Changes from PRIMER 6 to PRIMER 7

Wizards and major new analysis options

Dialog boxes generally are more ‘wizard’-based, unifying and simplifying parameter input and, importantly, there is a new **Wizards** menu (Section 10) with three items, as follows:

1. **Basic multivariate analysis** wizard which gathers together the main steps of a core PRIMER analysis in a single dialog box. It is aimed at new users who may initially have trouble formulating a reasonable choice of a sequence of routines. It covers pre-treatment options such as standardising and transforming samples, calculating resemblances, running group-average clustering, non-metric MDS, SIMPER, and either ANOSIM or (Type 1) SIMPROF for testing *a priori* or *a posteriori* groups, and offers robust defaults, tailored to data type (biotic or environmental) and availability of a factor with repeat levels. Importantly, the analysis sequence is fully laid out in the Explorer tree, allowing the new user to reconstruct the menu selections needed for this basic analysis.
2. **Matrix display** wizard which runs the new and comprehensive **Shade Plot** routine (Sections 4 & 10) which is an ‘image’ of the data matrix, with species abundances – or other quantity values – shown by depth/colour of shading. At a simple level, shade plots aid choice of transformation or other data pre-treatment but, when the species and/or sample axes are suitably grouped, clustered, seriated, non-linearly linked or subject to combinations of these, they become powerful tools for interpreting sample patterns (established by testing and seen in ordinations) in terms of individual species driving those patterns. The Matrix display wizard sets up the (rather involved) sequence of sample and species resemblance calculations, and clustering and seriation steps, to give a robust, initial shade plot, which the user can then refine by further ordering or constraining of axes.
3. **Coherence plots** wizard which uses a novel (Type 3) SIMPROF testing series on standardised or normalised biotic or environmental variables, to define e.g. groups of Coherent species having statistically indistinguishable patterns of response over the samples, within sets, and statistically different responses among sets. The wizard then displays the sets using the new **Line Plot** routine.
4. The **CLUSTER** menu now includes a new linkage option under its agglomerative hierarchical menu, in addition to UPGMA, single and complete linkage, namely flexible beta – a standard WPGMA extension – and a cophenetic distance matrix can be output to allow computation of cophenetic correlation (Section 6). More novel, however, is the introduction of the following new divisive and ‘flat’ clustering methods, designed for the non-parametric PRIMER framework:
5. **UNCTREE**, a binary divisive algorithm which is an unconstrained form of the v6 **LINKTREE** routine, e.g. successively dividing groups so as to maximise the ANOSIM R statistic between the two groups so formed. SIMPROF (Type 1 or 3) tests give a stopping rule for the binary divisions.

6. **kRCLUSTER**, a flat-form (non-hierarchical) method, based on the idea of *k*-means clustering in which the full set of samples are divided into a pre-specified number (*k*) of groups, minimising the within-group sums of squares ($\equiv$ within-group squared Euclidean distances). Generalising this to PRIMER's context, kRCLUSTER seeks to find a division into *k* groups which maximises the non-parametric ANOSIM R statistic, and is therefore definable for any resemblance measure, not just Euclidean distance. The routine is also able to choose *k* by computing *k*-R clusters for successively larger *k* until none of the groups is statistically heterogeneous, as seen by (Type 1 or 3) SIMPROF.
7. The **MDS** menu for non-metric MDS (*n***MDS**) now allows calculation of ordination axes in any number of dimensions (2 or more), not just 2 and 3, and any combination of the higher-d axes can be plotted in a 2- or 3-d plot (as for PCA in v6); a *Scree plot* shows the declining MDS stress for increasing dimensionality. Another important addition to *n*MDS is the ability to Fix collapse of the non-metric MDS plot when a sample (or samples) are sufficiently distant from remaining samples to cause an indeterminacy in the rank order information. This novel procedure is implemented by using both of the ideas in the following new additions to the MDS routine (Section 8):
8. **Metric MDS** (*m***MDS**) seeks to preserve the actual dissimilarities in the resemblance matrix as distances in the low-d ordination, rather than preserving only their rank order (as in *n*MDS). This can be a very successful alternative to *n*MDS when there are very few points (perhaps 4 or 5) and the rank orders do not carry enough information; this can happen easily for plots of group means. (Note, this is not PCO, as implemented in PERMANOVA+, which projects points into low-d space, whereas *m*MDS places points in that space, minimising stress in the linear Shepard plot).
9. Threshold metric MDS (*tm*MDS) which, instead of fitting the Shepard plot by a straight line through the origin (*m*MDS), fits a straight line with an intercept. It borrows from *n*MDS the ability to truncate small dissimilarities to effectively zero distances in the ordination (reflecting the fact that replicates from exactly the same condition are never 0% dissimilar because of sampling error) and from *m*MDS the preservation of linear distance additions with linear dissimilarity changes, where conditions differ. It can be a useful compromise for low species turnover among samples.
10. **Combined MDS** minimises an equal mixture of stress functions from two *n*MDS ordinations, which has potential application to produce a consensus view of among-sample relationships for two sets of variables which cannot be merged into a single matrix – perhaps needing different resemblance measures (e.g. biotic and abiotic; motile organism counts and colonial species areas). It is combined stress from *n*MDS with a small *m*MDS component that 'fixes' an *n*MDS collapse.
11. The **Bootstrap Averages** routine is another significant innovation in v7, providing a region estimate for each group mean in a 2-d (or 3-d) ordination plot from samples with an *a priori* one-way group structure (or a 2- or higher-way crossed group design flattened to simple groups with replicates). This bootstraps the samples in an m-dimensional *m*MDS space for which m is large enough for the among-sample distances to closely match the original dissimilarities (as judged by Pearson matrix correlation > 0.99 say) but small enough to avoid the unrepresentativeness of bootstrap samples in very high dimensions. The averages of repeated bootstrap samples for each group are ordinated into 2- or 3-d to form a region estimate for mean communities, which (in 2-d) is smoothed and marginally bias-corrected (but not formally a confidence region), Section

