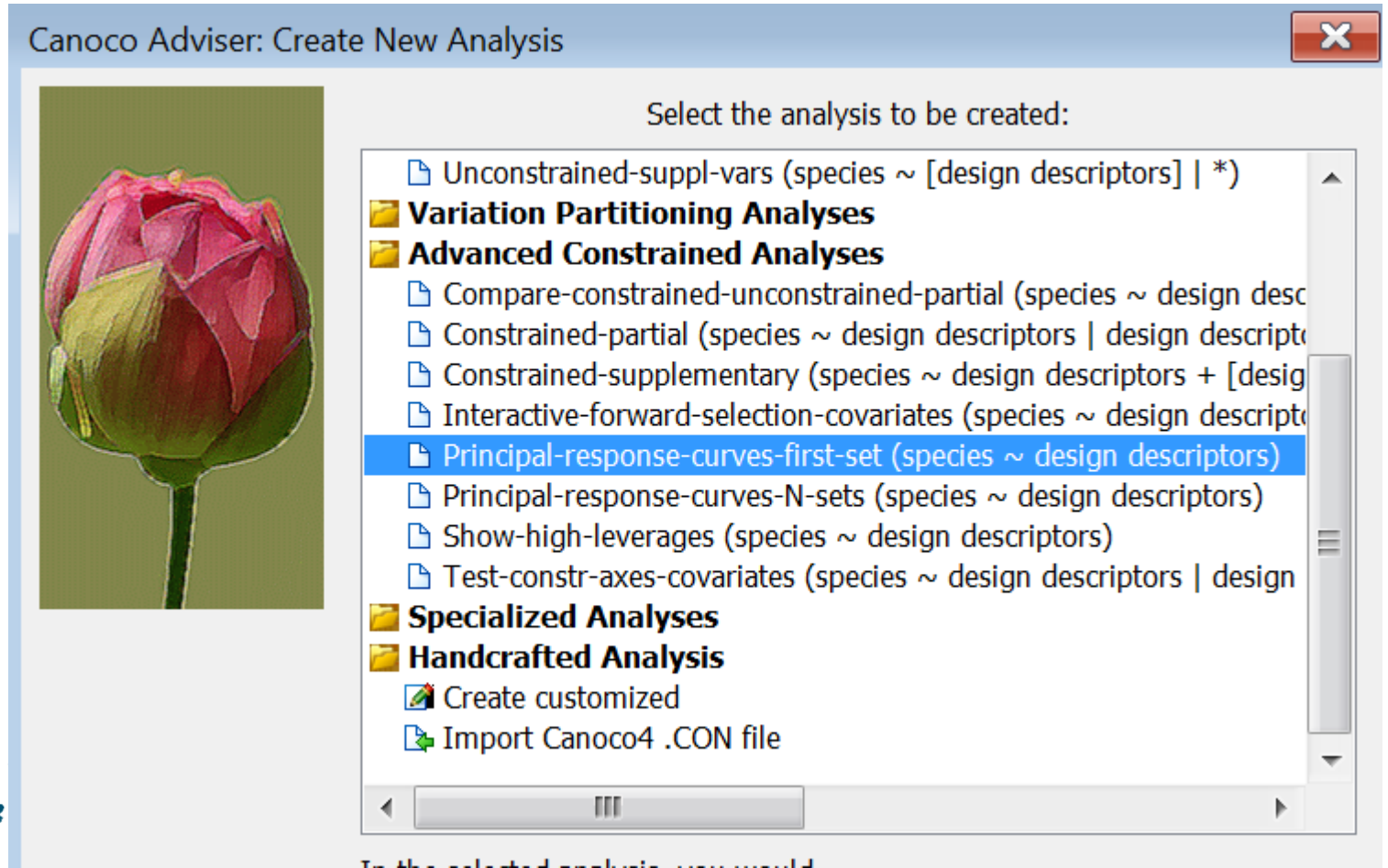


WAGENINGEN UR
For quality of life

Principal response curves (PRC)

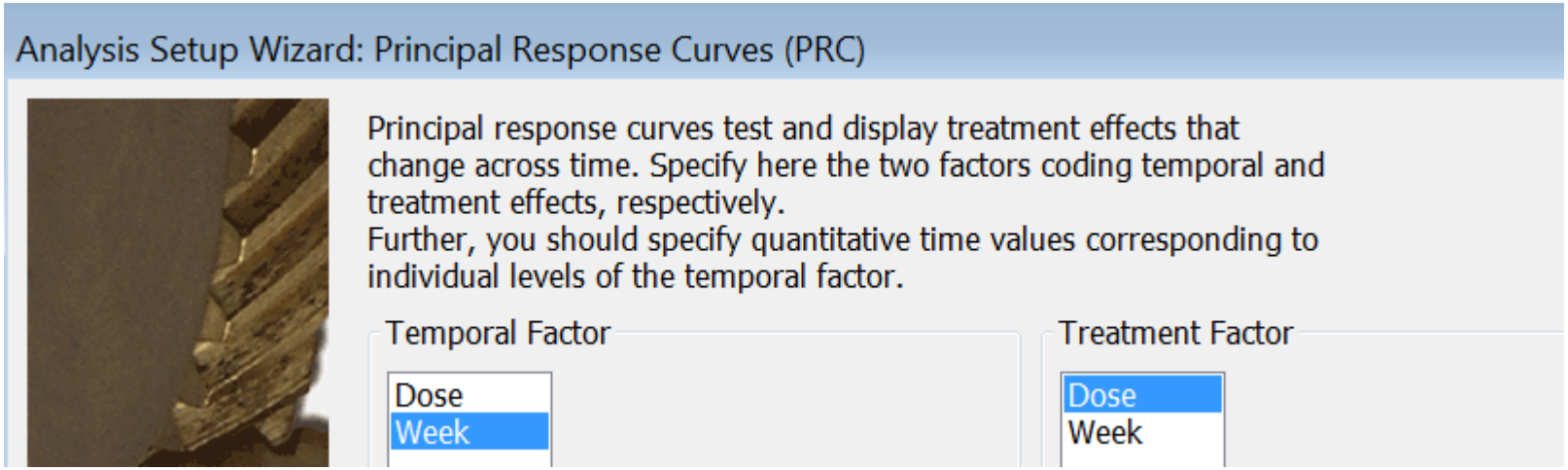
Method to show main effect + interaction

"How does the effect of a pesticide change over time?

