9. Group covariates

9.1 Why group covariables together?

There are situations where it may be useful or important to include one or more quantitative co-variables in a PERMANOVA model. For example, in our study of invertebrates inhabiting holdfasts of the kelp, *Ecklonia radiata*, it was not possible to standardise the sizes of the individual holdfasts that were sampled, but roughly, because every individual kelp that we sampled was naturally unique in size. Thus, we measured the volume of each collected holdfast using water displacement. This continous quantitative variable (*volume*) could then be used as a *covariable* in subsequent PERMANOVA analyses that compared the sizes of variance components at several hierarchical spatial scales (Anderson *et al.* (2005)).

PERMANOVA in PRIMER has, historically, always permitted the inclusion of more than one covariable at a time, *but* these were always treated as individual variables singly in PERMANOVA models. Specifically, each covariable is fitted linearly in the space of the resemblance measure and contributes 1 extra degree of freedom to the fitted model. An important limitation here, however, was that multiple covariables could not be 'kept together' and treated as a combined group in the analysis; so tests of their combined effect were not available. Of course, one could use the facility in DISTLM to fit a linear model on a combined set of quantitative (or other) variables, but in DISTLM, there is unfortunately no way to specify a complex ANOVA-style design with (say) random factors and/or nested/hierarchical terms.

In P8, it is now possible to *group multiple covariables together* using an indicator, so that a *set* of covariables are treated in a combined fashion in any PERMANOVA model. One or more sets of covariables can be specified in any given PERMANOVA model in the design file (along with one or more other factors that are identified in the usual way). Each *set* of covariables will then be included as a *single line* in the PERMANOVA output table, potentially contributing multiple degrees of freedom to the fitted model. In addition, one can choose either to include or exclude interactions of covariables (or sets of covariables) with either: (i) other factors, and/or (ii) other covariables (or sets of covariables).

This facility opens the door to a veritable plethora of new ways of formally analysing multivariate patterns and structures in the space of the chosen resemblance measure. It permits more complex conceptual spatio-temporal models to be examined formally with ease, such as spatial models with both latitude and longitude (that might also include multiple polynomials) or, notably, *periodic or cyclical models* (e.g., as would be used to capture seasonal patterns) and any other multi-dimensional model structures that might require multiple covariables (degrees of freedom) to model adequately within an ANOVA framework.

9.2 Periodic and cyclical models

Natural cycles in biology and ecology

Important situations where the treatment of multiple covariables as a single set would be desirable in PERMANOVA are cases of periodic or cyclical phenomena in biology or ecology. Examples might include:

- seasonal patterns (e.g., monthly or quarterly sampling at temperate latitudes, migrations)
- cyclical reproduction (e.g., spawning aggregations, flowering and fruiting cycles)
- lunar patterns (e.g., tidal cycles, hormonal cycles)
- global climate cycles (e.g., El Niño *vs* La Niña, North Atlantic Oscillations)
- circadian rhythms (e.g., 24-hr sleep cycles)

All such cyclical phenomena may generate cyclical patterns in multivariate data, and it would be very useful to be able to model these in the space of a resemblance measure, using PERMANOVA. However, as cyclical patterns are not linear, we (generally) need more than one dimension to model them appropriately.

Modeling cyclical phenomena using ANOVA/Regression

A straightforward way to model cyclical phenomena using ANOVA/regression is to use two variables, corresponding to the periodic sine and cosine functions (e.g., Bliss (1958)). Let's consider modeling a seasonal cycle. There are 12 months of the calendar year, and suppose we have sampled monthly and we label these months from 1 to 12 in our data set. However, we expect that samples taken in month 1 (January) will be more similar to samples taken in month 12 (December) than they will be to those taken, say, in month 6 (June).

Let k be the number of samples we have taken at equal intervals in our cycle.[¶] Here, k = 12. For our model, let's imagine that our k = 12 months occupy 12 equally spaced positions along the circumference of a circle with a radius of $r = 1$.[†] The circumference of a circle can be described as going from 0 all the way around to 2π (360 degrees) in radians. With each passing month, we will travel a distance $1/12^{th}$ of the way around this circle's circumference. So, the distance between each pair of consecutive months, in radians, is $c = 2\pi / k$. Thus, the model position of any given month, $t = 1 ,..., k$, along the circumference of the circle is easily found as $c \times t$.

We might consider using something like a sine (or a cosine) function to model periodic phenomena, as either one of these, on its own, will yield a wave-like pattern over the range from 0 to 360 degrees (that is, from 0 to 2π in radians; Fig. 9.1).

Fig. 9.1 Sine and cosine functions of $ct = 2\pi t / k$ versus twelve equally-spaced positions (i.e., months from $t = 1, \ldots, 12$) along the circumference of a circle of radius $r = 1$.

However, it is readily seen that using just one of these functions alone will not suffice for our purposes. For example, the value of the sine function is equal to zero at month 6 and also at month 12 (Fig. 9.1), but we obviously do not want the model to consider these as equivalent. Similarly, the value of the cosine function equates (for example) month 3 and month 9, which is also undesirable for us. Thus, any single wave function does not fully capture the full nature of cyclical phenomena.

If we were to take both functions simultaneously, however, we would indeed capture the full circle (Fig. 9.2).

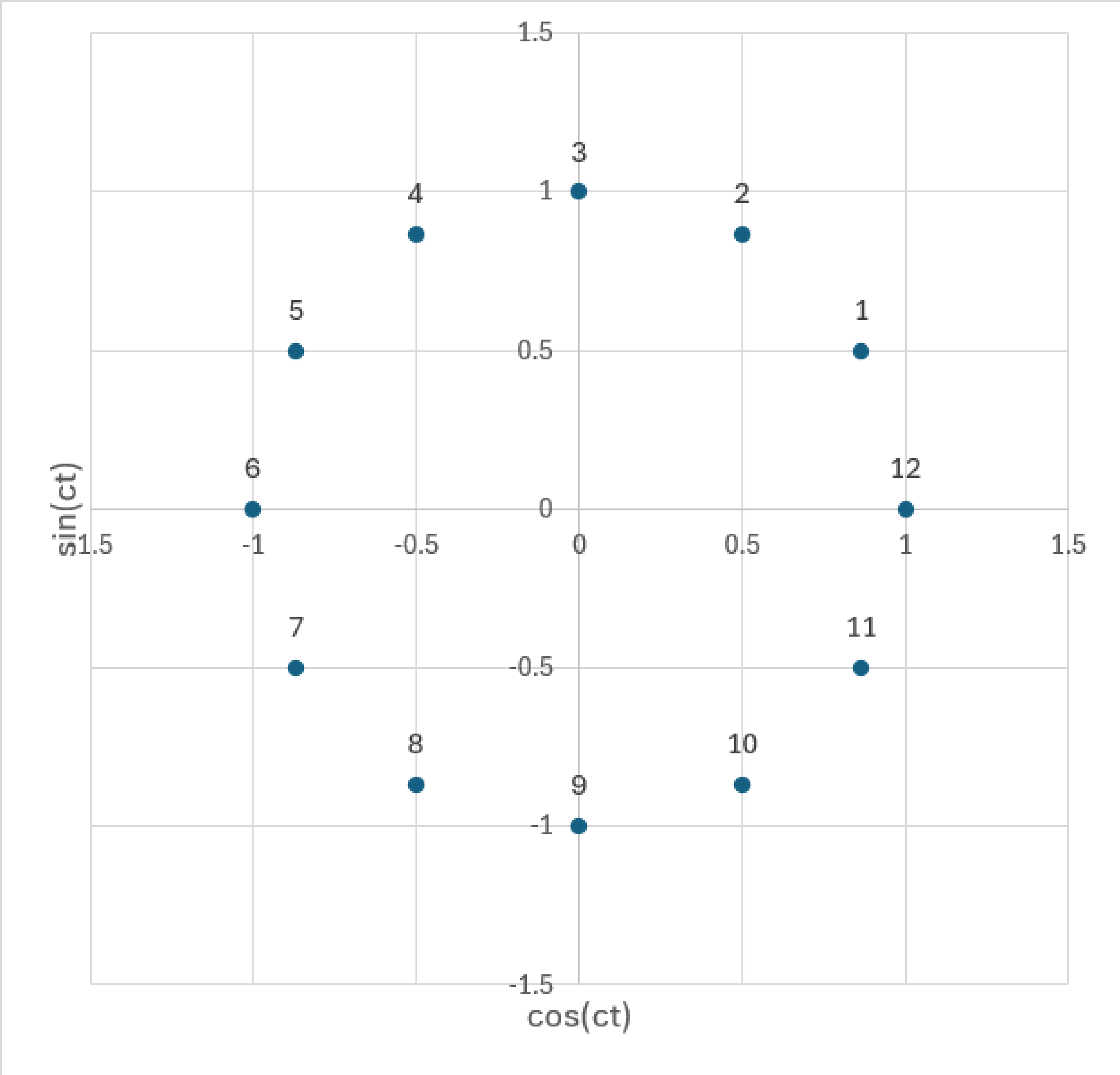


Fig. 9.2 Two-dimensional scatter plot of $\sin(ct)$ versus $\cos(ct)$, showing the 12 months (points) along the unit circle.