17.

12. The **ANOSIM** routine has been greatly expanded, firstly to include a concept of ordered group structure for one (or more) of the factors input to the ANOSIM test. This tests the null hypothesis of no group differences against a directed alternative in which the groups are in specified sequence (e.g. years under a time trend, spatial gradients, increasing impact conditions). It then permits a more powerful test, based on generalising the ANOSIM R statistic to an ordered R^0 (the slope of a regression of dissimilarity ranks against the ranks of a model 'seriation with replication' matrix) – a test which can also be run in the absence of replication, Section 9.
13. A second major extension of **ANOSIM** is to designs with three factors (A,B,C) in all feasible crossed and nested combinations – fully crossed A$\times$B$\times$C, fully nested C(B(A)), nested in crossed C(B$\times$A), and crossed with nested B$\times$C(A). All cases allow any factor to be ordered or not, and non-replicated models either exploit ordering or extensions of the previous approach to 2-way crossed designs without replication (e.g. inferring a B effect from commonality of B level patterns across the levels of A, as measured by a non-parametric matrix correlation ρ), Section 9.
14. The **RELATE** routine is extended (Section 14) to include a 2-way RELATE test, operating in a similar way to 2-way crossed ANOSIM. That is, the matrix-correlation matching statistic (ρ, now including Pearson as well as Spearman and Kendall) of the resemblances to any model (or biotic) matrix is calculated within the strata of a second (group) factor, and averaged. The permutations for the test are now similarly constrained within those strata, so that any effects of a second factor (e.g. site differences) are removed from the test of the first (e.g. an annual trend or seasonal cyclicity).
15. The **BEST** routine is similarly extended to a 2-way BEST procedure and 2-way BEST test, by choosing explanatory (e.g. environmental) variables which 'best explain' the multivariate pattern in the response (e.g. community) variables, simultaneously within strata of a secondary ('nuisance') factor, Section 13. E.g. this can remove location differences in base communities among oil-fields, when fitting contaminant variables to community patterns, simultaneously around several oil-fields.
16. The **SIMPROF** test (now Type 1 SIMPROF) for multivariate structure in (subsets) of samples, mainly used with agglomerative (CLUSTER) and constrained divisive clustering (LINKTREE), has been generalised to all four combinations of sample or variable resemblances, with permutation over samples or variables. E.g. Type 2 SIMPROF is a test for any association amongst species, and Type 3 is used with a cluster analysis (as with Type 1) to test for heterogeneity in associations over a subset of species, which would allow further sub-division of that subset (Section 10) – this is the core component of the Coherence plots wizard. Use of SIMPROF in CLUSTER and LINKTREE therefore now extends to Type 3 (on variables), and both Type 1 and 3 SIMPROF are also options within the new UNCTREE and kRCLUSTER clustering (Section 6).
17. Another new menu item is **Summary Stats**, a minor feature in analysis terms since it simply computes Min, Max, Average, Sum, SD, Variance, Range and number of Non-zero entries for every variable or for every sample, but it has widespread utility in preparing for other analyses, spotting outliers (run Max in both directions), identifying low occurrence species etc. (Section

3).