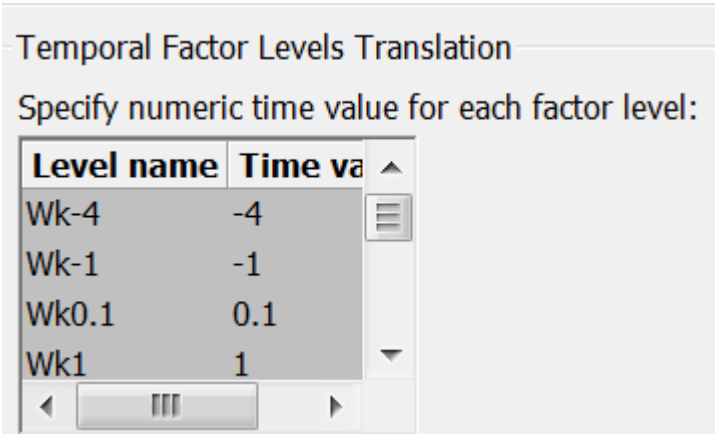


Principal response curves (PRC)(2)

- Specify Time and Treatment factors



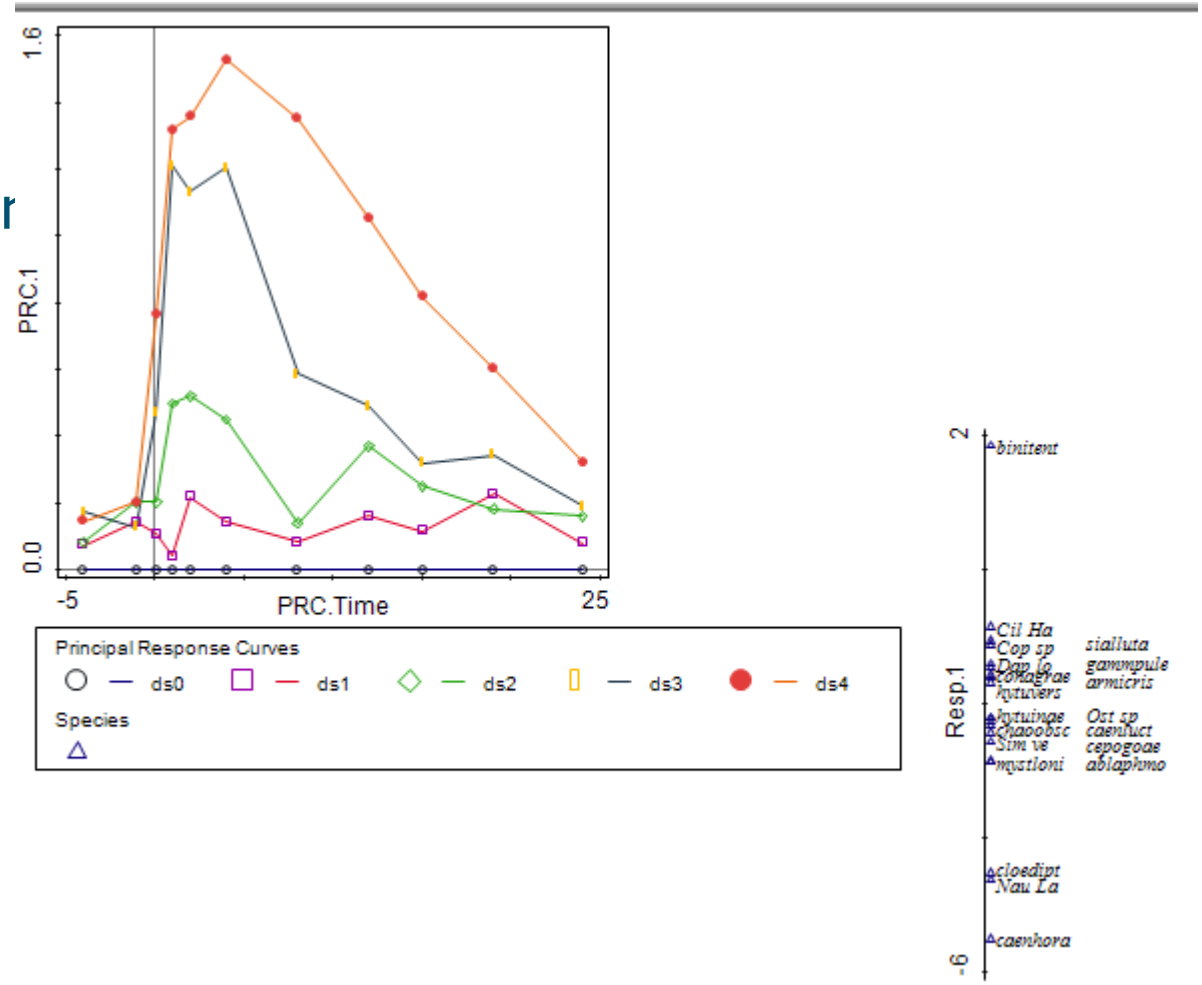
- Specify time values for horizontal axis (default often good)