In Figure 9.2, we can see that: (i) the 12th month apparently occurs at "3 o'clock" and (ii) the numbers run counter-clockwise, rather than clockwise around the circle. Clearly, neither of these things actually matters in the least for our purposes. We have just created a model of a cycle with 12 equally spaced positions. There is in fact no natural starting or finishing point - the value of 1 follows the value of 12 in the same way that the value of 12 follows the value of 11. Furthermore, and most importantly, the distance between any pair of points in this 2-D space directly reflects, monotonically, the distance in time that would occur between those two months for any 12-month period (cycle) we care to choose.

The most important point here is that we require *both* of these variables *together* to create this cyclical picture: namely, $\sin(ct)$ and $\cos(ct)$. This means that if we want to test a model of cyclicity using PERMANOVA, then we will need to include both of these variables as covariables *simultaneously*. We would treat them together as a single set (with 2 degrees of freedom) and we would want for them to appear as a *single line* in the PERMANOVA table of results (e.g., with a name like "Seasonal Cycle" or something similar), to permit a test for monthly seasonal cycles.

Testing either of these as individual covariables, alone, would clearly not achieve our aim.

We shall demonstrate this analytical approach using the new tool for grouping covariates available in the PERMANOVA routine for PRIMER 8. Our example is a data set with monthly sampling of intertidal macroalgae from Vancouver in British Columbia, Canada.

¶*The intervals do not need to be equally spaced. Any intervals can be modeled in this way (i.e., along the circumference of a circle) at whatever spacings are required.*

---

†*The radius is irrelevant here; choosing r = 1 is merely convenient.*

9.3 Example: Annual monthly cycles - B.C. macroalgae

Consider the study described by Schenk *et al.* (2025) consisting of regular surveys of macroalgal cover from a rocky intertidal area at Stanley Park, Vancouver in British Columbia, Canada. Macroalgal communities have been sampled monthly (since September 2021) along each of 3 permanent transects that run from a seawall towards the water, perpendicular to the shore (i.e., from high to low tidal heights). Individual 1 m^2 quadrats were placed at 5 m intervals along each transect (from 0 to 90 m) and the percentage cover for each of p = 74 macroalgal taxa was recorded.[¶]

The full dataset, updated regularly, is available from Borealis. Monthly macroalgal cover data from September 2021 through to April 2025, inclusive, are provided in the file named 'BC_Macroalgal_cover.pri', found inside the 'Examples_P8 > BC_macroalgae' folder. See the file 'BC_Covariables.pri' for associated covariables.

We expect spatial changes in macroalgal assemblages with changes in tidal height, but also cyclically, from month to month, through the calendar year at this temperate latitude (49.3027 $\degree$N). There may also, of course, be changes from year to year. In this example, we shall focus on the temporal factors (year and month) affecting the macroalgal communities at this site. Towards this aim, we will examine a subset of the data, made up of the monthly averages from January 2022 - December 2024 (hence, over three complete years) from quadrats at mid-tidal levels only (specifically, the distance classes from 30 m to 50 m from the seawall, inclusive) and from transects 3 and 4 only, as these have highly similar tidal height profiles. The subsetted and averaged data[§] are held in the file 'Macroalgae_subset.pri' in the 'Examples_P8 > BC_macroalgae' folder.

Set up covariables to code for cyclicity

To test for annual monthly cycles of change in the macroalgal assemblages through these years, we will first need to set up two covariables that code for this cyclicity model, as outlined on the previous page in section 9.2. Steps in the calculation are as follows:

- Let k be the total number of time-points (steps) in the cycle.
- If the time-points, $t = 1, ..., k$, are to be equally spaced, then calculate $c = 2\pi / k$.
- For samples occuring at each time-point t in the cycle, the corresponding values for the two covariables, x_1 and x_2, are given by:
  - $x_1 = \cos(ct)$; and
  - $x_2 = \sin(ct)$.

The calculations for this example are shown in Table 9.1. The two covariables of 'sin(ct)' and 'cos(ct)', which together code for an annual monthly cycle of change in these macroalgal assemblages are provided in the 'Covariables_subset.pri' data file for this example.

Table. 9.1. Construction of covariables that code for an annual monthly cycle over k = 12 months. For each month (t), we have: 'Prop. Dist.' - the proportional/fractional distance around the circumference of the circle (in twelfths) with each succeeding month; 'Dist. Rad' - the distance around the circumference of the circle in radians; and the two covariables that together code for the pattern of cyclicity in 2 dimensions: '$\sin(ct)$' and '$\cos(ct)$', where $c = 2\pi / k$.

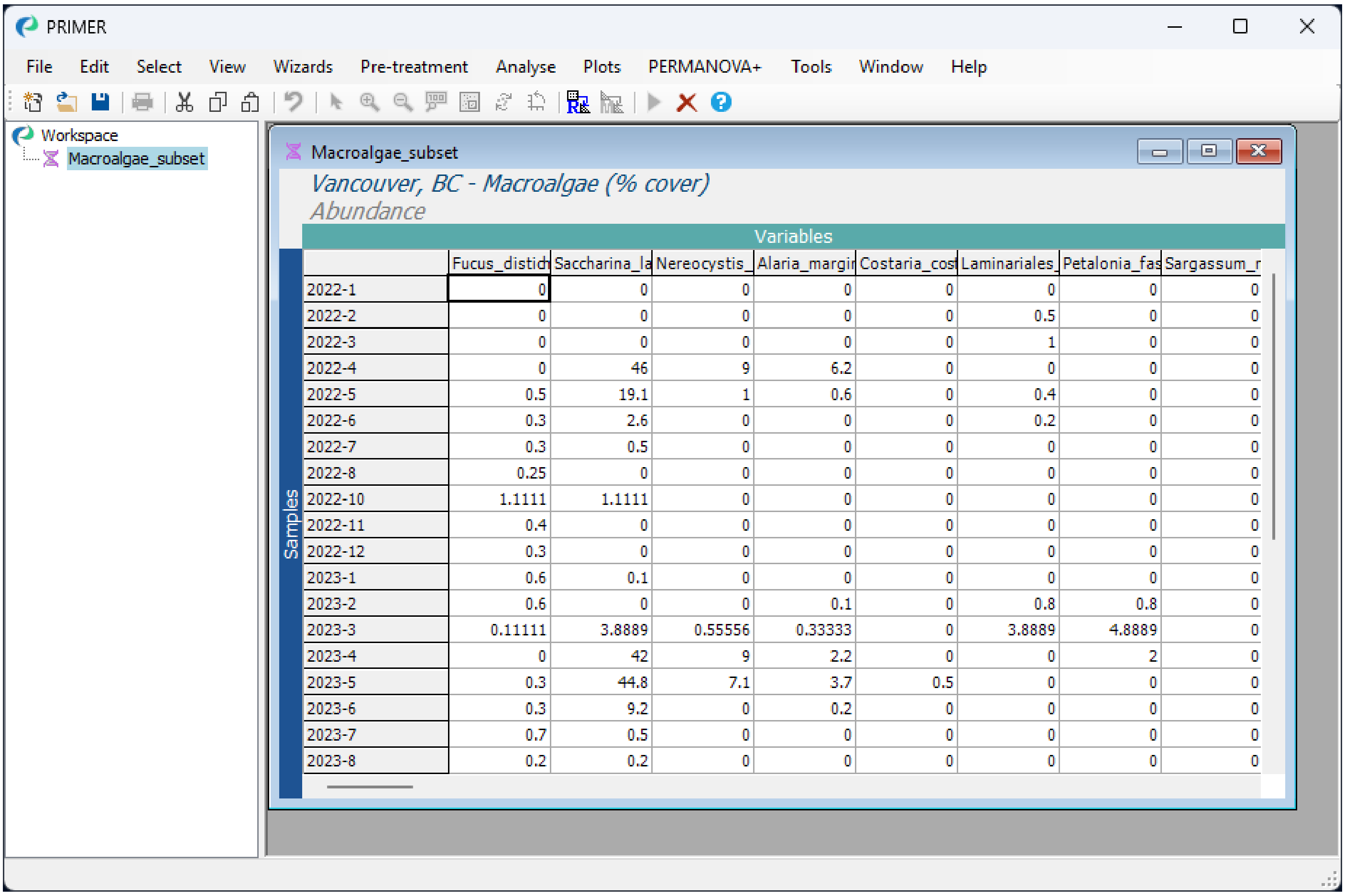
Month (t)	Prop. Dist. (t/k)	Dist. Rad. (ct)	$\cos(ct)$	$\sin(ct)$
1	0.08333	0.52360	0.86603	0.5
2	0.16667	1.04720	0.5	0.86603
3	0.25	1.57080	0	1
4	0.33333	2.09440	- 0.5	0.86603
5	0.41667	2.61799	- 0.86603	0.5
6	0.5	3.14159	- 1	0
7	0.58333	3.66519	- 0.86603	- 0.5
8	0.66667	4.18879	- 0.5	- 0.86603
9	0.75	4.71239	0	- 1
10	0.83333	5.23599	0.5	- 0.86603
11	0.91667	5.75959	0.86603	- 0.5
12	1	6.28319	1	0

Note that if some other, possibly unequal, positions along the circumference of the circle are required, each of these can be easily calculated directly from the *fractional distances* they take along the circumference of the circle. Thus:

- Let f_t be the fractional distance along the circumference of the circle for any given time-point t.
- The corresponding values for the two covariables, x_1 and x_2, at time-point t are then given by:
  - $x_1 = \cos(2\pi f_t)$; and
  - $x_2 = \sin(2\pi f_t)$.

