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.

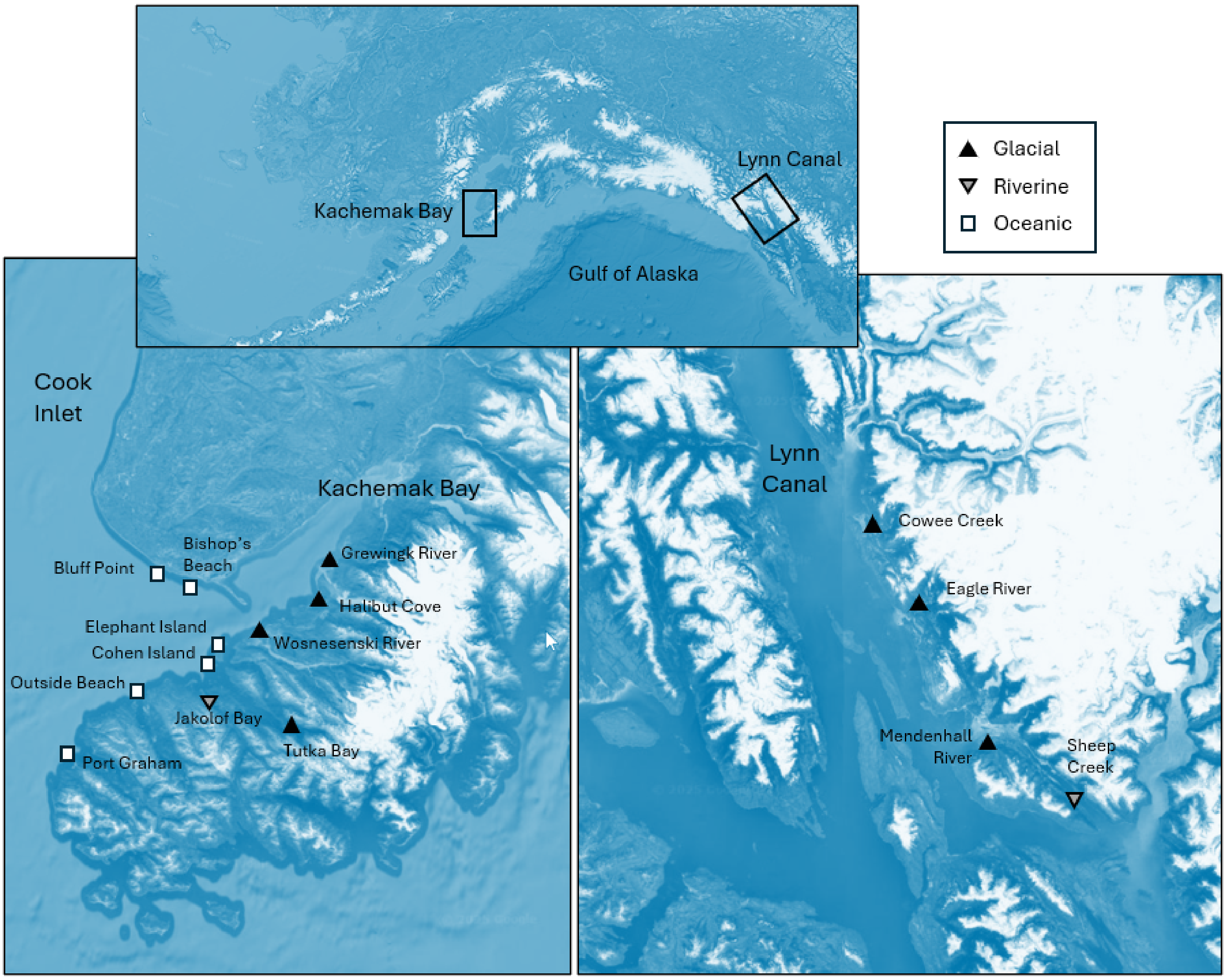
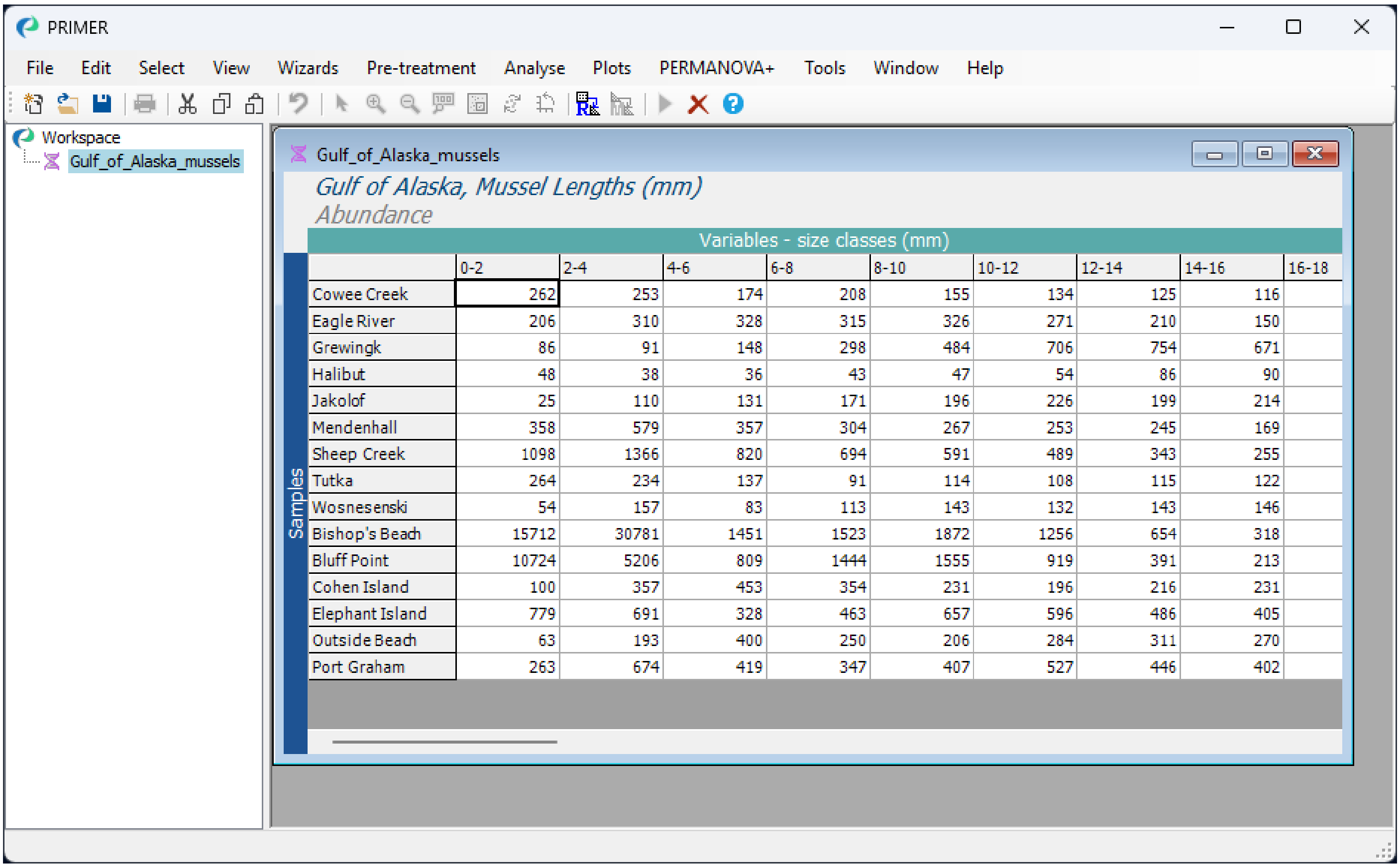