PRC diagram: Invertebrates~ treatment.time | time

- Example

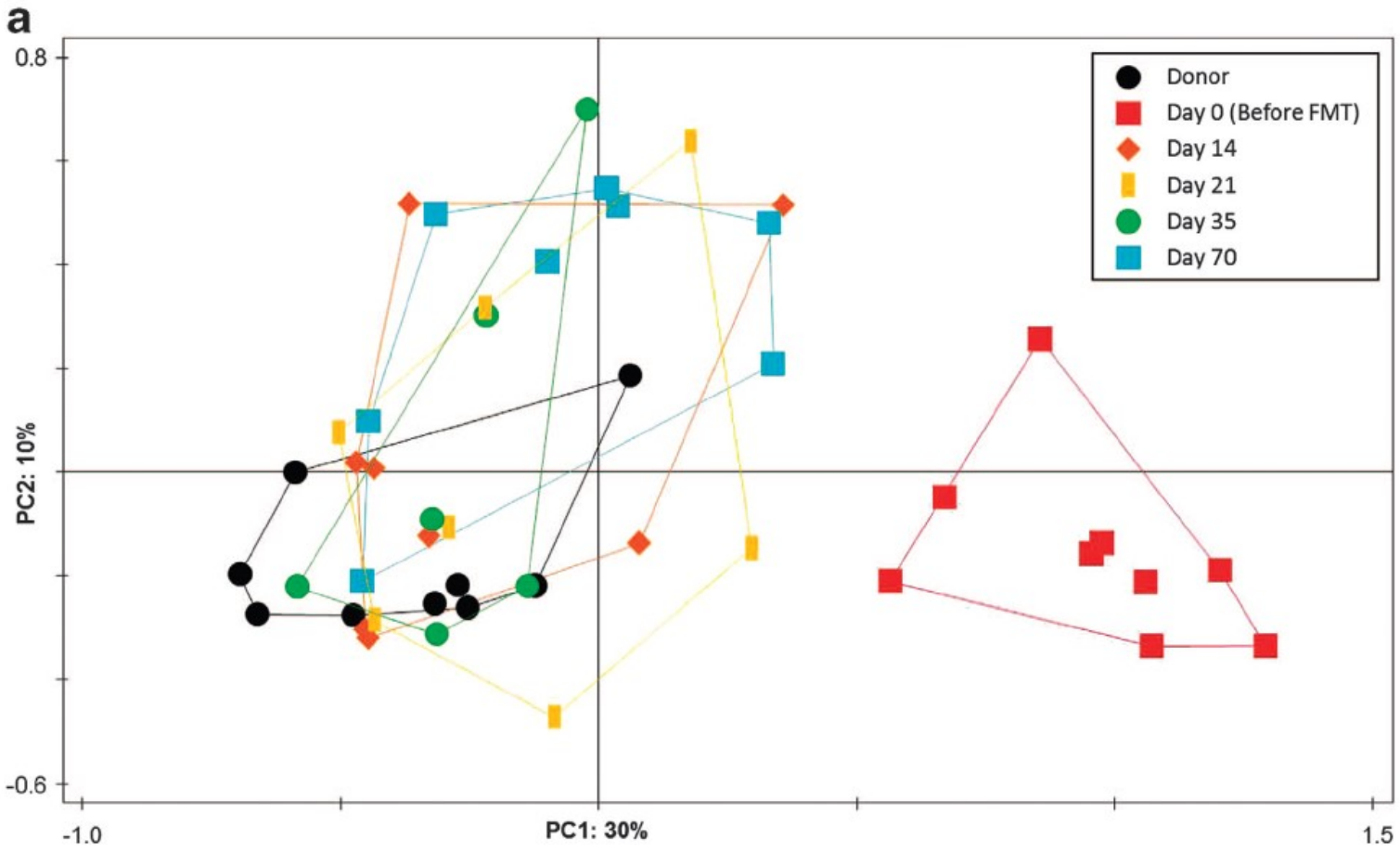
Van den Brink & ter Braak (1999)



Graph 1 in Canoco5\Samples\Advanced\PRC.c5p

Canoco in faecal transplant study (1)

Development of microbiota in patients: PCA

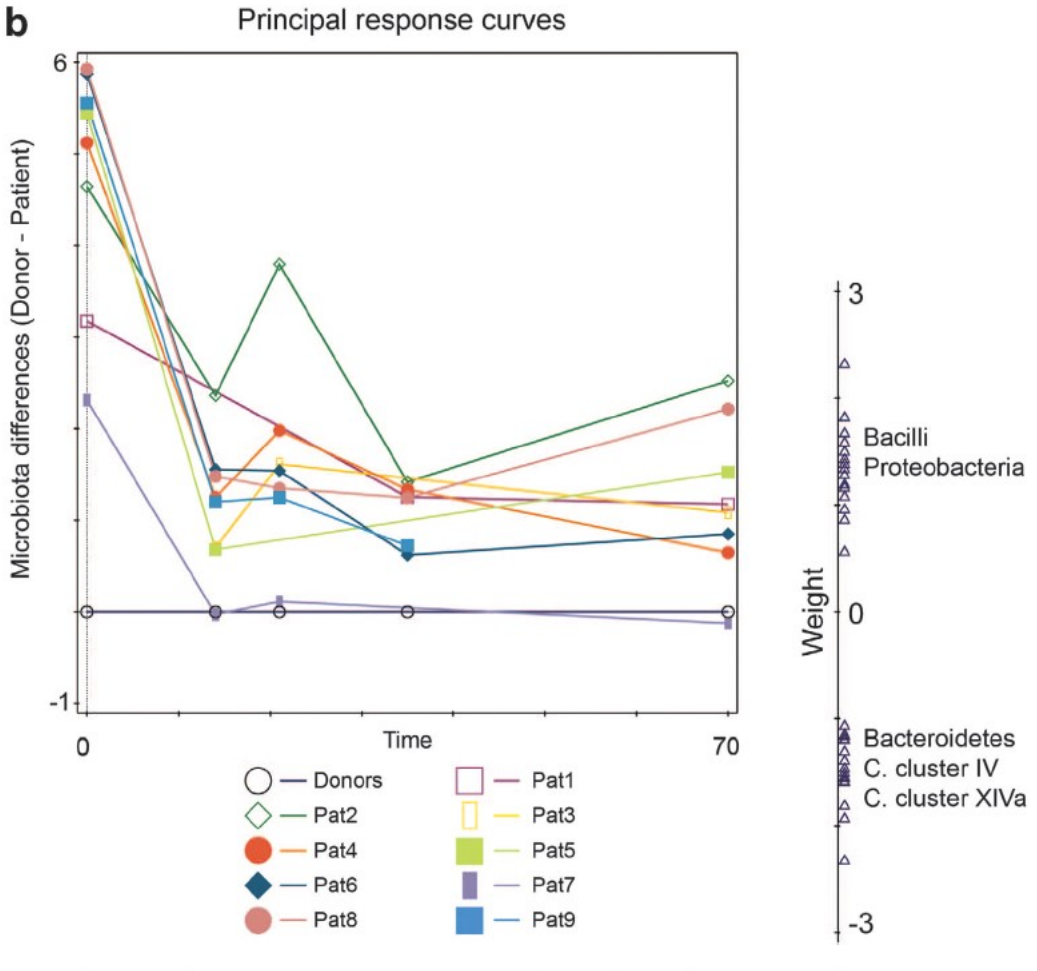


WAGENINGEN UR
For quality of life

Feces transplant study: Fuentes et al 214 ISME J.
Data: **microbiota taxa** (Susana Fuentes, W. de Vos)

