

Canoco practical (also for a brief general introduction)

Page numbers such as p.4 refer to the hand-out “Canoco 5 for Canoco 4 users”. For brevity, key sequences such as Alt-A-A-Enter are used. If you type such sequences slowly, you see onscreen what you would normally select by mouse clicks. Lower case letters work in these cases, so Alt-a-a-Enter gives the same result. Working in pairs helps: let one person dictate the sequences and both of you can see what happens.

1. Importing data and first analysis

We will work with the classic dune meadow data (Jongman et al., 1987). It is on the composition of plant communities on meadows, recorded in 20 plots, and how this composition is related to recorded characteristics of the environment and of grassland management. The data is in

C:\Program Files\Canoco5\Samples\DuneData.xlsx

in two sheets, Plants and Environment.

1.1 Optionally, first look in Excel at the data. Do not close Excel as you can use it again (copy and paste results in a new sheet).

1.2 Start Canoco5: double-click the Canoco 5 icon or via the Program Menu. When you run Canoco 5 for the first time, it opens a *Canoco5 First Steps* window. The final bullet point says how to create a new project using *File/Import project/from Excel...* Follow your hand-out *Introduction Canoco5* (p.5-9) to import both sheets, as separate data tables, perform the *Introductory Analysis* and save the project with a name of your choice.

1.3 Have a look at the data in Canoco5 to check whether all is OK. Check the Environment data and hover with the mouse over the cell C3 (above Management). What kind of variable is it?

1.4 Right-click on cell C3 and choose the one-before last option (*Expand to dummy variables*). Look at the result: the old Canoco 4 coding of factor variables. Click Ctrl-Z for undo; we want to keep the factor as it is.

1.5 Now hover with the mouse over the cell C4 (above Use Type). What kind of variable is it? But note that it has just three distinct values (1, 2 and 3) representing use for hay production (1) and grazing (3) and intermediate (2). Perhaps you want to treat Use Type as a factor with three levels. For this, right-click on cell C4 and again choose the one-before last option (*Convert ..to factor*) and accept. Inspect the result. Undo the transformation to stay in tune with the following.

1.6 Right-click on the top-left cell to see that you can copy and export (to Canoco format and CSV and TAB format a.o.), change the table kind to ‘compositional’ (that is, the current kind is ‘general’), transform the data and change name and row/column terms (in this table, column terms only; in first table both).

1.7 Click on the Plants tab (first table, depending on your input). Check that it is, at present, compositional. Or look at Data Tables (left top of Canoco 5 window).

1.8 Inspect the starting analysis by clicking on the first analysis ‘Unconstrained-suppl-vars’ on the left of the Canoco window. By this, you are ‘leaving’ the Data Notebook

with its table, and enter an 'Analysis Notebook'. Which percentage of variation is explained in two dimensions?

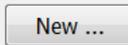
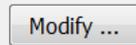
1.9 Inspect *Graph 1* by clicking on its tab. There are at least three types of scores in this plot. Which? If you have no idea, invoke 'Describe graph contents' . Note that the environmental variables are represented by arrows. But how is Management represented? What makes Management different from the other variables that it is represented in this way? (Hint: p.4, top panel).

1.10 To highlight the meadows in the graph, click on the circle of meadow 19 (for example), right-click and click *Select suchlike*. All meadow circles are now selected. Highlight these by clicking  which opens the Attribute editor. Click on Fill, choose the red color, click Apply, Close and click somewhere else in the graph. The graph has now red meadow points (with black outline, which you could also change to red).

1.11 Try to move a text item in the graph, for example, to position it even better. Try to move a meadow point. Why does this not work?

