Introduction Canoco 5 for Canoco 4.x users

Software for multivariable data analysis and visualization

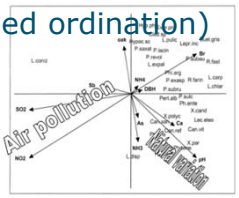
February 4/5, 2013, Cajo J.F. ter Braak and Petr Šmilauer

WAGENINGEN UR
For quality of life

CC BY

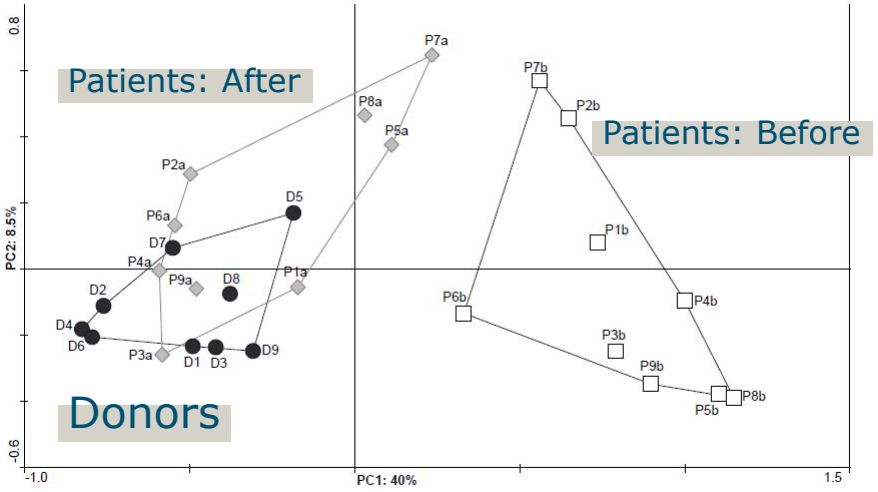
Recap:research questions and methods

- 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



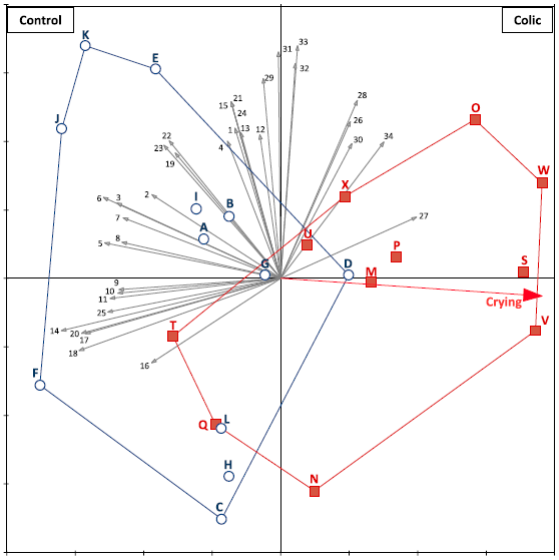
WAGENINGEN UR
For quality of life

Ex1: Comparison of three groups by PCA



Transplant study: van Nood et al. 2012 NEJM
Data: microbiota taxa (Susana Fuentes, W. de Vos)

Ex2: Extension of t-test (1)



- Comparison of two groups by RDA
- Horizontal (constrained) axis = difference of Control and Colic
- Vertical (unconstrained) axis = main residual pattern
- Correlation with Crying of babies

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

Ex2: Extension of t-test (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

WAGENINGEN UR
For quality of life

From Canoco 4.x to Canoco 5 (1)

Canoco 4: terms used	Canoco 5: terms used in manual and some help
Sample	Case
Species	Response
Environmental data	Explanatory data
	Supplementary data
Supplementary data	Supplementary data
Direct/ indirect analysis	Constrained/ Unconstrained ++++
If you wanted a PCA of soil properties: Enter soil data as 'species data': In output: species == soil property	Output uses the term you must define when entering the data. Above terms are used in manual and some help

WAGENINGEN UR
For quality of life

From Canoco 4.x to Canoco 5 (2)

	Canoco 4	Canoco 5
Project	One analysis	Data tables with analyses
Data from Excel	WCanoImp	Integrated
Plotting	Canodraw	Integrated
Solution in:	log and Canoco.sol	Analysis notebook
Factors	Dummy (1/0) variables	Factors with editing facilities
Factors	Define as nominal variables in CanoDraw	Automatic: classes plotted as centroid points
Change scaling of diagrams	Redo the whole analysis!	On the fly with & recreate graph

WAGENINGEN UR For quality of life

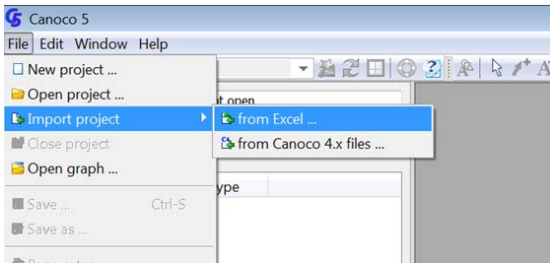
Possible roles of data tables

- 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

WAGENINGEN UR For quality of life

Starting a new Canoco project (1)

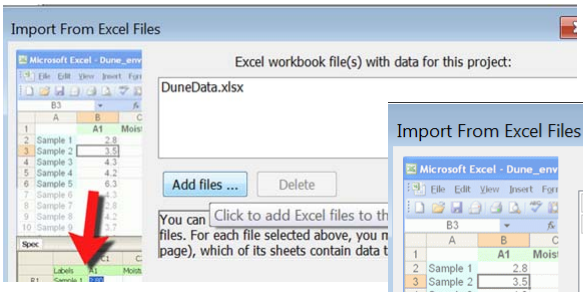
- Canoco 5 focuses on research questions on a set of data
- A Canoco 5 project thus consists of
  - one or more data tables
  - analyses on these data
- Easiest to start a new project with
  File|Import project|from Excel... (Alt-F-I-Enter)

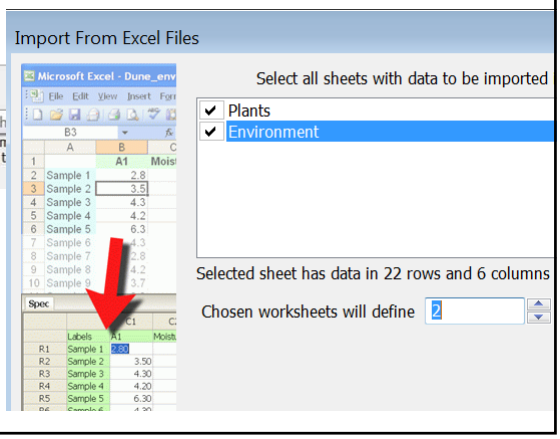


WAGENINGEN UR
For quality of life

Starting a new Canoco project (2)

