PAST - PAlaeontological STatistics, ver. 1.34

Øyvind Hammer, D.A.T. Harper and P.D. Ryan

March 17, 2005

1 Introduction

Welcome to the PAST! This program is designed as a follow-up to PALSTAT, an extensive package written by P.D. Ryan, D.A.T. Harper and J.S. Whalley (Ryan *et al.* 1995). It includes many of the functions which are commonly used in palaeontology and palaeoecology.

These days, a number of large and very good statistical systems exist, including SPSS, SAS and extensions to Excel. Why yet another statistics program?

- PAST is free.
- PAST is tailor-made for palaeontology. This means that it includes functions which are not found in off-the-shelf programs (for example cladistics, ordination, morphometry and biostratigraphy), and that it does not include functions which are of little use to palaeontologists and that only make the user interface more confusing.
- PAST is easy to use, and therefore well suited for introductory courses in quantitative palaeontology.
- PAST comes with a number of example data sets, case studies and exercises, making it a complete educational package.

Further explanations of many of the techniques implemented together with case histories are located in Harper (1999).

If you have questions, bug reports, suggestions for improvements or other comments, we would be happy to hear from you. Contact us at `ohammer@nhm.uio.no`. The PAST home page is

`http://folk.uio.no/ohammer/past`

2 Installation

The basic installation of PAST is easy: Just download the file 'Past.exe' and put it anywhere on your hard disk. Double-clicking the file will start the program. The data files for the case studies can be downloaded separately, or together in the packed file 'casefiles.zip'. This file must be unpacked with a program such as WinZip.

We suggest you make a folder called 'past' anywhere on your hard disk, and put all the files in this folder.

Please note: Problems have been reported for some combinations of screen resolution and default font size in Windows - the layout becomes ugly and it may be necessary for the user to increase the sizes of windows in order to see all the text and buttons. If this happens, please set the font size to 'Small fonts' in the Screen control panel in Windows. We are working on solving this problem.

PAST also seems to have problems with some printers. Postscript printers work fine.

When you exit PAST, a file called 'pastsetup'will be automatically placed in your personal folder (for example 'My Documents' in Windows 95/98), containing the last used file directories.

3 Entering and manipulating data

PAST has a spreadsheet-like user interface. Data are entered as an array of cells, organized in rows (horizontally) and columns (vertically).

Entering data

To input data in a cell, click on the cell with the mouse and type in the data. This can only be done when the program is in the 'Edit mode'. To select edit mode, tick the box above the array. When edit mode is off, the array is locked and the data cannot be changed. The cells can also be navigated using the arrow keys.

Any text can be entered in the cells, but almost all functions will expect numbers. Both comma (,) and decimal point (.) are accepted as decimal separators.

Absence/presence data are coded as 0 or 1, respectively. Any other positive number will be interpreted as presence. Absence/presence-matrices can be shown with black squares for presences by ticking the 'Square mode' box above the array.

Missing data are coded with question marks ('?') or the value -1. Unless support for missing data is specifically stated in the documentation for a function, the function will not handle missing data correctly, so be careful.

The convention in PAST is that items occupy rows, and variables columns. Three brachiopod individuals might therefore occupy rows 1, 2 and 3, with their lengths and widths in columns A and B. Cluster analysis will always cluster items, that is rows. For Q-mode analysis of associations, samples (sites) should therefore be entered in rows, while taxa (species) are in columns. For switching between Q-mode and R-mode, rows and columns can easily be interchanged using the Transpose operation.

Selecting areas

Most operations in PAST are carried only out on the area of the array which you have selected (marked). If you try to run a function which expects data, and no area has been selected, you will get an error message.

- A row is selected by clicking on the row label (leftmost column).
- A column is selected by clicking on the column label (top row).
- Multiple rows are selected by selecting the first row label, then shift-clicking (clicking with the Shift key down) on the additional row labels. Note that you can not 'drag out' multiple rows - this will instead move the first row (see below).
- Multiple columns are similarly marked by shift-clicking the additional column labels.

- The whole array can be selected by clicking the upper left corner of the array (the empty grey cell) or by choosing 'Select all' in the Edit menu.
- Smaller areas within the array can be selected by 'dragging out' the area, but this only works when 'Edit mode' is off.

Renaming rows and columns

When PAST starts, rows are numbered from 1 to 99 and columns are labelled A to Z. For your own reference, and for proper labelling of graphs, you should give the rows and columns more descriptive but short names. Choose 'Rename columns' or 'Rename rows' in the Edit menu. You must select the whole array, or a smaller area as appropriate.