18. Less noticeable, but with far-reaching ramifications, is that **Resemblance** calculations will all now operate in the presence of Missing! data entries, Section 5. Pairwise elimination of samples (or variables), with one or both entries missing, is undertaken separately for each calculated pair, and each coefficient corrected, where necessary, for their particular bias (arising from unequal numbers of terms in summations). Estimation of missing data by the EM algorithm – the previous **Missing** routine – is still available, and preferable where the rather strict model conditions for its use are met, but the simple bias corrections allow reasonable analyses in other cases with unavoidable and commonly occurring missing data (e.g. samples as questionnaire returns, variables are questions).

Additions to configuration (ordination) plots

PRIMER 7 has added several significant display features (Section 8) to its ordination plots, *n*MDS, *m* MDS and PCA (they also operate with PCO, dbRDA and CAP in the PERMANOVA+ add-on):

19. Bubble plot for a single variable – with values superimposed as circles of differing sizes on the sample points – can now have bubbles of different colours, dependent on the level of a group factor (colours are user-controlled through the usual symbol plotting mechanism).
20. A bubble colour saturation option makes labels plotted on bubbles more visible on a lighter background, and opacity control can make bubbles transparent, making hidden bubbles visible.
21. The key defining bubble sizes is now under user control, allowing specification of the number of bubble sizes drawn in the key, and the data values (actual or as a percentage of their range).
22. Bubble plots for one variable can be drawn with a single user-supplied image, e.g. of a relevant organism (as .jpg, .png etc.), displayed at different (rectangular) sizes in place of a (circular) bubble.
23. Bubble plots are now possible in 3-d configurations, utilising a ‘3-d effect’, giving a reasonable facsimile of a 3-d bubble. The 3-d effect can also be selected for 2-d plots, where it can be quite effective in making superimposed (preferably single-character) labels on each point stand out.
24. A new Segmented bubble plot construct is obtained by specifying more than one variable (*k*, say) to display – ideally on 2-d plots though it will operate in 3-d. Circles are divided into *k* equal sectors of different colours, and the sectors plotted at different sizes according to the data values for that point and variable. The colours, and variable order round the circle, are under user control.
25. The existing Spin option to rotate a 3-d configuration can now be captured as an animation file (in .mp4 or animated .gif format), along with any manual interventions which change the angle of view, etc. The sampling rate (frames per second) and image size are under user control. This should allow 3-d ordination plot rotations to be embedded in, for example, a Powerpoint presentation or as supplementary material for an on-line publication.
26. Ordination plots in which the points form a natural series (in time or space) can be displayed in animated form, in 2-d or 3-d, with points and/or the joining trajectory (or trajectories) fading in and out in this natural order. Sequence animations can again be

captured in an animation file; the speed of traverse through the series is under dynamic user control, and there is initial selection of decay speed, for fade-out of displayed components. It can be used, for example, in tracking natural or impact-induced temporal change (and perhaps recovery) in a longish time series, especially where the community is similar at different times and a static MDS plot of the whole series is cluttered.

27. A third animation option, which can again be captured in .mp4 or .gif format, is the evolution of the MDS iterative process. This is designed as a teaching tool, to see in action how an MDS configuration can sometimes get trapped in a local minimum, and thus the necessity for restarting the iteration from many different initial random configurations of the points.
28. There can now be Split trajectories joining points on an ordination, e.g. multiple time series trajectories drawn for a series of sites in different line types and colours (the latter determined by the symbol colour for the relevant points). Two factors are specified – a numeric factor defining the order in which points are joined and a categorical factor whose levels determine the separate trajectories. [In the authors' experience, this could be one of the most used of the new features!]
29. MDS diagnostics are enhanced by provision of the Minimum Spanning Tree. This is computed for the samples under study as the set of connections of samples to each other, on a single (though branched) route, such that the sum of all the connecting dissimilarities is minimised. It is drawn as a branched trajectory on the 2-d or 3-d MDS plot, so that if it visually departs from an MST that would clearly reduce the total connected length on the low-d plot, this is evidence of stress.
30. An alternative diagnostic is to Join pairs of points with similarity greater than (or dissimilarity less than) some supplied threshold value – in practice a series of threshold values, sequentially – and look for conflicts in the low-d representation given by the ordination (e.g., points close together but not joined, compared with points further apart but joined, i.e., with lower dissimilarity).
31. An Align option now rotates and reflects (and possibly shrinks/stretches, preserving the aspect ratio) the active configuration to best match another supplied ordination (Procrustes analysis). This could be done manually, and thus less precisely, in v6 but the main advantage here is in simplicity and speed when comparing several ordinations under different transforms, taxonomic levels, etc.
32. 2-d ordination plots can be visually merged with a full hierarchical cluster analysis using a Raised dendrogram plot, in which the cluster dendrogram is displayed in the third dimension and the whole structure able to be (manually) rotated, as usual.

