13. New standardisation options

13.1 Overview

PRIMER 8 offers a host of new options for standardising data (either samples or variables), *via* the menu item: **Pre-treatment** > **Standardise**. The new 'Standardise' dialog window in PRIMER 8, by comparison with that in PRIMER 7, is shown below (Fig. 13.1).

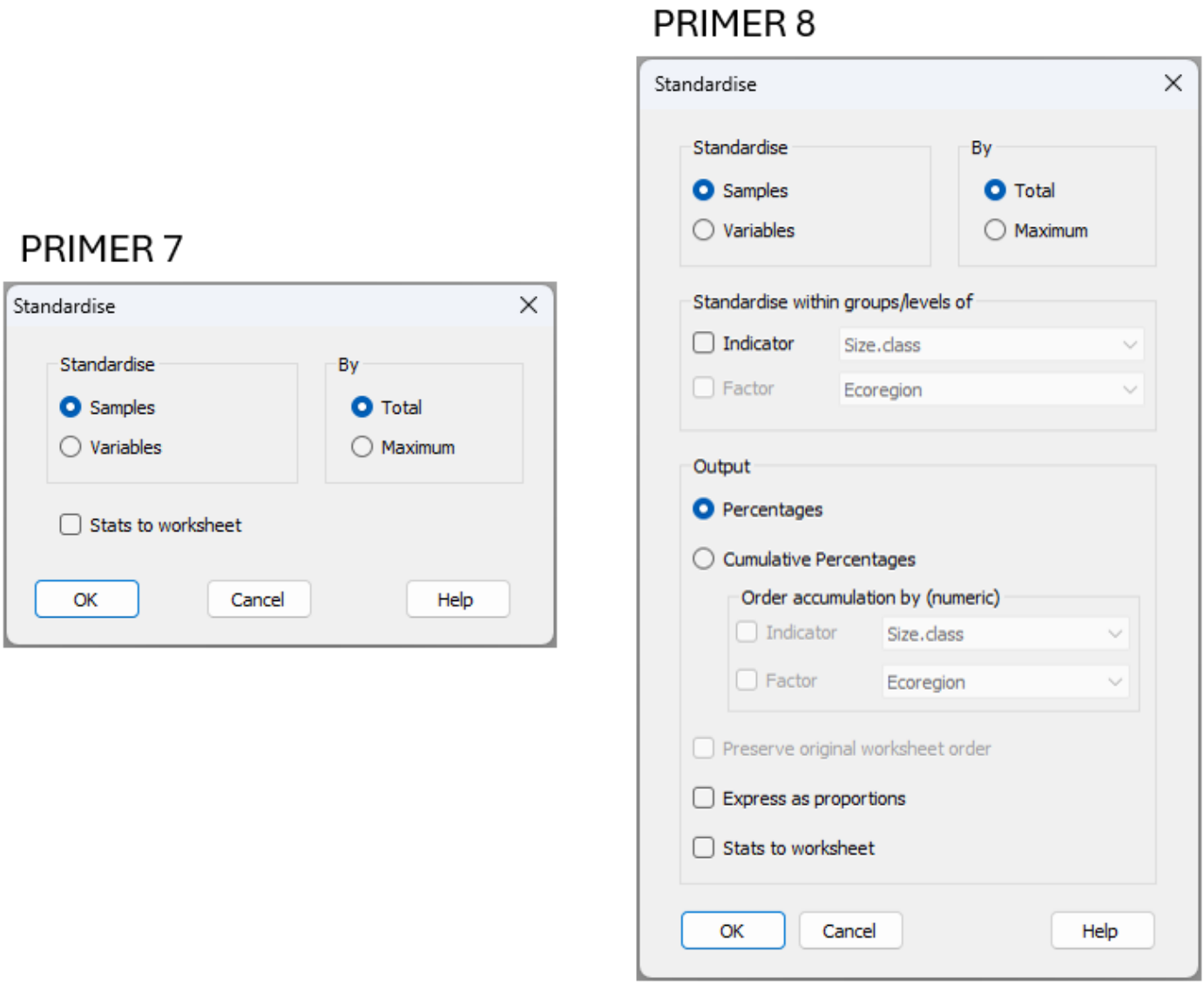


Fig. 13.1 Dialog for standardising samples or variables in PRIMER 8 by comparison with that available in PRIMER 7.

The two essential expansions to this tool in PRIMER 8 are:

1. You can output the standardised data either *directly* (as calculated) or *cumulatively* (ordered as in the data sheet, or optionally ordered by a factor/indicator) as either percentages or proportions; and
2. You can perform standardisations:
   - across **samples**, but done *separately for different sets of variables* identified by an Indicator; or

- across **variables**, but done *separately for different sets of samples* identified by a Factor.

Option (1) above enhances the ability of PRIMER 8 to handle datasets where *the variables themselves are ordered* in some fashion. For example, perhaps the multivariate variables are:

- size-class fractions in sediment samples;
- abundances of organisms in different size classes (i.e., size-frequency data);
- growth curves where individuals are monitored over time and their sizes recorded at different time points. So, individual time points are different variables, and these have a natural order.

Option (2) above enhances the ability of PRIMER 8 to handle various situations much more easily. For example, suppose you have multiple species, and in every sampling unit there are tallies for each of several size classes for each of the species. For this case, we might want to do the standardisation cumulatively across the size classes, but it makes sense to do so separately within each set of variables that belong to a particular species. In such a case, the species (to which each size-class variable belongs) can be given as an indicator. Another example could be sets of variables that consist of tallies across multiple categories, within each of several traits. In this case, we would want the standardisation to be done separately for each trait. In either of these examples, the standardisation task would be quite laborious were it not for the new PRIMER 8 standardisation tool.

13.2 Analysing cumulative standardised data

Rationale

Suppose we have data where the variables consist of different size classes of mussels (as we shall shortly see in a real example). In such cases, where the *variables have a natural order*, we may, of course, simply treat the variables multivariately just as they are, ignoring their intrinsic quantitative inter-relationships and natural ordering. However, note that Bray-Curtis (or Manhattan, or whatever measure we apply) will take no notice of the ordering of the variables. In other words, we could shuffle the order of the variables and it would make no difference at all to the calculated resemblance matrix among samples. However, we may want to permit the natural ordering of the variables to play a useful and meaningful role in the analysis itself.

If we use *cumulative percentage values* across each sample, then we can really see the sample as a 'profile' of size classes, from smallest to largest. This is different from viewing the sample as a set of individual size-class variables that bear no relationships to one another (i.e., where distances between those size classes would be of no importance). By standardising each sample to a *cumulative profile across the size classes*, we acknowledge that the variables are, themselves, structured, from smallest to largest; i.e., they are ordered. Of course, the variables will not be independent of one another following a standardisation like this, but we do not, typically, expect the variables to be independent of one another in multivariate analyses to begin with. It is the simultaneous action of all variables (in this case, the overall *shape of the profile*) that is relevant to us for comparative analysis among the samples.

Size-class data: a 'toy' example

To make this more concrete, consider the following 'toy' example dataset, where we have 5 samples (columns), and each of these have the following *raw percentages* of individuals of a given species belonging to each of 8 size classes (rows, which are the variables here) and these are ordered.

Toy example

Size class data - raw percentages

Abundance

Variables	Samples				
	1.More.small	2.Some.small	3.Even	4.Some.large	5.More.large
2-4 mm	50	0	12.5	0	0
4-6 mm	0	50	12.5	0	0
6-8 mm	50	0	12.5	0	0
8-10 mm	0	50	12.5	0	0
10-12 mm	0	0	12.5	50	0
12-14 mm	0	0	12.5	0	50
14-16 mm	0	0	12.5	50	0
16-18 mm	0	0	12.5	0	50

The sample labelled '3.Even' has a perfectly even distribution, with 12.5% of the individuals in every size class. The sample labelled '1.More.small' has 50% of its individuals in the smallest size-class (2-4 mm), and 50% in the 6-8 mm size-class. The sample labelled '2.Some.small' has 50% of its individuals in the 4-6 mm size-class, and 50% in the 8-10 mm size-class, and so on.

Now, the Manhattan distances between all pairs of samples here looks like this:

Resem1

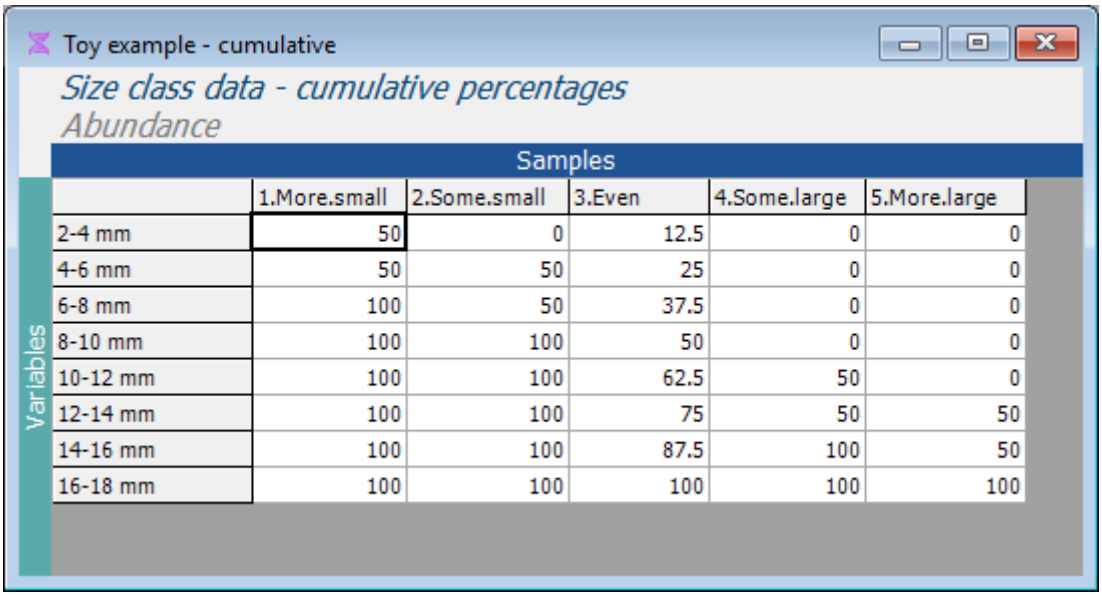
Size class data - raw percentages

Distance (0 to inf)

Samples	Samples				
	1.More.small	2.Some.small	3.Even	4.Some.large	5.More.large
1.More.small					
2.Some.small	200				
3.Even	150	150			
4.Some.large	200	200	150		
5.More.large	200	200	150	200	

Note that all of the samples are an equal distance away from the '3.Even' sample (150 units), and all other pairs of samples are considered to be 200 units away from each other, which is the maximum possible distance we could get using the Manhattan measure on these data. This is despite the fact that, biologically, we would rather prefer '2.Some.small' to be closer to '1.More.small' than it is to (say) '4.Some.large' or '5.More.large'. You can see that the above analysis completely ignores the fact that the variables themselves are ordered.

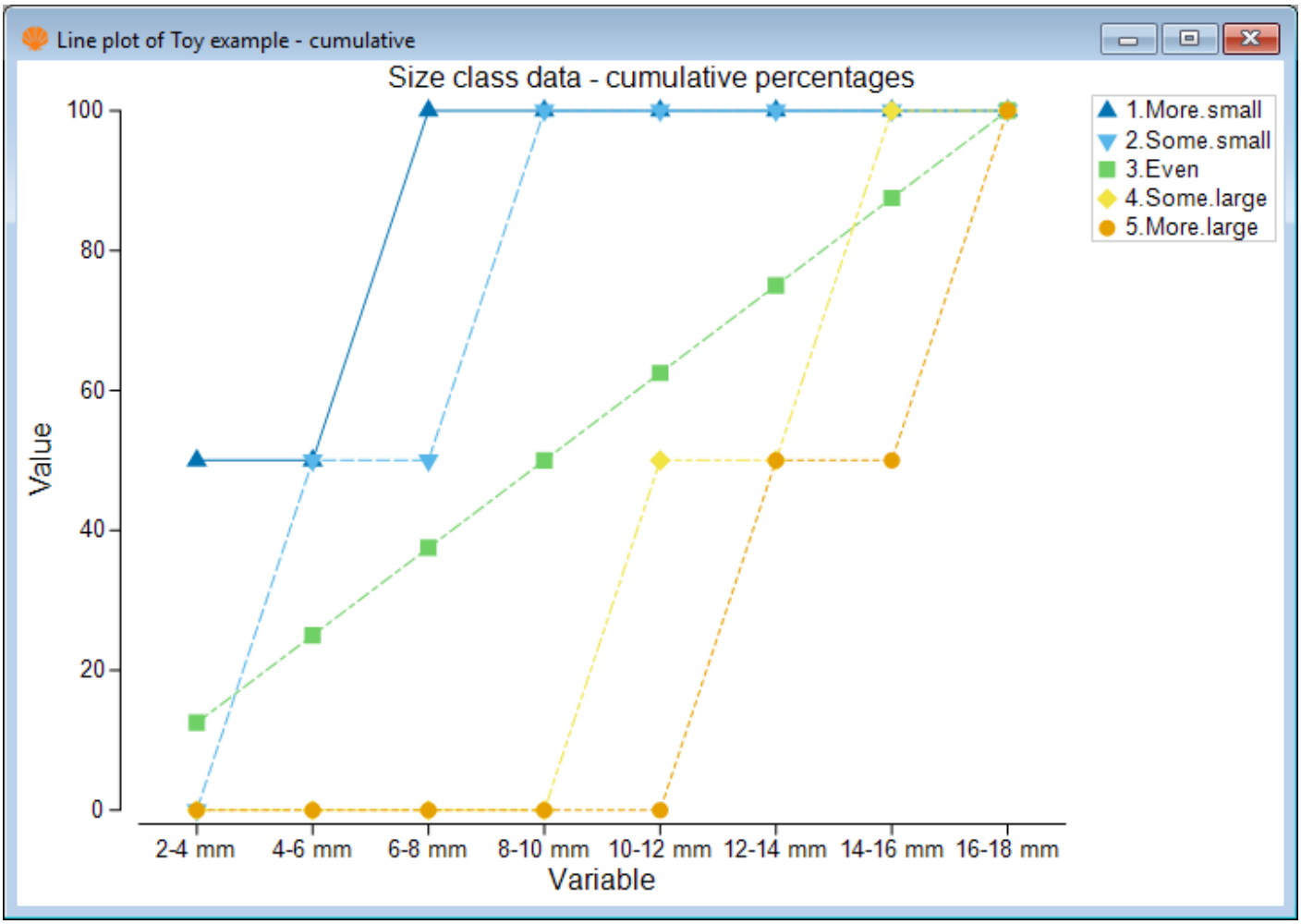
Now, let’s standardise the above raw percentages to a cumulative profile of percentages. Here, we cumulatively add the percentages in each size class across the sample so that, by the end of the list, we arrive at 100%. The cumulative percentages for this toy example look like this:



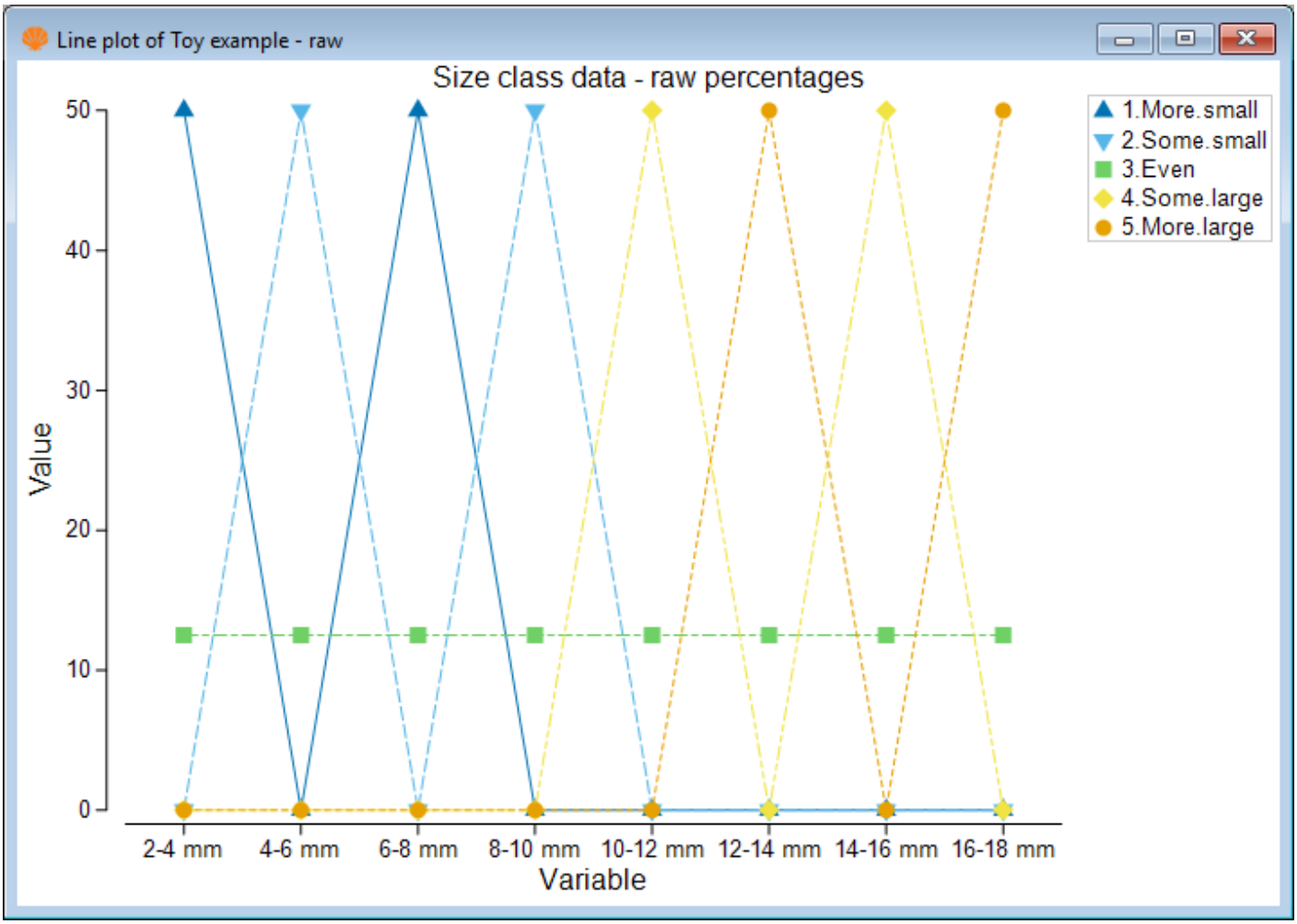
	1.More.small	2.Some.small	3.Even	4.Some.large	5.More.large
2-4 mm	50	0	12.5	0	0
4-6 mm	50	50	25	0	0
6-8 mm	100	50	37.5	0	0
8-10 mm	100	100	50	0	0
10-12 mm	100	100	62.5	50	0
12-14 mm	100	100	75	50	50
14-16 mm	100	100	87.5	100	50
16-18 mm	100	100	100	100	100

Let's think about what this transformation to cumulative percentages has done. Consider the '3.Even' column first. We start with 12.5% of the individuals in the 2-4 mm size class, then we add another 12.5% at the next step in the profile, 4-6 mm, which gets us to 25%, then we add another 12.5% at the next step in the profile for the 6-8 mm size class, which gets us to 37.5%, and so on.