2. New analysis: Constrained analysis, ranking the explanatory variables and embellish the graph.

In this exercise we want to model plant species as a function of environmental variables (plant species ~ environmental variables), but which ones are the most important? ¹

2.1 Add such an analysis by typing Alt-A-A-Enter (type slowly and follow on screen what you select in this sequence) or click New..   ]

and use p.12-14 to complete the analysis either by CCA (advised by the Adviser) or RDA. [If you changed in 1.6 the Environment data to compositional, no constrained analysis is offered, as compositional data normally do not take the role of explanatory data. So, change the table kind of Environment back to 'general']. Accept the defaults, but do not forget to change the default in *Test or Explore* to *Summarize effects of expl. variables*. Accept the default graphs suggested by the *Graph Wizard*.

2.2 How much percentage variation would be explained by *Manure Amount* if it were the only explanatory variable in the analysis? Is it the best variable in this sense? To see it, you may need to scroll down the Summary page of the Notebook of the second analysis, named *Constrained*. (By the way, you can change the name of the analysis to make clear, for example, whether you selected CCA or RDA).

2.3 Look up *Manure Amount* in the next table headed *Conditional effects*. You may need to scroll down the list (or *Copy* the table to the clipboard and paste it in a new Excel file or sheet). How much percentage variation is explained by *Manure Amount* and why does this percentage differ from that in the previous table headed *Simple*

¹ If you did not do exercise 1, you can start with
C:\Program Files\Canoco5\Samples\Chapter2\DunesFromExcel.c5p
and save this project under My Documents. To stay in tune, change the table items: Alt-W-1, right-click in top-left cell, and select last option, change row names to meadow (meadows) and species to plant species.

effects? Do you understand why is it so much lower and why is *Manure Amount* significant in *Simple effects* but not in *Conditional effects*?

2.4 At least how many explanatory variables are needed to explain to plant data; which variables?

2.5 Inspect *Graph 1* of this analysis (produced by the *Graph Wizard*) by clicking on its tab. (If there is no such graph yet, type Alt-G-A-A to invoke the *Graph Wizard* for the current analysis). Note that the default colours for explanatory variables differ from those in the unconstrained analysis. Otherwise, *Graph 1* appears as a slightly rotated version of that of the previous analysis; but it does not contain any meadow points.

2.6 Create a graph such as *Graph 1* but with meadows. Hint: Alt-G-T-Enter. (T= triplot).

2.7 In Canoco 4.x you had to choose a scaling for ordination axes right away. Now you get a graph, but which is its scaling? Find out by clicking *Describe contents* .

Just for fun, try to use another scaling in the graph: click  and change at the scaling of the model type you used in this analysis (Top: Linear; bottom: Unimodal models) and click . Undo the change (Ctrl-Z).

2.8 Triplots are often overcrowded, so let's plot meadows separately. Hint: Alt-G-S-Enter-↓-Enter. Which scaling is chosen by default?

2.9 Let's highlight the differences between the Management types in the graph. For this, we must first Classify the cases (meadows), starting with Alt-P-C-C-Enter-Enter. Click *From data*. Change *Data Table* from *Plants* to *Environment*. Click Management, OK, OK, check the *Use selected classification in graphs*, OK. Now recreate the graph (click .

2.10 We miss a legend in the graph. So let's ask to include a legend by Alt-E-S-↓-Enter, check the *Create Legend* box, OK. Also ask for envelopes around the management types by Alt-A-P- check the meadows box under Plot envelopes, OK.

Now recreate the graph (click ). Which two management classes appear most dissimilar according to the plot? [Note that you have two sets of options for graphs, the first option set used applies to all new analyses/graphs in the project; therefore it is under *Edit*. The second option set applies to graphs of a particular analysis; therefore it is under *Analysis*.]

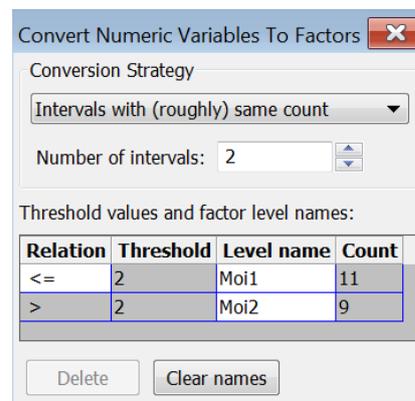
2.11 Store the project.