Fig. 13.2 (adapted from Fig. 1 in Dowling (2021)). Maps of the sites at which size-distributions of mussels were measured in each of two ecoregions: Kachemak Bay (left) and Lynn Canal (right). The top map shows the locations of these two ecoregions in Alaska. White squares represent ocean-influenced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

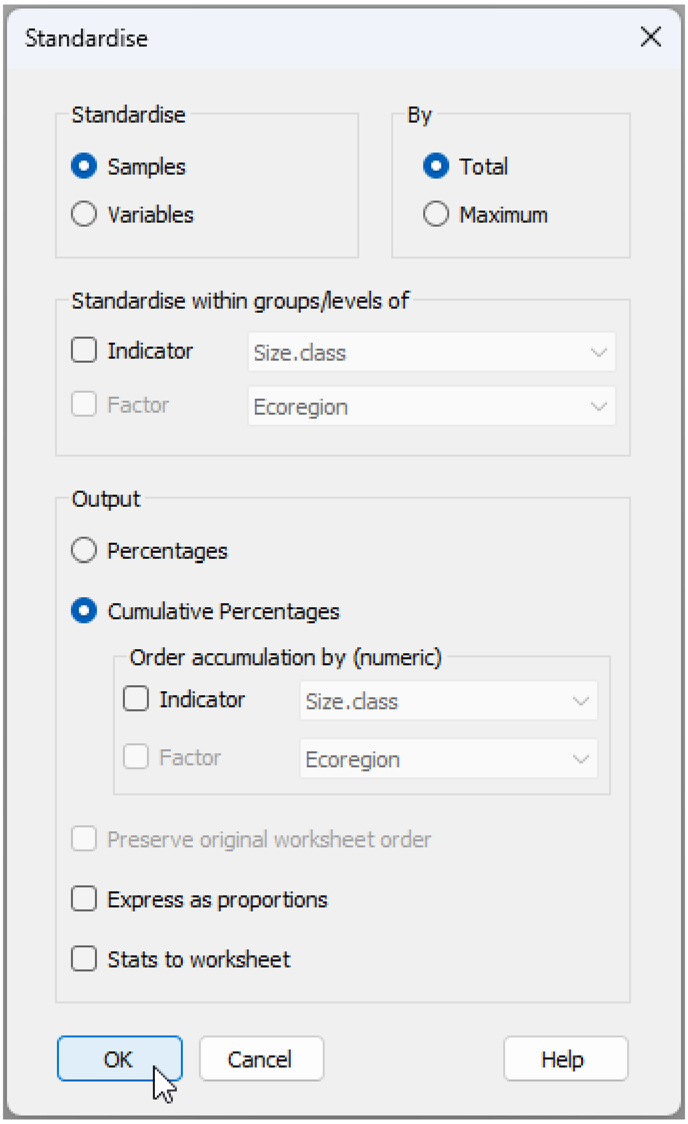
Input data and standardise

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).