Canoco in faecal transplant study (2)

Development of microbiota in patients



Feces transplant study: Fuentes et al 214 ISME J.
Data: **microbiota taxa** (Susana Fuentes, W. de Vos)

WAGENINGEN UR
For quality of life

Co-correspondence analysis

- How are two compositional data tables related?

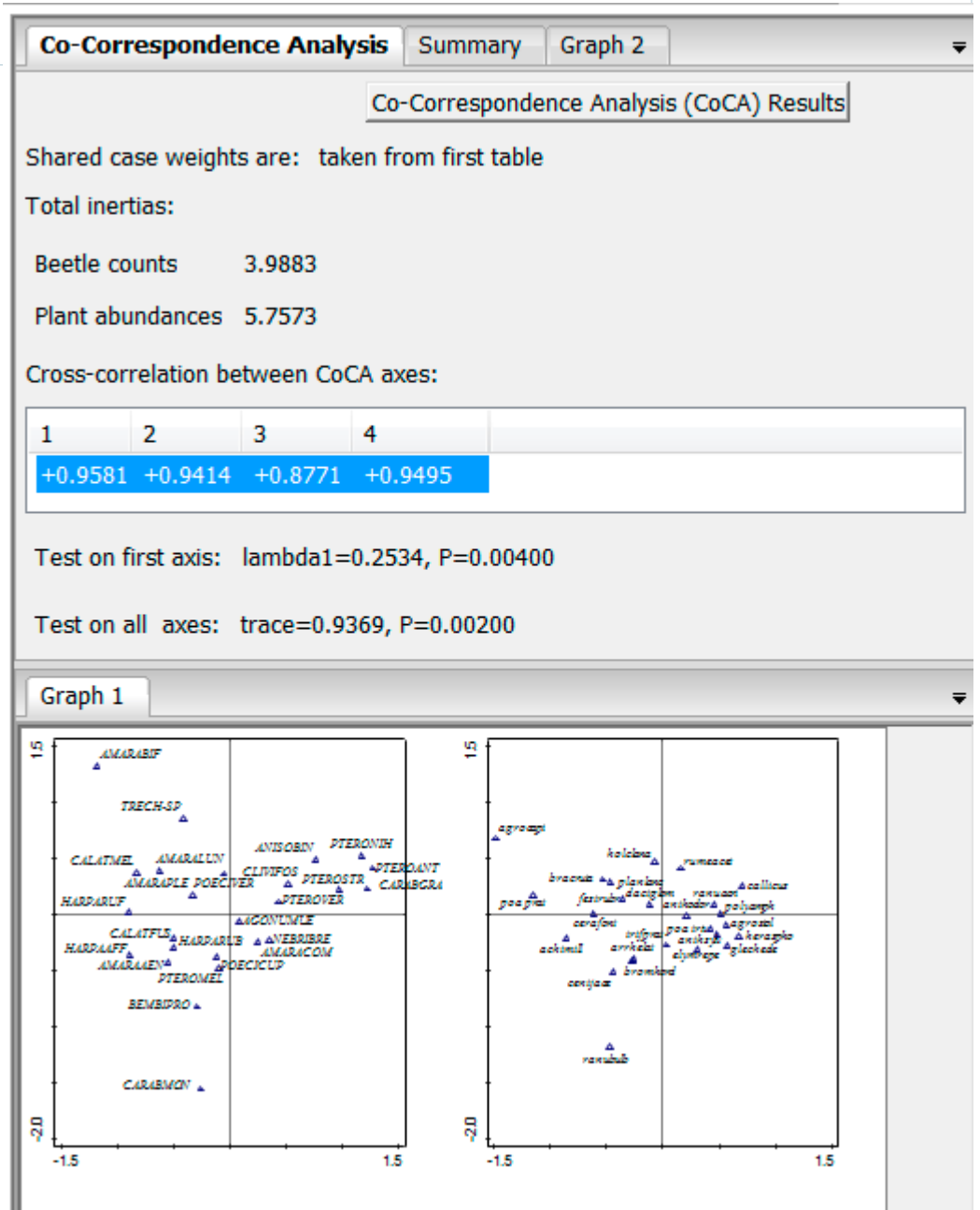
e.g.

plant and beetle communities

(Schaffers et al. 2008)