Visually, we have gone from a series of unrelated (unordered) numbers to a profile of numbers which increases from left to right along the size classes, from 0% to 100% of the sample. This is perhaps best visualised using a line plot¶, where the size classes are along the x-axis and the y-axis has the cumulative percentage values (see below). Of course, by the time we get to the final size class, we end up at 100% (and, for any given sample, we may well end up at 100% far before we reach the final size class, which is fine).

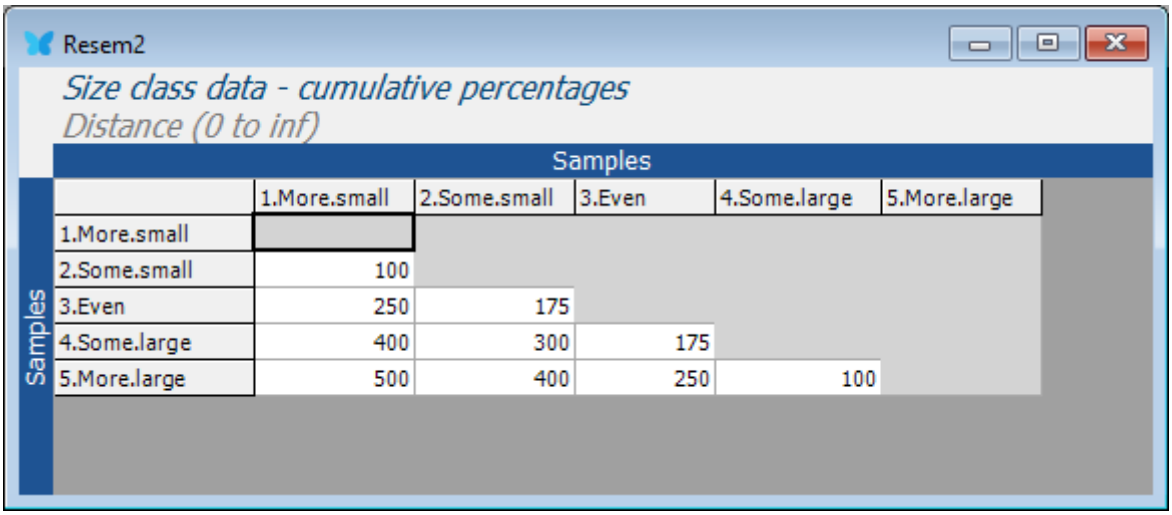


Now, in the above plot, we can see that the evenly distributed sample (in green) just marches along at equal step lengths from left to right, while the two samples that contain a large percentage of small individuals (in dark blue and light blue) reach 100% rather quickly, whereas the samples that contain larger individuals (in yellow and orange) do not increase towards 100% until the larger size classes are reached. This contrasts sharply with what the image looks like if we just plot the raw percentages:



The plot above is a bit more of a mess, and it ignores the important additional information we have about the size-class variables themselves; namely, that these variables are ordered!

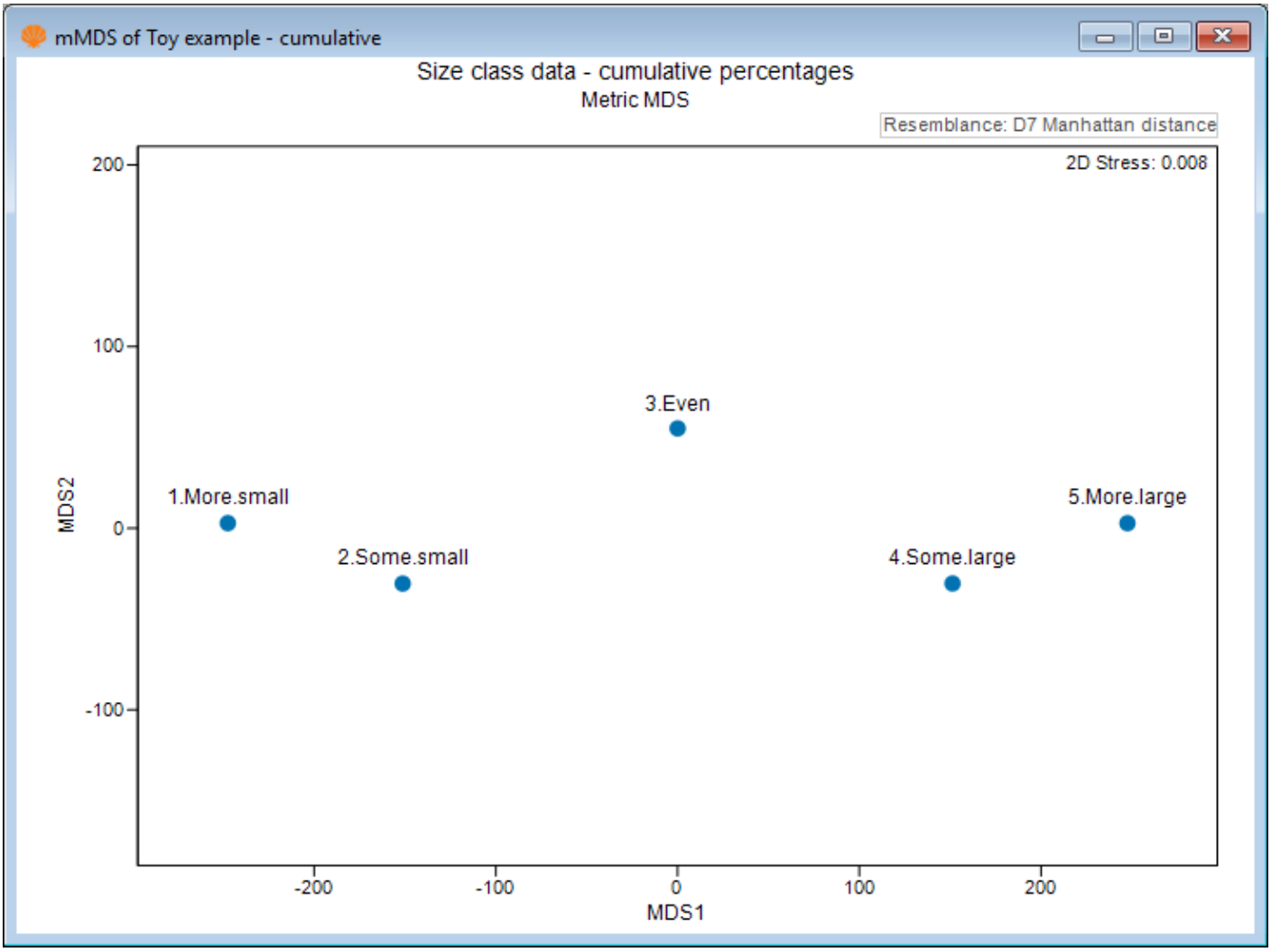
Next, if you calculate the Manhattan distances among each pair of samples using the *cumulative* percentages, it makes a lot more sense (see below).



	1.More.small	2.Some.small	3.Even	4.Some.large	5.More.large
1.More.small					
2.Some.small	100				
3.Even	250	175			
4.Some.large	400	300	175		
5.More.large	500	400	250	100	

For example, the largest distance (500 Manhattan units) is between the profile of '1.More.small' and '5.More.large'. Also, the distance between '1.More.small' and '2.Some.small' is less than that between '1.More.small' and either of '4.Some.large' or '5.More.large', and so on.

For even greater clarity, we can examine the metric MDS plot that we obtain based on the Manhattan distances calculated from raw percentages versus the one obtained from cumulative percentages (shown below).



Of course, the one using raw percentages really makes no sense at all, whereas the one based on cumulative percentages makes perfect sense!

Summary

Generally, if we are dealing with variables that are ordered, it make sense to treat them in a cumulative fashion and to analyse profiles, as demonstrated above. Another possibility would be to incorporate distances among variables in the calculation of the resemblances among samples. An example of this is taxonomic dissimilarity (Clarke *et al.* (2006b)), which includes the taxonomic or phylogenetic relationships among species in the calculation of resemblances among samples. One can, however, use any distance/dissimilarity matrix among the variables (whether they be species or not) within such a calculation. For example, Myers *et al.* (2021) used this approach to calculate functional dissimilarities among fish assemblages along depth and latitude gradients.

¶*In PRIMER 8, one has the flexibility to draw line plots either: (i) of samples (across variables, as done above) or (ii) of variables (across samples).*

13.3 Example: Mussel sizes in the Gulf of Alaska

To implement the new standardisation routine in PRIMER 8 and (simultaneously) demonstrate the utility of analysing cumulative percentages, we shall examine a study by Dowling (2021) , who measured the lengths of mussels (*Mytilus trossulus*) at two glacially influenced estuaries in the Gulf of Alaska, USA. Data were counts of mussels falling into each of 35 size classes (0-2 mm, 2-4 mm, etc. up to 90-92 mm) at each of 15 inter-tidal sites that were classified as being influenced predominantly by glacial, riverine or oceanographic hydrological conditions (Fig. 13.2). Each size class is a variable and these variables, themselves, have a natural quantitative ordering. We are interested in the shape of the distribution of size classes, and not in the total number of mussels at any given site. Are mussel size frequency distributions affected by the environmental conditions at these sites? What we want here is to standardise the data by sample totals (hence, transforming them to percentages), but – in addition – it makes sense to consider them as cumulative percentages (from the smallest to the largest mussels), as opposed to treating these variables as if they were unordered.

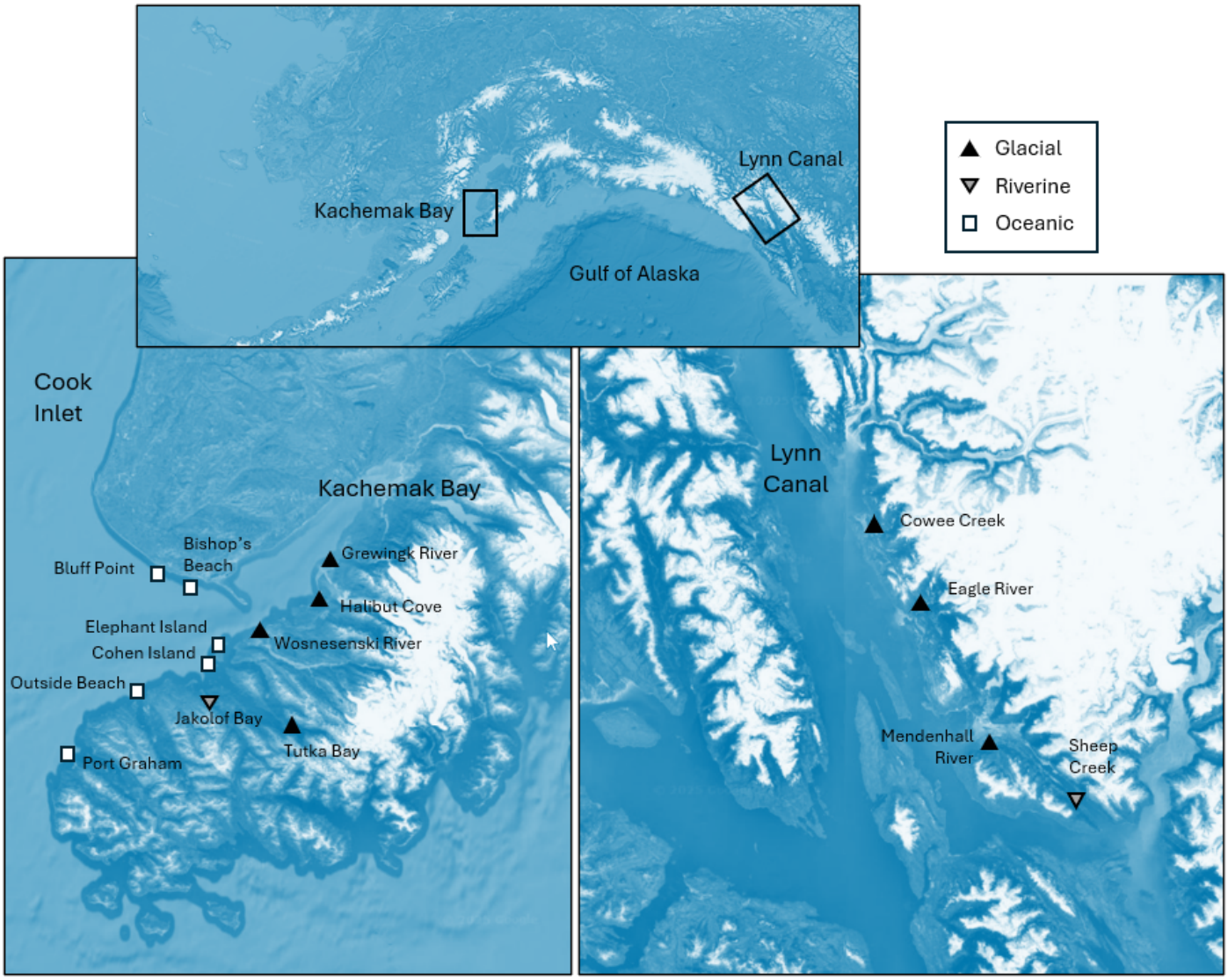
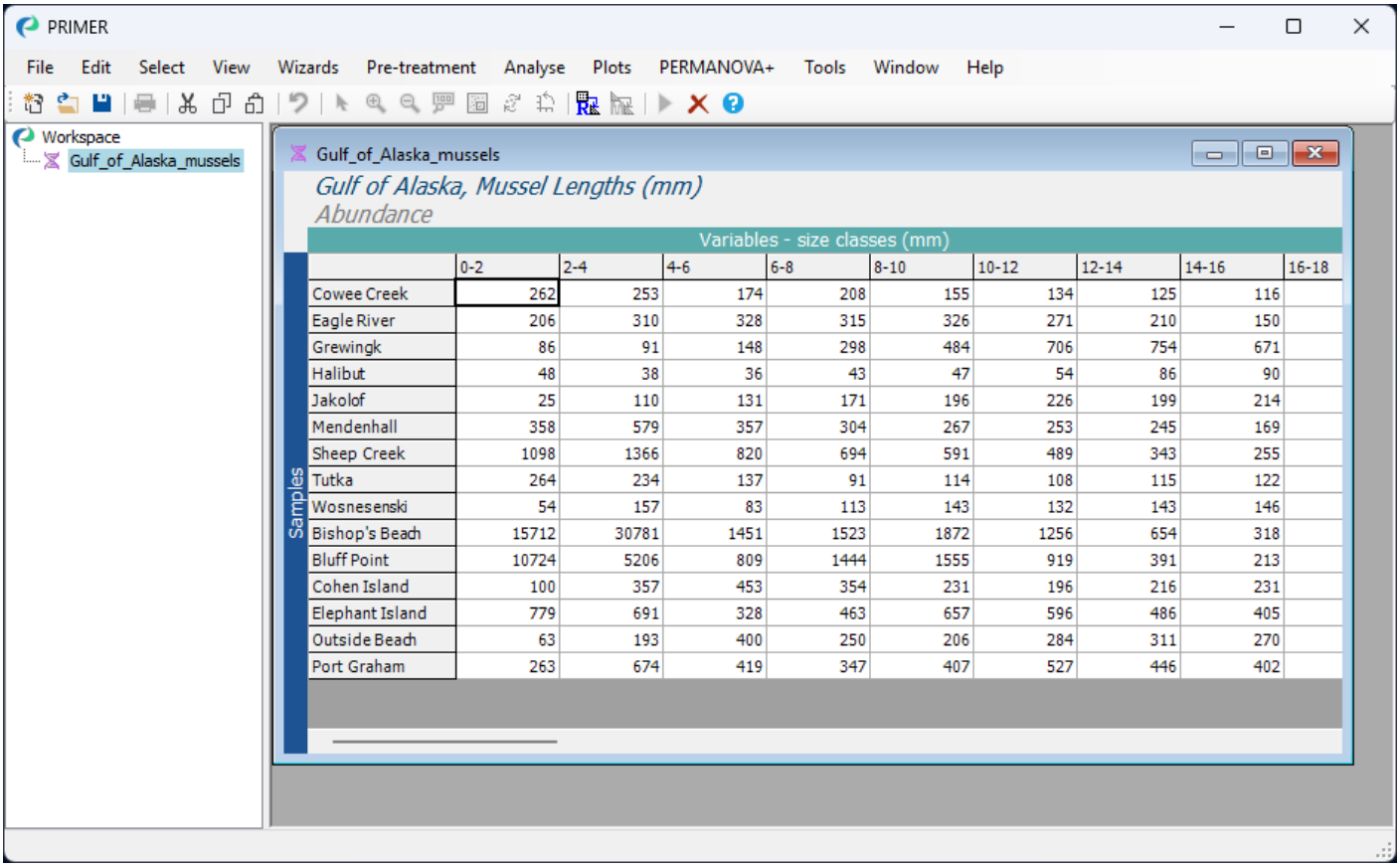