2.12 There are two straightforward extensions of the exercise: adding an analysis to test for the significance of each of the constrained axes and adding an analysis directly comparing the constrained analysis with the corresponding unconstrained analysis (p.13 top). But perhaps it is time for something simpler.

3. Comparing two groups; a multivariate t-test

This exercise makes a graph such as on p.2 bottom, but using the Dune Meadow data. To stay as close to this example and the t-test, we use redundancy analysis (RDA) in this example with a single explanatory variable, namely a factor with two levels (or a variable with just two values). There is no such factor in the Dune Meadow data so far, so let's create one. See the footnote on page 2 if you start here.

3.1 Go to the Environment data table (p.9 bottom, or type Alt-W-1, click Environment). Right-click in cell C2 (above *Moisture*) and select *Convert .. to factor* and select as *Conversion Strategy*, *Intervals with (rough) same count*, OK. Save the project under a new name, as the data was changed. [In real life, first copy *Moisture* to column C6, convert there, give the variable another name, because you do not want to lose (detail in) your data; but let's keep things simple for now]



3.2 Add an RDA analysis with as single explanatory variable the factor you just created (*Moisture*). [Tip: to do so quickly, select all environmental variables in the right box and move them to the left box. Now select *Moisture* in the left box and move it to the right box]. Accept the default graphs. You end up with a summary and three graphs.

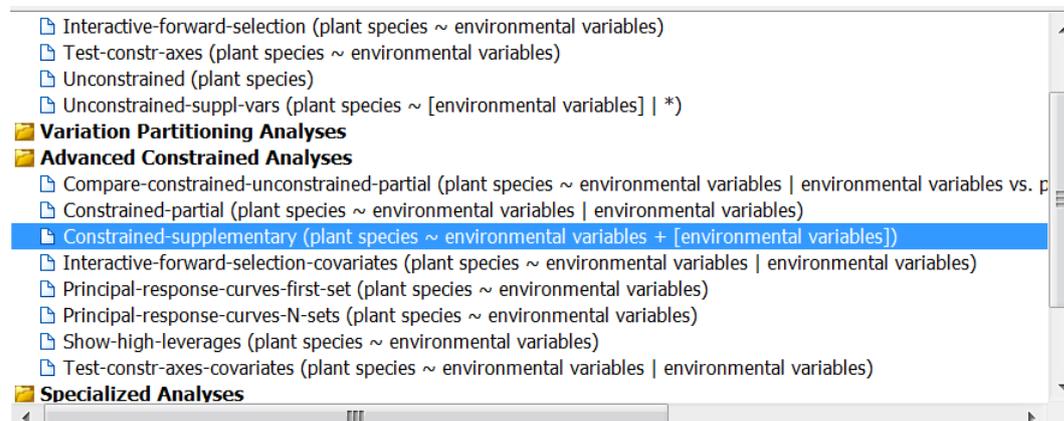
3.3 Is there a significant difference between the two moisture classes, or more precisely, do the two moisture classes differ significantly in terms of their plant species composition? [This test avoids the problem of multiple testing created by testing each plant species separately]. Click *Histogram* and *Help* in the window it generates so see what it represents.

3.4 On the basis of *Graph 1*: is plant species *Lol per* on average more abundant in the low moisture class than in the high moisture class?

3.5 *Graph 2* looks quite colourful, because Management type is still to be plotted in graphs. Where do SF (standard farming) plots lay compared to NM (nature management) plots? *Graph 3* is specialized for analyses with a single factor. How well are the classes separated? Which axis shows this? What might be visible vertically? (Hint: see also *Graph 2*).