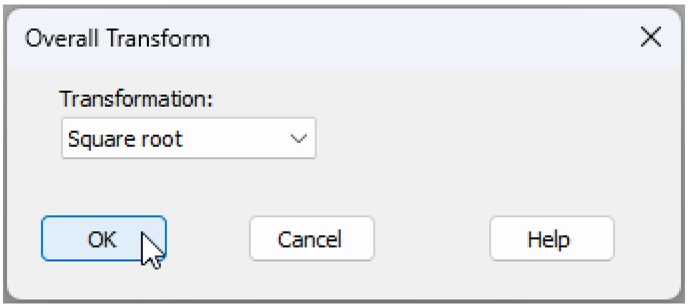
Open the file, transform the data and calculate resemblances

1. Open up PRIMER 8 and click **File** > **Open...** to open the data file named 'Macroalgae_subset.pri' (found inside the 'Examples_P8 > BC_macroalgae' folder).

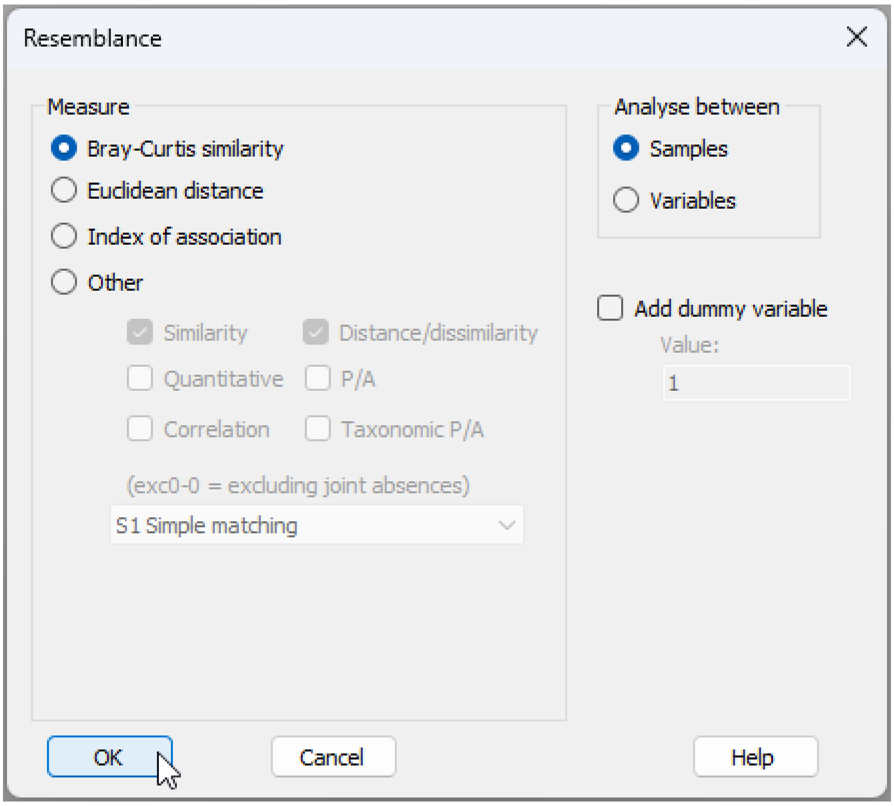


	Fucus_distich	Saccharina_la	Nereocystis_	Alaria_margir	Costaria_cost	Laminariales_	Petalonia_fas	Sargassum_r
2022-1	0	0	0	0	0	0	0	0
2022-2	0	0	0	0	0	0.5	0	0
2022-3	0	0	0	0	0	1	0	0
2022-4	0	46	9	6.2	0	0	0	0
2022-5	0.5	19.1	1	0.6	0	0.4	0	0
2022-6	0.3	2.6	0	0	0	0.2	0	0
2022-7	0.3	0.5	0	0	0	0	0	0
2022-8	0.25	0	0	0	0	0	0	0
2022-10	1.1111	1.1111	0	0	0	0	0	0
2022-11	0.4	0	0	0	0	0	0	0
2022-12	0.3	0	0	0	0	0	0	0
2023-1	0.6	0.1	0	0	0	0	0	0
2023-2	0.6	0	0	0.1	0	0.8	0.8	0
2023-3	0.11111	3.8889	0.55556	0.33333	0	3.8889	4.8889	0
2023-4	0	42	9	2.2	0	0	2	0
2023-5	0.3	44.8	7.1	3.7	0.5	0	0	0
2023-6	0.3	9.2	0	0.2	0	0	0	0
2023-7	0.7	0.5	0	0	0	0	0	0
2023-8	0.2	0.2	0	0	0	0	0	0

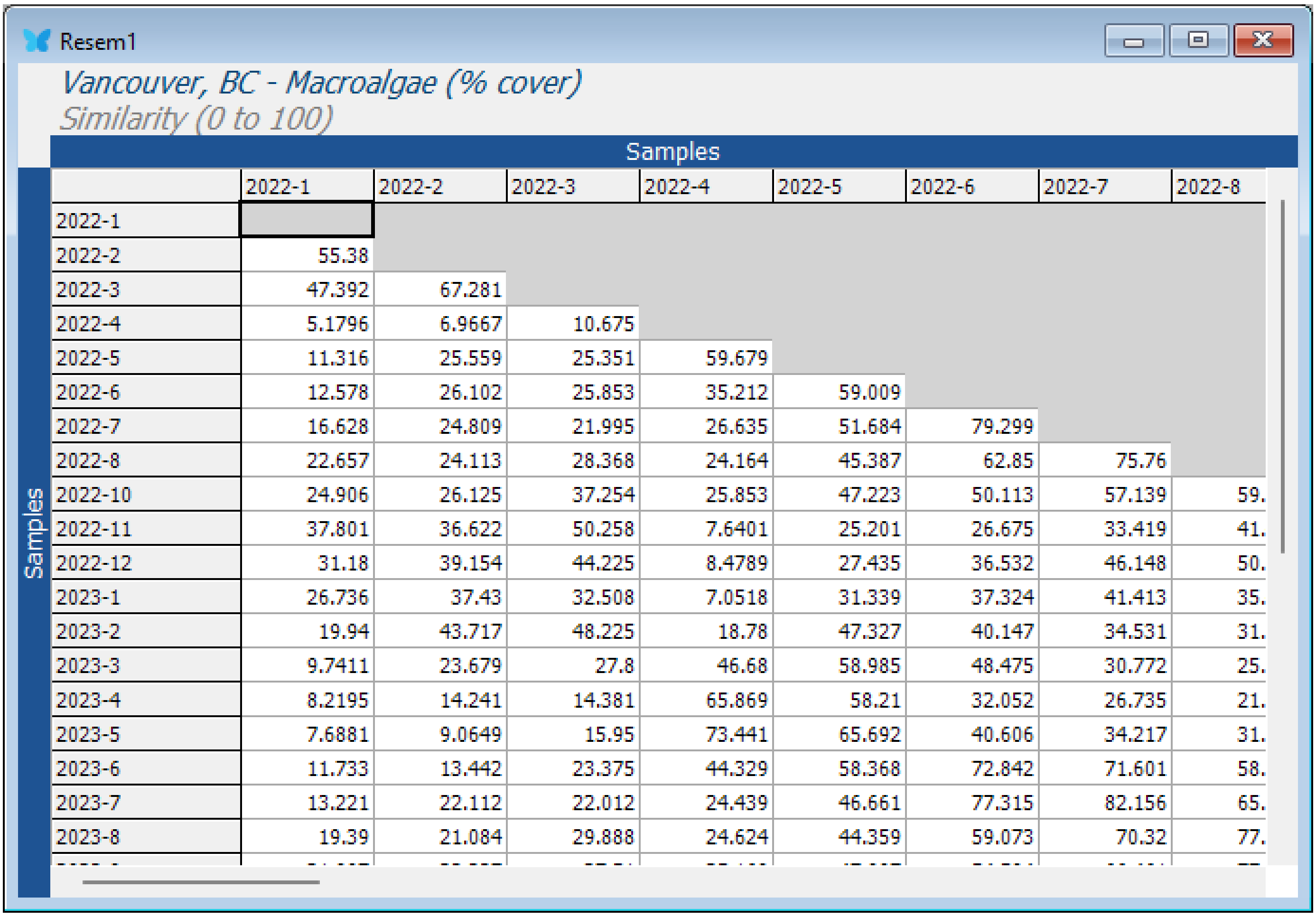
2. From the 'Macroalgae_subset' data sheet, click **Pre-treatment** > **Transform(overall)...**, choose '**Square root**' from the drop-down menu, then click '**OK**'.†



3. This will yield a data sheet of square-root transformed values called 'Data1'. From this, calculate Bray-Curtis similarities among the samples by clicking **Analyse** > **Resemblance...** and going ahead with the defaults in the resemblance dialog (as shown below) by clicking '**OK**'.