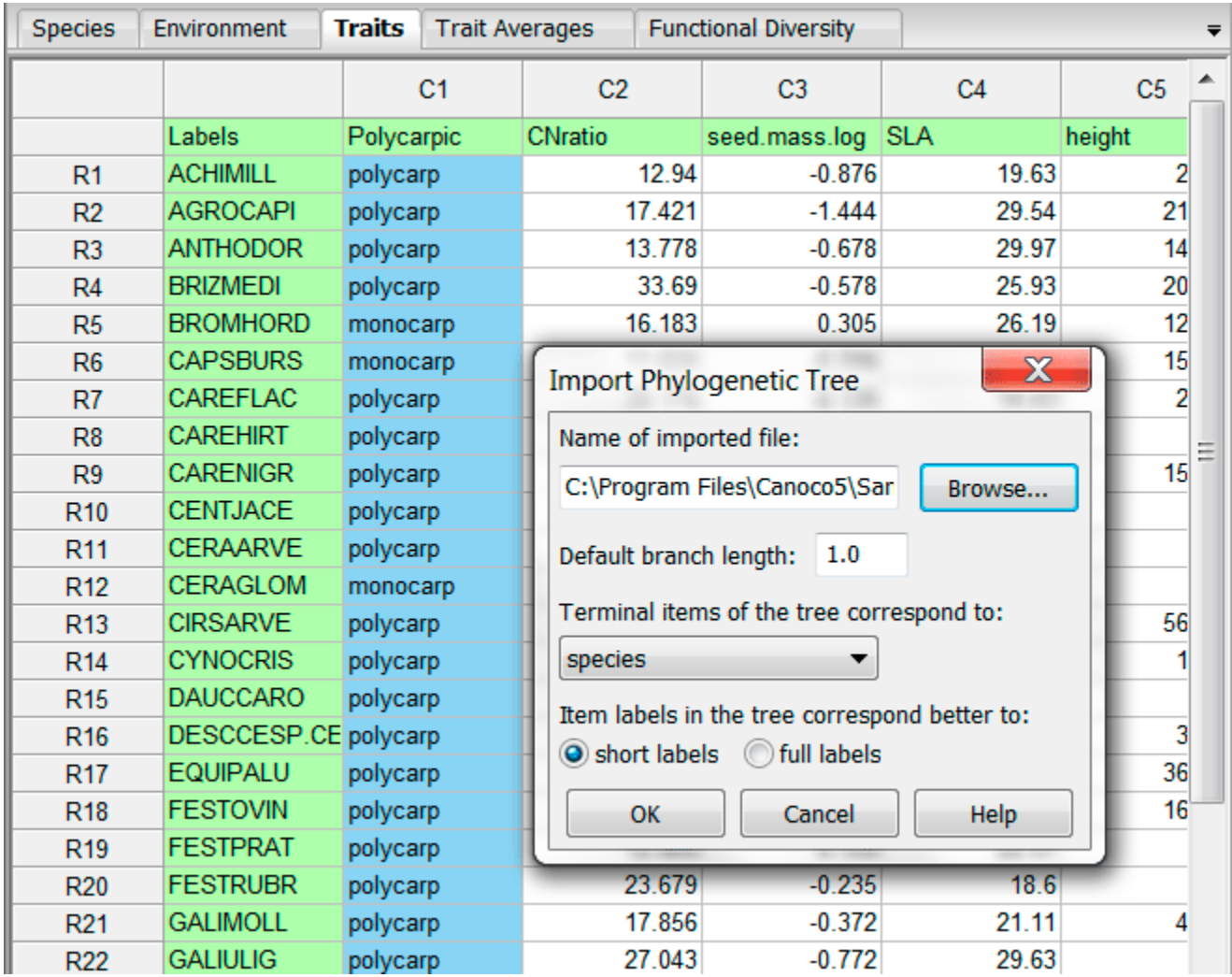
auto- and hetero-trophic microbial assemblages

(Alric et al. 2018, Mol Ecol Res)



Trait-based analyses and phylogenetic relations

- Trait averages
- Functional diversity
- Double-constrained correspondence analysis (dc-CA)
- Phylogenetic corrections



		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				
R9	CARENIGR	polycarp				15
R10	CENTJACE	polycarp				
R11	CERAARVE	polycarp				
R12	CERAGLOM	monocarp				
R13	CIRSARVE	polycarp				56
R14	CYNOCRIS	polycarp				1
R15	DAUCCARO	polycarp				
R16	DESCCESP.CE	polycarp				3
R17	EQUIPALU	polycarp				36
R18	FESTOVIN	polycarp				16
R19	FESTPRAT	polycarp				
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	

WAGENINGEN UR
For quality of life

dc-CA: ter Braak et al. 2018 http://rdcu.be/ETPh Peng et al. 2021 https://doi.org/10.1016/j.scitotenv.2020.142171

Generalized linear models (GLM)

Via

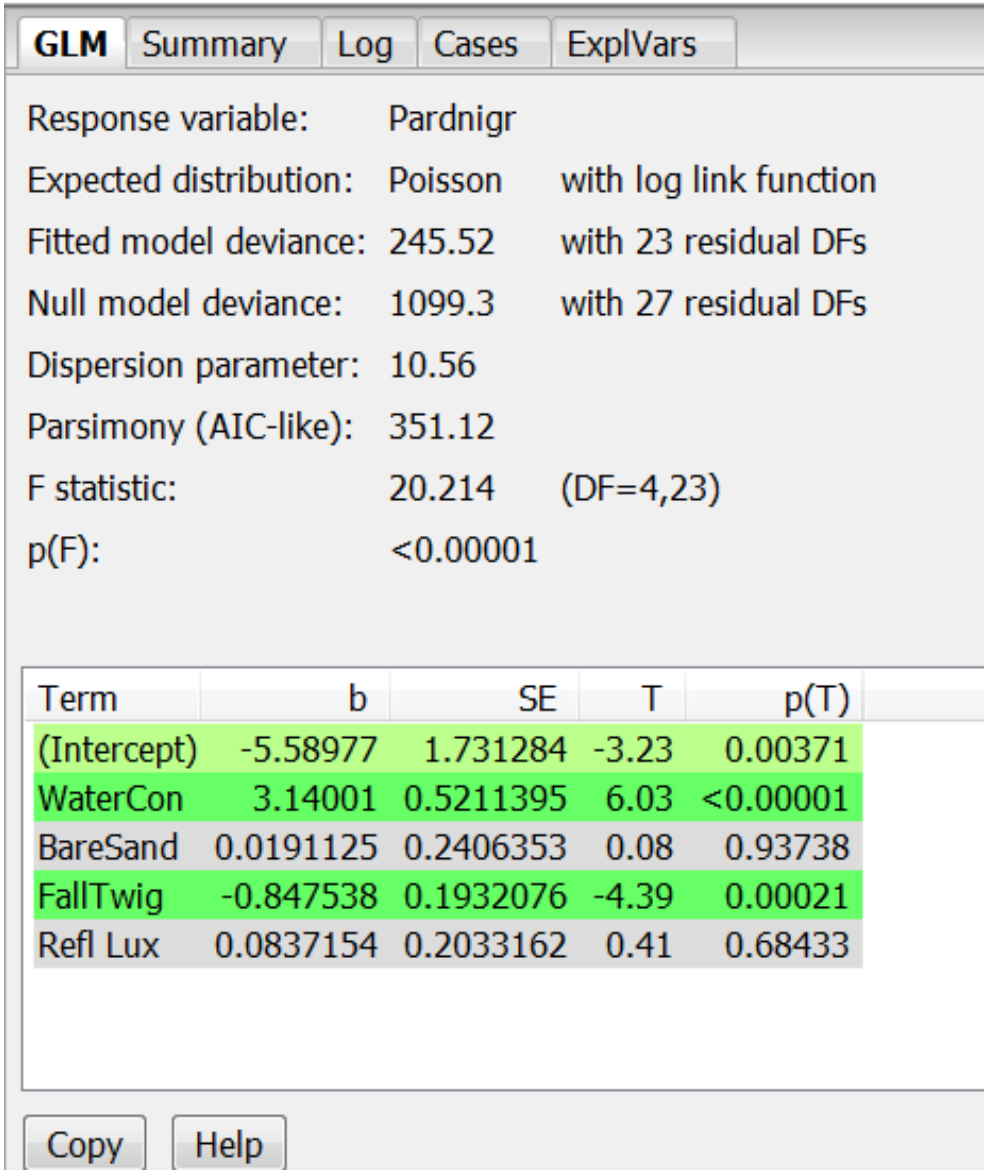
- GLM template for ≥ 1 predictors
- Graph| Attribute plots