Other new plots & plot features

There are many new plot types in PRIMER 7, greatly expanding the ability to view data structures (**Plot** menu, Section 7, but also introduced throughout this manual, wherever they find application).

33. A new *Multiplot* automatically generates groups of related plots together in a single window, to allow broad overview or to use as 'thumbnail' – clicking on individual plots makes them accessible for manipulation. These can all be plots of the same type (eg., a sequence of histograms or line plots) or different types (eg., MDS ordinations and their Shepard diagrams, in a range of dimensions, and a scree plot). Users can create their own

multiplot and fill this with any combination of (single) plots.

34. A major new **Shade Plot** routine images the data matrix, with numerous display combinations for ordering/grouping of both axes. It is the core of the Matrix display wizard – see (2) above – but can be run as a stand-alone routine (e.g., in Section 4 on aiding choice of transformation).
35. Standard **Box Plot**s and **Means Plot**s, for univariate data such as sets of diversity indices, allow respectively the usual non-parametric display of (medians, quartiles, ranges) and normality-based means and 95% confidence intervals for those means, for the supplied group factor of a one-way layout with replication (Section 15). Means plots can use common or separate variance estimates.
36. The histogram plots previously only output as null hypothesis distributions for multivariate tests are now available as a stand-alone **Histogram Plot**, e.g., to use in assessing individual needs for transformation of environmental variables. Histograms for all variables are put into a multiplot.
37. A **Line Plot** displays joined matrix values for a variable (y) along the matrix sample order (x), simultaneously for each variable (with different symbols and joining line colours) in a single plot. Supplying an indicator dividing the variables into groups results in several line plots, held in a multiplot. Such a Line Plot is the main display from the Coherence plots wizard – see (3) above.
38. By default, a **Bar Plot** is stacked, e.g., showing the breakdown of abundances over species for each sample (often for meaned samples then sample-standardised, to give % breakdown), but it can display individual species bars side-by-side in groups for each sample, and has a 3-d form.
39. A **Surface Plot** is relevant only to variables in a meaningful order, e.g., size-classes in particle size distributions or growth curve data, displaying a 3-d surface of (sample, species, data value).
40. **Scatter Plot** produces a single 2-d (x, y) or 3-d (x, y, z) plot of the sample values from 2 (or 3) specified variables which can be from different worksheets, eg., allowing a scatter plot of a diversity index or counts of a single species against an abiotic PC or single variable. Points can be labelled and a group factor can be used to give differing symbols/colours of points for the group levels.
41. An existing plot with a new display structure is the **LINKTREE** tree diagram, now by default more like a CLUSTER dendrogram, which greatly aids flexible identification of samples by labels or symbols – rather than just sample numbers – in the resulting SIMPROF groups, for example. The previous format is retained as an optional ‘classic’ layout. Also, the y axis can use equi-stepped binary divisions (A% scale) in addition to the previous scale using size of group separation (B%).
42. PRIMER 7 also adds a number of significant additional features to existing plots:
    a) The general Graph Options dialog applying to all plots and which controls display of titles, labels, symbols, axis scales, etc. now adds a tab for Variable symbols and labelling (e.g., in Shade Plots), and a new Key tab to allow control of key label/title sizes, selective suppression of keys etc.
    b) As for the Scatter Plot, a **Draftsman Plot** is now able to utilise a group factor to allow different symbol shapes/colours for the different levels of that factor, across the whole plot.

c) Individual points on a Shepard diagram from an MDS run may now be clicked on, to display the two sample labels this point represents. This can aid identification of outliers – samples which fit poorly into the low-d space. The dimension and stress value is now shown on the Shepard diagram.
d) **Cancel Zoom** is a new Graph menu item (and icon on the Tool Bar) to allow a quick return to the unmagnified plot; similar is a **Reset** option on the Graph menu which (in addition to cancelling a zoom) will, for example, restore an MDS plot to its original orientation before a manual rotation.
e) A Monochrome option on the General tab turns all colour displays to mono, and replaces colour fills with monochrome hatching, patterns for which can be chosen in the Key which defines colour.
f) Key dialogs, of whatever type, can be instantly accessed by clicking on the key in the plot. This extends to some other plot 'hot' areas – clicking on titles or axes brings up the relevant dialog.