The result will be a resemblance matrix among the samples, called 'Resem1', as shown below:

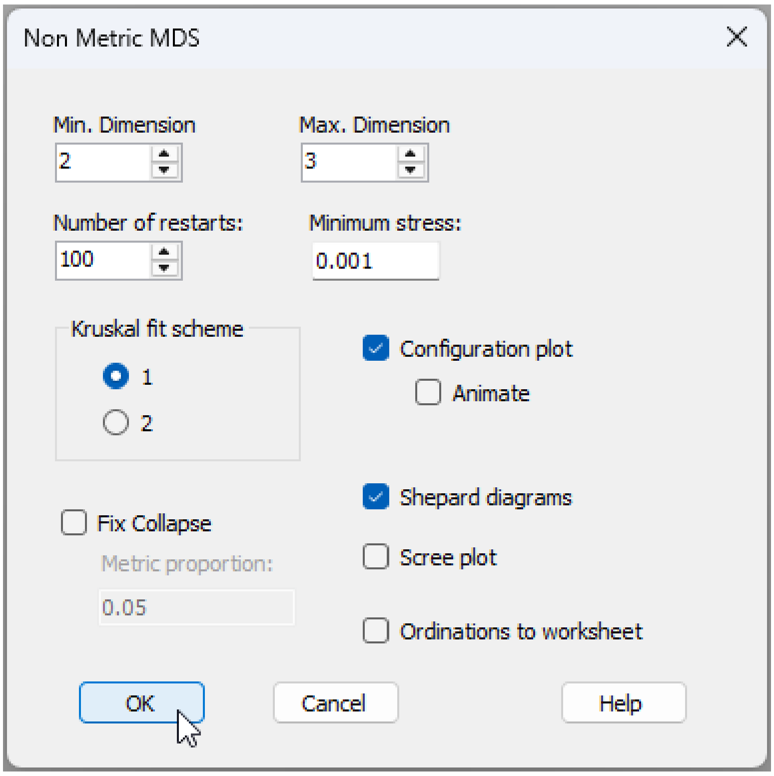


Samples	2022-1	2022-2	2022-3	2022-4	2022-5	2022-6	2022-7	2022-8
2022-1								
2022-2	55.38							
2022-3	47.392	67.281						
2022-4	5.1796	6.9667	10.675					
2022-5	11.316	25.559	25.351	59.679				
2022-6	12.578	26.102	25.853	35.212	59.009			
2022-7	16.628	24.809	21.995	26.635	51.684	79.299		
2022-8	22.657	24.113	28.368	24.164	45.387	62.85	75.76	
2022-10	24.906	26.125	37.254	25.853	47.223	50.113	57.139	59.
2022-11	37.801	36.622	50.258	7.6401	25.201	26.675	33.419	41.
2022-12	31.18	39.154	44.225	8.4789	27.435	36.532	46.148	50.
2023-1	26.736	37.43	32.508	7.0518	31.339	37.324	41.413	35.
2023-2	19.94	43.717	48.225	18.78	47.327	40.147	34.531	31.
2023-3	9.7411	23.679	27.8	46.68	58.985	48.475	30.772	25.
2023-4	8.2195	14.241	14.381	65.869	58.21	32.052	26.735	21.
2023-5	7.6881	9.0649	15.95	73.441	65.692	40.606	34.217	31.
2023-6	11.733	13.442	23.375	44.329	58.368	72.842	71.601	58.
2023-7	13.221	22.112	22.012	24.439	46.661	77.315	82.156	65.
2023-8	19.39	21.084	29.888	24.624	44.359	59.073	70.32	77.

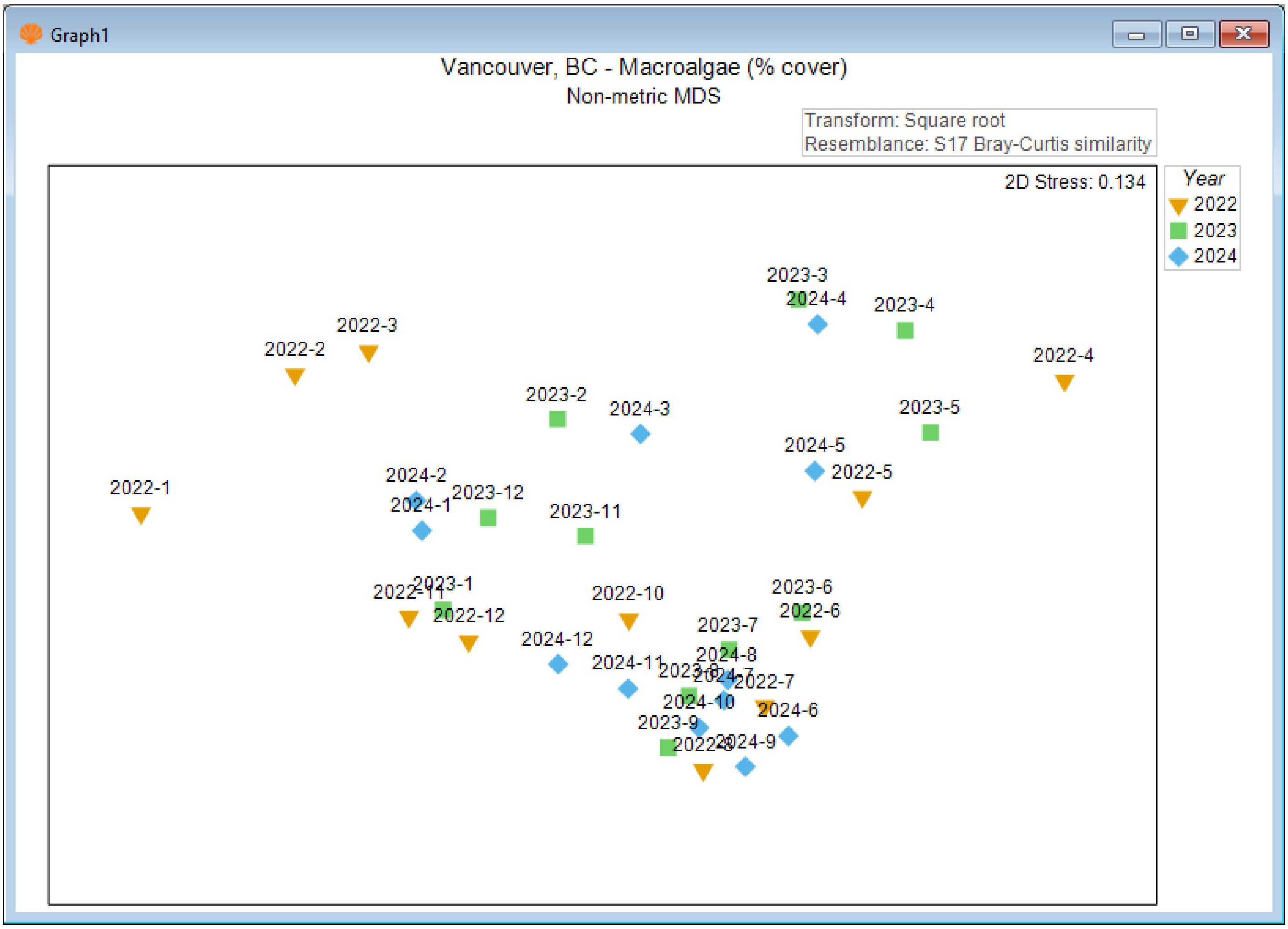
Create an ordination with temporal trajectories

We wish to visualise these relationships among the sampling units, which may be achieved using ordination *via* non-metric MDS. More specifically, we hope to visualise how macroalgal assemblages may change from month to month and/or from year to year. We wish, in particular, to discern whether a cyclical pattern of change in macroalgal assemblage structure is evident through each calendar year on our ordination plot, so a trajectory that connects the sequentially ordered months through each year would be helpful here.