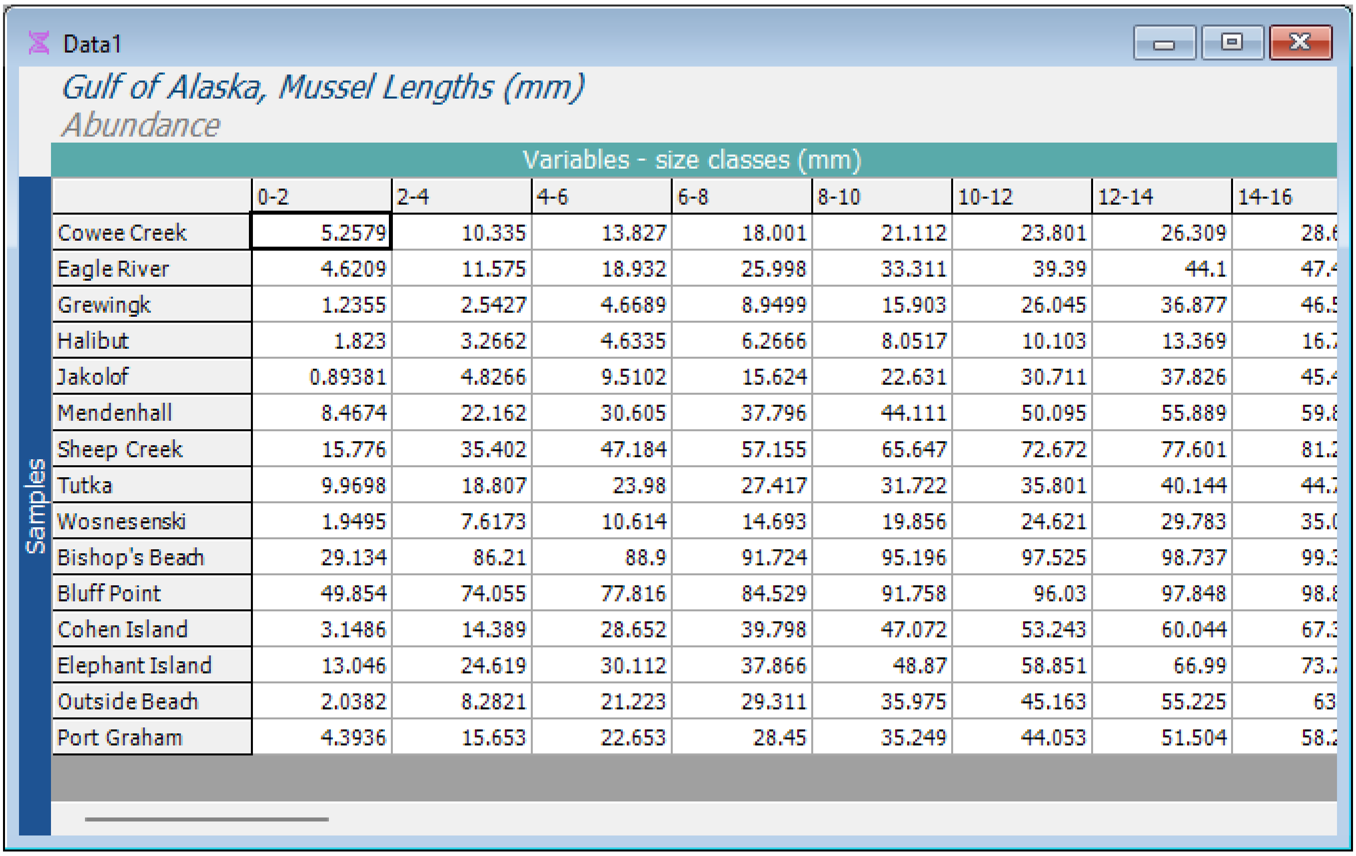


	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16	16-18
Cowee Creek	262	253	174	208	155	134	125	116	
Eagle River	206	310	328	315	326	271	210	150	
Grewingk	86	91	148	298	484	706	754	671	
Halibut	48	38	36	43	47	54	86	90	
Jakolof	25	110	131	171	196	226	199	214	
Mendenhall	358	579	357	304	267	253	245	169	
Sheep Creek	1098	1366	820	694	591	489	343	255	
Tutka	264	234	137	91	114	108	115	122	
Wosnesenski	54	157	83	113	143	132	143	146	
Bishop's Beach	15712	30781	1451	1523	1872	1256	654	318	
Bluff Point	10724	5206	809	1444	1555	919	391	213	
Cohen Island	100	357	453	354	231	196	216	231	
Elephant Island	779	691	328	463	657	596	486	405	
Outside Beach	63	193	400	250	206	284	311	270	
Port Graham	263	674	419	347	407	527	446	402	

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

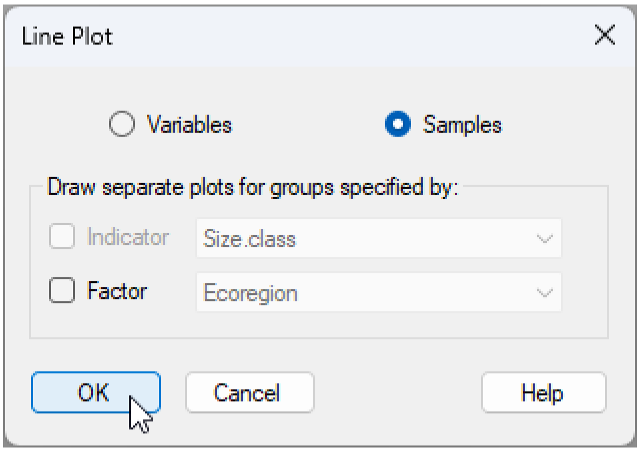
Gulf of Alaska, Mussel Lengths (mm)

Abundance

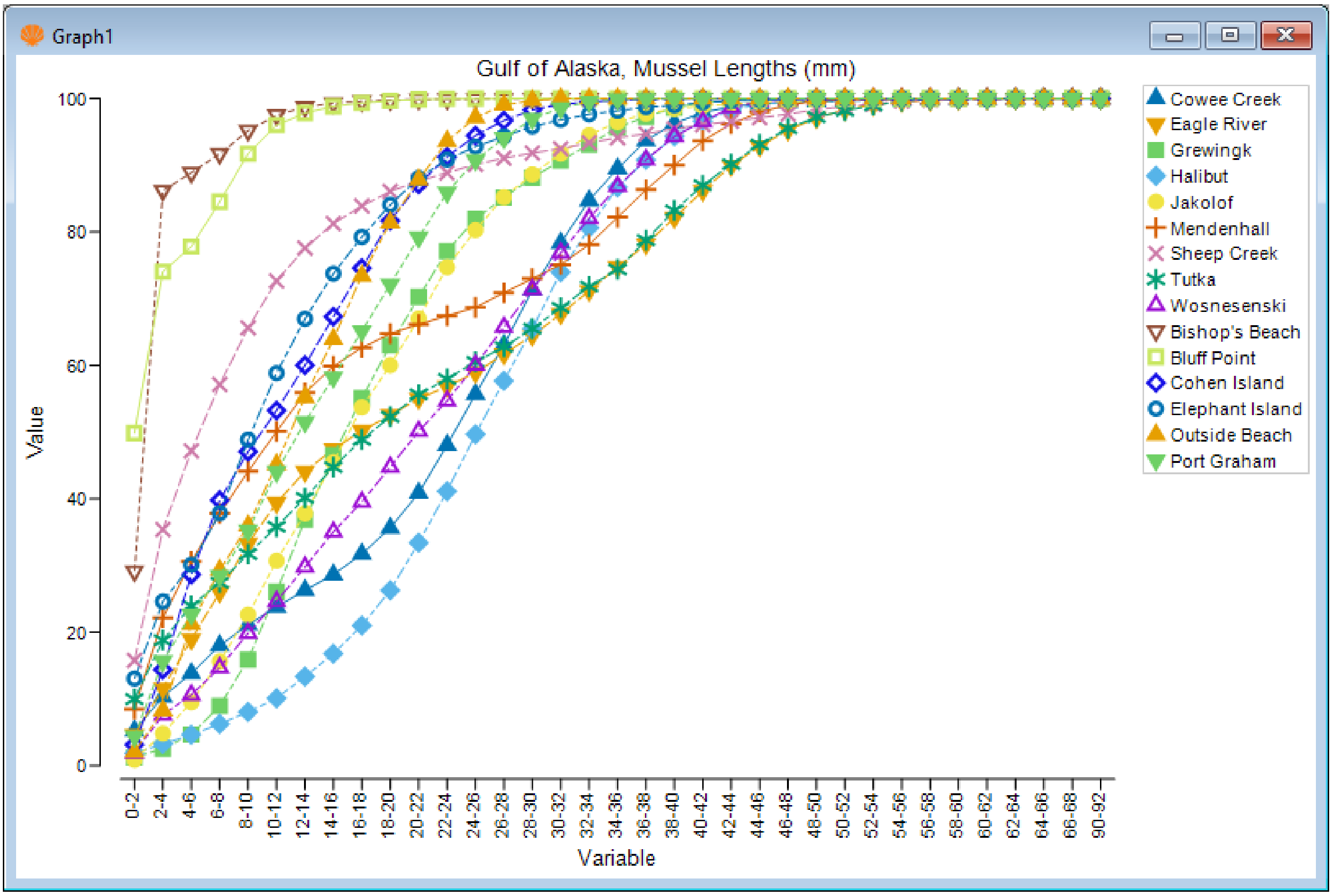
Samples	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16
	Variables - size classes (mm)							
Cowee Creek	5.2579	10.335	13.827	18.001	21.112	23.801	26.309	28.6
Eagle River	4.6209	11.575	18.932	25.998	33.311	39.39	44.1	47.4
Grewingk	1.2355	2.5427	4.6689	8.9499	15.903	26.045	36.877	46.5
Halibut	1.823	3.2662	4.6335	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.†