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



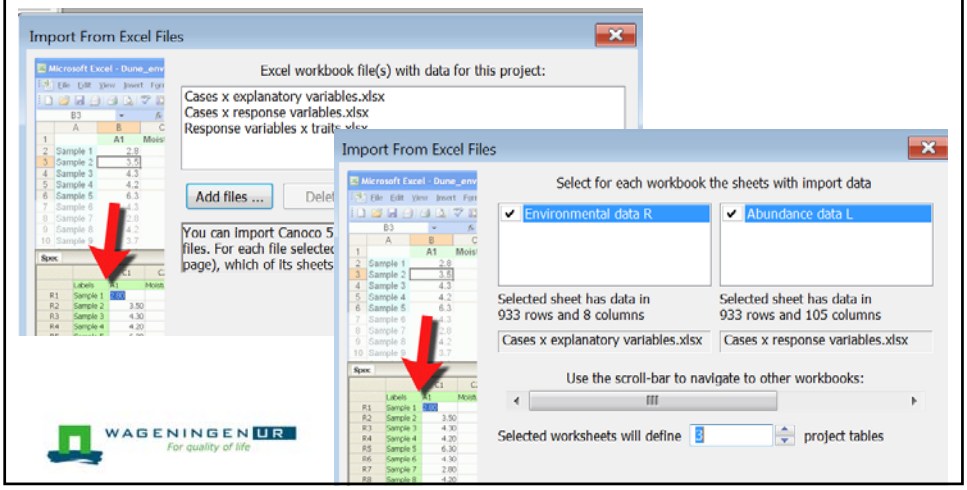


- Excel file can contain more than one sheet
- Each sheet can give ≥1 data tables

WAGENINGEN UR
For quality of life

Example with data in three Excel sheets

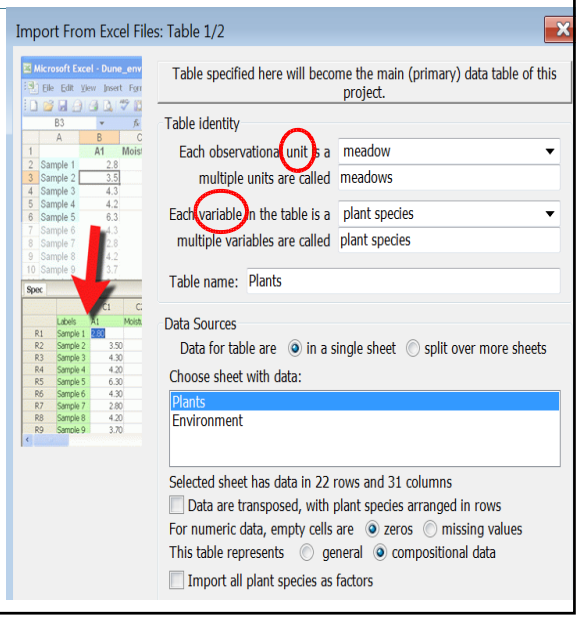
- Select one or more Excel files, here 3
- Select the number of project data tables, here 3



Starting a new Canoco project (3a)

Give names to YOUR units and variables

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



Starting a new Canoco project (3b)

Give names to YOUR units and variables

Empty cells: 0 or mis

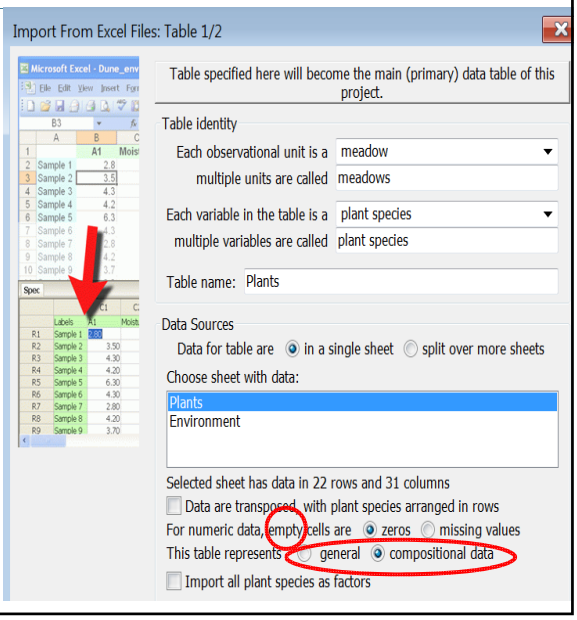
Data kind is

- General or
- Compositional:

-row sum has meaning

-variables measured on the same scale (≥0)

The right choice helps to select suitable methods



Starting a new Canoco project (3c)

Default kind:

- first data table -

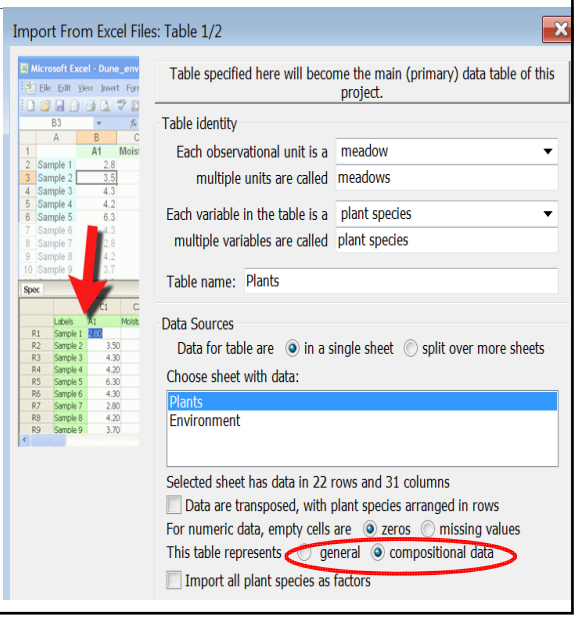
Compositional (e.g. species data)

- Later tables -

General (e.g. env. data/ study design)

Cannot do DCA or transformation on all columns (e.g. log) on a General table

Kind can be changed in table tab



Starting a new Canoco project (4)

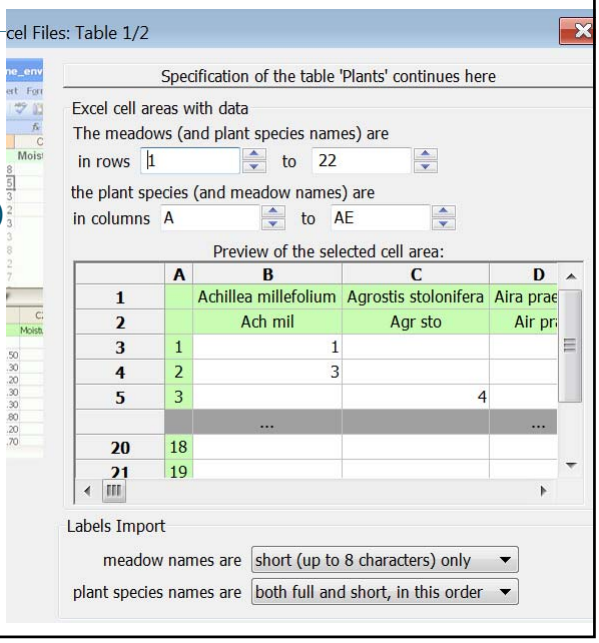
Names of row and column items:

-none

-short names (8 chars)

-full names (long)

-both



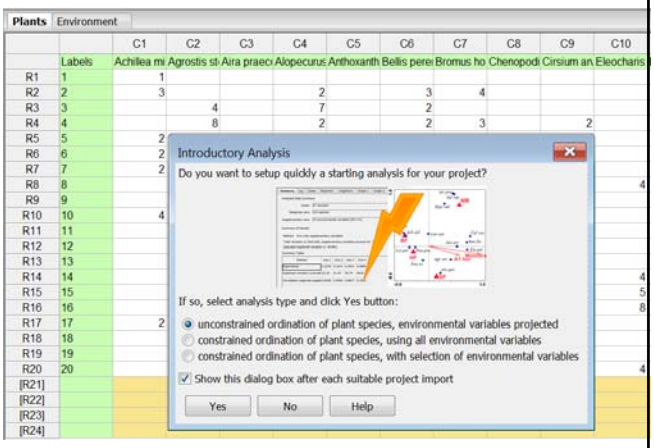
Starting a new Canoco project (5)

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.



Starting a new Canoco project (6)

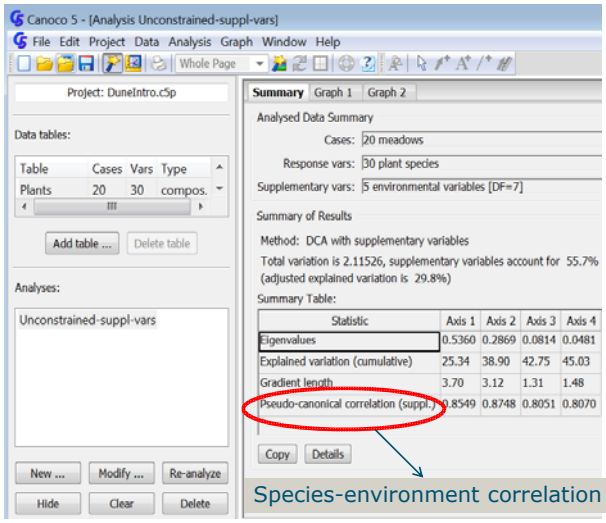
Accepting the offer and all default choices leads to

-Summary of DCA analysis

-Two graphs

Save your project!

- File Save.. or
- Ctrl-S



Starting a new Canoco project (6)

Accepting the offer and all default choices leads to

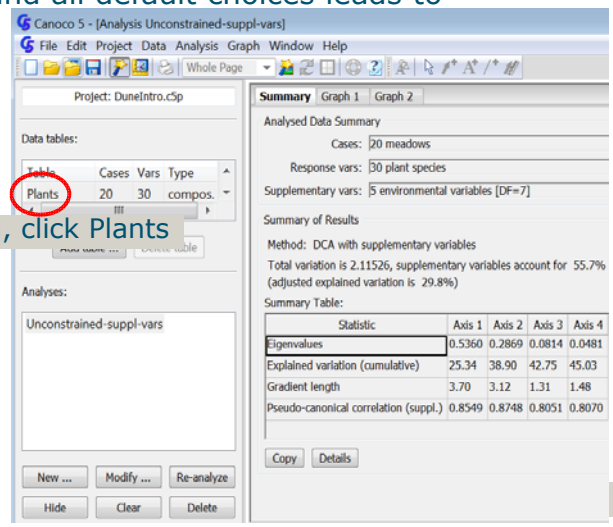
-Summary of DCA analysis

-Two graphs

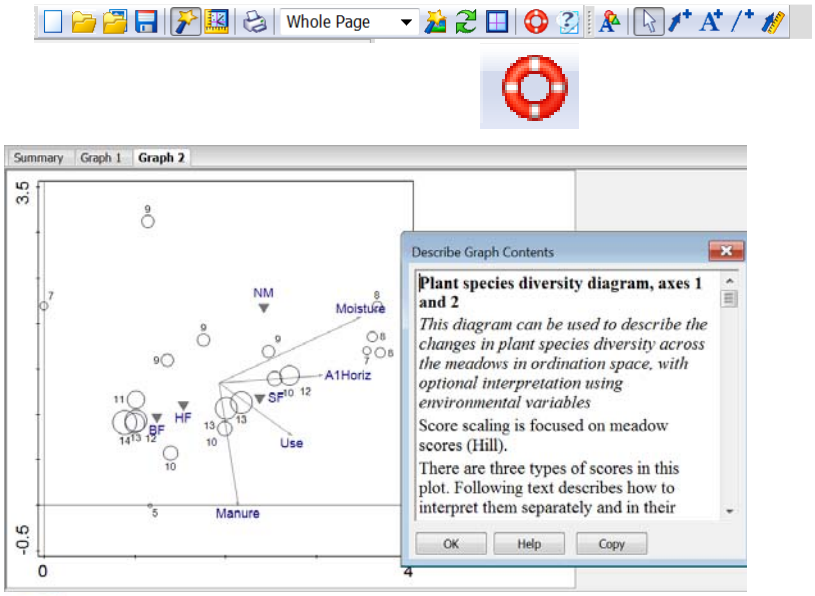
To view the data again, click Plants

Save your project!

- File Save.. or
- Cntr-S



Inspecting a graph with Describe Contents



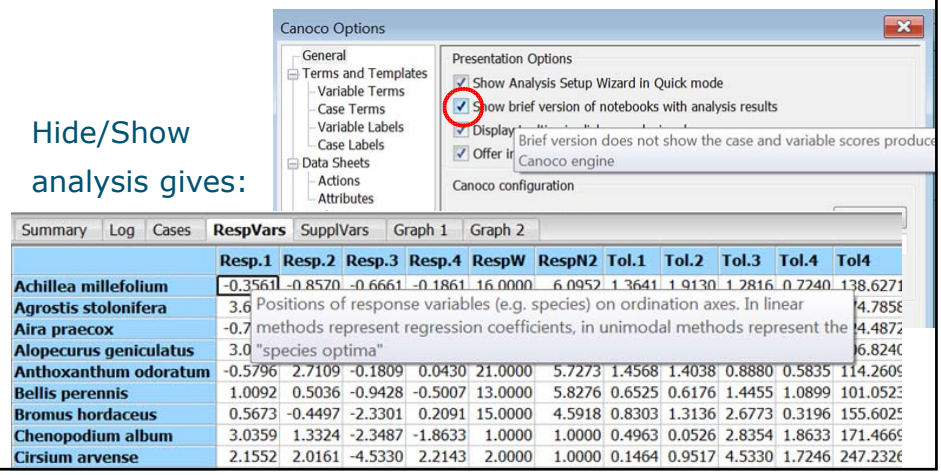
All scores are available too:

No separate Canoco.sol file anymore

Edit | Settings | Canoco5 Options:

- Uncheck *Show brief version of notebooks with ...*

Hide/Show analysis gives:

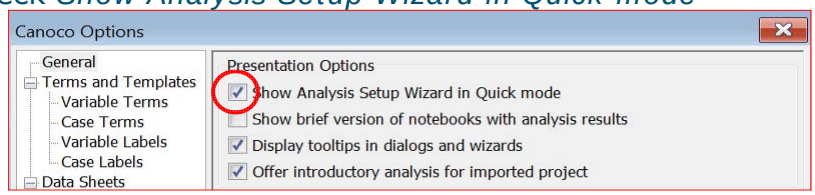


	Resp.1	Resp.2	Resp.3	Resp.4	RespW	RespN2	Tol.1	Tol.2	Tol.3	Tol.4	Tol4
Achillea millefolium	-0.3561	-0.8570	-0.6661	-0.1861	16.0000	6.0952	1.3641	1.9130	1.2816	0.7240	138.6271
Agrostis stolonifera	3.6										4.7858
Aira praecox	-0.7										4.4872
Alopecurus geniculatus	3.0										6.8240
Anthoxanthum 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

Canoco 5 Quick wizard mode

or Edit | Settings | Canoco5 Options:

Uncheck *Show Analysis Setup Wizard in Quick mode*



For:

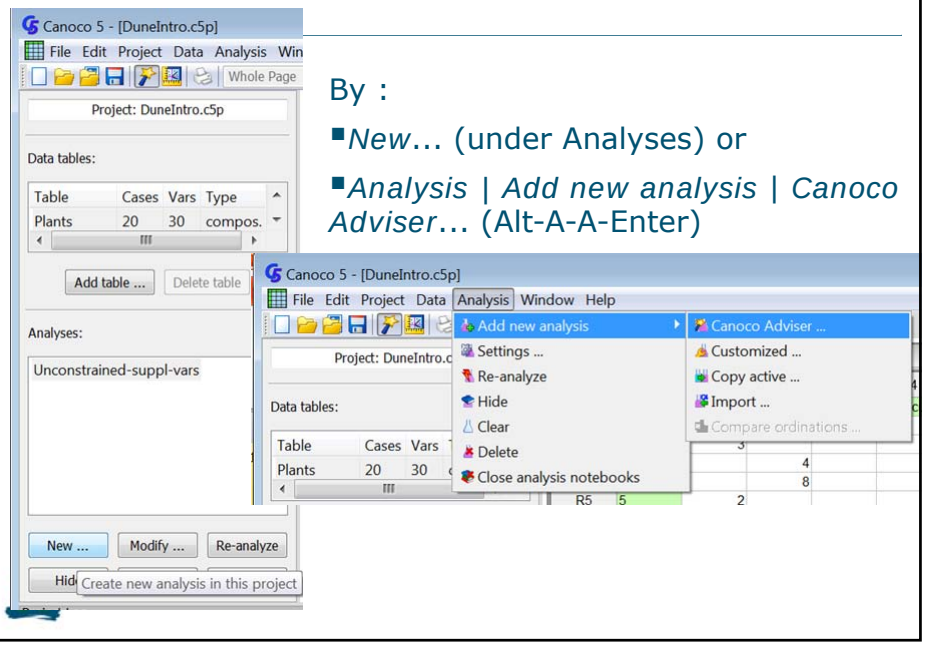
Weighting/deleting cases and response variables

Defining interactions between explanatory variables (can also be done in the data table, click two columns)

Covariate and supplementary variable page

WAGENINGEN UR
For quality of life

Adding a new analysis to the project (1)



By :

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

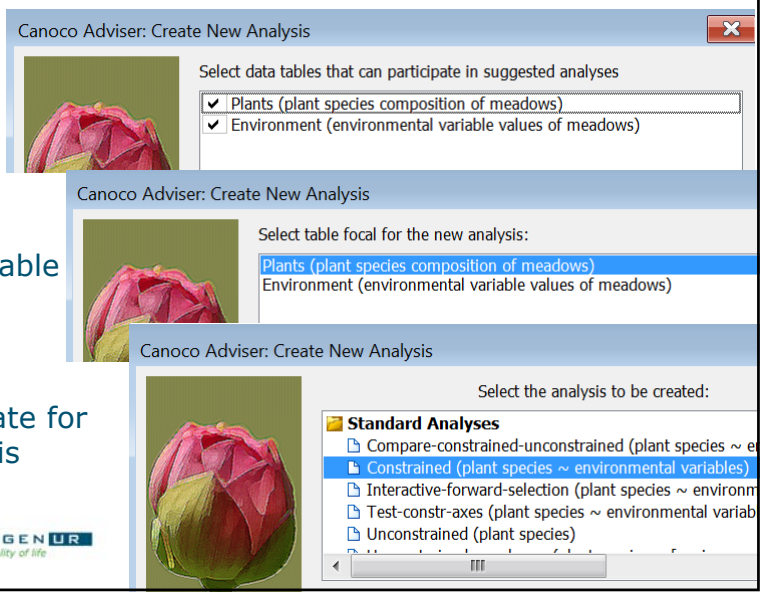
Adding a new analysis to the project (2)

Select:

1. Tables

2. Focal table

3. Template for analysis

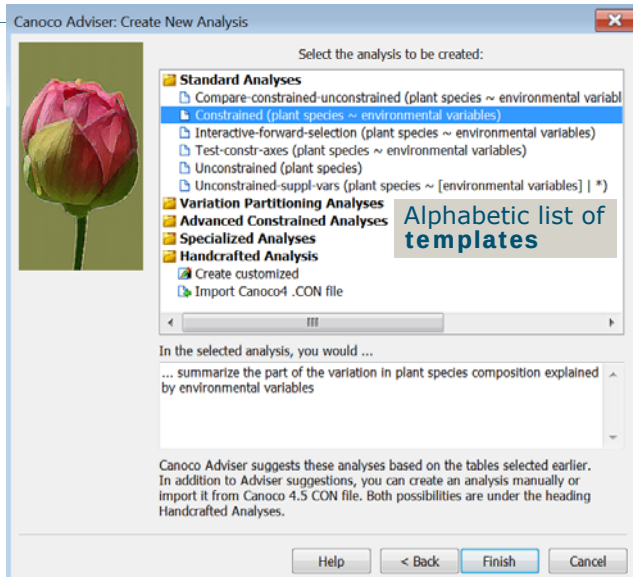


Adding a new analysis to the project (3)

3. Select **template**

-double click on bold terms to fold/unfold

(Can enlarge dialog window to see all)



Adding a new analysis to the project (4)

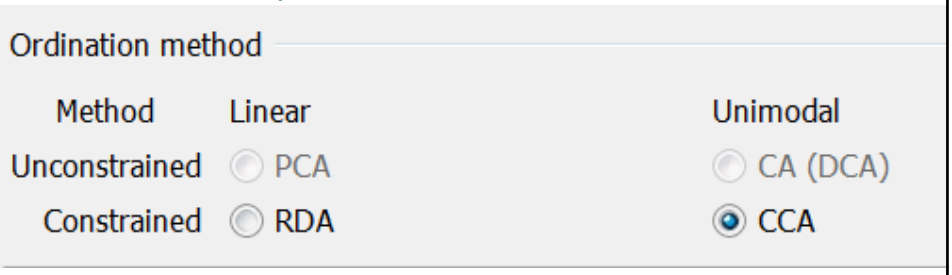
Standard analyses:

- 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:



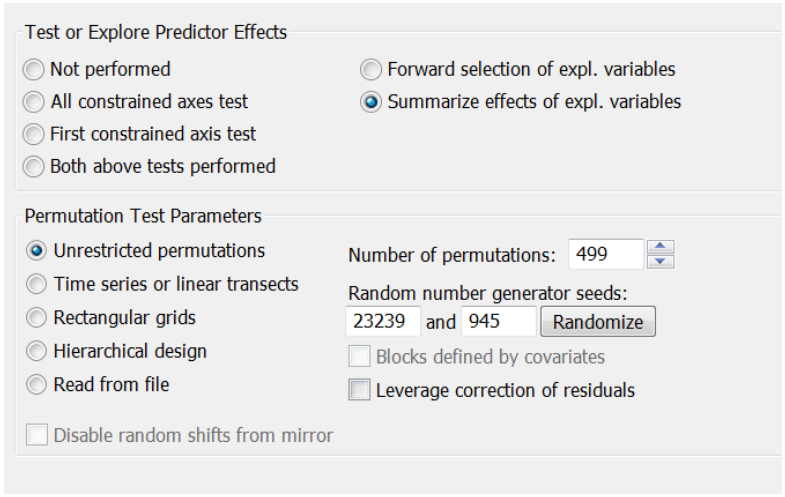
PCA: Principal component analysis

RDA: Redundancy analysis

CA (DCA): Correspondence analysis(Detrended)

CCA: Canonical correspondence analysis

Adding a new analysis to the project (6)



From Canoco 4.x to Canoco 5 (3)

Canoco 4	Canoco 5
Automatic forward selection	Summarize effects of expl. variables
Manual forward selection	Forward selection of expl. Variables (or via specialized template)
Terms in result:	
Marginal effect	Simple effects
Conditional effects	idem
lambda-1 and -A	Explains %
F-value	Pseudo-F
P-value	Added: P(adj) for multiple testing correction or false discovery rate (FDR)

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

Simple Effects

none
False discovery rate
Holm correction
Bonferroni correction

Name Ex P

Moisture 19.4 4.3 0.002

WAGENINGEN UR
For quality of life

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

Forward Selection Step

Candidate Terms

Name	Contribution %	F	P	P(adj)
Moisture	34.9	4.3	0.002	0.016
Management.NM	27.0	3.2	0.004	0.016
Manure Amount	20.3	2.3	0.016	0.04267
A1 horizont	19.1	2.1	0.032	0.064
Management.SF	16.7	1.8	0.058	0.0928
Management.HF	12.5	1.4	0.17	0.22667
Management.BF	11.9	1.3	0.228	0.26057
Use Type	10.9	1.2	0.294	0.294

Test Include Stop Help

Include whole factor

Term Contribution

All considered variables explain together 55.7% of total variation
Out of this variation, the highlighted term would contribute 34.9%

Selected Terms

Order	Name	P	P(adj)

P values correction: False discovery rate

WAGENINGEN UR
For quality of life

New: Canoco Adviser

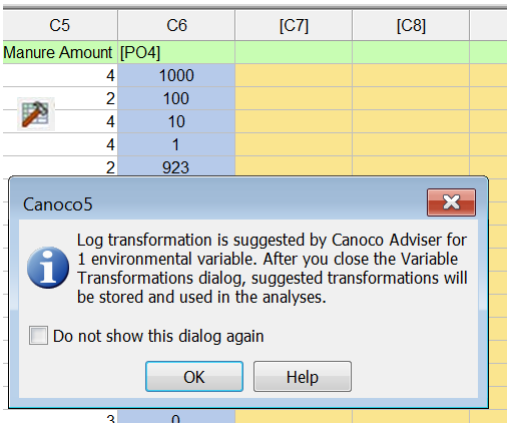
On the basis of the data properties the Adviser suggests

- Transformation and standardization of variables

right-click on top-left cell in data sheet

Or use

Data | Default transformation and …

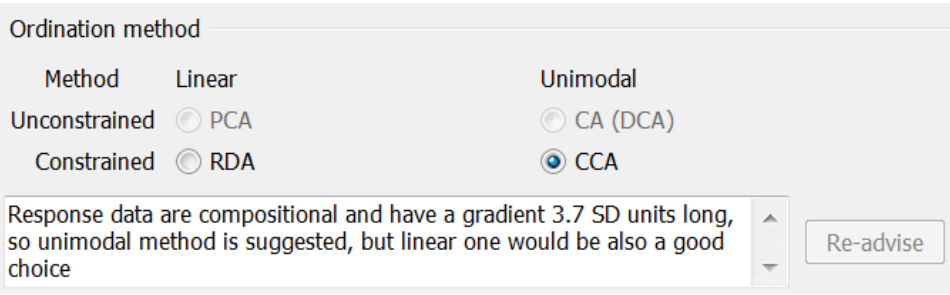


WAGENINGEN UR
For quality of life

New: 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



WAGENINGEN UR
For quality of life

New methods in Canoco 5 (1)

- Variation partitioning
- Distance-based methods
- Co-correspondence analysis
- Trait-based analyses
- Principal response curves (PRC) [via dedicated template]
- Generalized linear models (GLM) with permutation tests

(next two were available in CanoDraw 4)

- Response curves (GLM/GAMs with one predictor)
- Contour plots (GLM/GAM with two predictors)

WAGENINGEN UR
For quality of life

Variation partitioning

Which part of variation is due to

(a) Environment

and which to

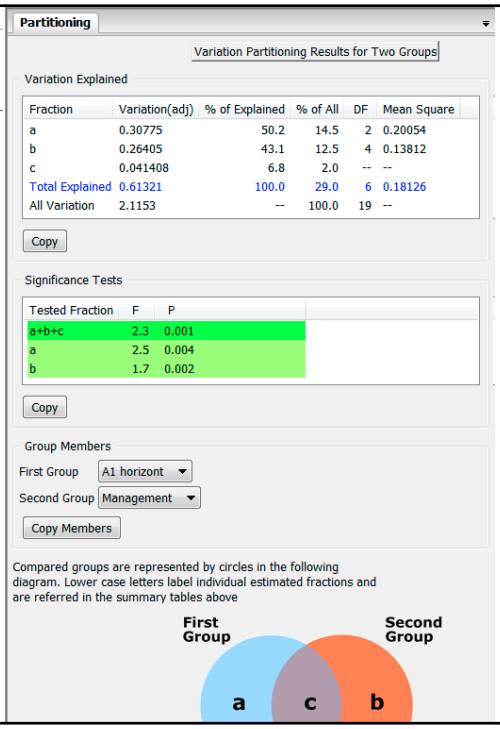
(b) Management

and which part is

(c) shared?