4. From the 'Resem1' matrix, click **Analyse** > **MDS** > **Non-metric MDS (nMDS)...**, go with the defaults in the 'Non Metric MDS' dialog, but choose 'Number of restarts: 100' (as shown below), then click '**OK**'.

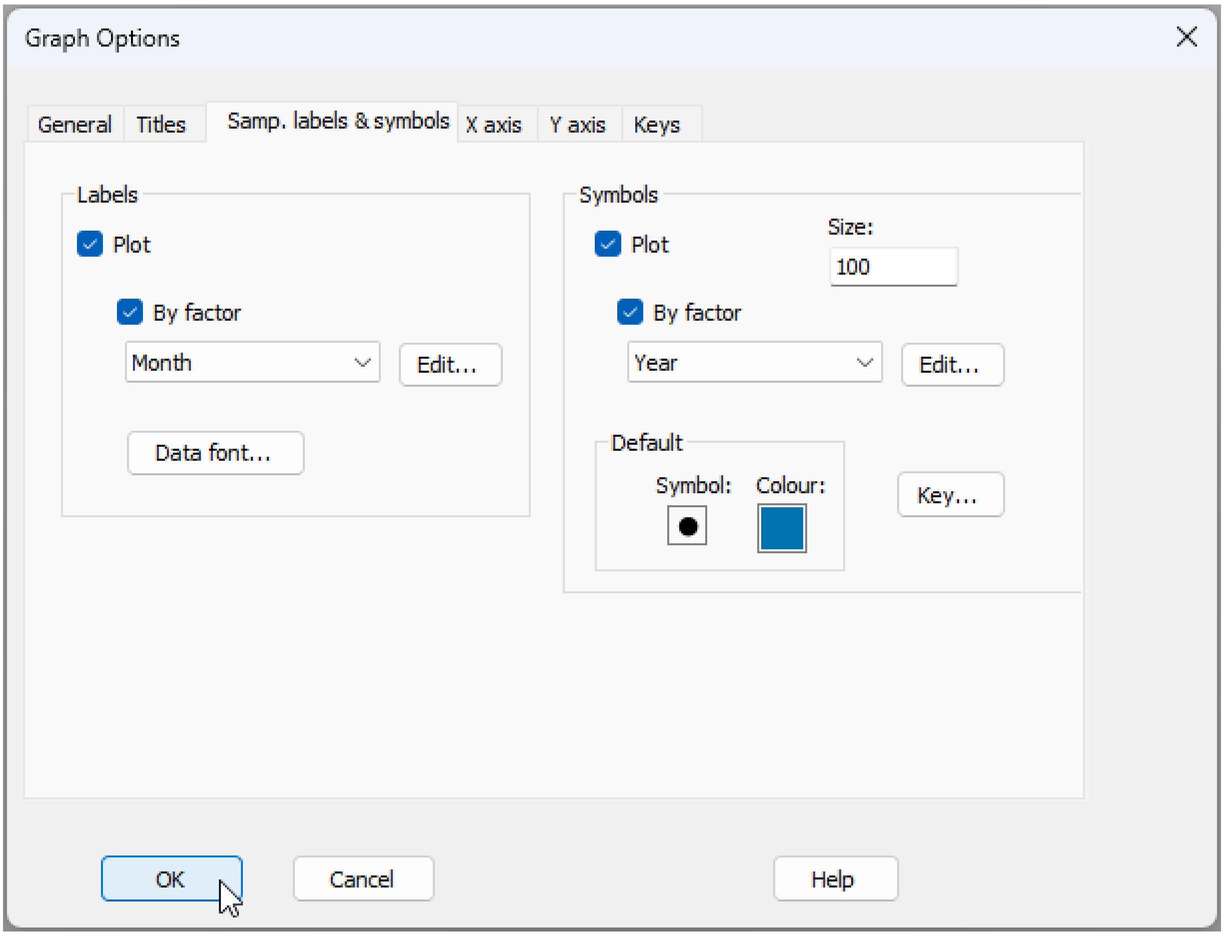


5. The lowest stress 2D nMDS that could be achieved by the algorithm will be provided in an item called 'Graph1' in the Explorer tree (as part of the 'MultiPlot1' output). It should look like this:[‡]

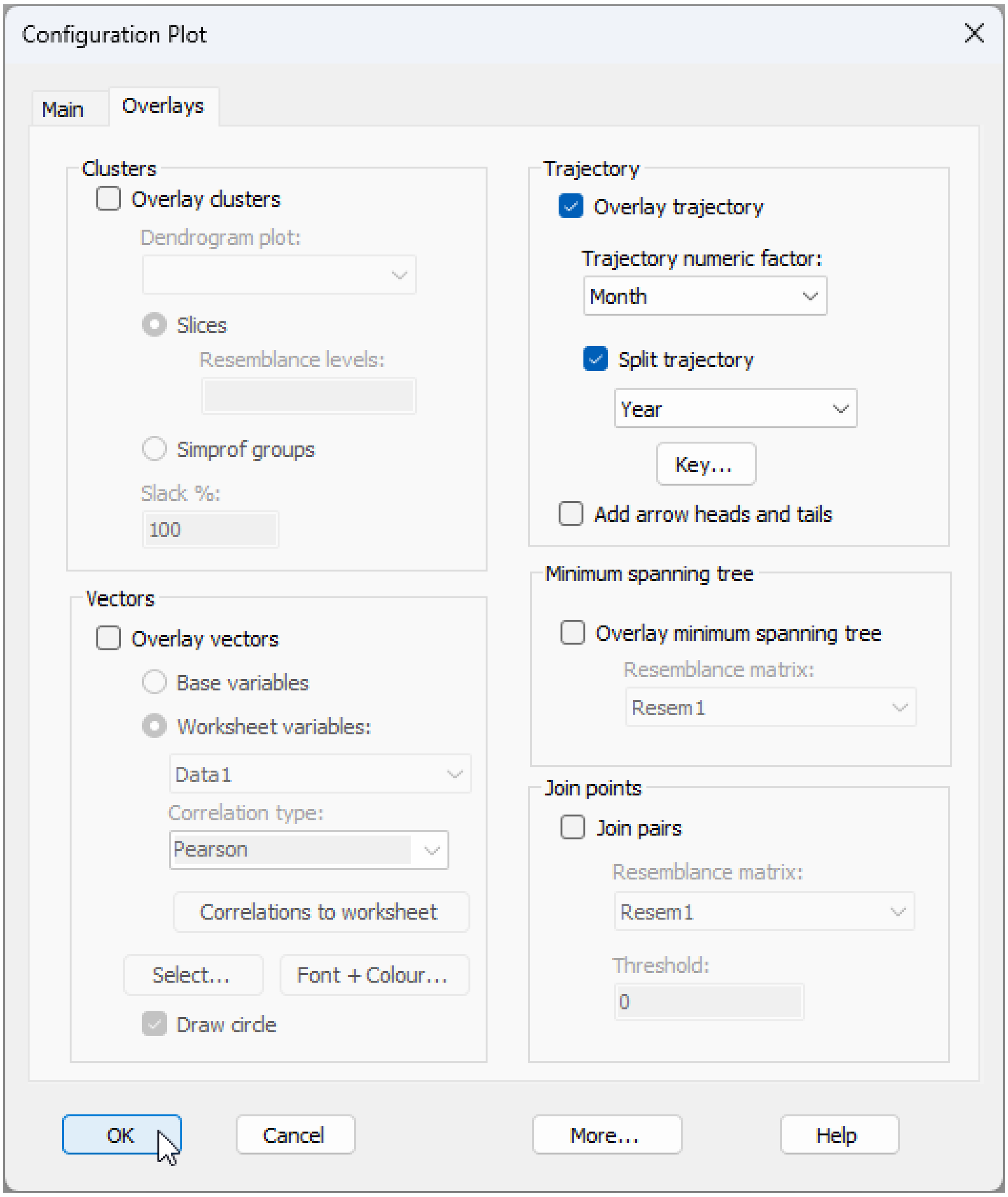


Let's now take two simple steps to make this graphic easier to interpret by reference to our temporal factors of interest.