3.6 *Graph 3* differs from that on p.2 bottom in that it lacks supplementary variables. Suppose we want to add the other variables in the Environment data table as Supplementary data. For this, we need to define a constrained analysis with

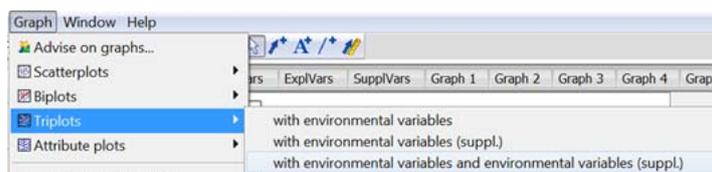
supplementary variables: response ~ explanatory factor [supplementary variables]. Such a template is available under Advanced Constrained Analyses. So, add a new analysis and in the New Analysis Wizard, expand the Advanced Constrained Analyses folder (by a double click; you may need to scroll down to see it) and select the third template, named Constrained-supplementary (plant species ~environmental variables + [environmental variables]).



Next, specify Moisture as only explanatory variable and the other variables as Supplementary variables. Accept the default graphs. *Graph 3* of this analysis is what we were seeking. Where is the point for SF (standard farming) and where is NM (nature management) and in which direction points Manure Amount? Guess how the triangle points of SF and NM are defined in terms of those of the meadow points belonging to these classes.

3.7 [Advanced]. Let's try to generate something like *Graph 3* ourselves from scratch. We want, for example, later include not only the means for the moisture classes, but also some other variables. Therefore we need to understand the details of *Graph 3*. To plot the meadows and moisture classes, type Alt-G-B-↓-↓-Enter (selecting Meadows + Environmental variables).

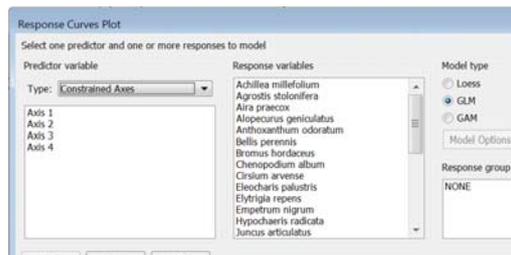
Perhaps to your surprise this gives not something like *Graph 3*. What do you see? Note that we did a kind of ANOVA which gives means; here two means, one for Moi1 and one for Moi2, as is visible in the graph. Meadows are plotted by way of these fitted values. To change this, use Alt-A-P Use CaseR scores (p.27 bottom) and recreate the graph. Now add envelopes for meadows belonging to each Moisture classes (see exercise 2.9). Recreate now gives a graph such as *Graph 3* but with added centres of the classes. To obtain a similar graph with supplementary variables use a triplot: Alt-G-T- ↓-↓-Enter (with environmental variables and environmental variables (suppl)).



4. Response curves, label and contour plots

In this exercise we return to the *Constrained* analysis of exercise 2 and will add some graphs with response curves and contour plots to its Analysis Notebook. Open the project you stored after exercise 2 (File | Recent Files| 1 or 2?) as we want Moisture to be quantitative...

4.1 Click first on analysis *Constrained* to show its Notebook (alternatively, type Alt-W-3-Enter). To draw response curves (under *Graph/ Attribute plots*), type Alt-G-A-Enter-↓-↓-Enter. You see



You can make response curves with respect to any variable in the data tables and results (select Type). Here we take Axis 1 of the Constrained Axes. Select it (name should change to blue) and also, for example, the first 6 species Achillea to Bellis. Now *Model Options* light up. Click on it and change *linear* to *quadratic* model and *Gaussian* to *Poisson* (if not yet set so). Check the Quasi box and OK, OK and then OK for each individual species. You end up with a summary of all fits and then a graph with six response curves. Note you can change the label of the horizontal axis from CaseE.1 to something you like.