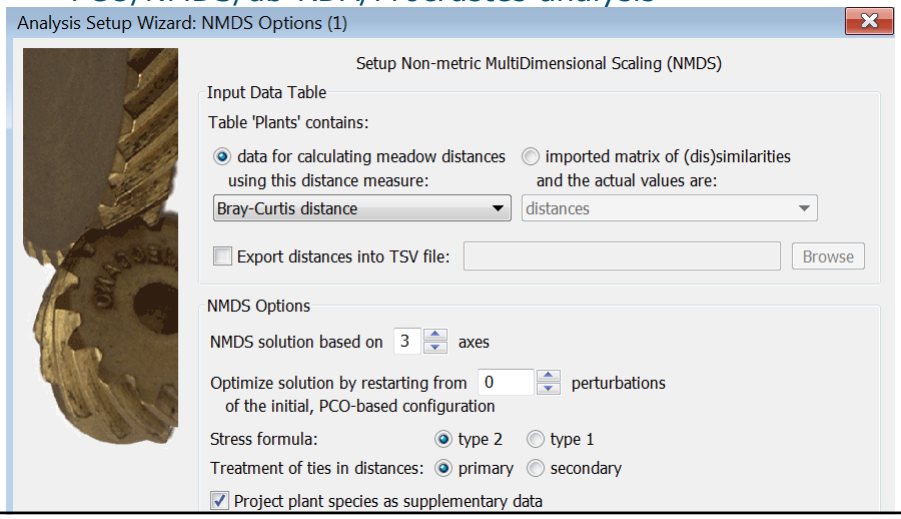
two or three groups of variables

WAGENINGEN UR
For quality of life



Distance-based methods

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

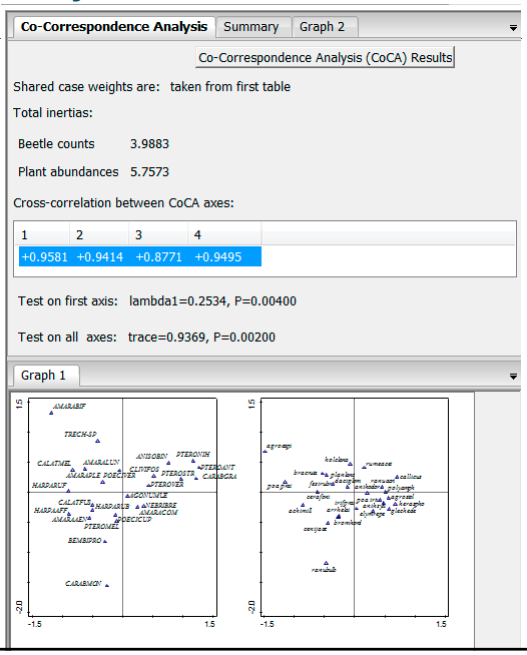


Co-correspondence analysis

- How are two compositional data tables related?

e.g. plant and beetle communities

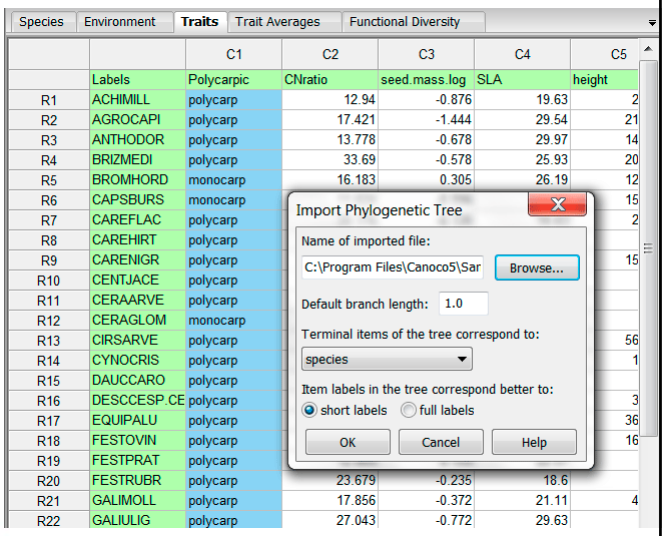
(Schaffers et al. 2008)



WAGENINGEN UR
For quality of life

Trait-based analyses and phylogenetic relations

- Trait averages
- Functional diversity
- RDA on community-mean traits
- 4th corner & RLQ (via Expand occurrences)
- Phylogenetic corrections

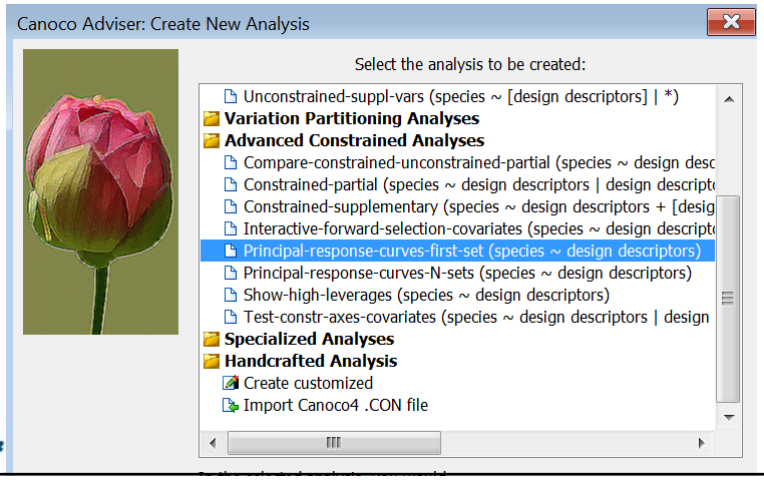


	Labels	C1 Polycarpic	C2 CNratio	C3 seed.mass.log	C4 SLA	C5 height
R1	ACHIMILL	polycarp	12.94	-0.876	19.63	2
R2	AGROCAPI	polycarp	17.421	-1.444	29.54	21
R3	ANTHODOR	polycarp	13.778	-0.678	29.97	14
R4	BRIZMEDI	polycarp	33.69	-0.578	25.93	20
R5	BROMHORD	monocarp	16.183	0.305	26.19	12
R6	CAPSBURS	monocarp				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	

Principal response curves (PRC)(1)

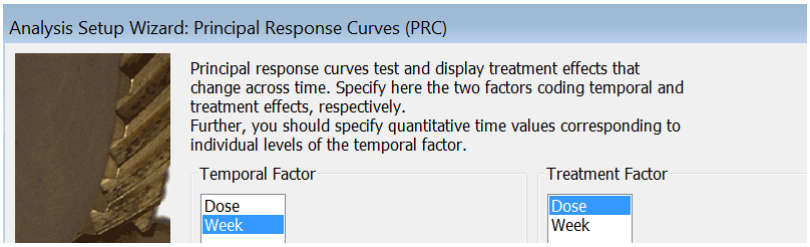
Template in Advanced...