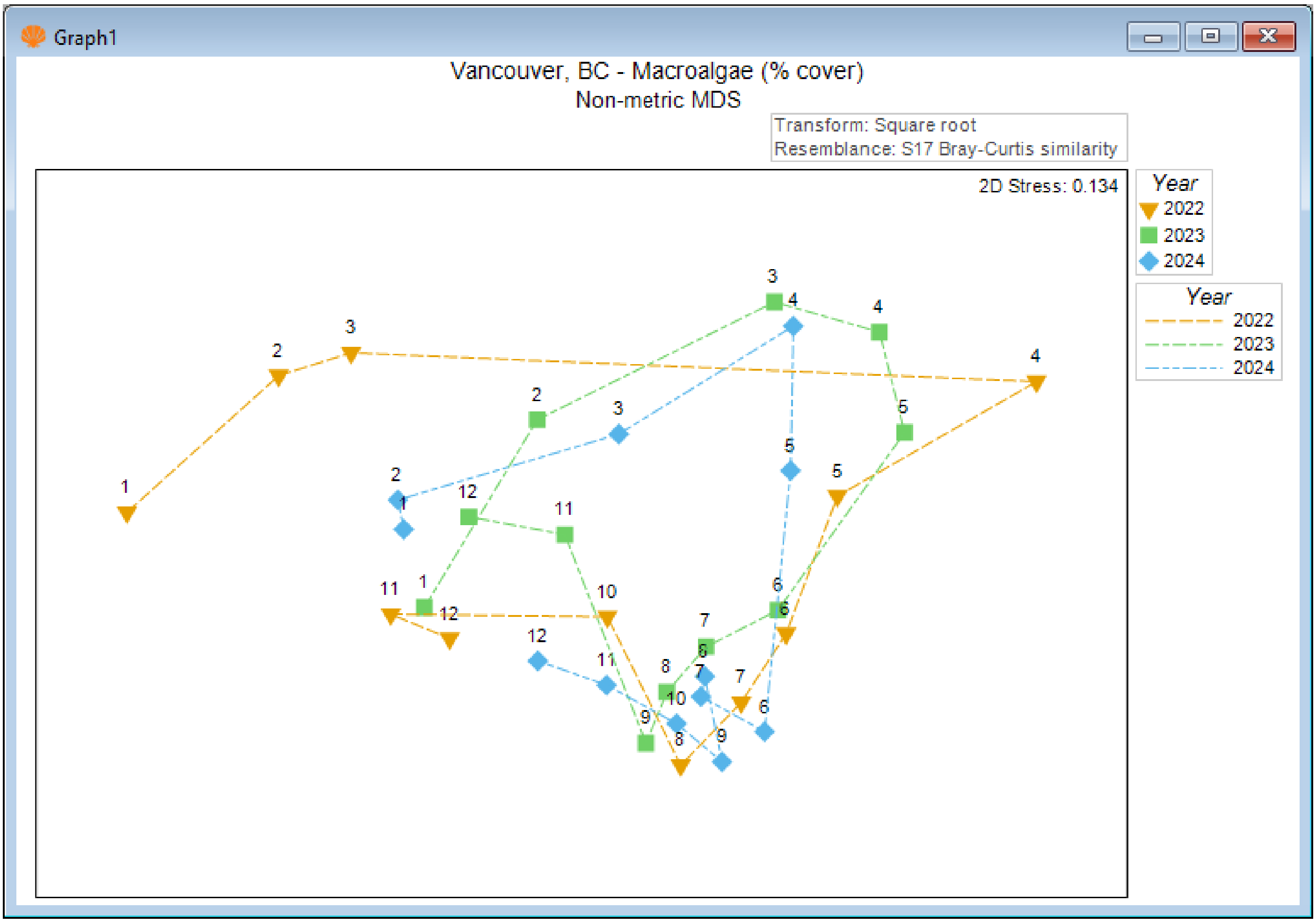
6. First, let's change the labels on the points to show the months. From 'Graph1', click **Graph > Sample Labels & Symbols...**, then choose to plot the 'Labels' $\checkmark$ **By factor** Month, (and leave it so that the Symbols are plotted $\checkmark$**By factor** Year), precisely as shown in the dialog below, then click **OK**.



7. Second, let's overlay a trajectory through the consecutive months within each calendar year. From 'Graph1', click **Graph** > **Special...**. Click on the '**Overlays**' tab, then choose to $\checkmark$**Overlay trajectory** by the numeric factor of Month. Choose also to $\checkmark$**Split trajectory** by Year, and untick the box in front of the words '$\Box$ **Add arrow heads and tails**', as shown in the dialog below:



The resulting nMDS plot will now look like the image below, which is much easier to interpret.

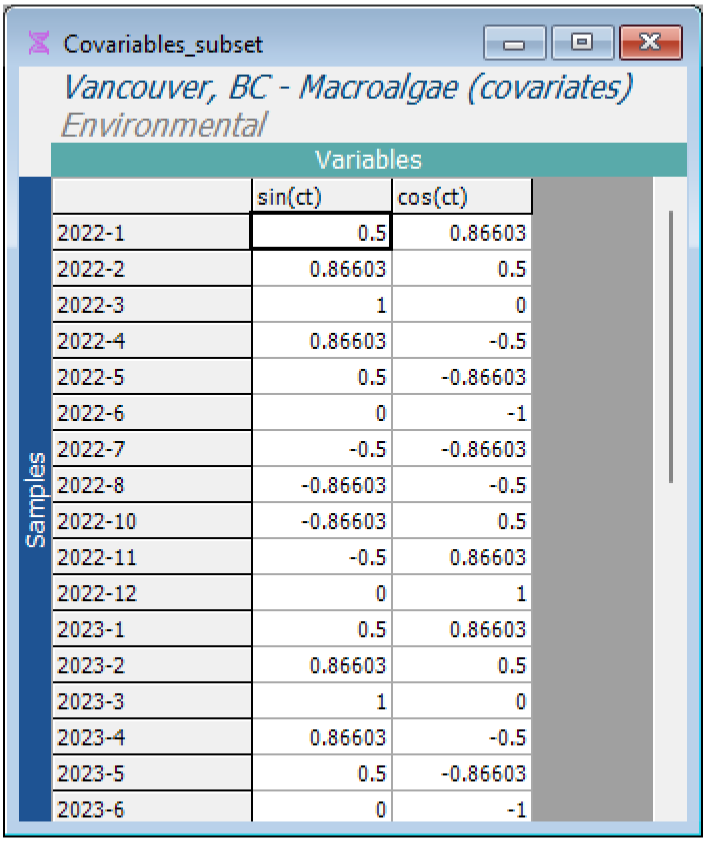


In each year, we can see a clear pattern of cyclical change in these macroalgal assemblages, from months 1 through 12. The greatest changes in assemblages appeared to occur in the early months of 2022 (months 1-4 for the amber symbols, from left to right in the plot), but the annual cycles of monthly change that occured in subsequent years (2023 and 2024, in green and blue) were broadly similar in both size and direction.

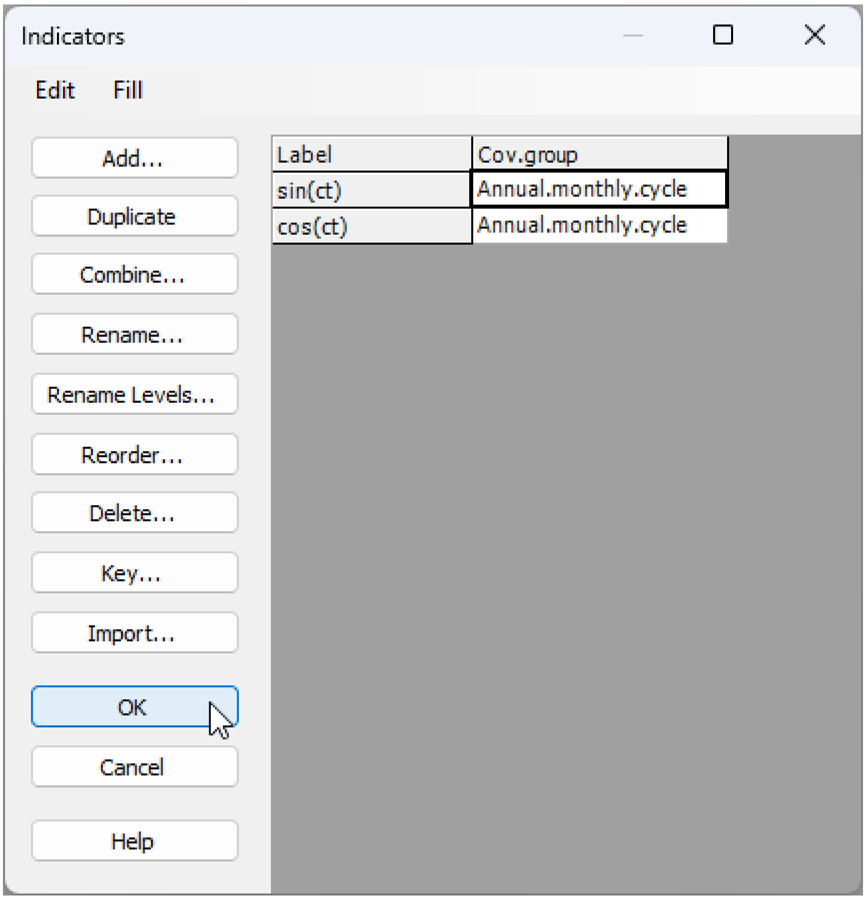
Create the PERMANOVA design file

Having seen these patterns, we are naturally keen now to perform a *formal test of cyclicity*. To do this, we need to bring in the covariables for these data and create an appropriate design file to run the PERMANOVA. This is where we will implement the new feature in PERMANOVA to *group covariables* together, as our cyclical model requires the simultaneous action of two covariables in order to adequately capture this two-dimensional pattern in the multivariate response.