1 predictor:

- Multiple response curves in single graph

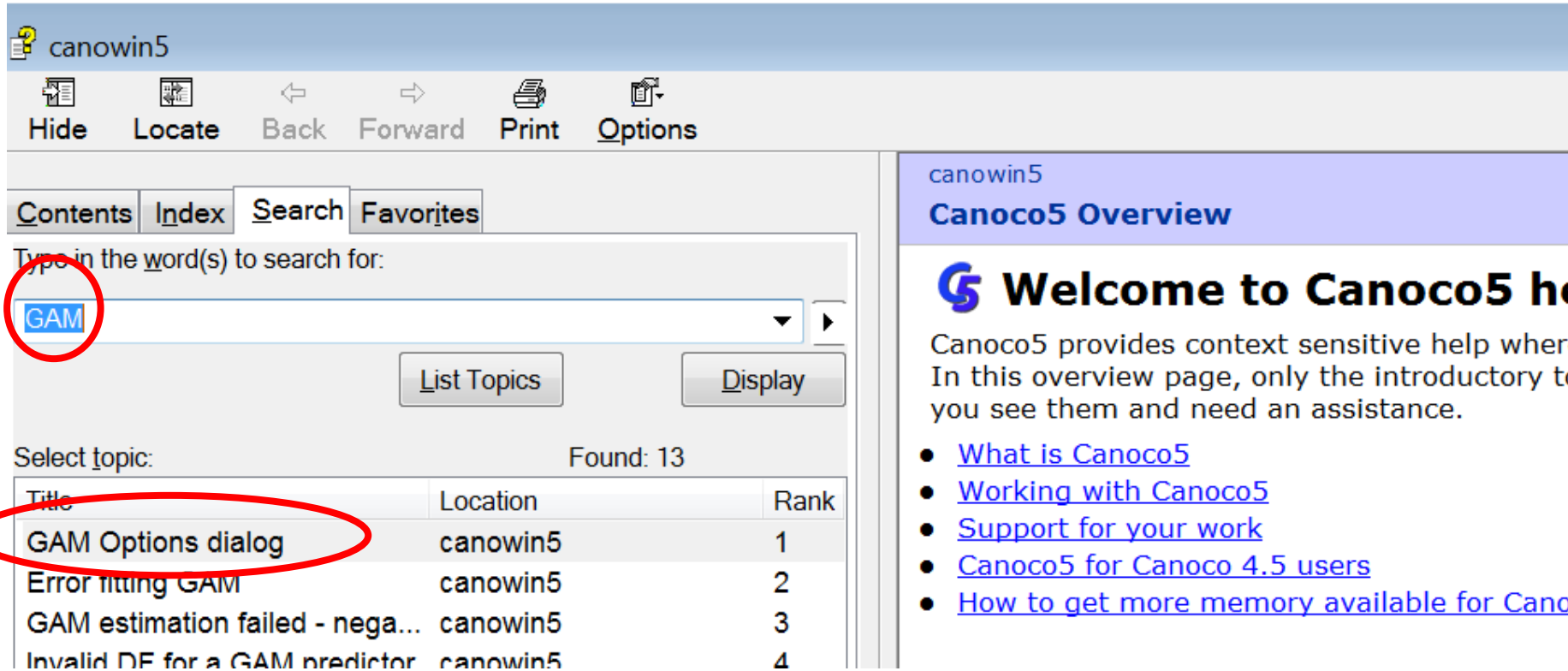
2 predictors:

- Contour plot



Find out how to get a method,eg. GAM (1)

- Help|Help contents (Alt-h-h) opens the help system
- Type GAM in search field, press Enter, gives



Find out how to get a method, eg. GAM (2)

Look in manual or use on-line help as follows:

- Help|Help contents (Alt-h-h) opens the help system
- Type GAM in search field, press Enter
- Click *GAM options dialog*
- Scroll down in the help page to find