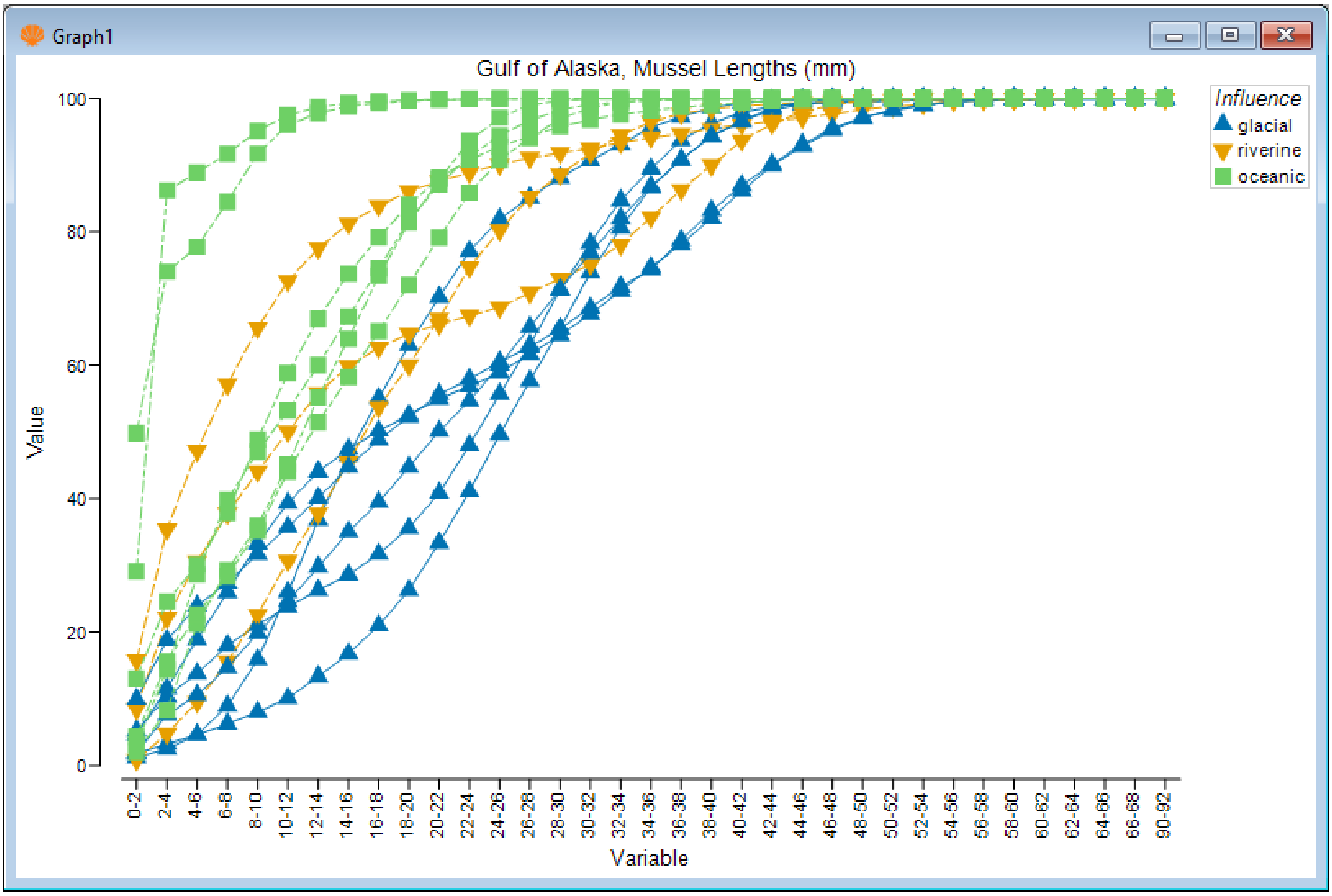
The resulting plot of cumulative size-distribution profiles for each site looks like this:



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**'.