Another way is to select the 'Edit labels' option above the spreadsheet. The first row and column are now editable in the same way as the rest of the cells.

Increasing the size of the array

By default, PAST has 99 rows and 26 columns. If you should need more, you can add rows or columns by choosing 'Insert more rows' or 'Insert more columns' in the Edit menu. Rows/columns will be inserted after the marked area, or at the bottom/right if no area is selected. When loading large data files, rows and/or columns are added automatically as needed.

Moving a row or a column

A row or a column (including its label) can be moved simply by clicking on the label and dragging to the new position.

Cut, copy, paste

The cut, copy and paste functions are found in the Edit menu. Note that you can cut/copy data from the PAST spreadsheet and paste into other programs, for example Word and Excel. Likewise, data from other programs can be pasted into PAST.

Remember that local blocks of data (not all rows or columns) can only be marked when 'Edit mode' is off.

All modules giving graphic output have a 'Copy graphic' button. This will place the graphical image into the paste buffer for pasting into other programs, such as a drawing program for editing the image. Note that graphics are copied using the 'Enhanced Metafile Format' in Windows. This allows editing of individual image elements in other programs. When pasting into Coreldraw, you have to choose 'Paste special' in the Edit menu, and then choose 'Enhanced metafile'. Some programs may have idiosyncratic ways of interpreting EMF images - beware of strange things happening.

Remove

The remove function (Edit menu) allows you to remove selected row(s) or column(s) from the spreadsheet. The removed area is not copied to the paste buffer.

Grouping (colouring) rows

Selected rows (data points) can be tagged with one of 12 attractive colors using the 'Tag rows' option in the Edit menu. Each group is also associated with a symbol (dot, cross, square, diamond, plus, circle, triangle, line, bar, filled square, star, oval). This is useful for showing different groups of data in plots, and is also required by a number of analysis methods.

The 'Numbers to colors' option in the Edit menu allows the numbers 1-9 in one selected column to set corresponding colours (symbols) for the rows.

Transpose

The Transpose function, in the Edit menu, will interchange rows and columns. This is used for switching between R mode and Q mode in cluster analysis, principal components analysis and seriation.

Grouped columns to multivar

Converts from a format with multivariate items presented in consecutive groups of N columns to the PAST format with one item per row and all variates along the columns. For $N = 2$, two specimens and four variables $a - d$, the conversion is from

$a_1\ b_1\ a_2\ b_2$
$c_1\ d_1\ c_2\ d_2$

to

a_1	b_1	c_1	d_1
a_2	b_2	c_2	d_2

Grouped rows to multivar

Converts from a format with multivariate items presented in consecutive groups of N rows to the PAST format with one item per row and all variates along the columns. For $N = 2$, two specimens and four variables $a - d$), the conversion is from

a_1	b_1
c_1	d_1
a_2	b_2
c_2	d_2

to

a_1	b_1	c_1	d_1
a_2	b_2	c_2	d_2

Samples to events (UA to RASC)

Given a data matrix of occurrences of taxa in a number of samples in a number of sections, as used by the Unitary Associations module, this function will convert each section to a single row with orders of events (FADs, LADs or both) as expected by the Ranking-Scaling module. Tied events (in the same sample) will be given equal ranking.

Loading and saving data

The 'Open' function is found in the File menu. PAST uses an ASCII file format, for easy importing from other programs (e.g. Word) and easy editing in a text editor. The format is as follows:

.	columnlabel	columnlabel	columnlabel
rowlabel	data	data	data
rowlabel	data	data	data
rowlabel	data	data	data

Empty cells (like the top left cell) are coded with a full stop (.). Cells are separated by white space, which means that you must never use spaces in row or column labels. 'Oxford Clay' is thus an illegal column label which would confuse the program.

If any rows have been assigned a colour other than black, the row labels in the file will start with an underscore, a number from 0 to 8 identifying the colour (symbol), and another underscore.

In addition to this format, PAST can also detect and open files in the following formats:

- Nexus format (see below), popular in systematics.
- TPS format developed by Rohlf (only the landmark, id and scale fields are supported, other fields are ignored).
- BioGraph format for biostratigraphy (SAMPLES or DATUM format). If a second file with the same name but extension ".dct" is found, it will be included as a BioGraph dictionary.
- RASC format for biostratigraphy. You must open the .DAT file, and the program expects corresponding .DIC and .DEP files in the same directory. The decimal depths format is not supported.

The 'Insert from file' function is useful for concatenating data sets. The loaded file will be inserted into your existing spreadsheet at the selected position (upper left). Other data sets can thus be inserted both to the right of and below your existing data.