Requires at least two factors in explanatory data to show up

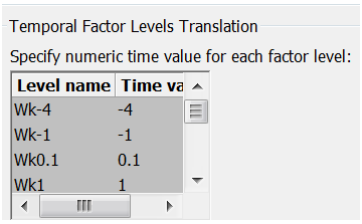


Principal response curves (PRC)(2)

- Specify Time and Treatment factors



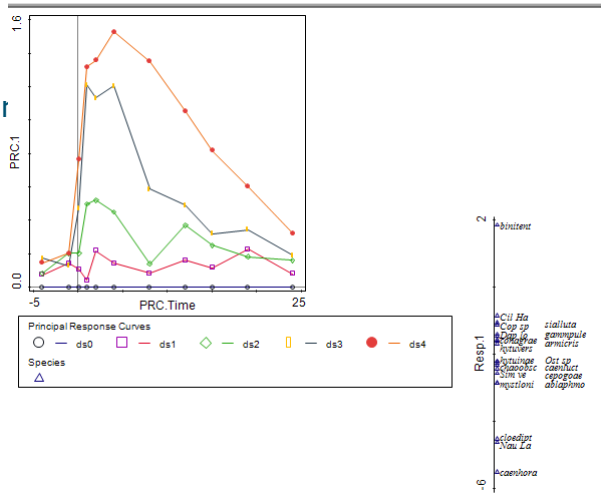
- Specify time values for horizontal axis (default often good)



PRC diagram: Invertebrates~ treatment.time | time

- Example

Van den Brink & ter Braak (1999)



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

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

GLM | Summary | Log | Cases | ExplVars

Response variable:	Pardnigr	
Expected distribution:	Poisson	with log link function
Fitted model deviance:	245.52	with 23 residual DFs
Null model deviance:	1099.3	with 27 residual DFs
Dispersion parameter:	10.56	
Parsimony (AIC-like):	351.12	
F statistic:	20.214	(DF=4,23)
p(F):	<0.00001	

Term	b	SE	T	p(T)
(Intercept)	-5.58977	1.731284	-3.23	0.00371
WaterCon	3.14001	0.5211395	6.03	<0.00001
BareSand	0.0191125	0.2406353	0.08	0.93738
FallTwig	-0.847538	0.1932076	-4.39	0.00021
Refl Lux	0.0837154	0.2033162	0.41	0.68433

Copy Help

Nonlinear response curves via GLM or GAM

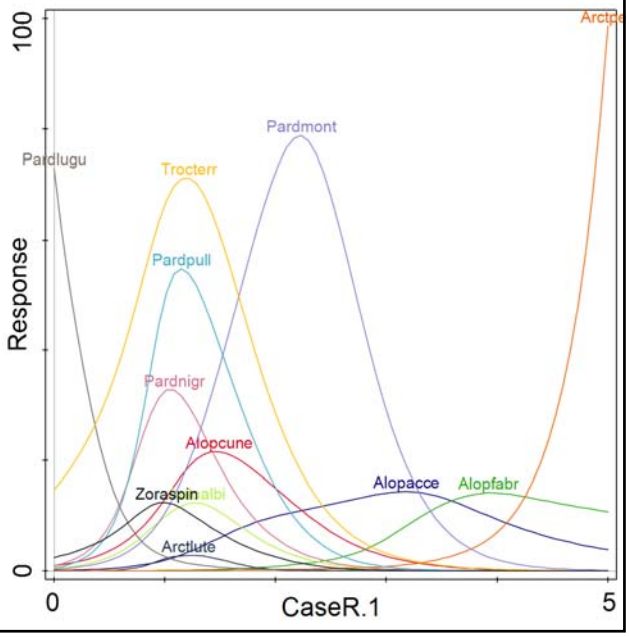
Fitted Generalized additive models

Predictors	CaseR.1
Distribution	quasi-Poisson
Link function	log

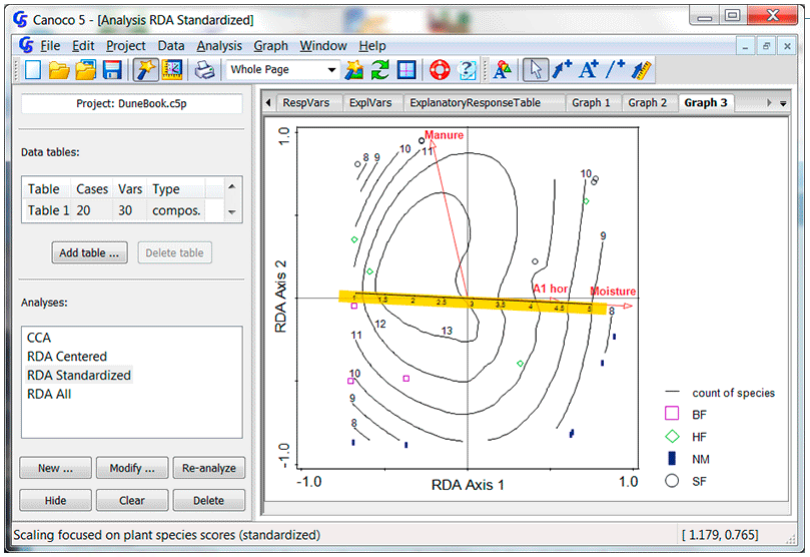
GAM fitted for 12 response variables:

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



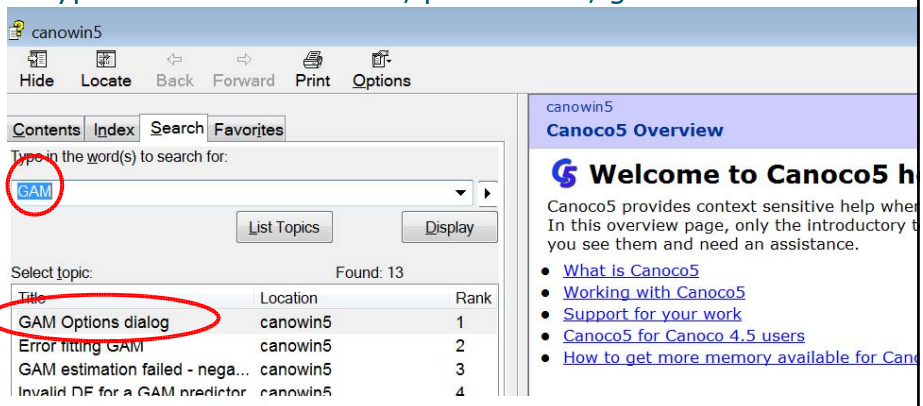
GAMs or GLMs with two predictors



WAGENINGEN UR
For quality of life

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



WAGENINGEN UR
For quality of life

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*

WAGENINGEN UR
For quality of life

New methods in Canoco 5 (2)

- **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...