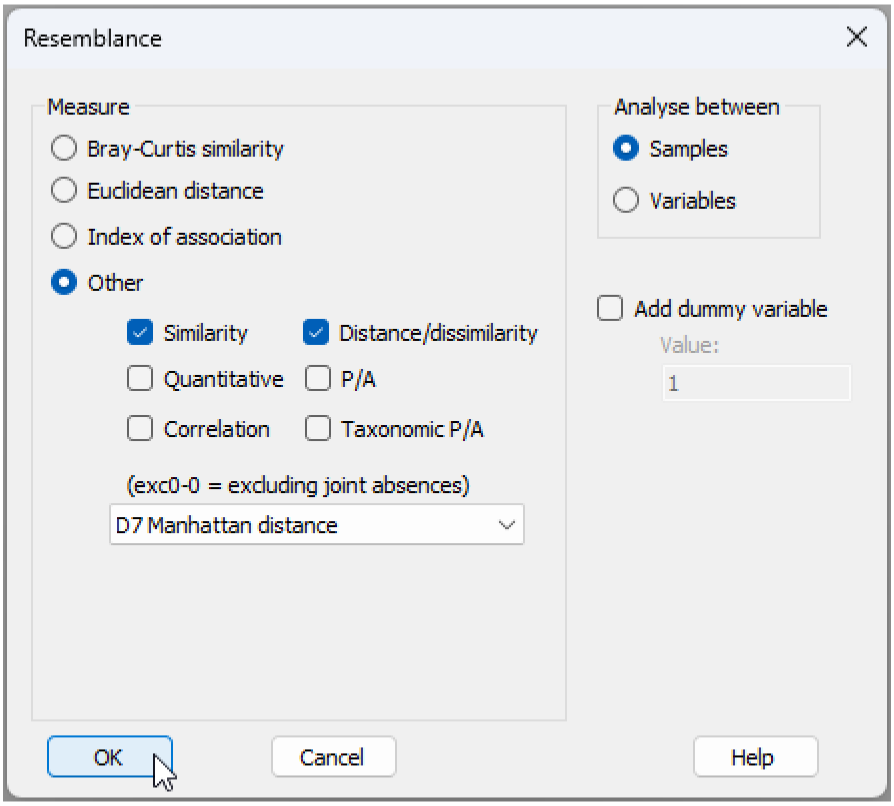
Our plot of cumulative size-distribution profiles, now showing the different influences, looks like this:



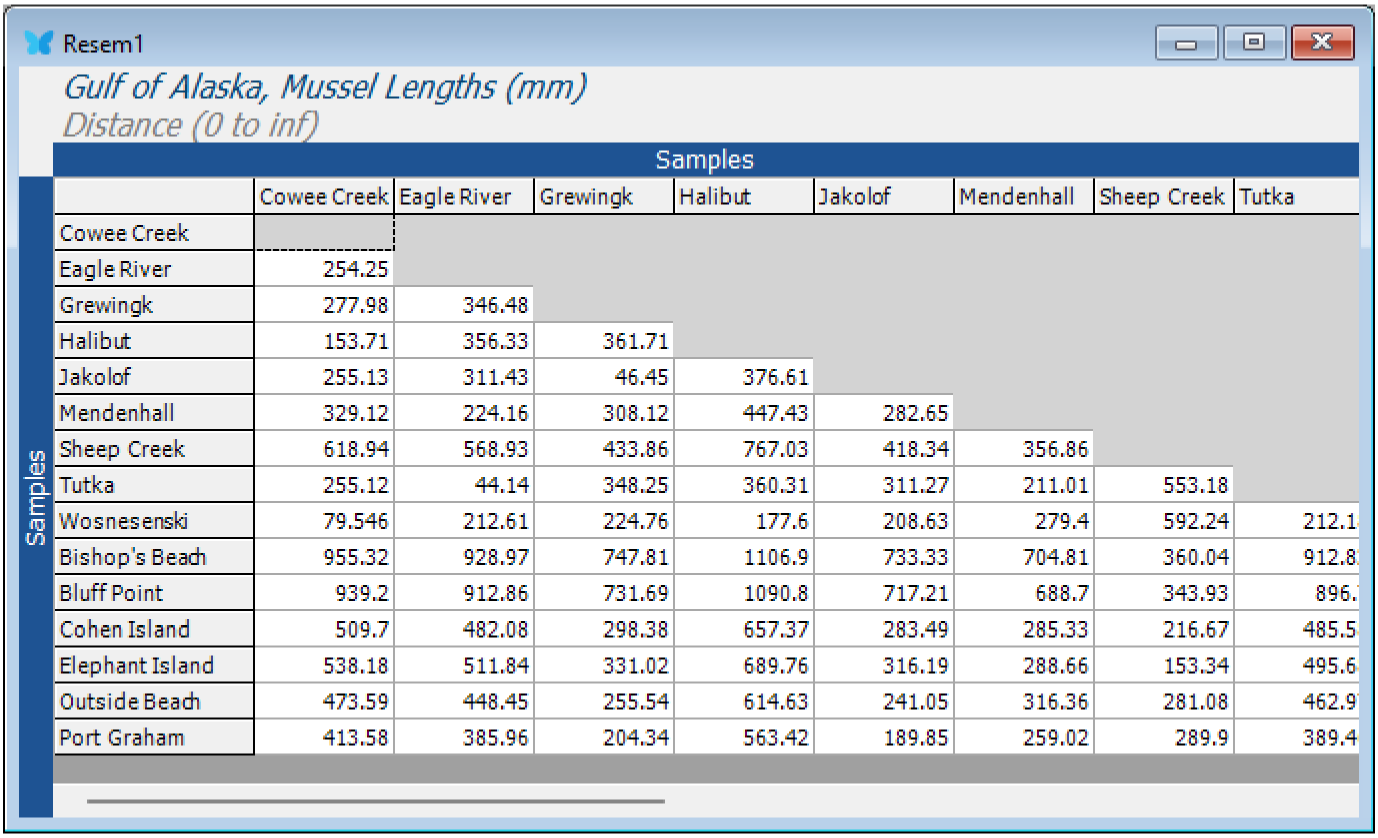
Ordination and analysis of size-frequency distributions