General & miscellany

43. PRIMER 7 is now a downloadable product, with the same download serving for trial, licenced and update purposes (with/without the PERMANOVA+ add-on). Full functionality requires the relevant key to be purchased, and needs authentication automatically *via* an internet connection (of course the software will then operate off-line). There is an off-line authentication process, but this should be avoided if at all possible – it needs manual code exchanges with the PRIMER-e office.
44. Maintenance updates to v7 will be offered automatically on release, the next time PRIMER is opened and there is an internet connection – these are quickly implemented. This facility can be switched off by **Options** (under **Tools**), and a manual check made for updates via the **Help** menu.
45. PRIMER 7 now has an **Undo** facility on the **Edit** menu so that any changes to data entries can be (multiply) reversed – this does not apply to operations on other menus which create new sheets (e.g. **Pre-treatment**) since these can always be deleted in the Explorer tree and re-run. An **Undo workspace** on the **File** menu reverses deletions, name changes etc. on the Explorer tree.
46. At PRIMER-e, we are not keen on upgrades which simply 'shuffle the furniture' but a small number of routines have been moved to different menu positions, for important reasons.
a) **Pre-treatment** becomes a main menu in its own right, since it is an almost inevitable first step. **Transform (individual)** has logically been moved into this menu from the **Tools** menu.
b) The active sheet for **BEST** is now the (usually biotic) resemblance matrix not, as previously, the explanatory (usually abiotic) variables sheet. This is because the resemblance matrix defines the specific 'response' samples to be analysed and the explanatory variables could be from a larger look-up table covering environmental data for a larger region or time. This change also makes the logical link with DISTLM in the PERMANOVA+ software – a more precise analogue of multiple linear regression which, as with all the add-on routines, has the resemblance matrix as active sheet.
c) The same reason of defining samples to be analysed as those in the (biotic) resemblance matrix makes this the active sheet for **LINKTREE**, which has logically been moved under **CLUSTER** – it is a constrained version of the unconstrained **UNCTREE**,

which must start from a resemblance.

47. A more flexible variable information worksheet replaces the old aggregation file (which can store trait and score information in addition to taxonomy), e.g. species can now be selected, ordered etc. using indicators on variable information. This is significant for taxonomic distinctness work, as is the alternative entry of species similarity matrices to **TAXDTEST** and **DIVERSE** (Taxdisc tab).
48. By default, the **BEST** results window now lists variable names (numbers are still optional) and also gives a summary table of the best solutions for each number of explanatory variables. Results are output continuously rather than at the end – as in all routines now.
49. Some routines are faster: the compute-intensive SIMPROF divides computation over multi-core processors; taxonomic distinctness will sample from large trees; and PCA is now SVD based.
50. Other minor changes are aimed at improving usability, convenience or speed of analysis.
    a) The **Fill** and **Value** (or **Pattern**) menu, used when creating factor entries, now fills a partly blank and highlighted column in which the only needed entries to type in are the first one at each change.
    b) The facility to open a very large workspace with its branches ‘rolled up’ in the Explorer tree – rather than all worksheets displayed (as at the last Save operation) – will allow a much more rapid start to a session, with only the analysis lines required then being unrolled.
    c) Other changes include an improved **Print** dialog and the ability to restore the ‘factory defaults’ from the **Options** menu. Added Notes (right-click on the Explorer tree) now have fuller editing operations, e.g. for font type and colour, and can include images – another new worksheet type. **Duplicate** will now allow the copy to be placed directly below the original in the Explorer tree.
    d) And there are new routines not even mentioned yet! (e.g. see **Variability weighting**, Section 4, a new pre-treatment option, and **Expand**, Section 14, to fill out model matrices to match biotic ones).

Revision #46
Created 12 March 2024 22:34:15 by Arden
Updated 7 January 2025 22:36:39 by Abby Miller