8. Bring in the file containing the relevant temporal covariables that are associated with the subsetted and averaged dataset. Click **File** > **Open...** and open the data file named 'Covariables_subset.pri' (found inside the 'Examples_P8 > BC_macroalgae' folder).

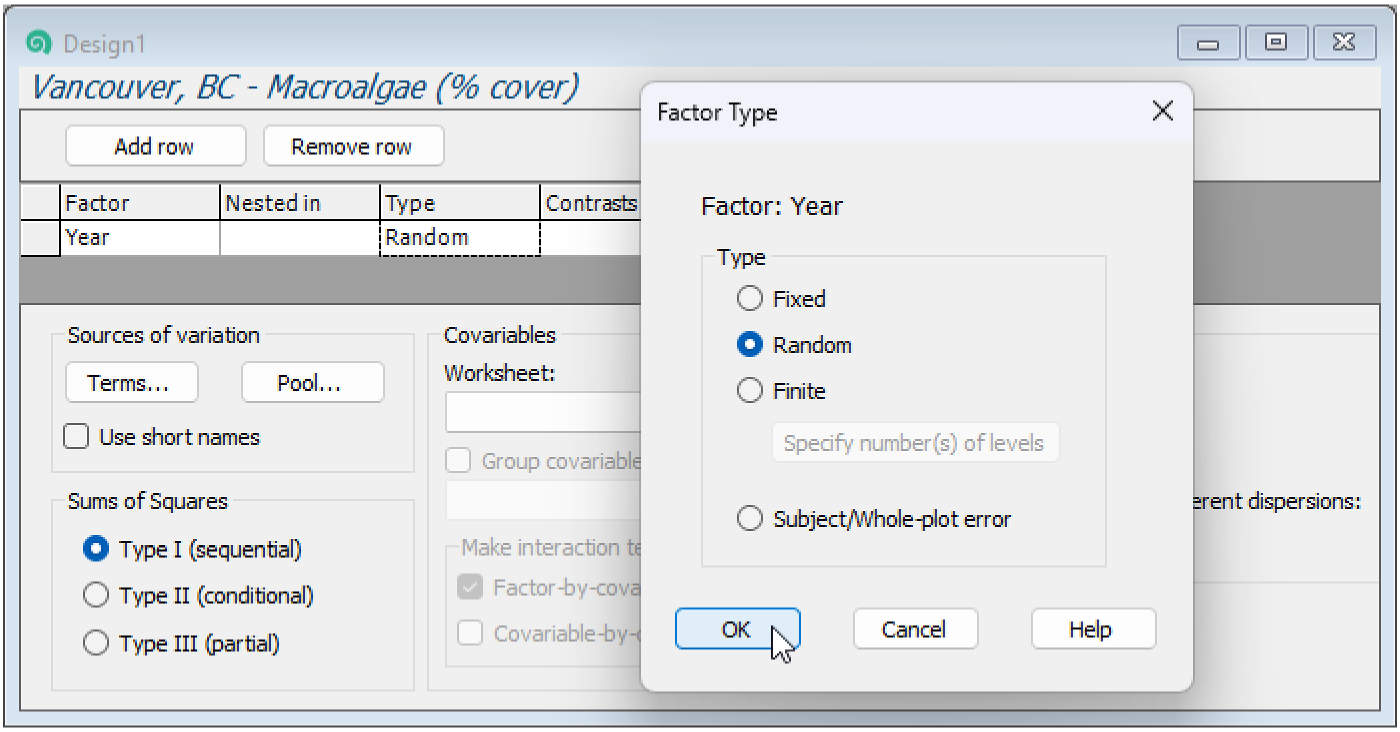


From the 'Covariables_subset' sheet now in the Explorer tree, click **Edit** > **Indicators...** and you will see that there is an indicator called 'Cov.group' for this dataset, showing that both of the variables 'sin(ct)' and 'cos(ct)' belong to a single group called 'Annual.monthly.cycle'.