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

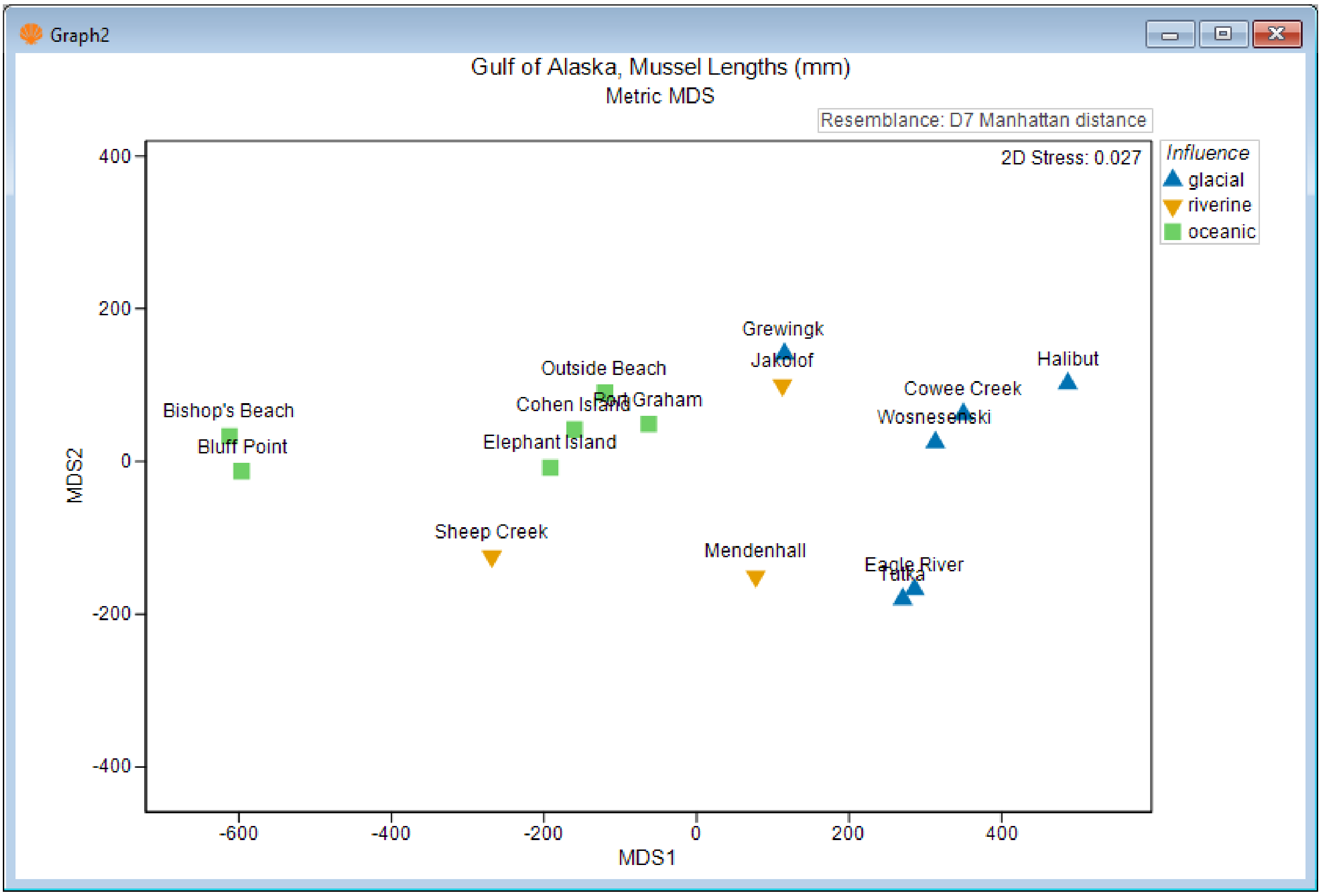
Gulf of Alaska, Mussel Lengths (mm)

Distance (0 to inf)

	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

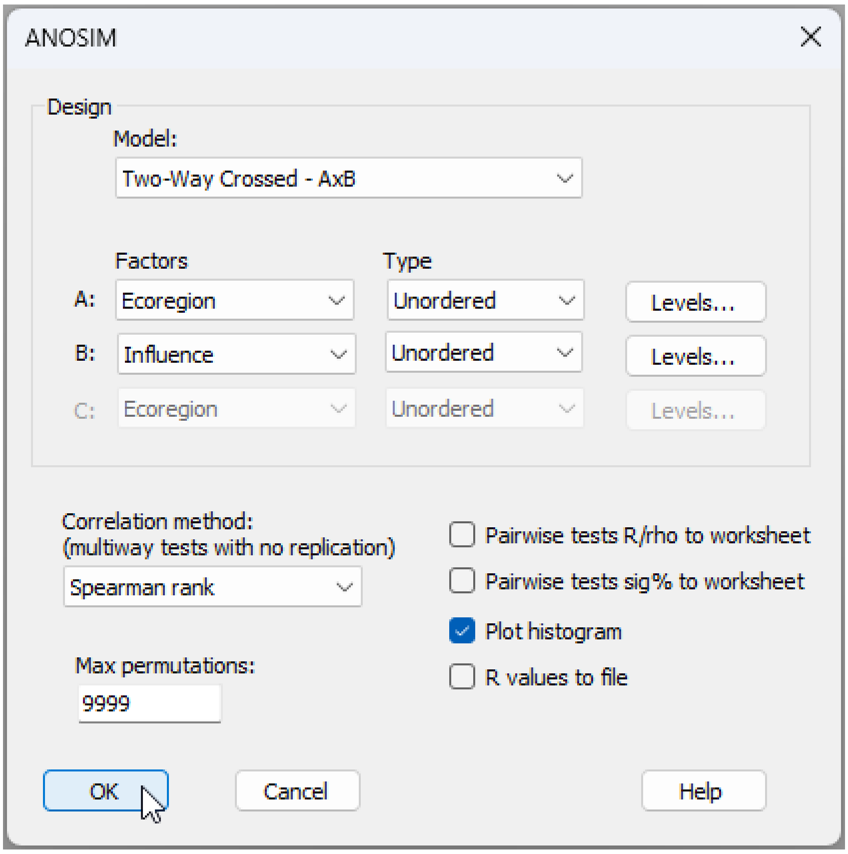
6. Create a metric MDS plot. Note that, for this example, a metric MDS can be done quite successfully.¶ From the 'Resem1' matrix, click **Analyse** > **MDS** > **Metric MDS (mMDS / tmMDS)...** > (Choice of intercept: $\bullet$Metric MDS (zero intercept)), **OK**. Put symbols on the resulting MDS plot (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