Data from Excel

Data from Excel can be imported in two ways:

- Copy from Excel and paste into PAST. Note that if you want the first row and column to be copied into the label cells in PAST, you need to switch on the "Edit labels" option.
- Make sure that the top left cell in Excel contains a single dot (.) and save as tab-separated text in Excel. The resulting text file can be opened directly in PAST.

Reading and writing Nexus files

The Nexus file format is used by many cladistics programs. PAST can read and write the Data (character matrix) block of the Nexus format. Interleaved data are not supported. Also, if you have performed a parsimony analysis and the 'Parsimony analysis' window is open, all shortest trees will be written to the Nexus file for further processing in other programs (e.g. MacClade or Paup).

4 Transforming your data

These routines subject your data to mathematical operations. This can be useful for bringing out features in your data, or as a necessary preprocessing step for some types of analysis.

Logarithm

The Log function in the Transform menu log-transforms your data using the natural logarithm (base e):

$$y = \ln(x + 1)$$

This is useful, for example, to compare your sample to a log-normal distribution or for fitting to an exponential model. Also, abundance data with a few very dominant taxa may be log-transformed in order to downweight those taxa.

Subtract mean

This function subtracts the column mean from each of the selected columns. The means cannot be computed row-wise.

Remove trend

This function removes any linear trend from a data set (two columns with X-Y pairs). This is done by subtraction of a linear regression line from the Y values. Removing the trend can sometimes be a useful operation prior to spectral analysis.

Procrustes coordinates, Normalize size, Burnaby size removal

For description of these functions, see 'Geometrical analysis'.

Sort ascending and descending

Sorts the marked area, each column independently. Note that this procedure will reorder the contents of rows, so that row labels will no longer refer to the 'correct' rows.

The 'Sort descending' function is useful, for example, to plot taxon abundances against their ranks (this can also be done with the Abundance Model module).

Column difference

Simply subtracts two selected columns, and places the result in the next column.

Evaluate expression

This powerful feature allows flexible mathematical operations on the selected array of data. Each selected cell is evaluated, and the result replaces the previous contents. A mathematical expression must be entered, which can include any of the operators +, -, *, /, ^(power), and mod (modulo). Also supported are brackets (), and the functions abs, atan, cos, sin, exp, ln, sqrt, sqr, round and trunc.

The following variables can also be used:

- x (the contents of the current cell)
- l (the cell to the left if it exists, otherwise 0)
- r (the cell to the right)
- u (the cell above, or up)
- d (the cell below, or down)
- mean (the mean value of the current column)
- min (the minimum value)
- max (the maximum value)
- n (the number of cells in the column)
- i (the row index)
- j (the column index)
- random (uniform random number from 0 to 1)
- normal (Gaussian random number with mean 0 and variance 1)
- integral (sum of current column)
- stdev (standard deviation of current column)

Examples:

sqrt(x)	Replaces all numbers with their square roots
(x-mean)/stdev	Mean and standard deviation normalization, column-wise
x-0.5*(max+min)	Centers the values around zero
(u+x+d)/3	Three-point moving average smoothing
x-u	First-order difference
i	Fills the column with the row numbers (requires non-empty cells, such as all zeros
sin(2*3.14159*i/n)	Generates one period of a sine function down a column (requires non-empty cells)
5*normal+10	Normally distributed random number, mean 10 and standard deviation 5

5 Plotting functions

Graph

Plots one or more columns as separate graphs. The x coordinates are set automatically to 1,2,3,... There are three plot styles available: Graph (lines), bars and points. The 'X labels' options sets the x axis labels to the appropriate row names.

XY graph

Plots one or more pairs of columns containing x/y coordinate pairs. The 'log Y' option log-transforms your Y values (if necessary, a constant is added to make the minimum log value equal to 0). The curve can also be smoothed using 3-point moving average.

95 percent confidence ellipses can be plotted in most scatter plots in PAST, such as scores for PCA, CA, DCA, PCO, NMDS, and relative and partial warps. The calculation of these ellipses assumes a bivariate normal distribution.

Convex hulls can also be drawn in the scatter plots, in order to show the areas occupied by points of different 'colours'. The convex hull is the smallest convex polygon containing all points.

Histogram

Plots histograms (frequency distributions) for one or more columns. The number of bins is 10 by default, but can be changed by the user. The "Fit normal" option draws a graph with a fitted normal distribution (Parametric estimation, not Least Squares).