WAGENINGEN UR
For quality of life

New/better graphs in Canoco 5

Integrated! Graphs require at least one analysis

Graph options: -Edit | Settings (application wide) AND
-Analysis | plot creation options

- Better name placing in ordination diagrams

Examples of new graphs:

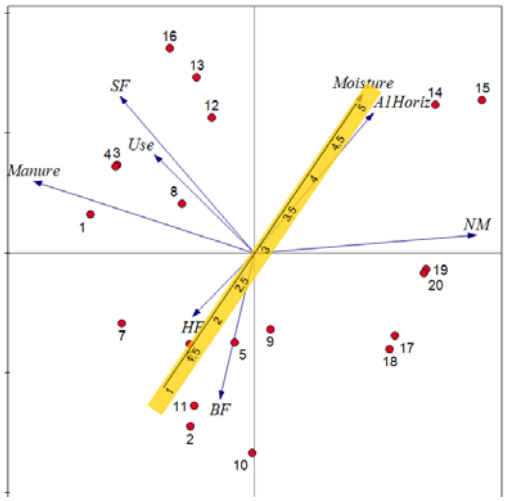
Calibration of arrows

(Graffelman & Van Eeuwijk, 2005)

- E.g. PCA on

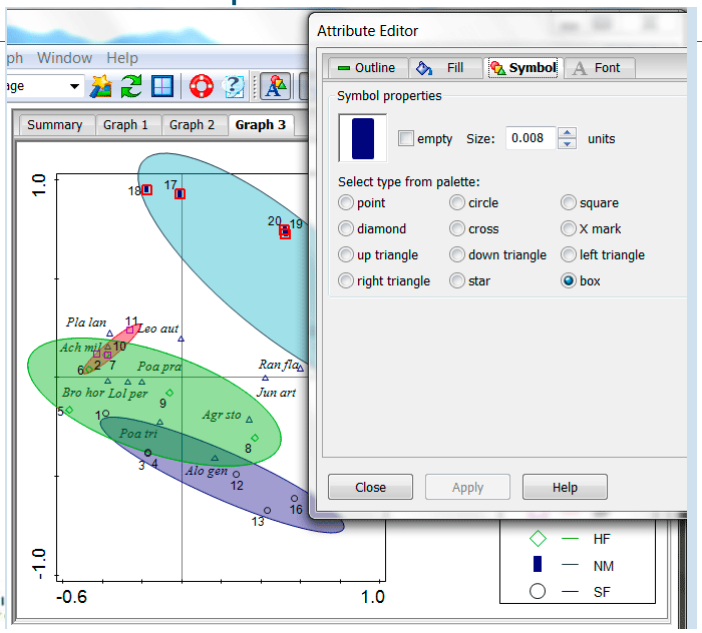
Environment data of Dune Meadows

Arrow for Moisture calibrated



Management automatic expanded to dummies

Ellipses and transparent colours

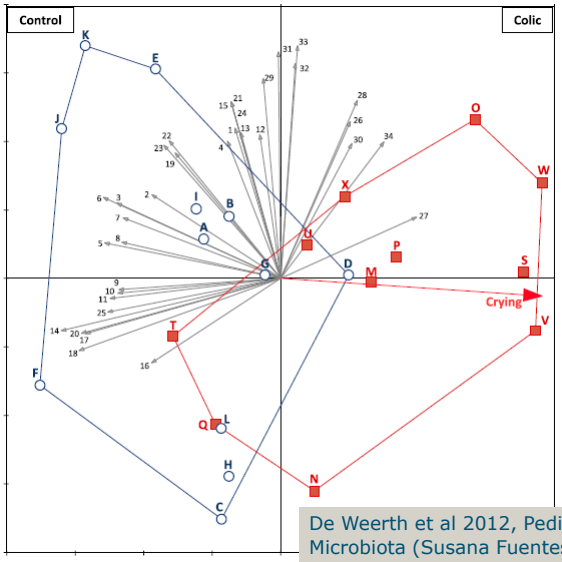


Resources/help

- Canoco 5 Tutorial under Programs
- Canoco 5 manual: ~500 pp
  - Look in WUR Library catalogue to see where it is available on loan or for sale
  - On sale now in tea break from 35€ for 25€
- Support site with Discussion list: www.canoco5.com
- Ask help from Biometris (often me...) English preferred
- Demo and practical

Ex2: Extension of t-test (1)

RDA or CCA: response ~ factor Advice Graphs: ex.3



- Comparison of two groups by RDA
- Horizontal (constrained) axis = difference of Control and Colic
- Vertical (unconstrained) axis = main residual pattern
- Correlation with Crying of babies

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

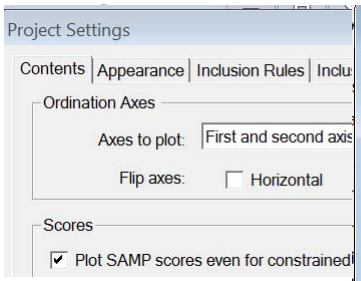
From Canoco 4.x to Canoco 5 (4)

RDA or CCA: response ~ factor

Canoco 4

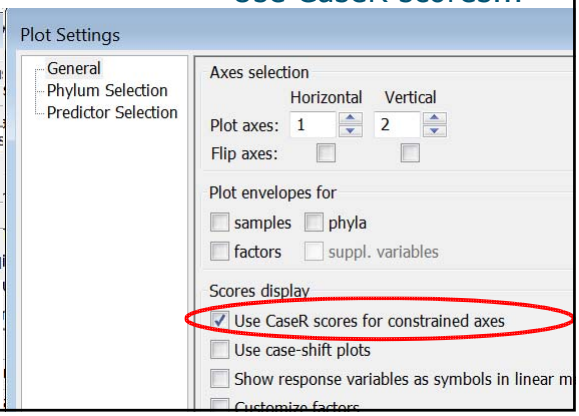
- Canodraw|Project |Settings

Plot Samp scores even



Canoco 5

- Analysis | Plot creation options (Alt-A-P)
  - Use CaseR scores...



WAGENINGEN UR
For quality of life

Canoco 5: partial RDA/CCA

Groups avoid
one variable taking both roles!

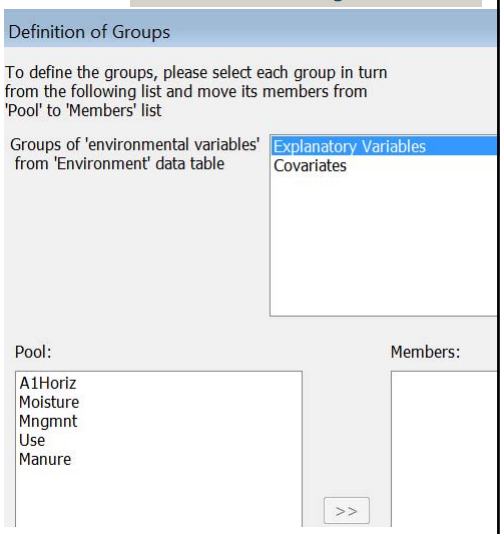
Via Advanced constrained analyses

Division of variables in one table in:

- Explanatory variables (First group)
- Covariates (Second group)

Use of 'grouped' in:

Template and own multistep analyses



WAGENINGEN UR
For quality of life

Thank you!

Resources:

www.canoco.com

www.canoco5.com

Overview/Tips/Issues

Mailing list of Canoco users

WAGENINGEN UR
For quality of life