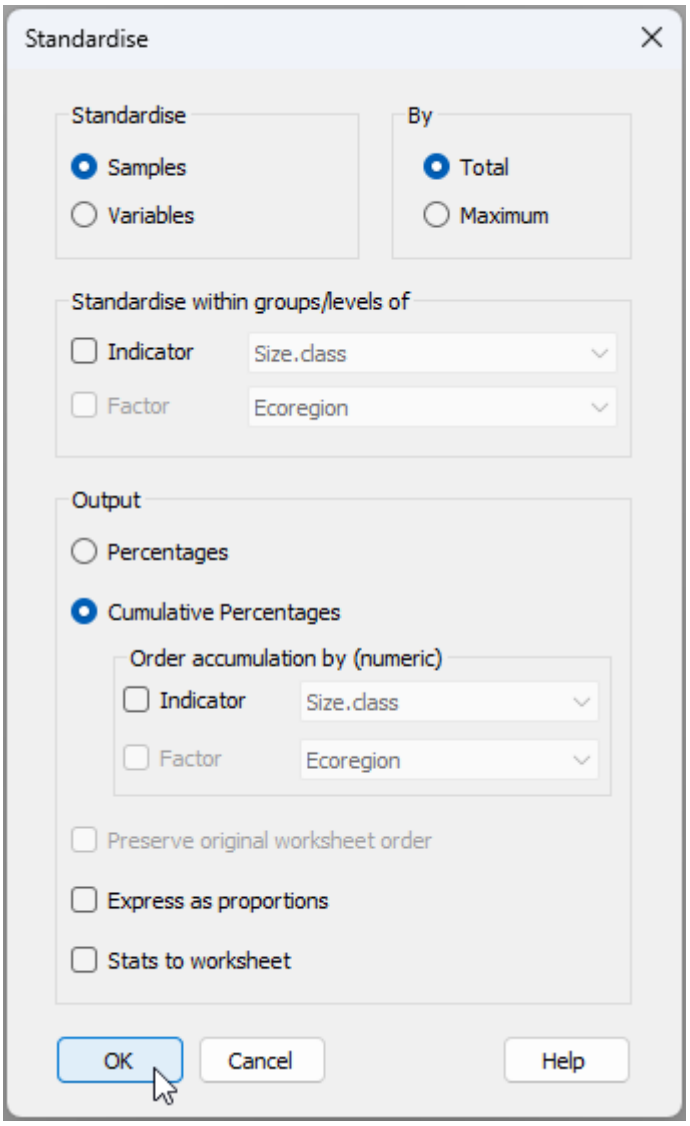
Fig. 13.2 (adapted from Fig. 1 in Dowling (2021)). Maps of the sites at which size-distributions of mussels were measured in each of two ecoregions: Kachemak Bay (left) and Lynn Canal (right). The top map shows the locations of these two ecoregions in Alaska. White squares represent ocean-influenced sites; solid black triangles represent glacially-influenced sites, and upside-down grey triangles represent freshwater-influenced sites. Satellite images: Google Earth.

Input data and standardise

1. Open up the data file 'Gulf_of_Alaska_mussels.pri' (located in the folder 'Examples_P8' > 'Gulf_of_Alaska_mussels'), and notice that the names of the variables here are the (numerical) size classes, listed in order. The values in the data sheet are the raw abundances (counts) of the numbers of mussels in each size class at each of the above sites (Fig. 13.2).



2. From the 'Gulf_of_Alaska_mussels' data sheet, click **Pre-treatment** > **Standardise...**. Choose to standardise samples by total and output cumulative percentages, as shown in the dialog below.



The resulting datasheet of cumulative percentages (called 'Data1') will look like this:

Data1

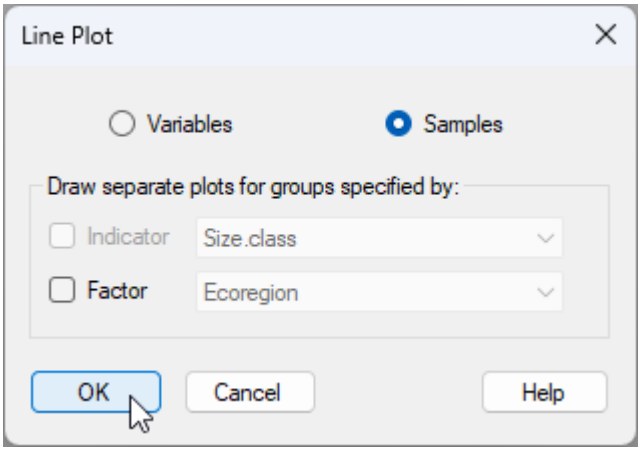
Gulf of Alaska, Mussel Lengths (mm)

Abundance

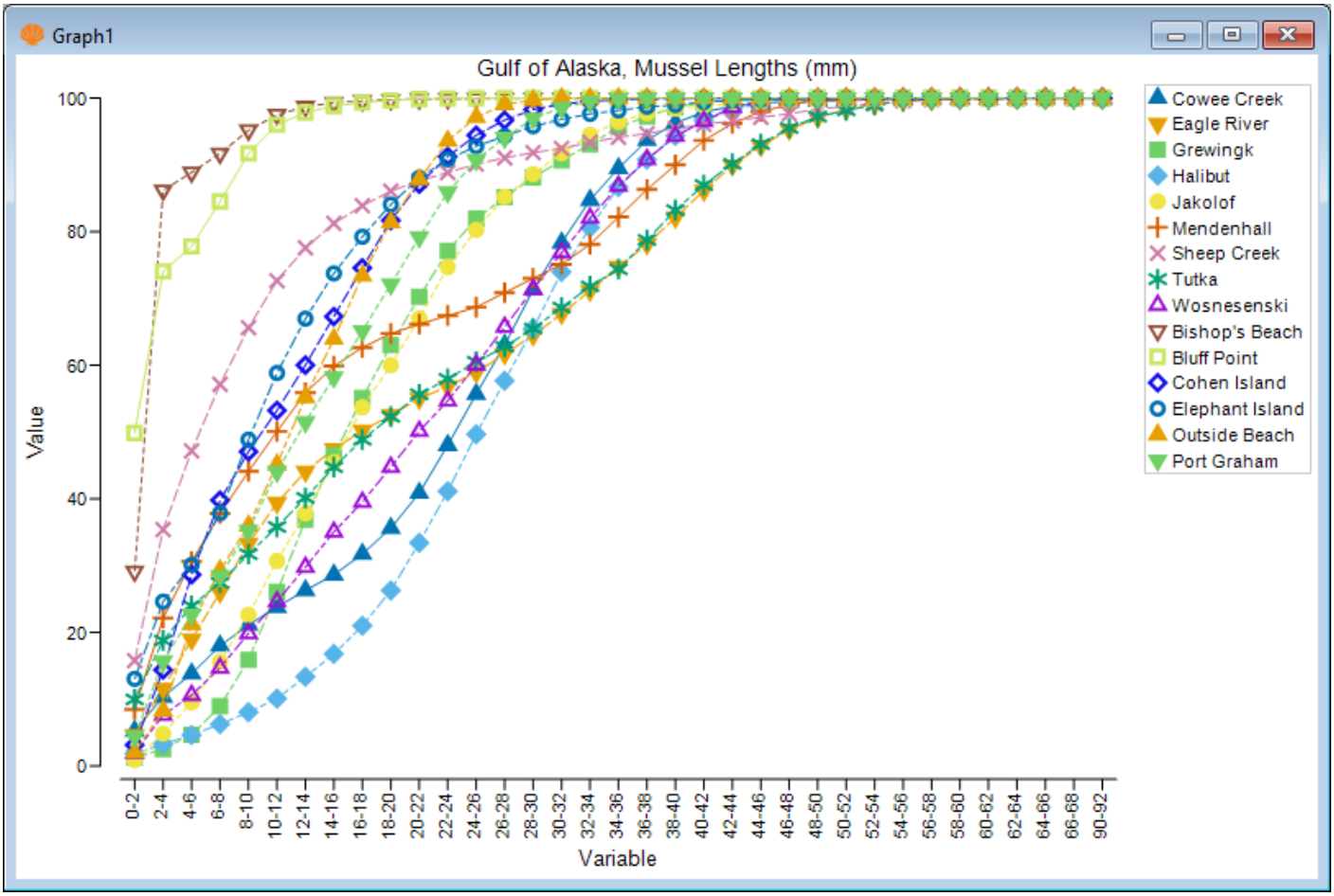
Samples	Variables - size classes (mm)							
	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16
Cowee Creek	5.2579	10.335	13.827	18.001	21.112	23.801	26.309	28.6
Eagle River	4.6209	11.575	18.932	25.998	33.311	39.39	44.1	47.4
Grewingk	1.2355	2.5427	4.6689	8.9499	15.903	26.045	36.877	46.5
Halibut	1.823	3.2662	4.6335	6.2666	8.0517	10.103	13.369	16.7
Jakolof	0.89381	4.8266	9.5102	15.624	22.631	30.711	37.826	45.4
Mendenhall	8.4674	22.162	30.605	37.796	44.111	50.095	55.889	59.8
Sheep Creek	15.776	35.402	47.184	57.155	65.647	72.672	77.601	81.2
Tutka	9.9698	18.807	23.98	27.417	31.722	35.801	40.144	44.7
Wosnesenski	1.9495	7.6173	10.614	14.693	19.856	24.621	29.783	35.0
Bishop's Beach	29.134	86.21	88.9	91.724	95.196	97.525	98.737	99.3
Bluff Point	49.854	74.055	77.816	84.529	91.758	96.03	97.848	98.8
Cohen Island	3.1486	14.389	28.652	39.798	47.072	53.243	60.044	67.3
Elephant Island	13.046	24.619	30.112	37.866	48.87	58.851	66.99	73.7
Outside Beach	2.0382	8.2821	21.223	29.311	35.975	45.163	55.225	63
Port Graham	4.3936	15.653	22.653	28.45	35.249	44.053	51.504	58.2

View size-distribution profiles

3. We can use a line plot to view the size-distribution profiles across the sites. From 'Data1' click **Plots...** > **Line Plot...** and choose to plot ($\bullet$Samples), then click '**OK**', as shown below.†



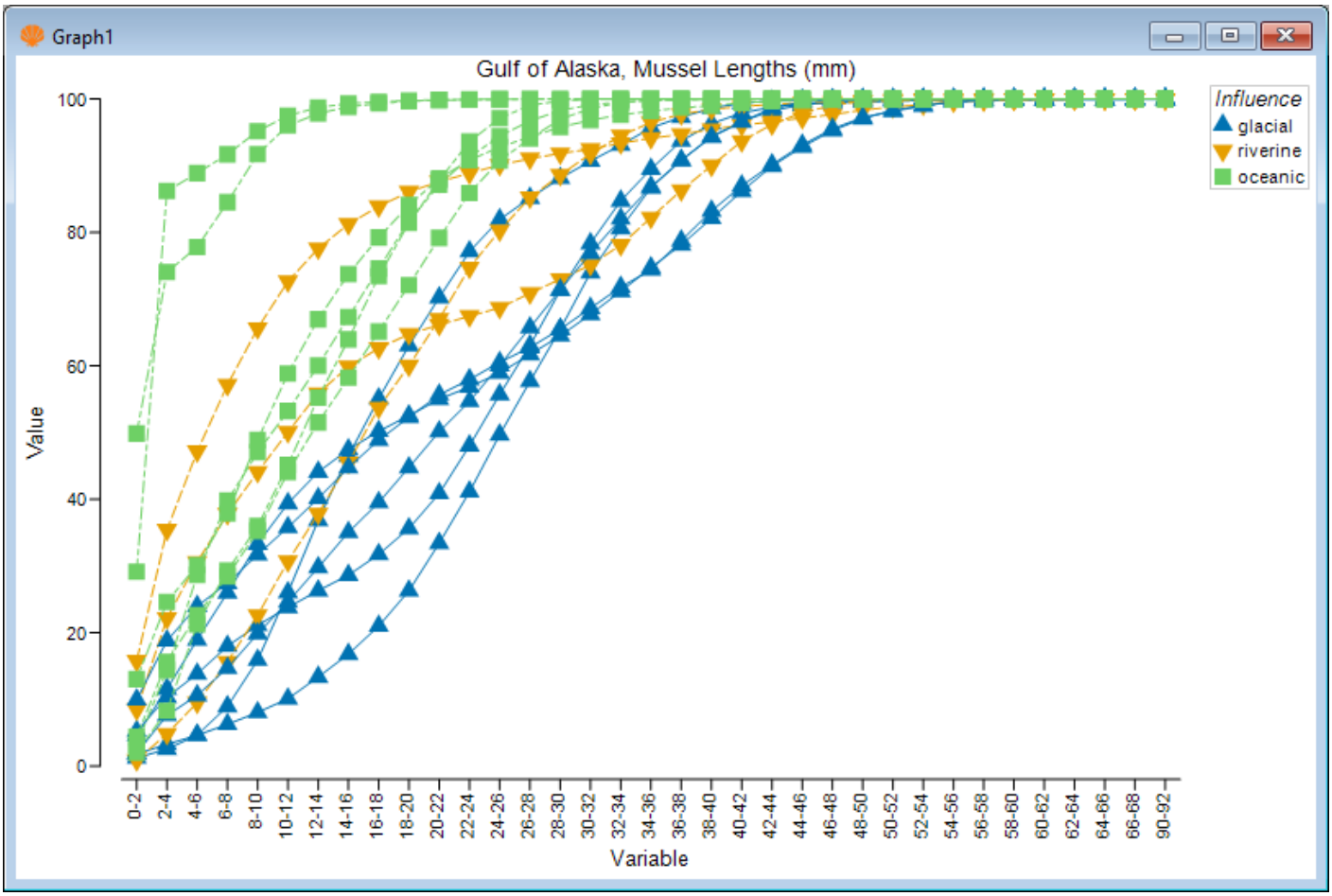
The resulting plot of cumulative size-distribution profiles for each site looks like this:



Some sites, such as Bishop's Beach and Bluff Point, are dominated by small-sized mussels (new recruits), whereas other sites (such as Halibut) tend to have a greater percentage of larger mussels.

4. It would be helpful to see these lines in different colours corresponding to the different influences (glacial, riverine or oceanic). From the plot produced at step 3 (called 'Graph1'), click **Graph** > **Sample Labels & Symbols...** and choose to plot Symbols: $\checkmark$ By factor Influence, then click '**OK**'.