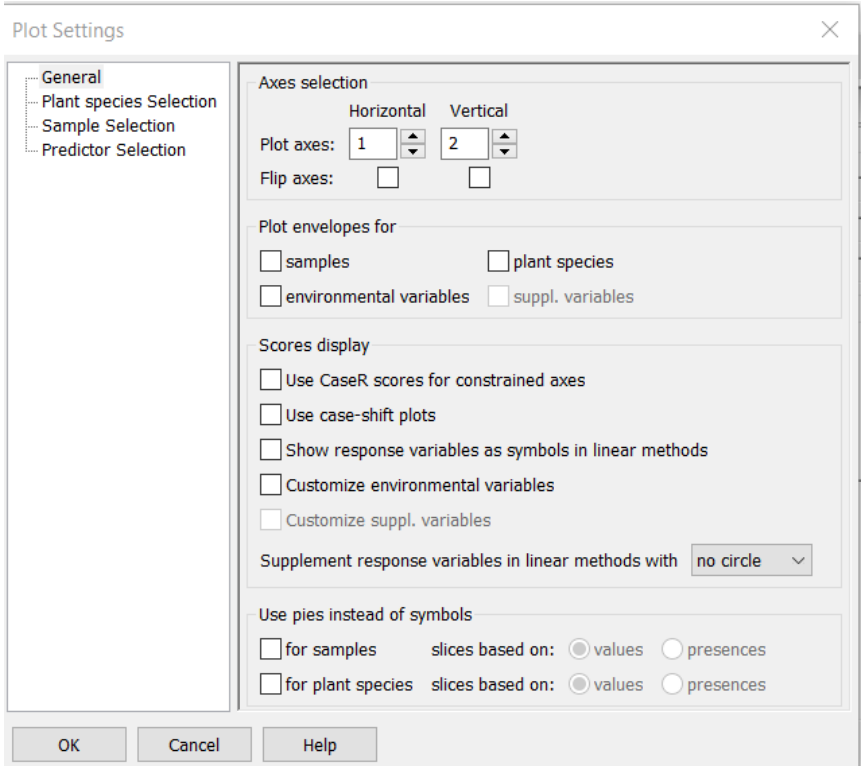
where it says:

Use one of the commands in *Graph / Attribute plots* submenu (use the *Model Options* button)

Type: response curves → topic Response curves plot → Getting Here: use *Graph / Attribute plots / <Col-term> response curves*

Options for ordination graphs

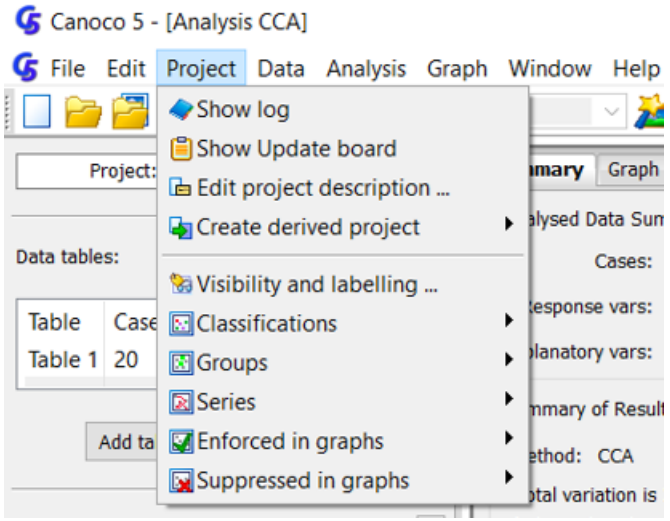
- In Canoco, graphs belong to an analysis
- Options for graphs can therefore be found under Analysis
- Click: Analysis|Plot creation options..



Classification and groups

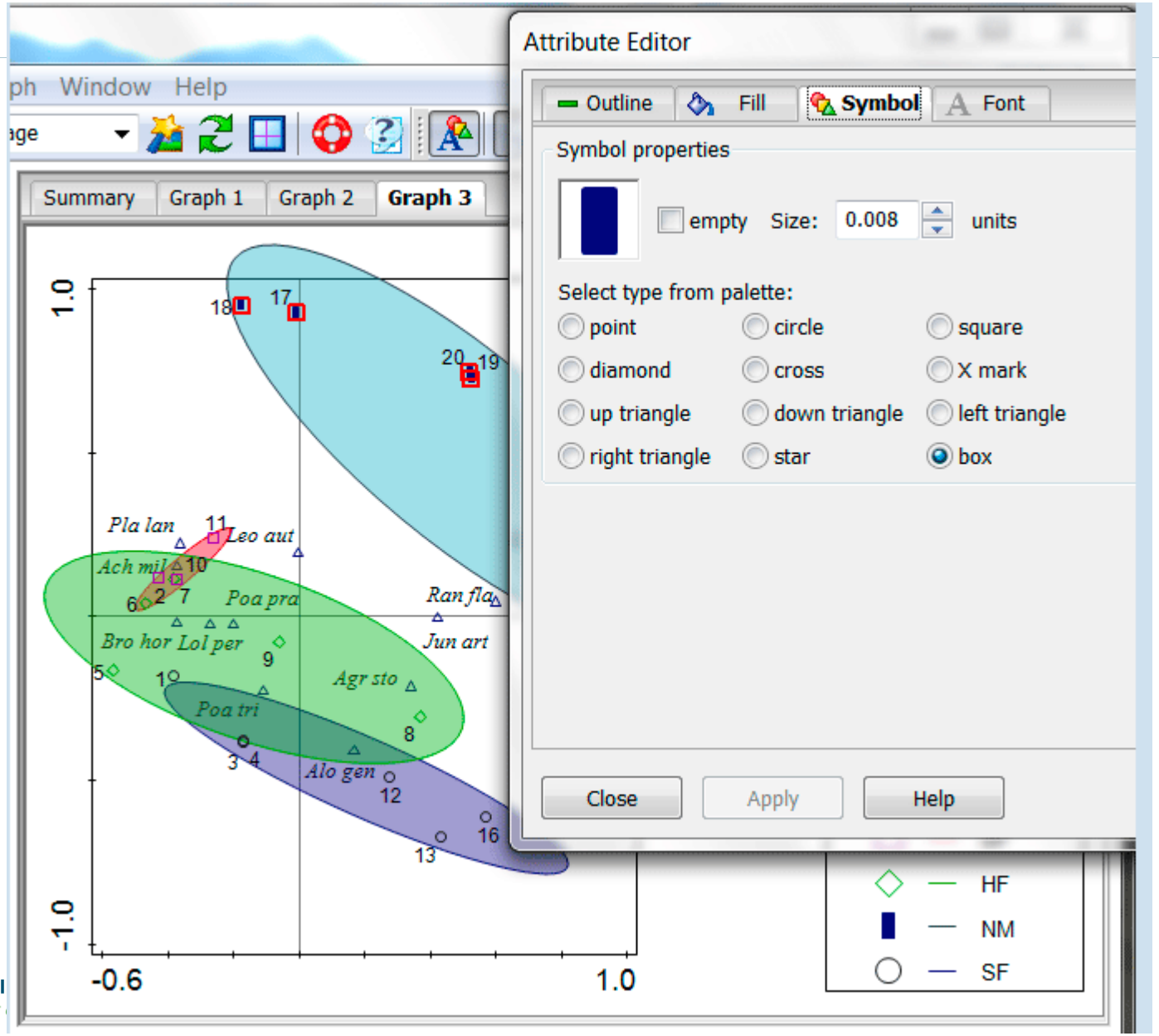
Click Project on the toolbar

- to create classifications, groups or series[1] of samples and species
  - for use in plotting (e.g. symbols or colours per class)
  - to plot a subset
  - etc.
- See example on next slide



