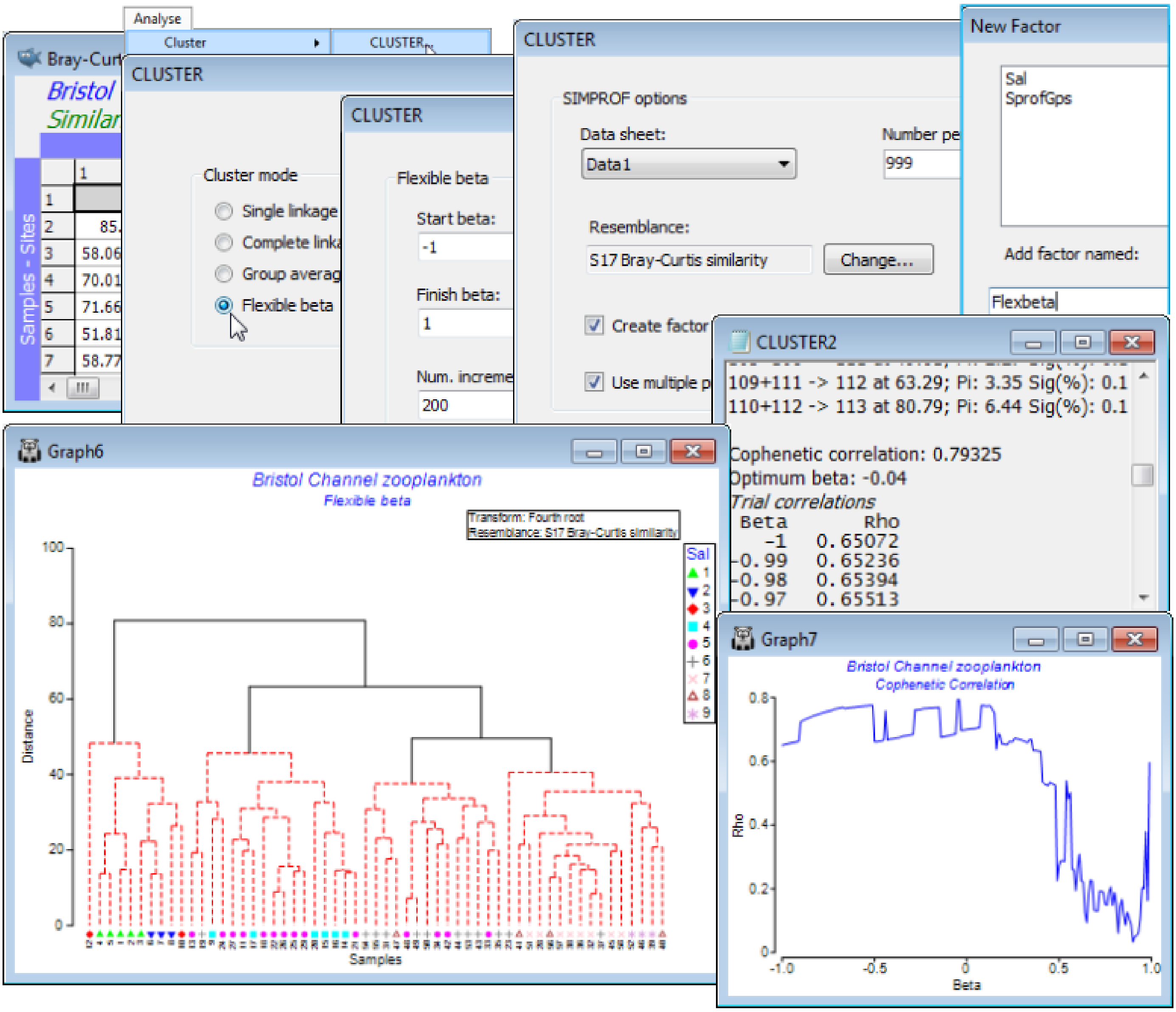
Linkage by flexible beta method

There are four possible Cluster mode choices within the **Analyse>Cluster>CLUSTER** dialog box, distinguished by the way they redefine the among-group dissimilarities at each proposed step of the agglomerative process. The *linkage options* are: •Single (/nearest neighbour) linkage, which has a tendency to produce unhelpful ‘chaining’ of groups, with many steps adding just a single sample to an existing group; •Complete (/furthest neighbour) linkage, which tends to have the opposite ‘over-grouped’ effect; •Group average (Unweighted Pair Group Method with Arithmetic mean UPGMA) which is the option shown in all the above plots and is widely used; and •Flexible beta, introduced by Lance GN & Williams WT 1967, *Comp J* 9: 373-380, a generalisation of a WPGMA method in which a range of options is controlled by choice of a parameter β. Chapter 3 of CiMC gives precise definitions of all these options, e.g. for flexible beta see the footnote on p3-4. Choice of is made automatically to maximise the *cophenetic correlation* ρ (rho) between the dissimilarities/distances in the resemblance matrix and distances through the dendrogram between the matching pairs of samples – this idea was met near the beginning of this section – and a plot of ρ vs. β displayed.

Remove the selection on the fourth-root transformed data matrix Data1, by **Select>All** (and **Edit> Clear Highlight**, though this is not essential) then with the active sheet as the similarity matrix calculated from Data1, take **Analyse>Cluster>CLUSTER**>(✓SIMPROF test) & (Cluster mode• Flexible beta)>(Start beta: -1) & (Finish beta: 1) & (Num. increments: 200). These are the defaults, meaning that the cophenetic correlation is computed and graphed for β in increments of 0.01, with the optimum β (maximum ρ) given in the Cluster results window, and this value used to calculate the dendrogram. Note that β does need to be in the range (–1, 1) but negative values (or zero) make better sense theoretically, as is seen here in the line plot of the cophenetic correlation ρ vs. β, so there is a case for restricting to (Start beta: -1)&(Finish beta: 0)&(Num. increments: 100). If a fixed value of β is preferred (Lance & Williams suggest β= –0.25), as it might be for repeated clustering, then take, for example (Start beta: -0.25)&(Finish beta: -0.25)&(Num. increments: 1). You will also need to specify a factor for the SIMPROF groups, e.g. Add factor named: Flexbeta, which gives a Multi-plot (see next section) of the dendrogram and the line plot of ρ vs. β.



	1
1	
2	85.
3	58.06
4	70.01
5	71.66
6	51.81
7	58.77

CLUSTER

Cluster mode

○ Single linkage
○ Complete linka
○ Group averag
● Flexible beta

CLUSTER

Flexible beta

Start beta: -1

Finish beta: 1

Num. increme 200

CLUSTER

SIMPROF options

Data sheet: Data1

Number pe 999

Resemblance: S17 Bray-Curtis similarity Change...

☑ Create factor

☑ Use multiple p

New Factor

Sal
SprofGps

Add factor named: Flexbeta

CLUSTER2

```
109+111 -> 112 at 63.29; Pi: 3.35 Sig(%): 0.1
110+112 -> 113 at 80.79; Pi: 6.44 Sig(%): 0.1

Cophenetic correlation: 0.79325
Optimum beta: -0.04
Trial correlations
 Beta      Rho
   -1   0.65072
-0.99   0.65236
-0.98   0.65394
-0.97   0.65513
```

Revision #4
Created 19 June 2024 04:42:49 by Arden
Updated 19 June 2024 22:54:54 by Arden