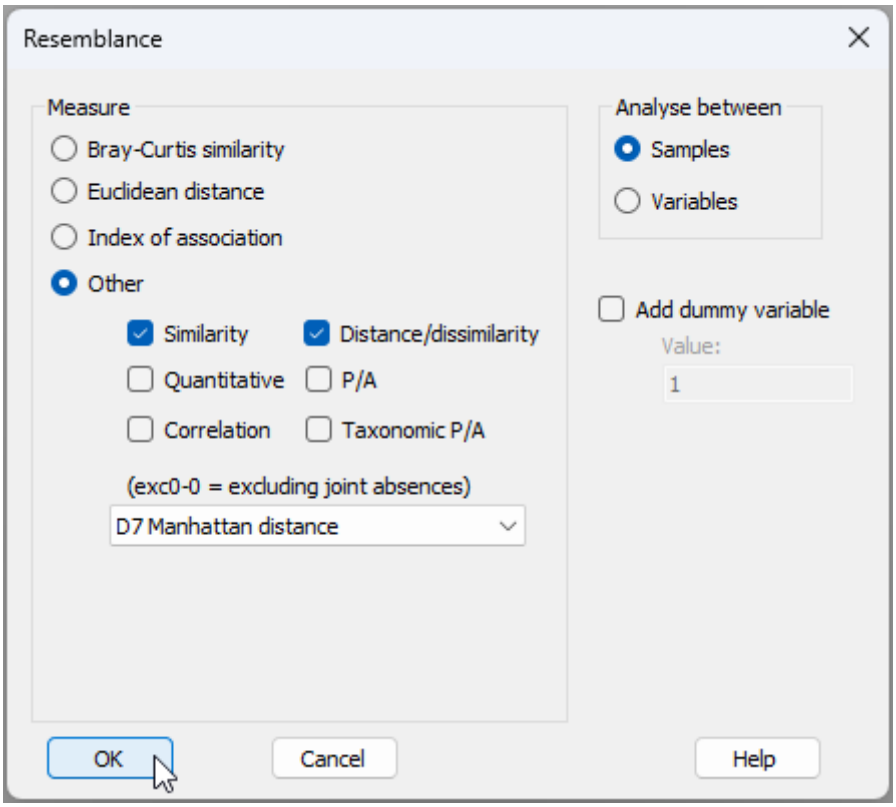
Our plot of cumulative size-distribution profiles, now showing the different influences, looks like this:



Ordination and analysis of size-frequency distributions

Let's look at an MDS plot of Manhattan distances between all pairs of cumulative curves here to get a multivariate representation of the relationships among these size frequency distributions. We consider the Manhattan distances here to be a really useful and straightforward way of calculating relationships among these cumulative profile curves. Specifically, to get the Manhattan distance between two cumulative profiles, we calculate the absolute difference between the two curves at each size class, then sum these absolute differences up across all of the size classes. So, the larger the 'gaps' are between any two profiles, the bigger the Manhattan distance.

5. Calculate Manhattan distances among the profiles. From the standardised cumulative data in 'Data1', click **Analyse** > **Resemblance...** > (Measure $\bullet$Other > D7 Manhattan distance).



The resulting matrix will be called 'Resem1'.

Resem1

Gulf of Alaska, Mussel Lengths (mm)

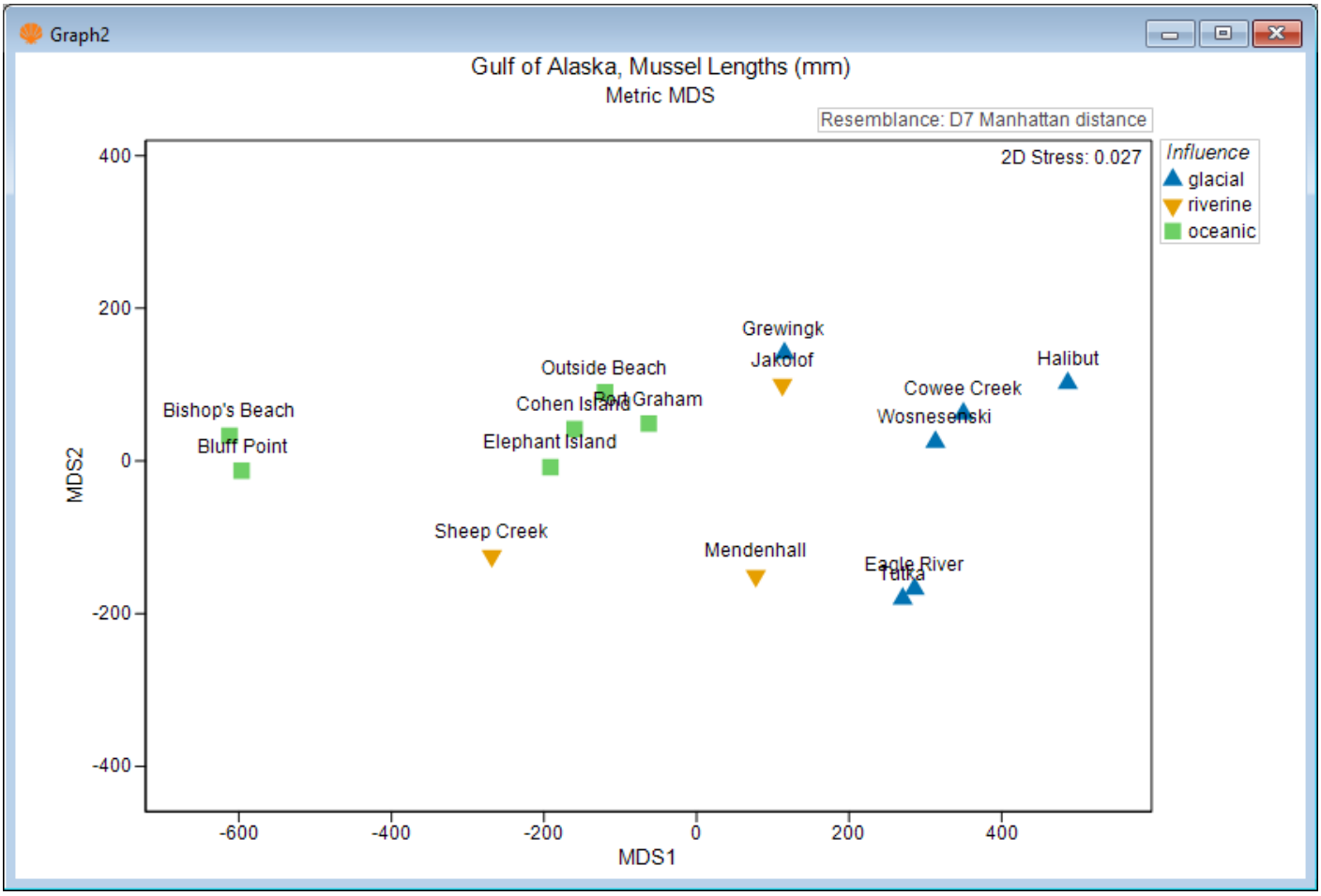
Distance (0 to inf)

Samples

Samples	Cowee Creek	Eagle River	Grewingk	Halibut	Jakolof	Mendenhall	Sheep Creek	Tutka
Cowee Creek								
Eagle River	254.25							
Grewingk	277.98	346.48						
Halibut	153.71	356.33	361.71					
Jakolof	255.13	311.43	46.45	376.61				
Mendenhall	329.12	224.16	308.12	447.43	282.65			
Sheep Creek	618.94	568.93	433.86	767.03	418.34	356.86		
Tutka	255.12	44.14	348.25	360.31	311.27	211.01	553.18	
Wosnesenski	79.546	212.61	224.76	177.6	208.63	279.4	592.24	212.1
Bishop's Beach	955.32	928.97	747.81	1106.9	733.33	704.81	360.04	912.8
Bluff Point	939.2	912.86	731.69	1090.8	717.21	688.7	343.93	896.
Cohen Island	509.7	482.08	298.38	657.37	283.49	285.33	216.67	485.5
Elephant Island	538.18	511.84	331.02	689.76	316.19	288.66	153.34	495.6
Outside Beach	473.59	448.45	255.54	614.63	241.05	316.36	281.08	462.9
Port Graham	413.58	385.96	204.34	563.42	189.85	259.02	289.9	389.4

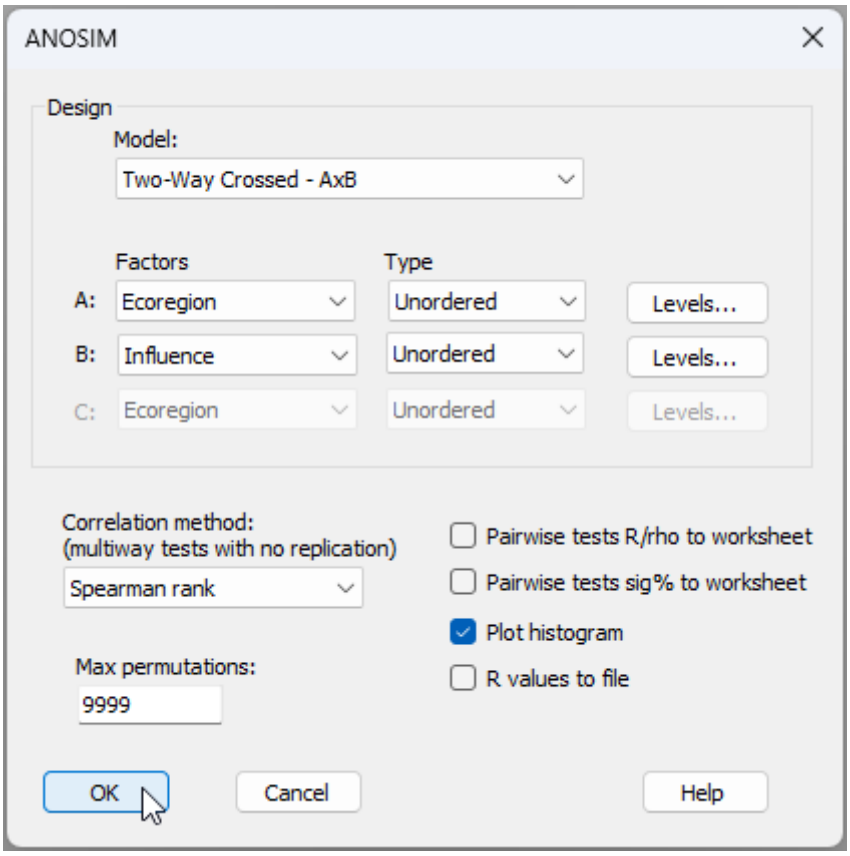
6. Create a metric MDS plot. Note that, for this example, a metric MDS can be done quite successfully.¶ From the 'Resem1' matrix, click **Analyse** > **MDS** > **Metric MDS (mMDS / tmMDS)...** > (Choice of intercept: $\bullet$Metric MDS (zero intercept)), **OK**. Put symbols on the resulting MDS plot (called 'Graph2' in the Explorer tree) corresponding to the factor 'Influence' by clicking **Graph** > **Sample Labels & Symbols...**, and the resulting

ordination graphic will look like the one shown below.



This plot shows fairly clear differences in the size-class distributions for mussels growing under the influence of glacial, riverine and oceanic conditions. There is clearly a gradient across the plot in the size-class distributions of mussels, from sites on the left (typically under oceanic influences) having proportionately more small-sized mussels (e.g., Bishop's Beach and Bluff Point), through to those on the right (typically experiencing more glacial influences) which have a greater proportion of large-sized mussels (e.g., Cowee Creek and Halibut). We can formally test the effects of influence on size-class distributions of mussels at these sites formally using ANOSIM, also taking into account potential differences in the two ecoregions (Kachemak Bay and Lynn Canal).

7. Do a two-way crossed ANOSIM for the factors of 'Ecoregion' and 'Influence' on mussel size-class distributions. From the 'Resem1' matrix, click **Analyse** > **ANOSIM...**, then choose Design > Model: Two-way Crossed - AxB, with the two factors being A: Ecoregion (unordered) and B: Influence (also unordered), leave all other options as the defaults and click **OK**, like so:



The ANOSIM results are shown below.

ANOSIM1

ANOSIM
Analysis of Similarities

Two-Way Crossed - AxB

Resemblance worksheet
Name: Resem1
Data type: Distance
Selection: All

Factors

Place	Name	Type	Levels
A	Ecoregion	Unordered	2
B	Influence	Unordered	3

Ecoregion levels
Lynn Canal
Kachemak Bay

Influence levels
glacial
riverine
oceanic

Tests for differences between unordered Ecoregion groups
(across all Influence groups)
Global Test
Sample statistic (Average R): -0.286
Significance level of sample statistic: 82.2% (P-Value: 0.8222)
Number of permutations: 45 (All possible permutations)
Number of permuted statistics greater than or equal to Average R: 37

Tests for differences between unordered Influence groups
(across all Ecoregion groups)
Global Test
Sample statistic (Average R): 0.483
Significance level of sample statistic: 0.7% (P-Value: 0.0071)
Number of permutations: 6930 (All possible permutations)
Number of permuted statistics greater than or equal to Average R: 49

Pairwise Tests

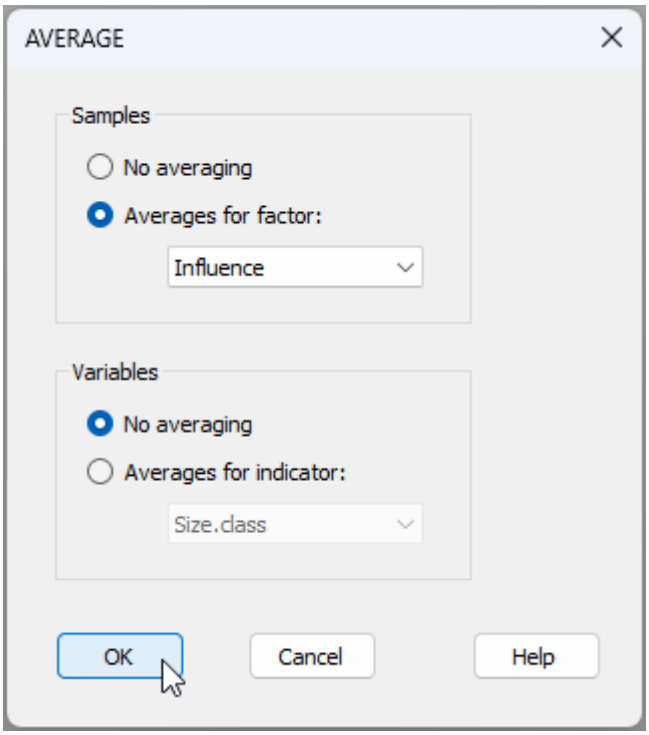
Groups	R Statistic	Significance Level %	P-Value	Possible Permutations	Actual Permutations	Number >= Observed
glacial, riverine	-0.07	66.7	0.6667	15	15	10
glacial, oceanic	0.679	0.5	0.0048	210	210	1
riverine, oceanic	0.289	42.9	0.4286	7	7	3

Outputs
Plot: Graph6
Plot: Graph7

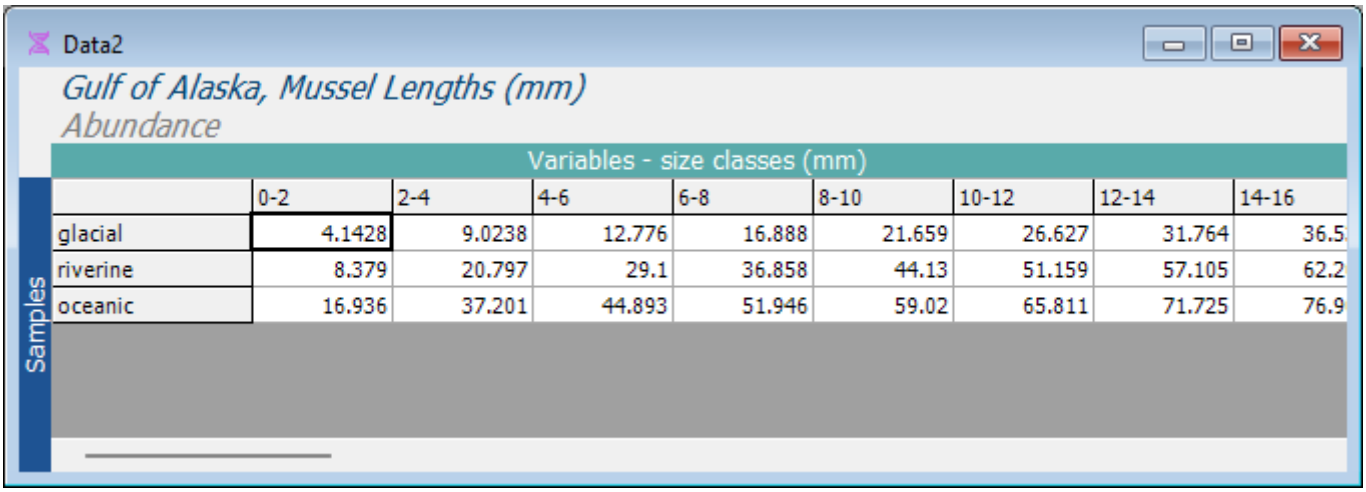
These results are quite clear. There is no effect of Ecoregion (ANOSIM R = -0.286, P > 0.80), but Influence has a statistically significant effect on the size-class distributions of these mussels (ANOSIM R = 0.483, P < 0.01). Moreover, the pair-wise tests show significantly different size-class distributions for mussels at sites under glacial *vs* oceanic influences (ANOSIM R = 0.679, P < 0.005), but distributions of mussels at sites under riverine influences (amber symbols in the mMDS plot) apparently lie somewhere in-between these two and they do not differ significantly from either of them (P > 0.40 for both tests).