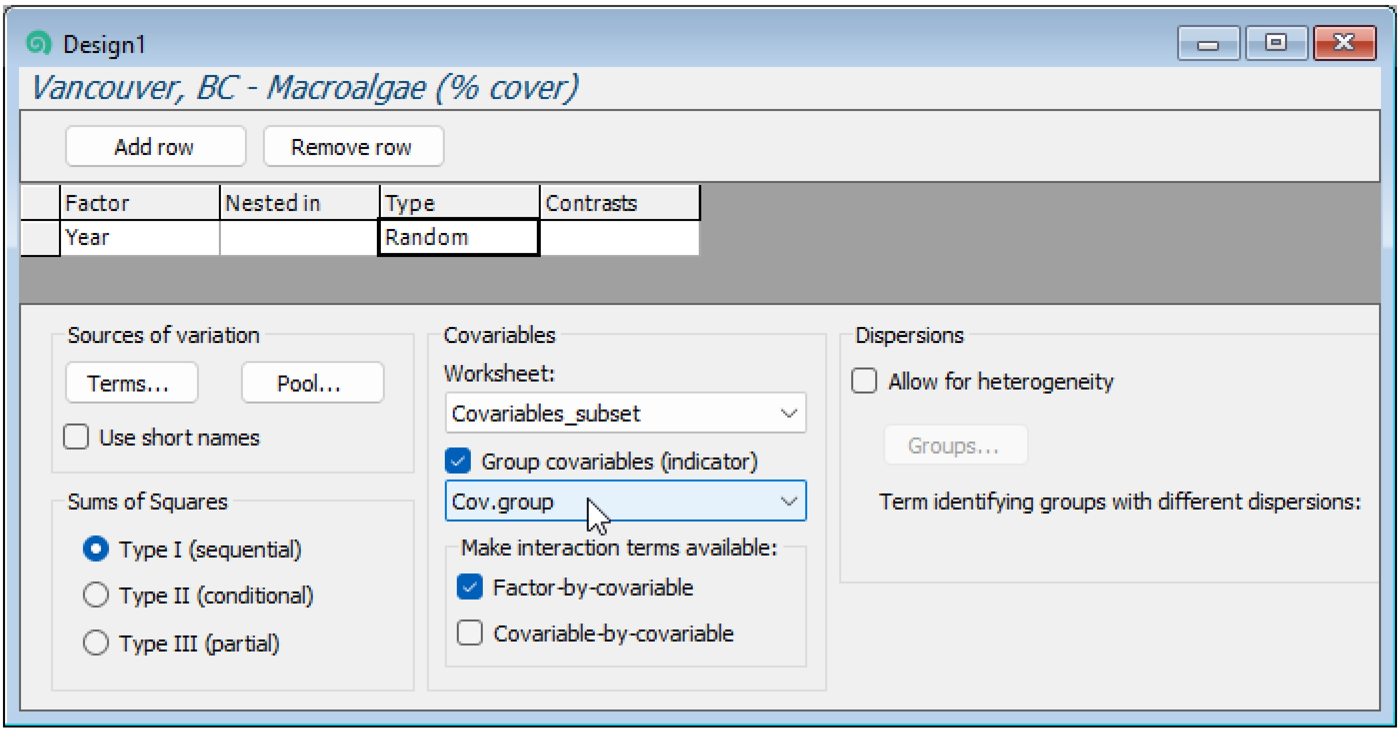


9. From the 'Resem1' resemblance matrix, click **PERMANOVA+** > **Create PERMANOVA design...**, then do the following:

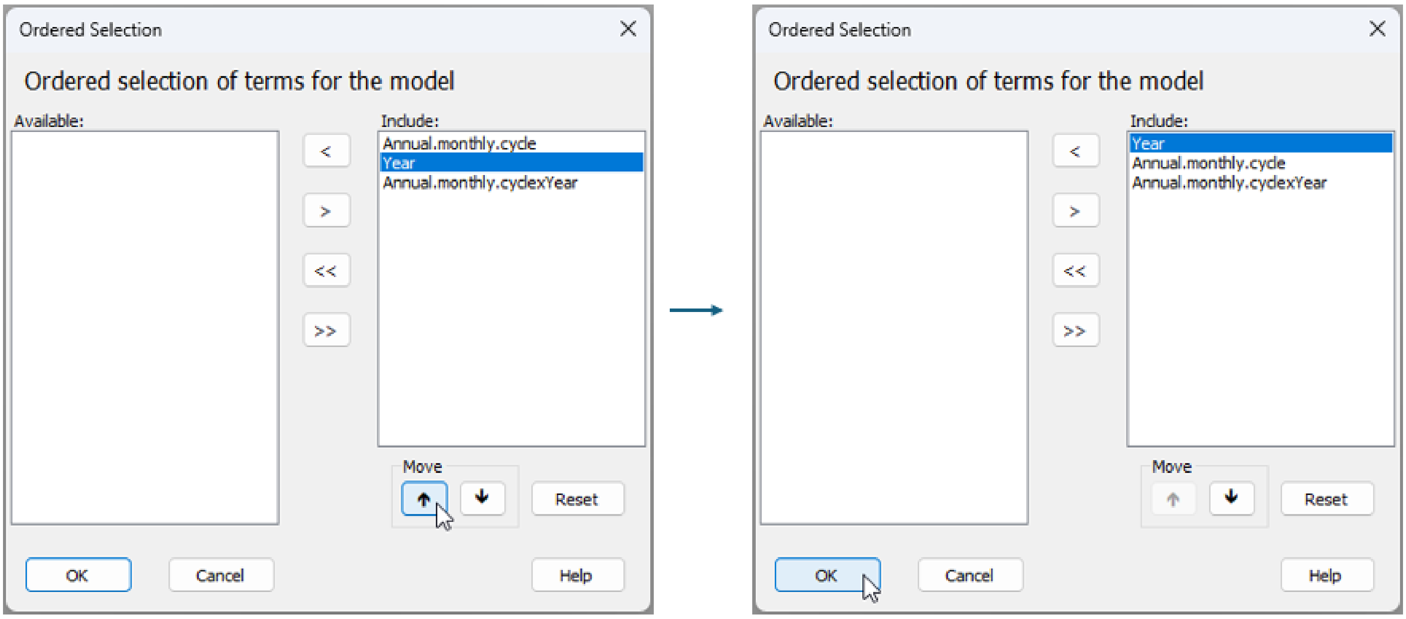
- Double-click the first blank cell (under the word 'Factor'), and choose the factor of 'Year' from the drop-down menu so that it shows in the first cell (row 1, column 1) of the design file. Also, specify that the 'Type' of this factor is 'Random' (double-click the cell in row 1 column 3 to bring up that dialog).



- Next, in the section of the dialog entitled '**Covariables**', use the drop-down menu to choose the '**Worksheet:**' as 'Covariables_subset', then tick the box to $\checkmark$ **Group covariables (indicator)** and choose the indicator 'Cov.group' for this, as shown below:



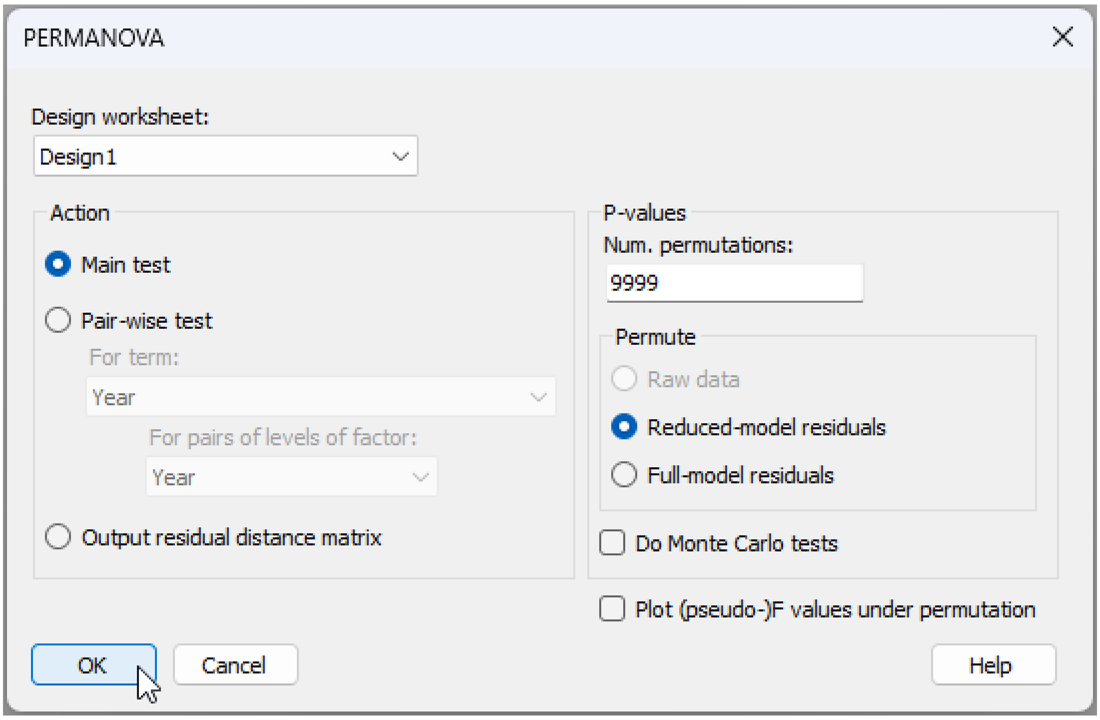
If you now click on the 'Terms' button (Terms...) under the words '**Sources of variation**' in this design file, you will see that the terms in the model include 'Year', 'Annual.monthly.cycle', and their interaction. Notice that the default in PERMANOVA is to put the covariable(s) first. For this analysis, however, it might seem somewhat more natural to fit 'Year' first, as the largest temporal scale. To order the terms in this way, click on the word 'Year' inside the '**Include:**' box and then on the upwards arrow under the word '**Move**' in the dialog, then click '**OK**', as shown below.



Now we are ready to run the PERMANOVA model on the basis of this finalised design file.

Run the PERMANOVA analysis

10. From the 'Resem1' matrix, click **PERMANOVA+** > **PERMANOVA...**. Ensure that the correct design worksheet file is nominated here ('Design1'), keep all of the other defaults, as shown below, then click '**OK**'.



The PERMANOVA output file and table of results are very clear. The annual monthly cycle is a term with 2 degrees of freedom, and this is highly statistically significant ($F_{2,25}$ = 14.491, P = 0.0001). There are also significant differences among the three years ($F_{2,25}$ = 3.8407, P = 0.0001), but the sizes of yearly effects are not as large as the effects of monthly seasonal changes within each year (*cf.* the sizes of their respective estimated components of variation). Interestingly, there is also a significant interaction term ($F_{4,25}$ = 1.5997, P = 0.0465), indicating that the nature of the monthly cycles differs somewhat among the years. These results align well with the earlier patterns seen in the nMDS ordination; specifically, clear cyclical patterns and also some deviations in the size of the cyclical pattern for 2022 compared to the other two years.

PERMANOVA1

PERMANOVA table of results

Source	df	SS	MS	Pseudo-F	P(perm)	Unique perms
Year	2	7079.8	3539.9	3.8407	0.0001	9925
Annual.monthly.cycle	2	26712	13356	14.491	0.0001	9944
Annual.monthly.cyclexYear	4	5897.4	1474.4	1.5997	0.0465	9912
Res	25	23042	921.67			
Total	33	62731				

Estimates of components of variation

Source	Estimate	Sq.root
V(Year)	231.22	15.206
S(Annual.monthly.cycle)	366.52	19.145
V(Annual.monthly.cyclexYear)	49.01	7.0007
V(Res)	921.67	30.359

Details of the expected mean squares (EMS) for the model

Source	EMS
Year	1*V(Res) + 11.324*V(Year)
Annual.monthly.cycle	1*V(Res) + 33.926*S(Annual.monthly.cycle)
Annual.monthly.cyclexYear	1*V(Res) + 11.277*V(Annual.monthly.cyclexYear)

Construction of Pseudo-F ratio(s) from mean squares

Source	Numerator	Denominator	Num.df	Den.df
Year	1*Year	1*Res	2	25
Annual.monthly.cycle	1*Annual.monthly.cycle	1*Res	2	25
Annual.monthly.cyclexYear	1*Annual.monthly.cyclexYear	1*Res	4	25

[§]*There were no data available from transects 3 and 4 at these mid-tidal levels (30 m - 50 m distances) in September 2022 and in October 2023, due to the tide not being sufficiently low on the dates of sampling, so these two data points are missing from this example.*

[¶] *If the estimated percentage cover was between 0% and 5%, then a percentage cover value was estimated to the nearest 1%. If the estimated percent cover was over 5%, then percentage cover was estimated to the nearest 5%. Therefore, the cover values were: 0, 1, 2, 3, 4, 5, 10, 15, 20, ..., 95, 100. Note also that the total percentage cover measured within a single quadrat could exceed 100%, due to there being multiple layers of different species of macroalgae.*

[†] *The square-root transformation is typically a good one to use when dealing with percentage cover data, as the raw values tend to range from 0 to ~100 percent, hence the square-root values will*

tend to range from 0 to 10. This places different taxa on a fairly even footing, yet retains information regarding differences in relative percentage cover values.

---

‡*You may find that your particular nMDS solution has the same stress, but is 'reflected' across the X and/or Y axes. From 'Graph1, you can click Graph > Flip X and/or Graph > Flip Y in order to check this and (optionally) change it.*