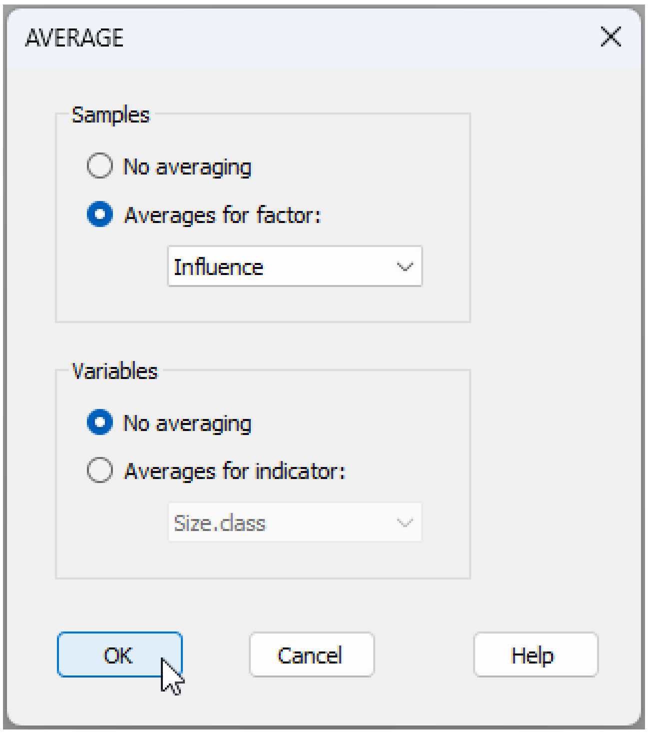
Groups	R Statistic	Significance Level %	P-Value	Possible Permutations	Actual Permutations	Number >= Observed
glacial, riverine	-0.07	66.7	0.6667	15	15	10
glacial, oceanic	0.679	0.5	0.0048	210	210	1
riverine, oceanic	0.289	42.9	0.4286	7	7	3

Outputs
Plot: Graph6
Plot: Graph7

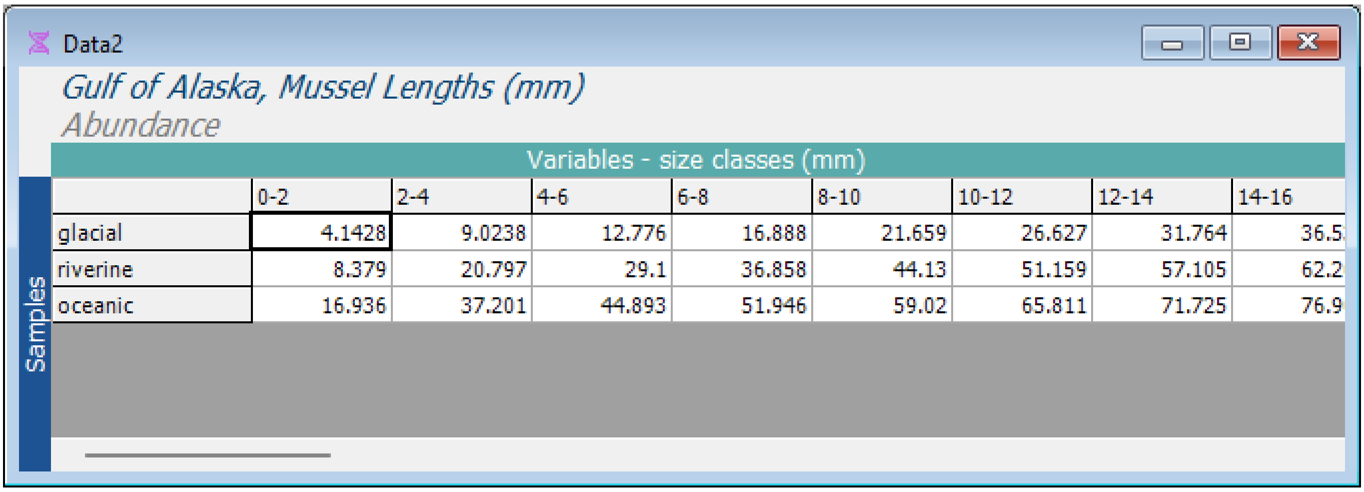
These results are quite clear. There is no effect of Ecoregion (ANOSIM R = -0.286, P > 0.80), but Influence has a statistically significant effect on the size-class distributions of these mussels (ANOSIM R = 0.483, P < 0.01). Moreover, the pair-wise tests show significantly different size-class distributions 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:

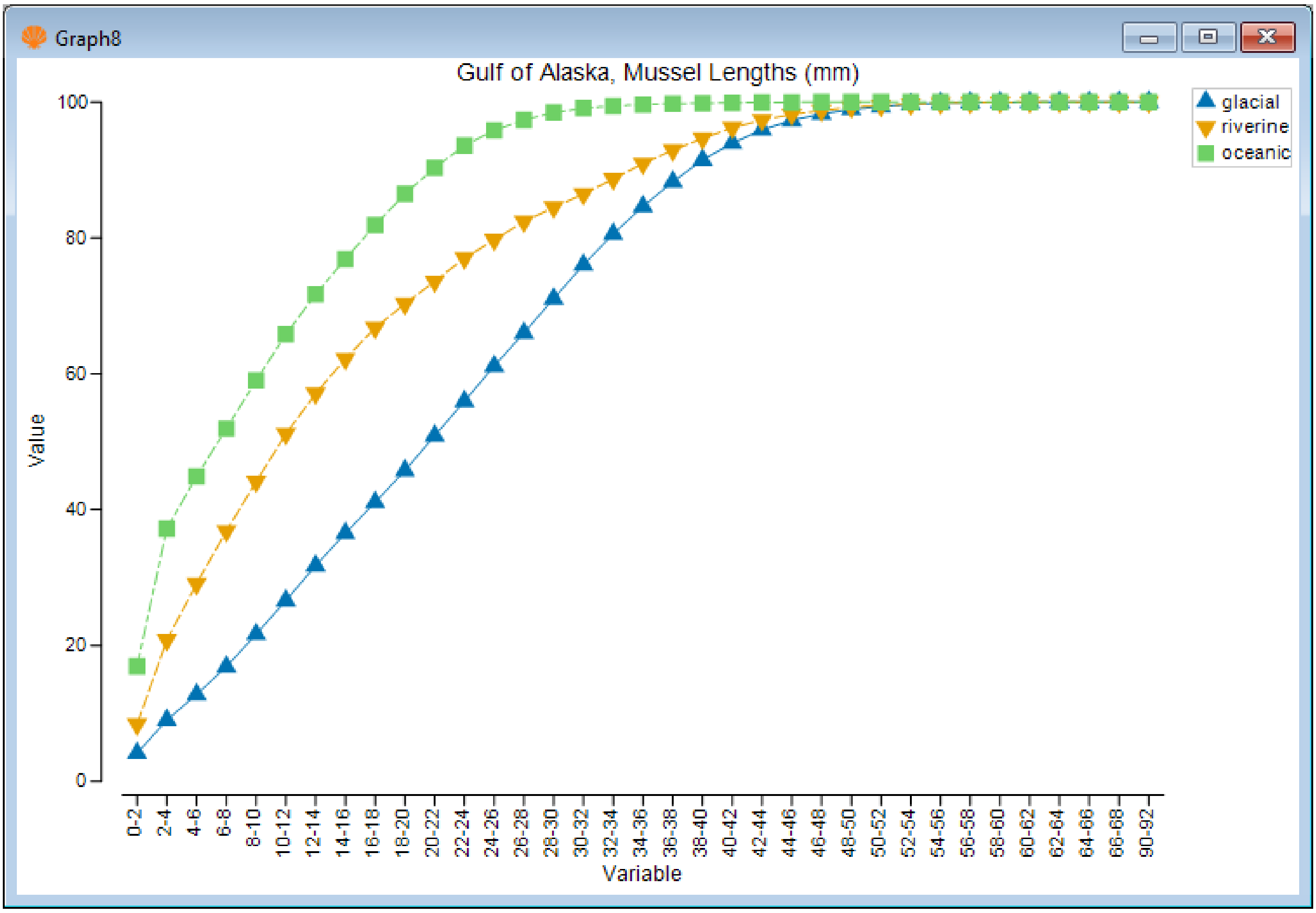


This will produce a data file with just three profiles in it, called 'Data2':



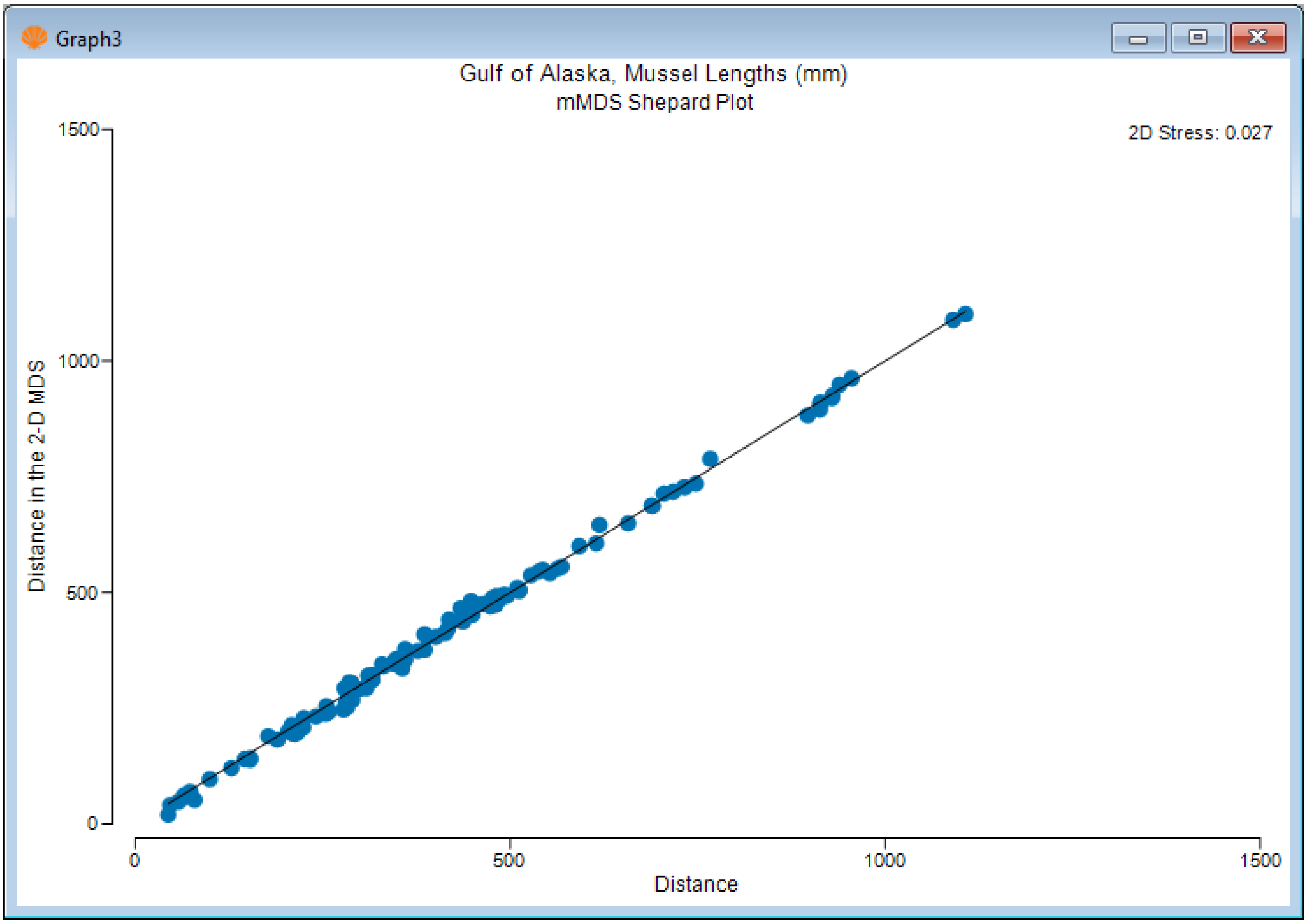
	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16
glacial	4.1428	9.0238	12.776	16.888	21.659	26.627	31.764	36.5
riverine	8.379	20.797	29.1	36.858	44.13	51.159	57.105	62.2
oceanic	16.936	37.201	44.893	51.946	59.02	65.811	71.725	76.9

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.*



Revision #37
Created 7 July 2025 02:02:03 by Marti
Updated 5 December 2025 22:13:13 by Marti