Plot mean size-distribution profiles

8. We can show the mean size-distribution profiles for each of these three 'Influence' groups, as follows. From the full set of cumulative profiles for all of the sites held in 'Data1', click **Tools** > **Average...** and choose (Samples $\bullet$Averages for factor: Influence), then click '**OK**', like so:

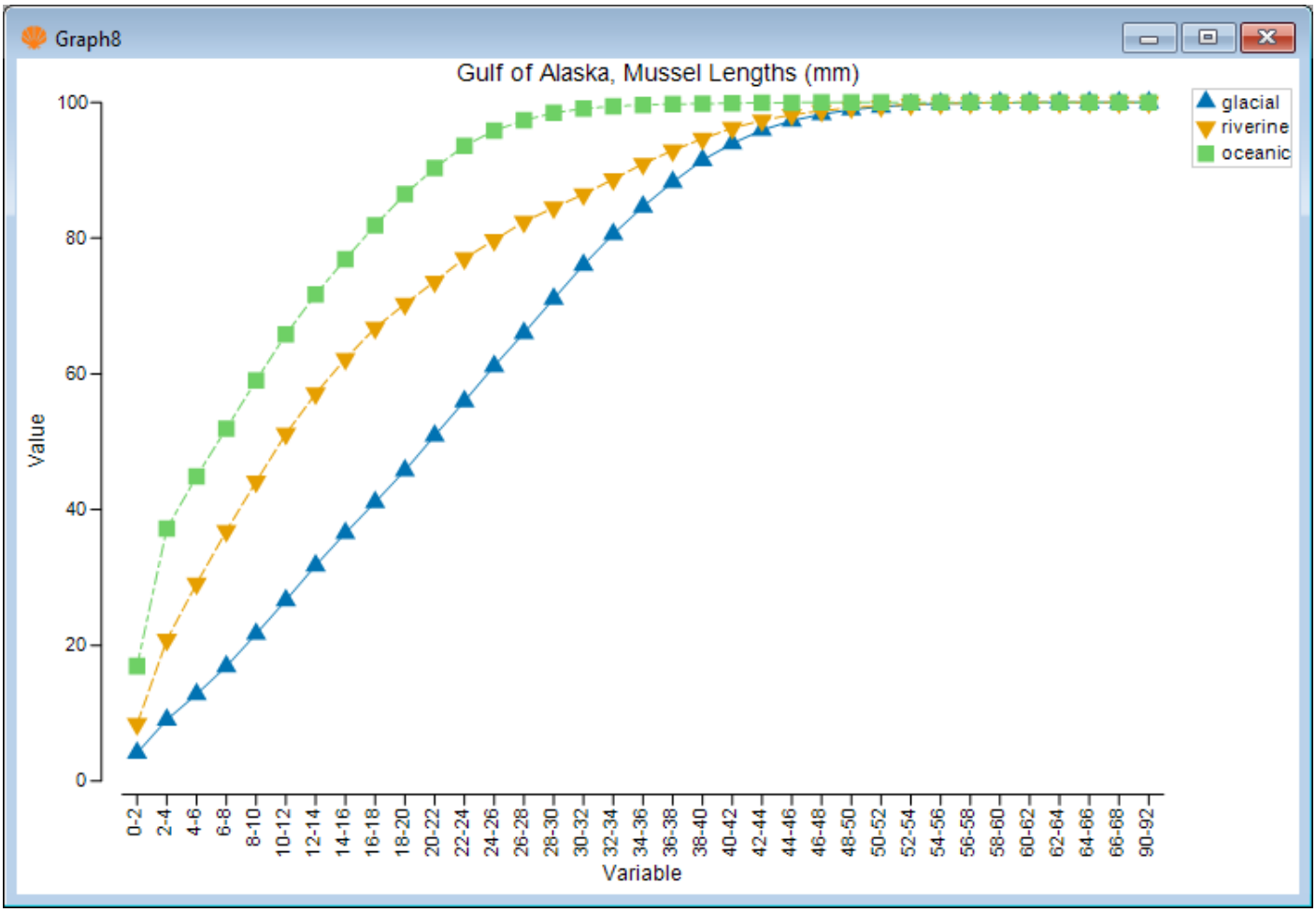


This will produce a data file with just three profiles in it, called 'Data2':



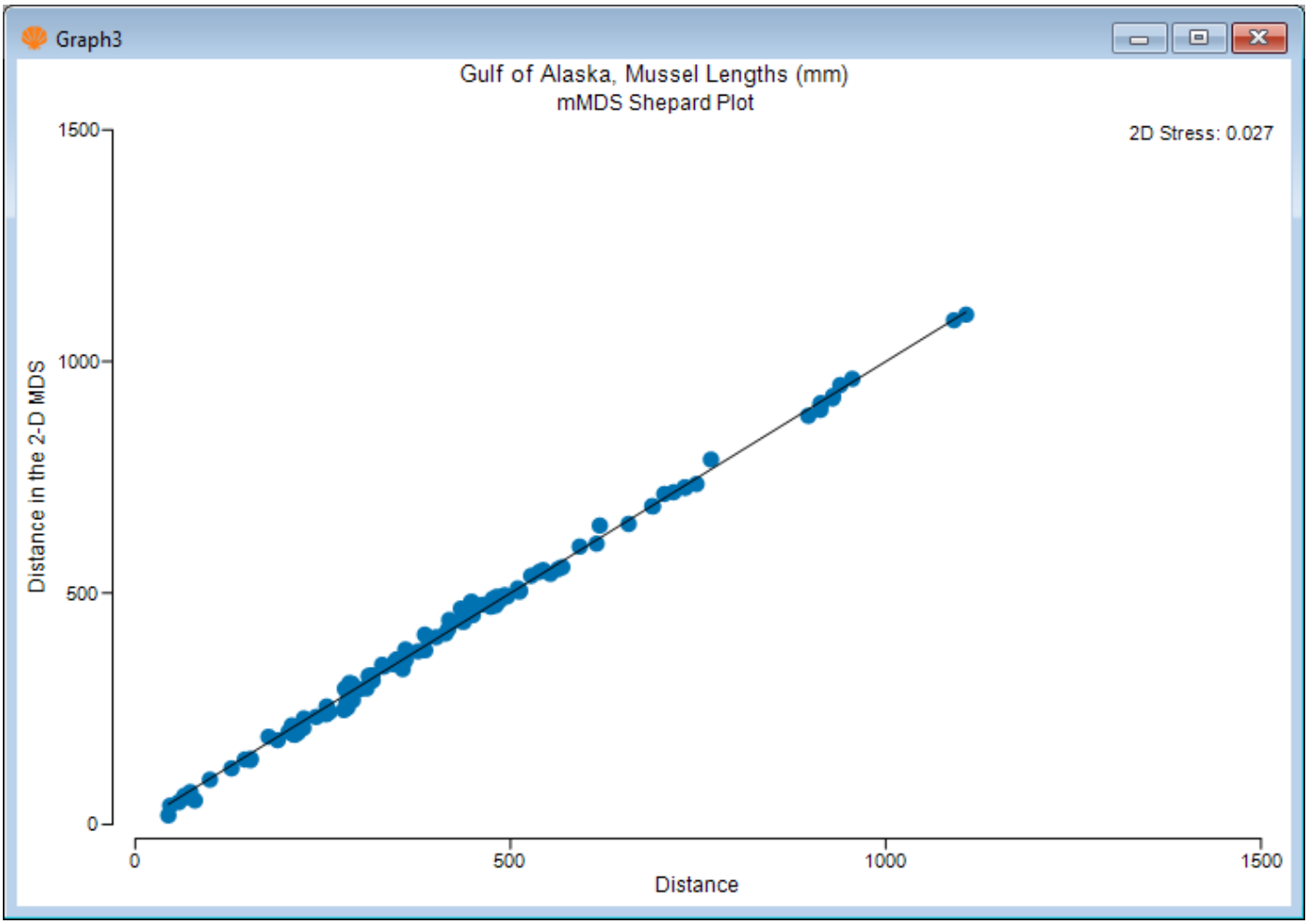
	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16
glacial	4.1428	9.0238	12.776	16.888	21.659	26.627	31.764	36.5
riverine	8.379	20.797	29.1	36.858	44.13	51.159	57.105	62.2
oceanic	16.936	37.201	44.893	51.946	59.02	65.811	71.725	76.9

From 'Data2', click **Plots** > **Line Plot...**, choose to plot ($\bullet$Samples), then click '**OK**', and we have the following graphical output ('Graph8').



† *It is a new feature in PRIMER 8 to be able to create a line plot of samples (across variables). PRIMER 7 only permitted line plots to be drawn of variables (across samples).*

¶ *We would typically use non-metric MDS for most cases, but in situations where the Shepard diagram shows an approximately linear relationship between the Euclidean distances in the MDS plot and the original dissimilarities, we can move towards using a metric MDS. Although threshold metric MDS is often then our next 'go-to' tool, we can actually go for a fully metric MDS in cases where a zero intercept is also feasible yet without incurring too much stress, as in the present case. We accumulate advantages for interpretation (i.e., distances on the MDS plot = original dissimlarities) the further we can get down the path from nMDS → tmMDS → mMDS. The Shepard diagram for the present case ('Graph3'), demonstrating both linearity and the perfect suitability of a zero intercept, is shown below.*


Graph3
Gulf of Alaska, Mussel Lengths (mm)
mMDS Shepard Plot
2D Stress: 0.027
Distance in the 2-D MDS
1500
1000
500
0
0
500
1000
1500
Distance

13.4 Example: Gulf of Maine invertebrates - functional resemblance

There are many situations where the standardisation of samples is required as a pre-treatment prior to analysis, but which needs to be done *separately within groups* of variables that may be identified by an indicator.[¶] Here, we shall consider a dataset comprised of occurrences of p = 91 macroinvertebrate species recorded from intertidal areas at each of N = 12 exposed rocky headlands (surveyed in 2012) in the Gulf of Maine (Fig. 13.3, Trott (2022) [†]). For each species, we have additional information in the form of trait data. The trait data for each species consists of presence/absence information for a series of q = 93 individual trait variables. The trait variables are organised into 14 groups. These groups of traits are: Form, Position, Lifestyle, Trophic Relation, Body Shape, Body Support, Flexibility, Ecoengineer, Fertilization, Development, Size, Body Plan, Asexual Reproduction and Regeneration. The number of categories (trait variables) within each trait group varies, and any given species can be recorded as a '1' (indicating that they belong) to one or more of these categories within each trait group.

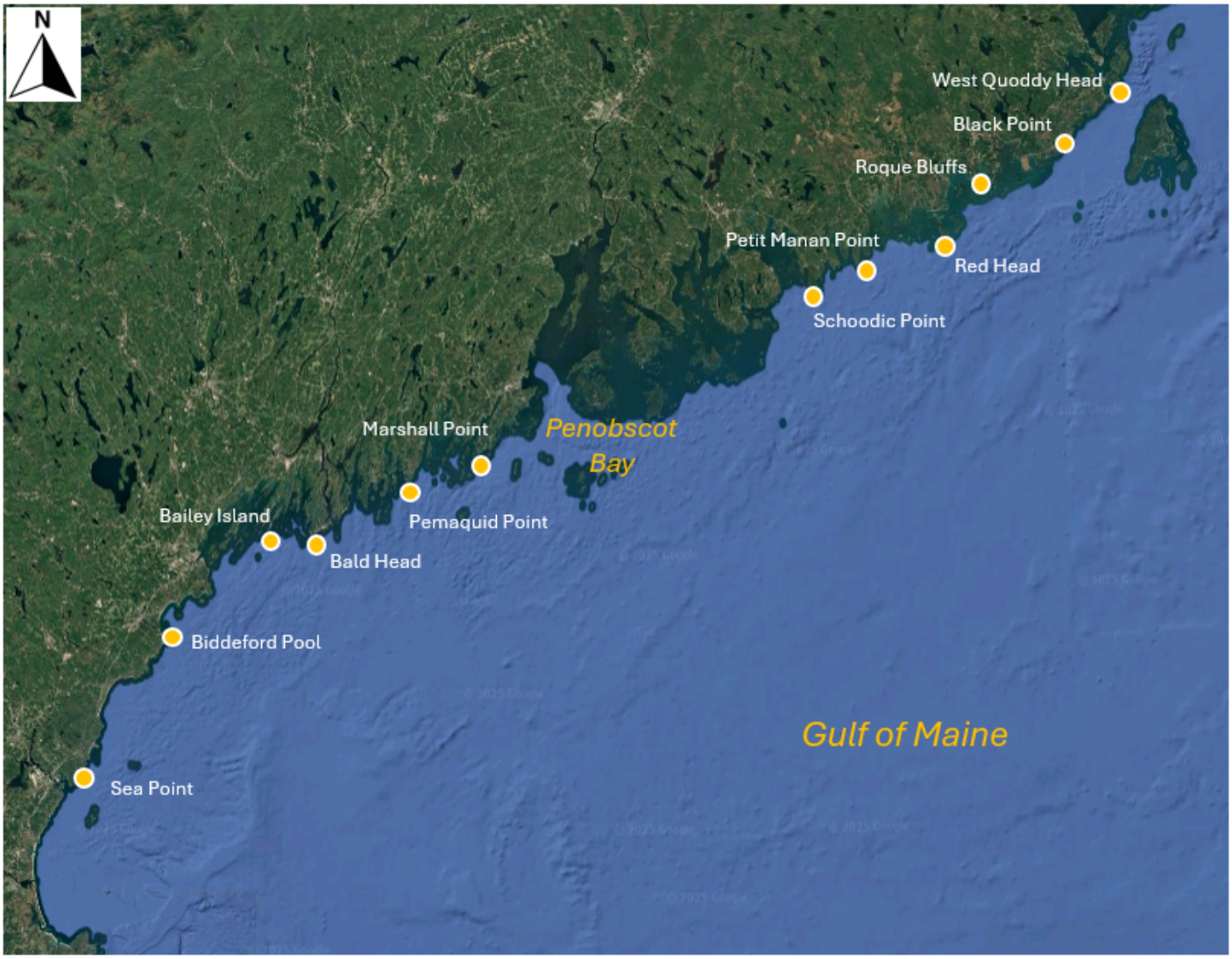


Fig. 13.3 Map of the Gulf of Maine showing 12 sites where macroinvertebrates were sampled by Trott (2022) . Satellite image: Google Earth.

Our initial agenda here may be to consider the relationships among the sites based on the species they contain in the usual way, e.g., using the Bray-Curtis (or Sørensen) resemblance measure directly on the species $\times$ site matrix. However, we may wish to nuance this calculation further by incorporating relationships among the species, based on their traits. In other words, if two sites do not share any species in common, they might nevertheless share two species that have similar traits. We can exploit the method of calculating taxonomic resemblance ('**Gamma+**', Clarke *et al.* (2006b)) in order to calculate, instead, a *functional resemblance*, using a *matrix of inter-species relationships built from the trait data* (e.g., Myers *et al.* (2021)).