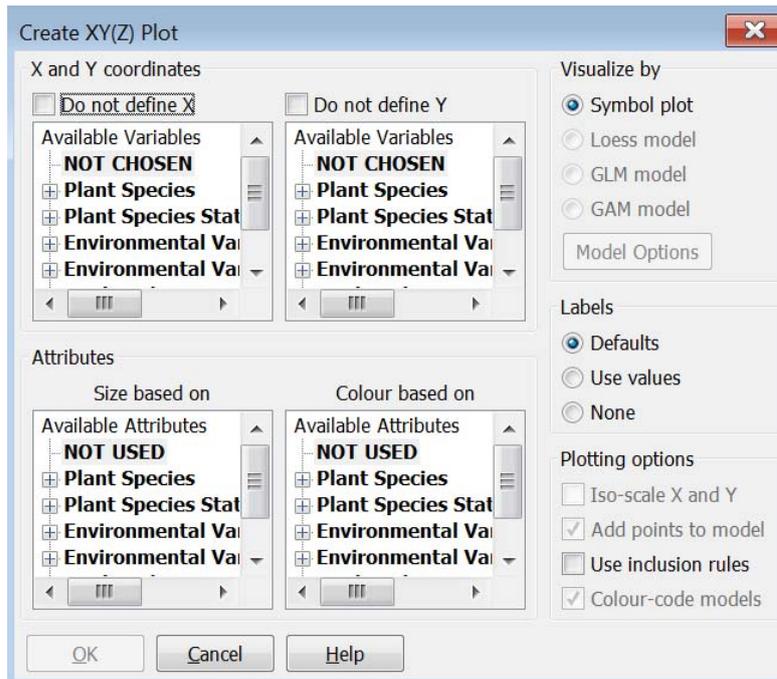
These are all quadratic fits, but perhaps the quadratic term is not always needed. For this you can check the *Stepwise select* box in *Model Options*. This works as well for craftsmanship's curves using GAM with e.g. 4 df (degrees of freedom).

4.2 Make a similar set of curves with respect to Moisture.

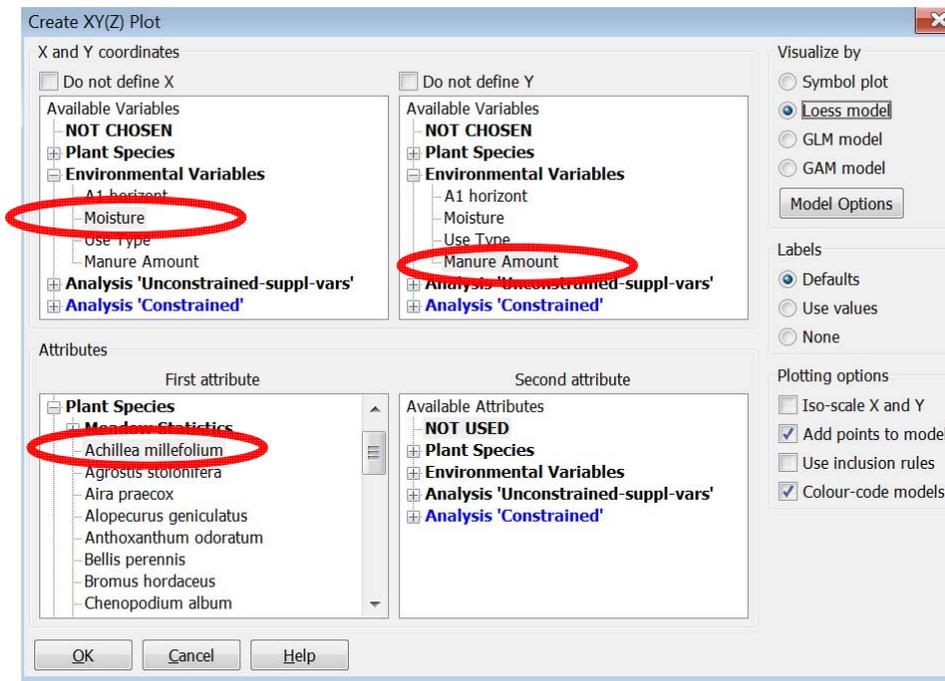
4.3 We now want to plot the amounts of Achillea in ordination space (under *Attribute plots*). Type Alt-G-A-Enter-Enter. Expand Plant species in the left box and select Achillea and click OK. The amounts of Achillea are visualized both by the size of the circles (with a small + for a zero amount/absence) and the amounts are also shown numerically.