[1] See https://doi.org/10.6084/m9.figshare.13259534.v1 for more example

Ellipses and transparent colours



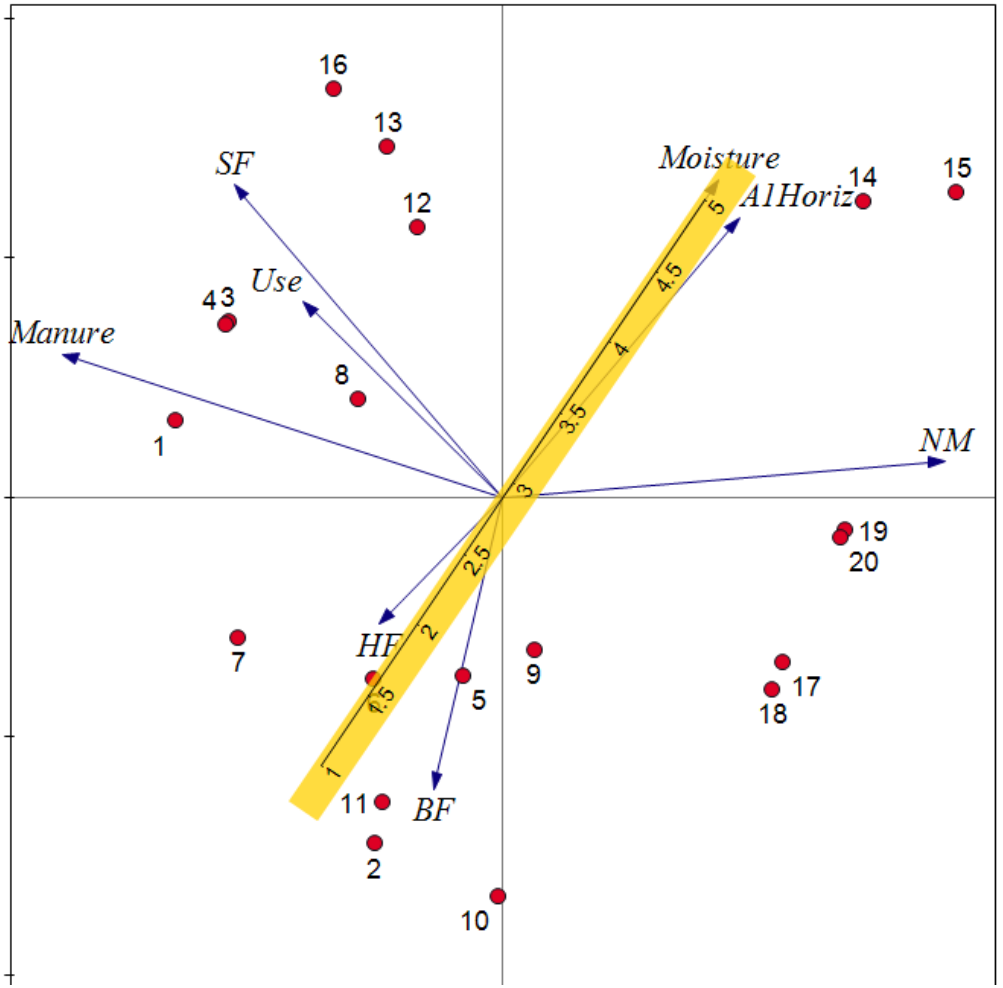
Calibration of arrows

(Graffelman & Van Eeuwijk, 2005)

- E.g. PCA on

Environment data of Dune Meadows

Arrow for Moisture calibrated

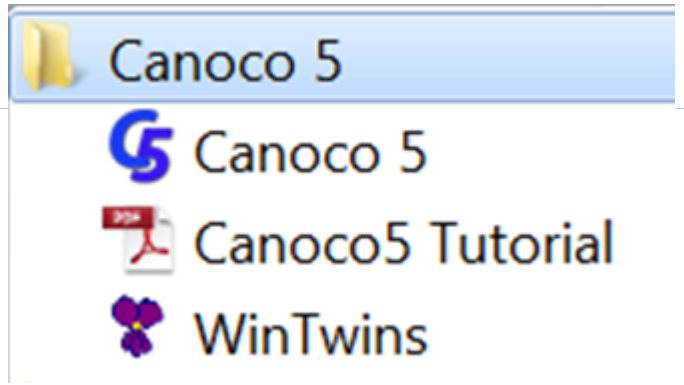


Management automatic expanded to dummies

Also available in Canoco 5

- **Predicted** and fitted **response values** for constrained methods, via *Data* | *Add new table* | *Predict..*; Alt-d-a-p
- Calibration - **predicted explanatory values**; imputing of missing explanatory values on basis of constrained meth. via *Advanced constrained template*
- Diversity indices, via *Data* | *Add new table* | *Statistics*; Alt-d-a-s
- Functional diversity via Alt-d-f
- Indicator values of species for a grouping
- Multiple testing and FDR
- Multi-step analyses and more...

Resources/help



- Canoco 5 Tutorial under Programs
- Canoco 5 manual: ~500 pp
- Support site with Discussion list: www.canoco5.com
- Demo and practical

WAGENINGEN UR
For quality of life

For Canoco projects of published analyses
https://doi.org/10.6084/m9.figshare.13259534

Thank you!

WAGENINGEN UR
For quality of life