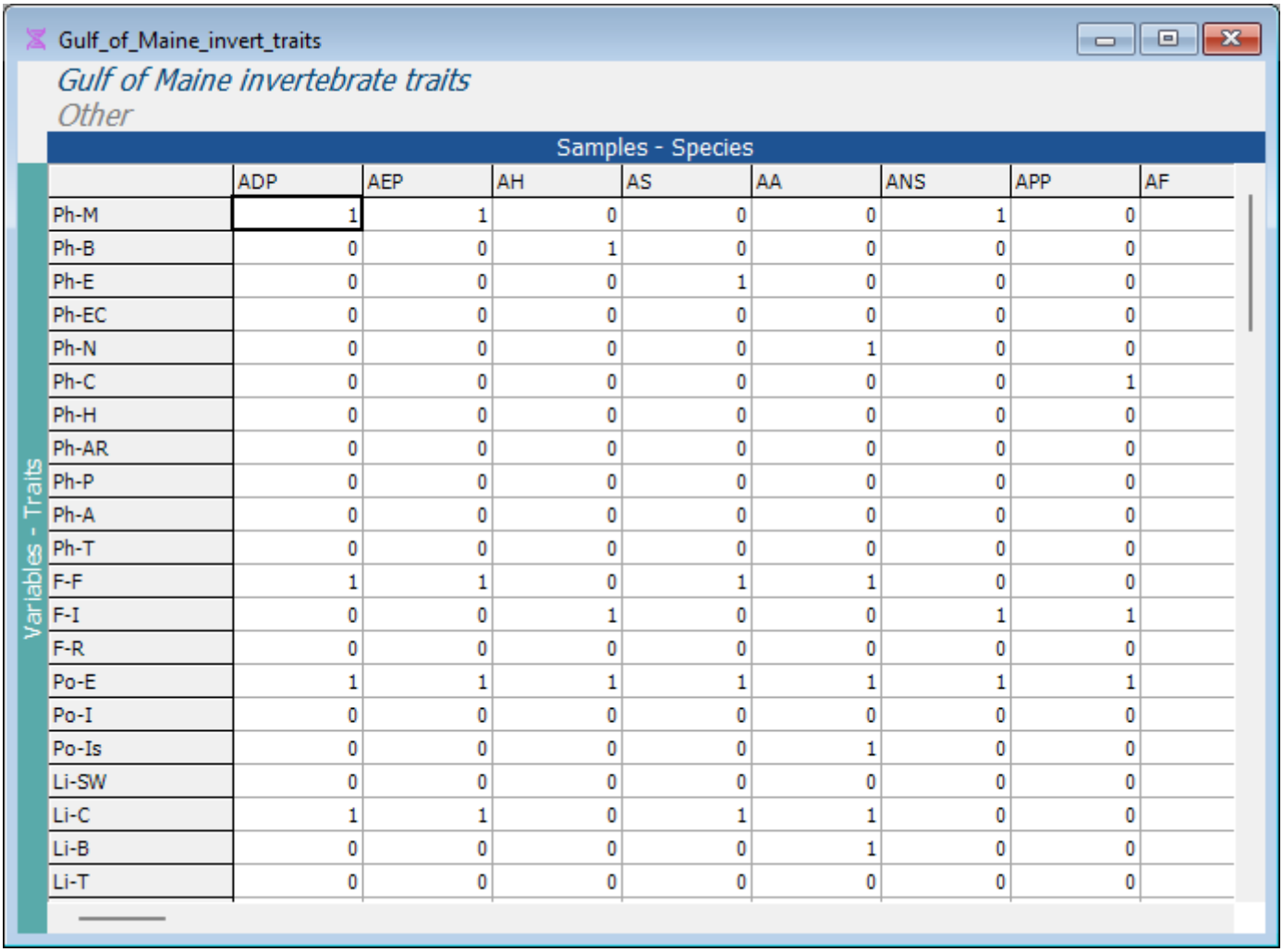
Further to this aim, we shall consider our trait data matrix such that the classical role of 'samples' is given to the species (among which we shall calculate the resemblances), while traits are given the role of 'variables'. However, we want each trait *group* to be weighted equally in the calculation. To make sure the sum of values for each species within each trait group sums to 100 (so gets equal weight), we will want to apply a pre-treatment to standardise across the trait variables for each species ('sample'), but separately within each trait grouping. This standardisation would not be necessary if every species only had one trait within a group, but that is not the case here. For example, the mollusc species, *Adalaria proxima* (abbreviated ADP) is in two trophic categories: it grazes on algal fronds and blades (Tr-P-GF), and also grazes macro prey on the substratum (Tr-P-GSM).

Our analysis pathway looks like this:

- Open the trait data matrix in PRIMER 8.
- Standardise the 'samples' (species here) across the trait variables, separately within each trait grouping, using our new tool in P8, **Pre-treatment** > **Standardise...**.
- Calculate functional resemblances among species using Bray-Curtis, based on the standardised trait data.
- Examine the inter-species functional relationships in an ordination (nMDS).
- Calculate functional resemblances among the 12 sites, by incorporating these species inter-relationships.
- Examine the functional relationships among the sites in an ordination (nMDS).

Open the trait data matrix

1. In PRIMER 8, click **File** > **Open...** and open up the file named 'Gulf_of_Maine_invert_traits.pri' (found in the 'Gulf_of_Maine_inverts' folder inside the 'Examples_P8' folder), as shown below.

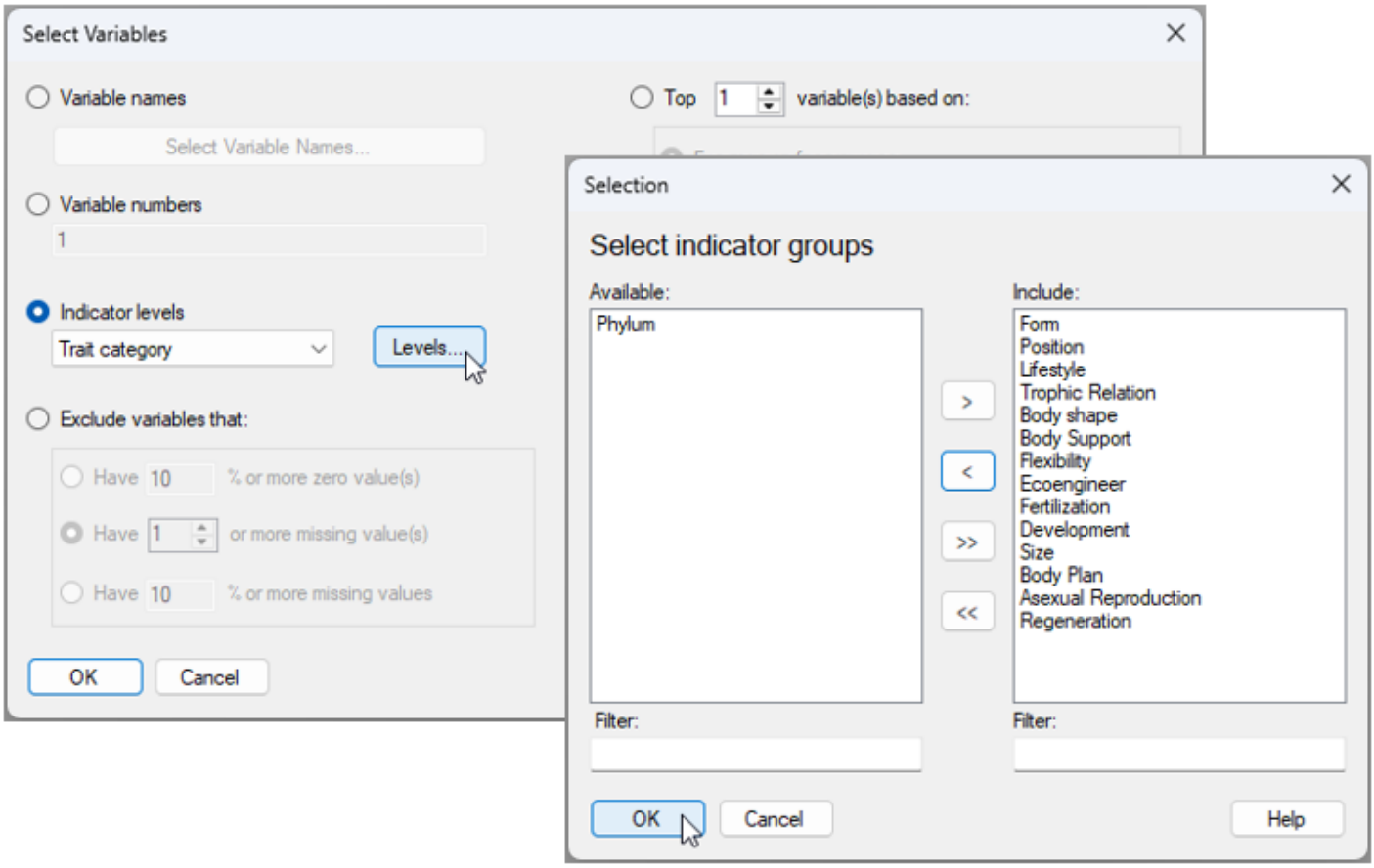


	ADP	AEP	AH	AS	AA	ANS	APP	AF
Ph-M	1	1	0	0	0	1	0	
Ph-B	0	0	1	0	0	0	0	
Ph-E	0	0	0	1	0	0	0	
Ph-EC	0	0	0	0	0	0	0	
Ph-N	0	0	0	0	1	0	0	
Ph-C	0	0	0	0	0	0	1	
Ph-H	0	0	0	0	0	0	0	
Ph-AR	0	0	0	0	0	0	0	
Ph-P	0	0	0	0	0	0	0	
Ph-A	0	0	0	0	0	0	0	
Ph-T	0	0	0	0	0	0	0	
F-F	1	1	0	1	1	0	0	
F-I	0	0	1	0	0	1	1	
F-R	0	0	0	0	0	0	0	
Po-E	1	1	1	1	1	1	1	
Po-I	0	0	0	0	0	0	0	
Po-Is	0	0	0	0	1	0	0	
Li-SW	0	0	0	0	0	0	0	
Li-C	1	1	0	1	1	0	0	
Li-B	0	0	0	0	1	0	0	
Li-T	0	0	0	0	0	0	0	

Note here that the 'Samples' are actually the species (the names are abbreviated). More information about the species can be seen by clicking **Edit** > **Factors...**. Note also that the traits are variables. More information about the traits (also abbreviated) can be seen by clicking **Edit** > **Indicators...**.

The first trait group is actually the name of the Phylum for each species. These are not actually traits, so let's start by selecting all of the traits that do not belong to the 'Phylum' group.

2. From the 'Gulf_of_Maine_invert_traits' matrix, click **Select** > **Variables...**, then choose ($\bullet$Indicator levels Trait category) and click the 'Levels...' button (Levels...). In the Selection dialog, first click the double right arrows button (>>) to move all of the groups to the 'Include' column on the right. Then click on 'Phylum' and the left arrow button (<), so that it appears back in the 'Available' column on the left (as shown below), then click '**OK**' (in both dialog windows).



This will create a dataset that is re-coloured with a blue background, showing only the selected subset of traits (excluding 'Phylum'), like this:

Gulf_of_Maine_invert_traits

Gulf of Maine invertebrate traits

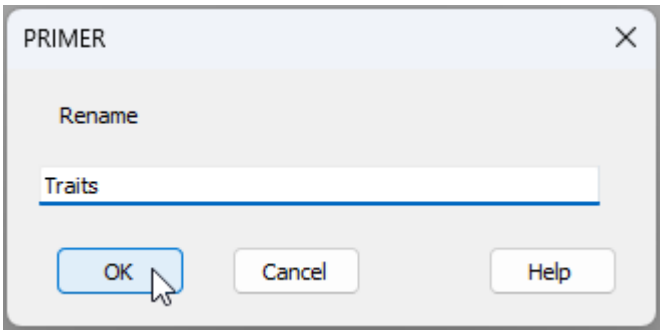
Other

Samples - Species

Variables - Traits

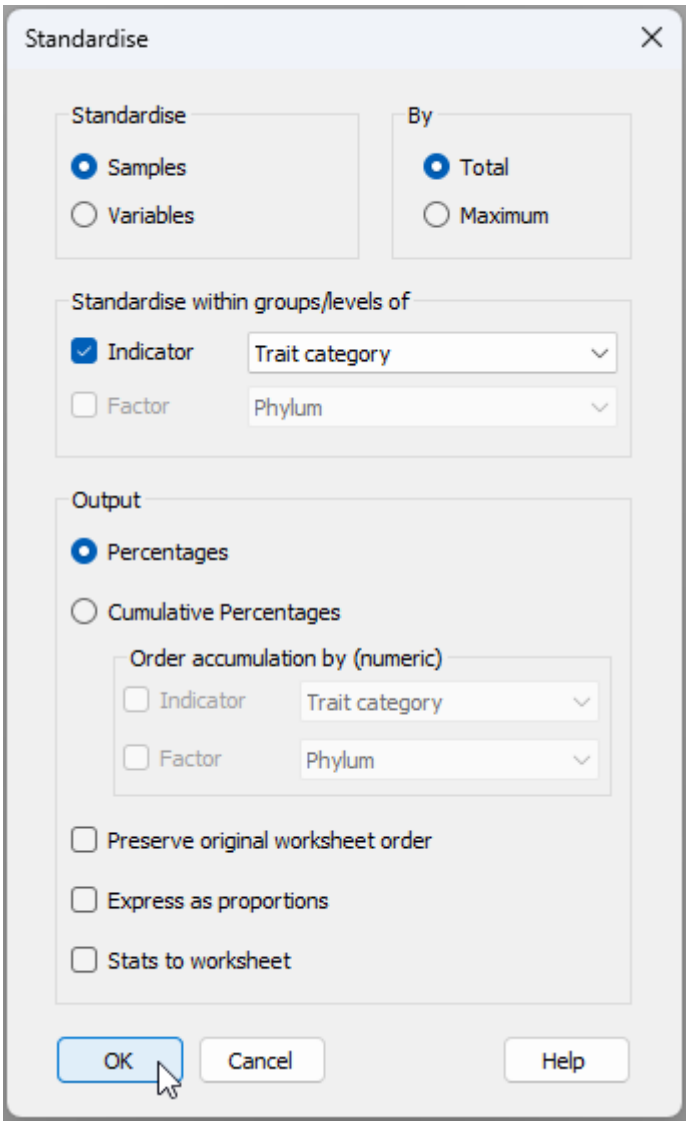
	ADP	AEP	AH	AS	AA	ANS	APP	AF
F-F	1	1	0	1	1	0	0	
F-I	0	0	1	0	0	1	1	
F-R	0	0	0	0	0	0	0	
Po-E	1	1	1	1	1	1	1	
Po-I	0	0	0	0	0	0	0	
Po-Is	0	0	0	0	1	0	0	
Li-SW	0	0	0	0	0	0	0	
Li-C	1	1	0	1	1	0	0	
Li-B	0	0	0	0	1	0	0	
Li-T	0	0	0	0	0	0	0	
Li-Sh	0	0	0	0	0	0	0	
Li-PA	0	0	1	0	0	1	1	
Li-TA	0	0	0	0	0	0	0	
Tr-Sa	0	0	1	0	0	1	1	
Tr-Sp	0	0	0	1	0	0	0	
Tr-P-SAW	0	0	0	0	0	0	0	
Tr-P-S	0	1	0	0	1	0	0	
Tr-P-P	0	0	0	0	0	0	0	
T-P-GP	0	0	0	0	0	0	0	
Tr-P-GF	1	0	0	0	0	0	0	
Tr-P-GS	0	0	0	1	0	0	0	

Any analyses from this matrix will now only be done on the sub-setted data. To keep things tidy, click **Tools** > **Duplicate** and the subsetted data will now be provided in the Explorer tree as a new matrix called 'Data1'. We can re-name this to 'Traits' by clicking **File** > **Rename Data** and typing in this desired new name, then clicking '**OK**', as shown below.



Standardise the trait data

3. From the 'Traits' data sheet, click **Pre-treatment** > **Standardise...** and choose (Standardise $\bullet$Samples) & (By $\bullet$Total) & (Standardise within groups/levels of $\checkmark$Indicator Trait category) & (Output $\bullet$Percentages), then click '**OK**', as shown below.