4.4 We make a contourplot of Achillea in ordination space (under *Attribute plots*). Type Alt-G-A-Enter-Enter again, select Achillea as before and now select loess model to obtain a smooth surface.

4.5 Attribute plots in ordination space have predefined axes. You can choose your own axes in Graph|Attribute plots| XY(Z) diagrams. Type Alt-G-A-Enter-↓- Enter to get (you can enlarge this box):



Compared to exercise 4.4 we have two more lists (the top two), where we can define our own variables. Select for X Moisture (in the first Environmental Variables list) and for Y Manure, and select Loess model and Achillea in First attribute and OK (see snapshot)

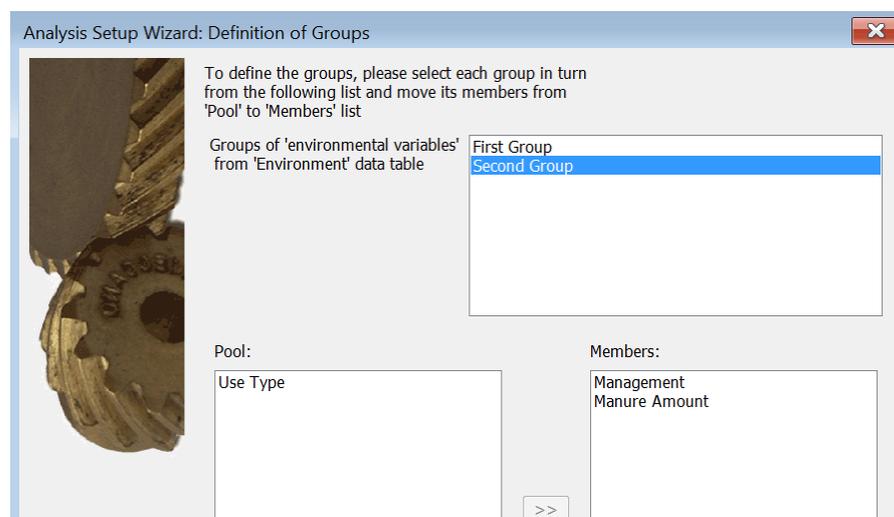


The positions of the cases and their names are also plotted.

5. Variation partitioning

In this exercise we try to find out which part of the variation in the plant species data is due to the soil variables (Moisture and A1 horizon) and which ones to Management including Manure and what their shared part is.

5.1 Start a new analysis as in exercise 2 and now open the *Variation Partitioning Analyses* folder in the New Analysis Setup wizard (p. 12 bottom) by double clicking (to see Variation ... you may have to enlarge the window). The first template is OK; it tests Conditional effect, the other tests Simple effects of the groups of explanatory variables. Accept the use of Adjusted variation and then you assign variables to each of two groups. Select First Group and assign Moisture and A1 horizon to it; then select Second Group and assign Management and Manure Amount to it (snapshot).



Accept all further defaults. See p.17 bottom for the result. See that the Analysis Notebook has now a dedicated first page, called Partitioning. This page collects main results from the three steps needed to obtain the partitioning result. In Canoco 4.x you had to define and carry out these three analysis and then compute the results yourself. The individual analyses (in Canoco 5 called steps) are still visible in the Summary page, which says that the analysis has 3 steps. See also section 1.7 of the tutorial.pdf.

There are already three graphs in this Analysis Notebook (created by Advise Graph). If you want to make your own graphs, you must select the Step by Alt-A-S-S-Enter. See section 1.6 the tutorial.pdf ; I recommend to make custom-made graphs of a single step analysis for simplicity; so first, newly create yourself the analysis you want custom graphs for.

6. Principal response curves

In C:\Program Files\Canoco5\Samples\Advanced there are two example projects for PRC. Open one of them, select an analysis and click *Modify...* to see how it is defined.