The resulting data sheet (called 'Data1') can be renamed (using **File** > **Rename Data**, as we have done before) to (say) 'Standardised Data', for clarity. The standardised trait data sheet will look like this:

Standardised Data

Gulf of Maine invertebrate traits

Other

Samples - Species

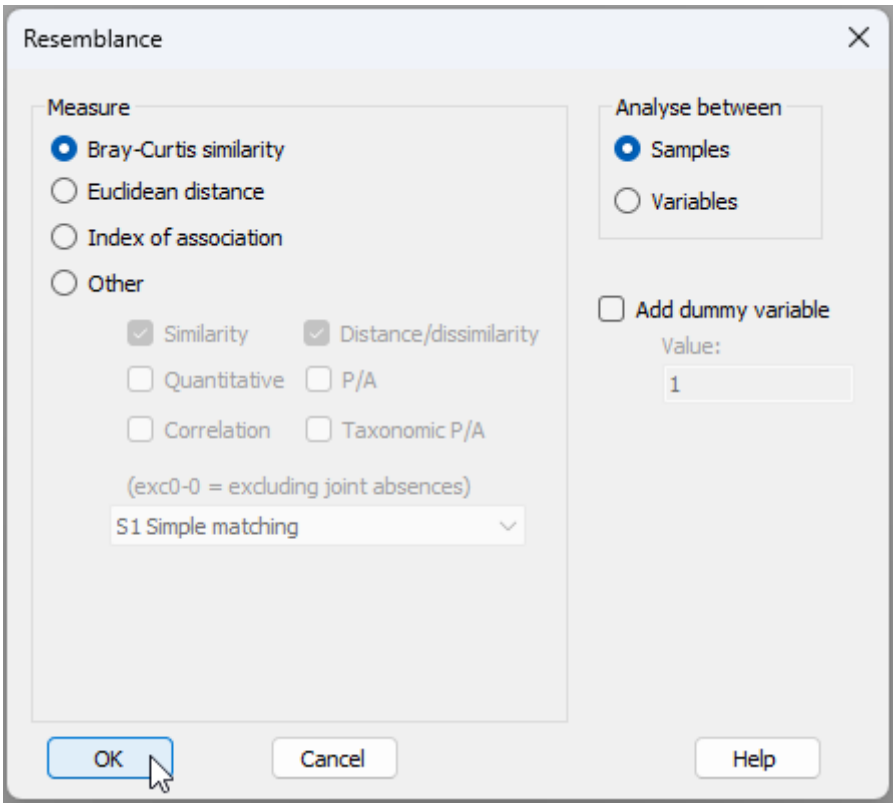
Variables - Traits

	ADP	AEP	AH	AS	AA	ANS	APP	AF	AR	BB
F-F	100	100	0	100	100	0	0	0	0	
F-I	0	0	100	0	0	100	100	100	100	
F-R	0	0	0	0	0	0	0	0	0	
Po-E	100	100	100	100	50	100	100	100	100	
Po-I	0	0	0	0	0	0	0	0	0	
Po-Is	0	0	0	0	50	0	0	0	0	
Li-SW	0	0	0	0	0	0	0	0	0	
Li-C	100	100	0	100	50	0	0	100	100	
Li-B	0	0	0	0	50	0	0	0	0	
Li-T	0	0	0	0	0	0	0	0	0	
Li-Sh	0	0	0	0	0	0	0	0	0	
Li-PA	0	0	100	0	0	100	100	0	0	
Li-TA	0	0	0	0	0	0	0	0	0	
Tr-Sa	0	0	100	0	0	100	100	0	0	
Tr-Sp	0	0	0	14.286	0	0	0	0	0	
Tr-P-SAW	0	0	0	0	0	0	0	0	0	
Tr-P-S	0	50	0	0	100	0	0	0	0	
Tr-P-P	0	0	0	0	0	0	0	50	33.333	
T-P-GP	0	0	0	0	0	0	0	0	0	
Tr-P-GF	50	0	0	0	0	0	0	0	0	
Tr-P-GS	0	0	0	14.286	0	0	0	0	0	
Tr-P-GSM	50	50	0	14.286	0	0	0	0	0	
Tr-S-Fa	0	0	0	14.286	0	0	0	50	33.333	
Tr-S-Fl	0	0	0	0	0	0	0	0	0	
Tr-D-S	0	0	0	14.286	0	0	0	0	33.333	
Tr-D-SS	0	0	0	0	0	0	0	0	0	
Tr-H-F	0	0	0	0	0	0	0	0	0	
Tr-H-GP	0	0	0	0	0	0	0	0	0	
Tr-H-GF	0	0	0	0	0	0	0	0	0	
Tr-H-GS	0	0	0	14.286	0	0	0	0	0	

Note how the distribution of traits within a species for any particular trait group (in this example, the trait groups are identifiable by reference to the letters appearing before the dash '-' in the variable names, e.g., 'Po', 'Li', 'Tr', etc.) will sum to 100.

Calculate relationships among species based on traits

4. From the 'Standardised Data' sheet, click **Analyse** > **Resemblance...** > (Measure $\bullet$Bray-Curtis similarity) & (Analyse between $\bullet$Samples), as shown below.



The resulting resemblances among the species, based on their traits is shown below.

Resem1

Gulf of Maine invertebrate traits

Similarity (0 to 100)

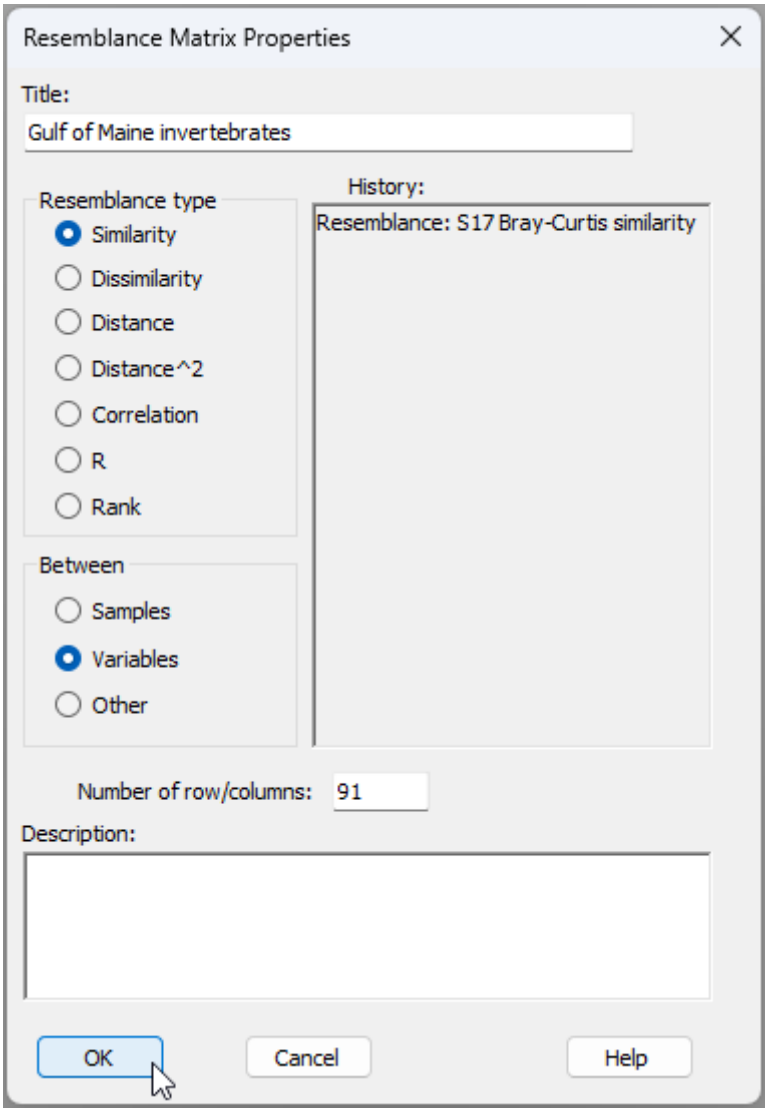
Samples - Species

Samples - Species	ADP	AEP	AH	AS	AA	ANS	APP	AF
ADP								
AEP	86.905							
AH	24.691	24.691						
AS	68.878	67.092	22.222					
AA	64.286	63.095	35.802	46.071				
ANS	30.357	30.357	38.889	38.929	26.786			
APP	25	25	79.012	21.429	20.238	30.357		
AF	39.286	37.5	44.444	49.592	46.429	46.429	42.857	
AR	39.286	37.5	44.444	50.612	46.429	46.429	42.857	97.6
BB	38.095	38.095	43.21	47.092	20.238	51.786	48.81	23.2
BE	47.619	40.476	54.321	39.286	29.762	51.786	61.905	32.1
BV	27.381	28.571	79.012	21.429	23.81	37.5	89.286	35.7
BS	27.381	35.714	71.605	21.429	23.81	37.5	82.143	28.5
BU	46.429	57.143	24.691	45.663	27.381	37.5	32.143	33.9
BT	32.143	39.286	76.543	28.571	27.381	30.357	75	28.5
C	23.81	23.81	64.198	30.357	20.238	44.643	66.667	23.2
CA	38.095	39.286	59.259	47.092	35.714	21.429	57.143	

We mustn't forget that the role of 'samples' is being taken by the species here! (The traits were the variables that created this matrix). However, in subsequent analyses to follow, the species will be variables, and sites will be samples. So let's go ahead and change the properties of this resemblance matrix now.

From 'Resem1', click **Edit** > **Properties...** and change the following:

- Title 'Gulf of Maine invertebrates'; and
- Between $\bullet$Variables then click '**OK**' (as shown below).



We now have a resemblance matrix among the species which are clearly identified as variables, hence that is ready for ensuing analyses (see below).

Resem1

Gulf of Maine invertebrates

Similarity (0 to 100)

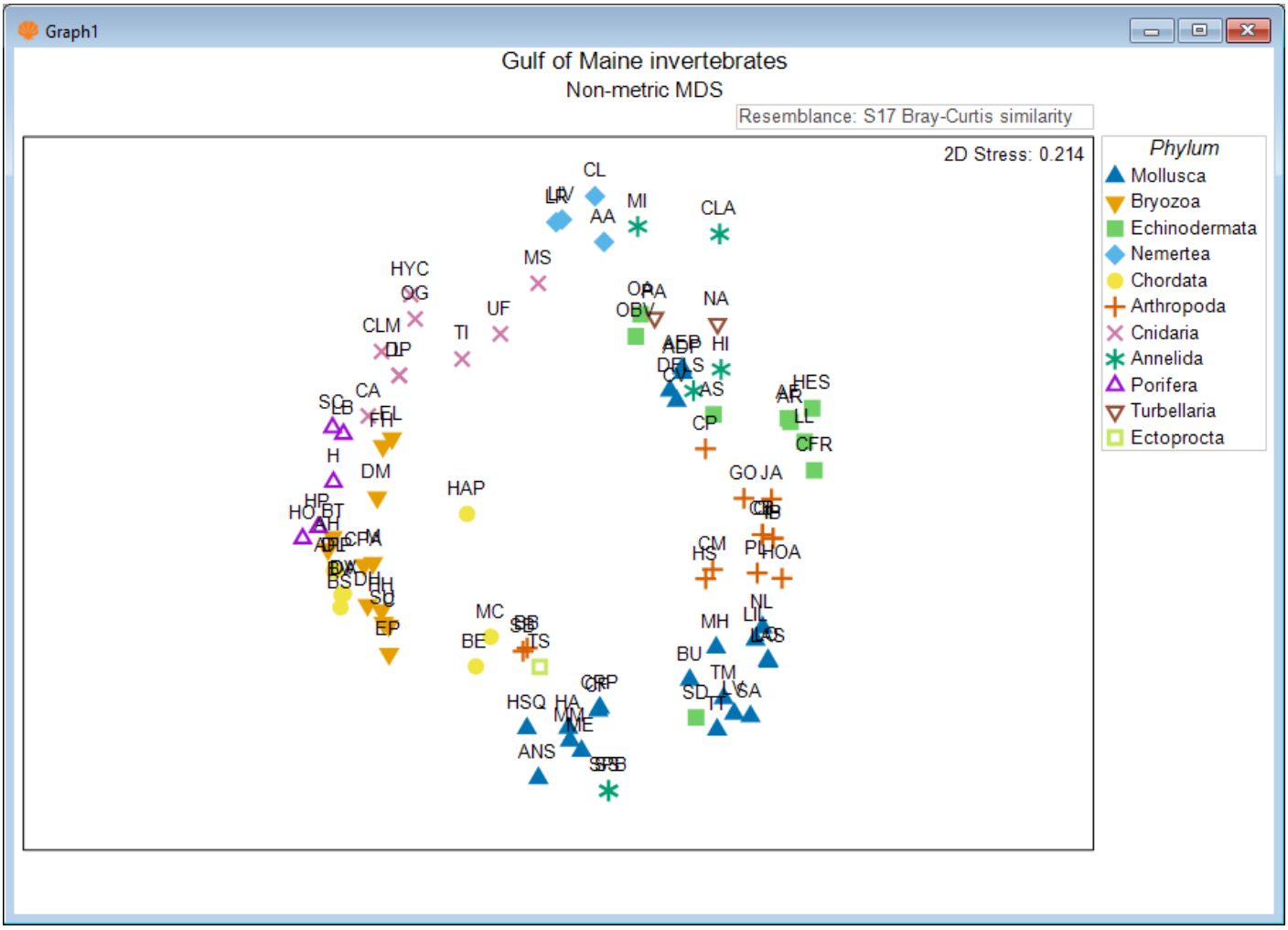
Variables - Species

Variables - Species	ADP	AEP	AH	AS	AA	ANS	APP	AF	AR	BE
ADP										
AEP	86.905									
AH	24.691	24.691								
AS	68.878	67.092	22.222							
AA	64.286	63.095	35.802	46.071						
ANS	30.357	30.357	38.889	38.929	26.786					
APP	25	25	79.012	21.429	20.238	30.357				
AF	39.286	37.5	44.444	49.592	46.429	46.429	42.857			
AR	39.286	37.5	44.444	50.612	46.429	46.429	42.857	97.619		
BB	38.095	38.095	43.21	47.092	20.238	51.786	48.81	23.214	23.214	
BE	47.619	40.476	54.321	39.286	29.762	51.786	61.905	32.143	32.143	
BV	27.381	28.571	79.012	21.429	23.81	37.5	89.286	35.714	35.714	
BS	27.381	35.714	71.605	21.429	23.81	37.5	82.143	28.571	28.571	
BU	46.429	57.143	24.691	45.663	27.381	37.5	32.143	33.929	32.738	
BT	32.143	39.286	76.543	28.571	27.381	30.357	75	28.571	28.571	
C	23.81	23.81	64.198	30.357	20.238	44.643	66.667	23.214	23.214	
CA	38.095	39.286	59.259	47.092	35.714	21.429	57.143	25	25	
CB	51.786	55.357	31.481	57.092	35.714	42.857	44.643	66.071	64.881	
CI	51.786	55.357	31.481	57.092	35.714	42.857	44.643	66.071	64.881	
CP	73.214	66.071	25.556	70.561	51.786	37.143	24.643	37.5	38.929	

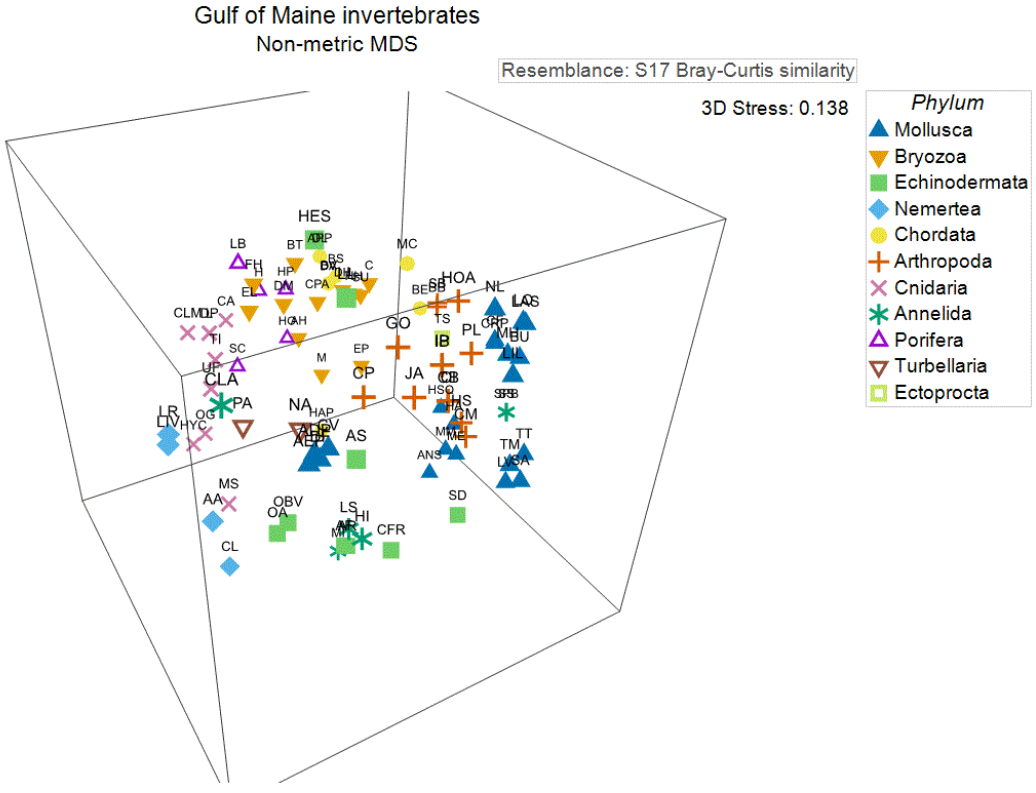
As a next step, let's aim to visualise similarities among species, based on their traits, using a non-metric MDS ordination.

Visualise inter-species trait-based relationships

5. From 'Resem1', click **Analyse** > **MDS** > **Non-metric MDS (nMDS)...**, and go with all of the default options in the dialog (just click '**OK**'). The resulting 2D ordination ('Graph1' in the Explorer tree) is shown below. Different symbols/colours show the different phyla in which individual species (labeled by their abbreviations) belong.



This ordination has rather high stress (0.214), so we might opt to examine the 3D graphic ('Graph1' in the Explorer tree), which has a lower stress (0.138), as shown below. In essence, this ordination shows us the relationships among these invertebrate species in multi-dimensional trait space.



Next, we shall use these trait-based relationships among the species to inform the analysis of functional turnover among the sites, based on the species they contain.

Calculate functional resemblances (Gamma+) among sites

6. First, we need to open up the species $\times$ site matrix of presence/absence (occurrence) data in the same workspace (click **File** > **Open...**). This file is called 'Gulf_of_Maine_invert_occurrences.pri' (also found in the 'Gulf_of_Maine_inverts' folder inside the 'Examples_P8' folder).

Gulf_of_Maine_invert_occurrences

Gulf of Main invertebrate occurrences

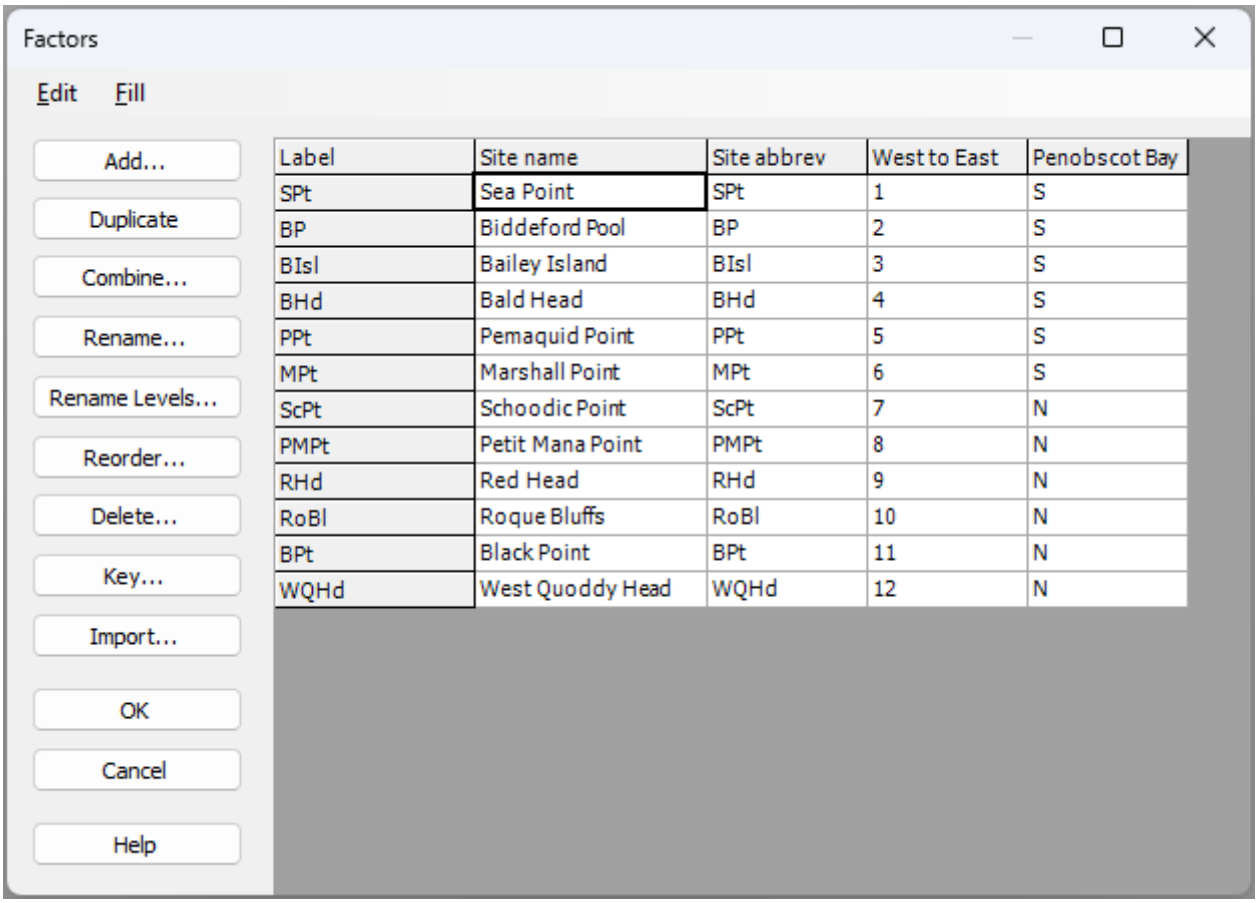
Abundance

Samples - Sites

Variables - Species

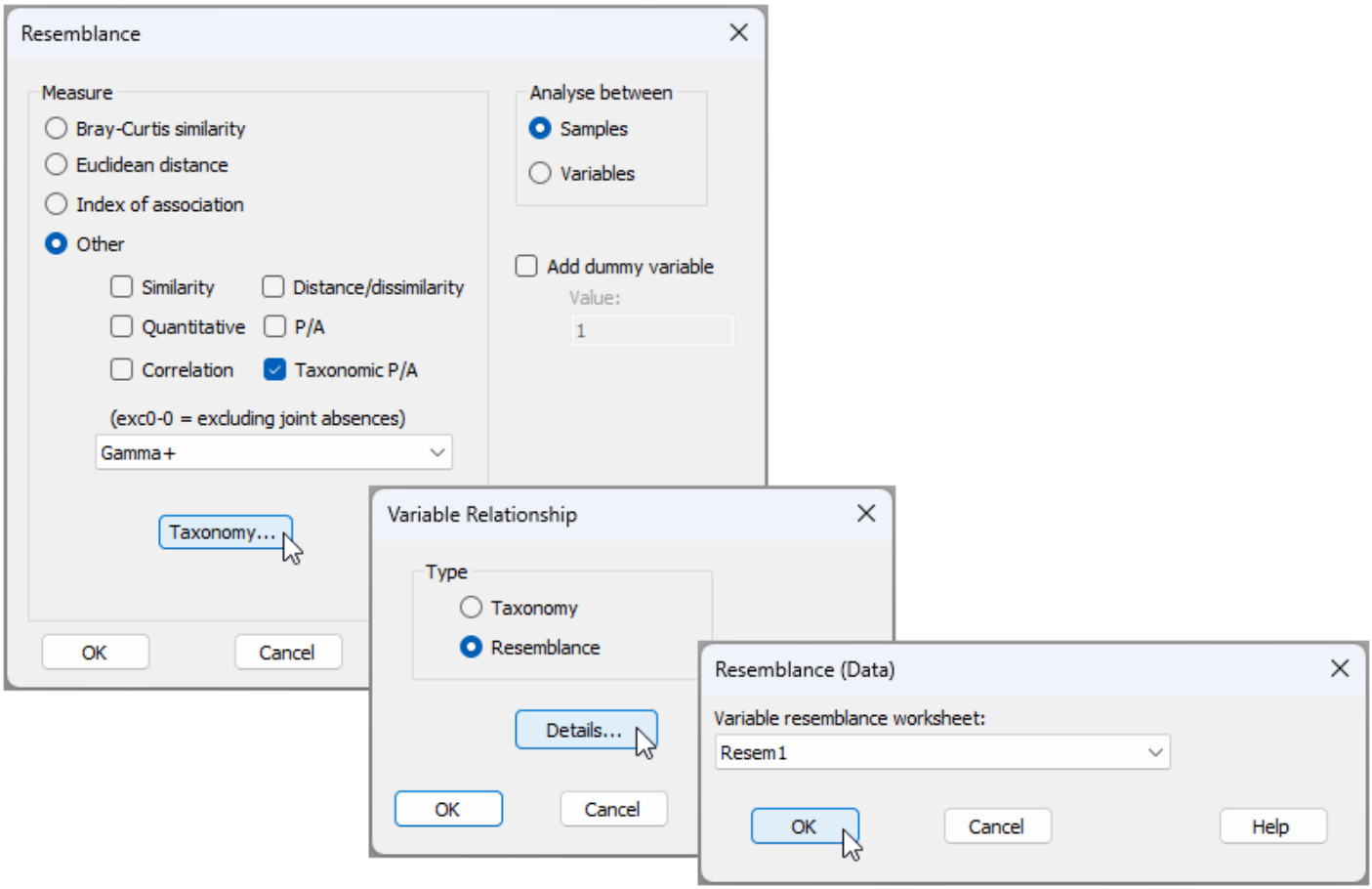
	SPt	BP	BIsl	BHd	PPt	MPt	ScPt	PMPt	RHd	R
AA	0	0	0	0	0	0	0	0	1	
ADP	1	0	0	0	0	0	1	0	0	
AEP	0	0	0	0	1	1	1	0	0	
AF	1	1	0	1	1	1	1	0	0	
AH	0	0	1	1	0	0	0	0	0	
ANS	0	0	0	0	1	1	0	0	0	
APP	1	1	1	1	1	1	1	1	0	
AR	1	1	0	1	1	0	1	1	1	
AS	1	0	0	0	0	1	1	1	0	
BB	1	1	1	1	1	1	1	0	1	
BE	0	0	0	0	0	0	0	0	0	
BS	1	1	1	0	0	1	0	0	0	
BT	0	0	1	0	0	0	0	1	0	
BU	0	0	0	0	0	0	0	0	1	
BV	1	1	1	1	1	1	1	0	0	
C	1	1	1	0	0	0	1	1	0	
CA	0	0	0	1	0	1	0	0	0	
CB	0	1	1	1	1	1	1	1	1	
CF	1	1	1	1	1	1	0	0	0	
CFR	0	0	0	0	0	0	1	1	0	
CI	1	1	0	1	1	1	1	1	0	

The sites have abbreviated names. Full names (as shown in the map in Fig. 13.3 above) can be seen by clicking on **Edit** > **Factors...**. You will also see that the sites have been given a rank ordered value for their position along the Maine coastline ('West to East'), and have also been classified as occurring either north or south of Penobscot Bay, viz:



Label	Site name	Site abbrev	West to East	Penobscot Bay
SPt	Sea Point	SPt	1	S
BP	Biddeford Pool	BP	2	S
BIsl	Bailey Island	BIsl	3	S
BHd	Bald Head	BHd	4	S
PPt	Pemaquid Point	PPt	5	S
MPt	Marshall Point	MPt	6	S
ScPt	Schoodic Point	ScPt	7	N
PMPt	Petit Mana Point	PMPt	8	N
RHd	Red Head	RHd	9	N
RoBl	Roque Bluffs	RoBl	10	N
BPt	Black Point	BPt	11	N
WQHd	West Quoddy Head	WQHd	12	N

7. Now we shall calculate functional (trait-based) resemblances (Gamma+) among the sites. From the 'Gulf_of_Maine_invert_occurrences' worksheet, click **Analyse** > **Resemblance...**, and in the 'Resemblance' dialog window:
   - Choose (Analyse between $\bullet$Samples) & (Measure $\bullet$Other > $\checkmark$ Taxonomic P/A > Gamma+), and click the 'Taxonomy...' button (Taxonomy...).
   - In the 'Variable Relationship' dialog window, choose (Type $\bullet$Resemblance), and click the 'Details...' button (Details...).
   - In the 'Resemblance (Data)' dialog, choose (Variable resemblance worksheet: Resem1).
   - Click '**OK**' (3 times, once for each successive dialog window) to complete the operation. The dialog windows involved in this operation (i.e., to calculate Gamma+ on the basis of a resemblance matrix, which in our case is a trait-based resemblance matrix) are shown below:



The resulting resemblance matrix of Gamma+ values among the sites (called 'Resem2') is shown below.

Resem2

Gulf of Main invertebrate occurrences

Dissimilarity (0 to 100)

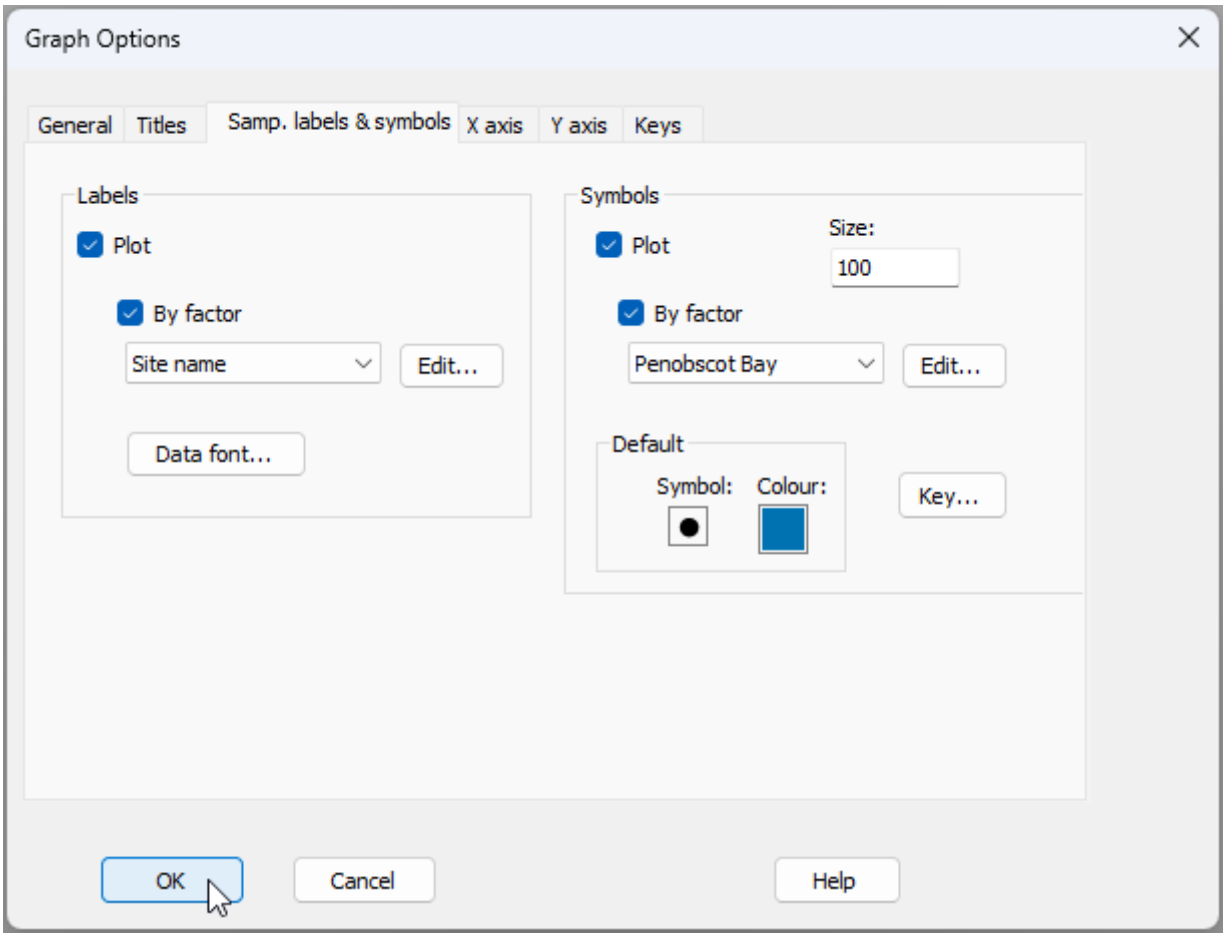
Samples - Sites

Samples - Sites	SPt	BP	BIsl	BHd	PPt	MPt	ScPt	PMPt	RHd	RoBl	BPt	WQ
SPt												
BP	5.3744											
BIsl	5.5122	5.1881										
BHd	4.2373	4.4618	3.9962									
PPt	6.21	2.8273	4.8778	2.9953								
MPt	4.9333	4.1816	5.8148	3.99	3.8353							
ScPt	4.1539	5.7213	7.7288	5.5054	5.6803	3.8588						
PMPt	5.7521	7.5095	6.1953	6.6509	8.0954	5.059	6.0106					
RHd	8.0944	8.4376	8.2215	6.7051	7.3283	8.6338	7.1213	7.7042				
RoBl	6.5	5.0083	7.8067	6.7359	6.3122	5.0294	5.0944	5.8051	6.6361			
BPt	8.2854	6.2704	7.8388	6.1272	7.0697	6.6552	4.8497	6.2225	5.8001	6.7433		
WQHd	6.1395	7.5173	7.0454	6.1653	7.1065	5.5993	5.1322	4.8605	6.7192	6.9964	5.1302	

Visualise functional turnover among sites

8. We can now easily do an ordination (or indeed a cluster analysis or any other resemblance-based analysis) of the trait-based turnover among the sites. From 'Resem2', click **Analyse** > **MDS** > **Non-metric MDS (nMDS)**, just keep all of the defaults and click '**OK**'.

9. From the resulting 2D nMDS ordination graphic (called 'Graph5' in the Explorer tree), click **Graph** > **Sample Labels & Symbols...** and choose to display labels by the factor of 'Site name', and symbols by the factor of 'Penobscot Bay', as shown in the dialog below, then click '**OK**'.



The resulting nMDS graphic looks like this:

This plot shows there is a clear shift in functional traits of intertidal assemblages from sites located to the south (blue symbols) *vs* those located to the north (amber symbols) of Penobscot Bay. We can do statistical tests of relevant hypotheses about this using ANOSIM (from 'Resem2', click **Analyse** > **ANOSIM...**). Doing this shows that:

- This shift is statistically significant (one-way unordered ANOSIM of the factor 'Penobscot Bay', R = 0.546, P = 0.0022); and
- There is a significant sequential change in trait-based relationships among these communities along the coastline (one-way ordered ANOSIM of the factor 'West to East', R^O = 0.422, P = 0.0026).

¶*Equally, and in a directly analogous fashion, there are situations where the standardisation of variables is required as a pre-treatment prior to analysis, but which needs to be done separately within groups of samples that may be identified by a factor.*

---

†*See also the following abstract by Thomas J. Trott, highlighted in the 4th World Conference on Marine Biodiversity: 'Traits matter: when rarity means more than abundance to functional diversity'*

.