Box plot

Box plot for one or several columns (samples) of univariate data. For each sample, the 25-75 percent quartiles are drawn using a box. The median is shown with a horizontal line inside the box. The minimal and maximal values are shown with short horizontal lines ('whiskers').

Ternary

Ternary plot for three columns of data, normally containing proportions of compositions.

Survivorship

Survivorship curves for one or more columns of data. The data will normally consist of age or size values. A survivorship plot shows the number of individuals

which survived to different ages. Assuming exponential growth (highly questionable!), size should be log-transformed to age. This can be done either in the Transform menu, or directly in the Survivorship dialogue.

Landmark plot

This function is very similar to the 'XY graph', the only difference being that all XY pairs on each row are plotted with the appropriate row colour and symbol. It is well suited for plotting landmark data.

Landmarks 3D

Plotting of points in 3D (XYZ triples). Especially suited for 3D landmark data, but can also be used e.g. for PCA scatter plots along three principal components. The point cloud can be rotated around the x and the y axes (note: left-handed coordinate system). The 'Perspective' slider is normally not used. The 'Stems' option draws a line from each point down or up to a plane centered along the y axis, which can sometimes enhance 3D information. 'Lines' draws lines between consecutive landmarks within each separate specimen (row). 'Axes' shows the three coordinate axes with the centroid of the points as the origin.

Normal probability plot

Plots a normal probability (normal QQ) plot for one column of data. A normal distribution will plot on a straight line. For comparison, an RMA regression line is given, together with the Probability Plot Correlation Coefficient.

Matrix

Two-dimensional plot of the data matrix, using a grayscale with white for lowest value, black for highest. Can be useful to get an overview over a large data matrix.

6 Basic statistics

Univariate statistics

Typical application	Assumptions	Data needed
Quick statistical description of a univariate sample	None, but variance and standard deviation are most meaningful for normally distributed data	Single column of measured or counted data

Displays the following statistics: Number of entries (N), smallest value (Min), largest value (Max), mean value (Mean), standard error of the estimate of the mean (Std. error), population variance (that is, the variance of the population estimated from the sample), population standard deviation (square root of variance), median, skewness (positive for a tail to the right) and kurtosis (positive for a peaked distribution).

Missing data (?) are supported.

Comparing data sets

There are many different standard tests available for comparing two distributions. Here is the standard disclaimer: You can never prove that two distributions are the same. A high probability value is only consistent with a similar distribution, but does of course give an indication of the similarity between the two sample distributions. On the other hand, a very low probability value does show, to the given level of significance, that the distributions are different.

Chi-square (one sample vs. normal)

Typical application	Assumptions	Data needed
Testing for normal distribution of a sample	Large sample (N>30)	Single column of measured or counted data

Tests whether a single distribution (one selected column) is normal, by binning the numbers in four compartments. This test is generally inferior to the Shapiro-Wilk test, and should only be used for relatively large populations (N>30). See Brown & Rothery (1993) or Davis (1986) for details.

Missing data (?) are supported.

Shapiro-Wilk (one sample vs. normal)

Typical application	Assumptions	Data needed
Testing for normal distribution of a sample	Minimum 3, maximum 5000 data points	Single column of measured or counted data

Tests whether a single distribution (one selected column) is normal. This test is designed for populations with $3 \leq N \leq 5000$.

Missing data (?) are supported.

F and t tests (two samples)

Typical application	Assumptions	Data needed
Testing for equality of the variances and means of two samples	Normal or almost normal distribution (apart from the permutation test)	Two columns of measured or counted data

Two columns must be selected. The F test compares the variances of two distributions, while the t test compares their means. The F and t statistics, and the probabilities that the variances and means of the parent populations are the same, are given. The F and t tests should only be used if you have reason to believe that the parent populations are close to normally distributed. The Shapiro-Wilk test for one distribution against a normal distribution can give you an idea about this.

Also, the t test is really only applicable when the variances are the same. So if the F test says otherwise, you should be cautious about the t test. An unequal variance t statistic (Welch test) is also given, which should be used in this case.

The permutation t test compares the observed t statistic (normalized difference between means) with the t statistics from 1000 random pairs of replicates from the pooled data set. This test will be more accurate than the normal t test for non-normal distributions and small samples.

Sometimes publications give not the data, but values for sample size, mean and variance for two populations. These can be entered manually using the 'F and t from parameters' option in the menu.

See Brown & Rothery (1993) or Davis (1986) for details.

Missing data (?) are supported.

How do I test lognormal distributions?

All of the above tests apply to lognormal distributions as well. All you need to do is to transform your data first, by taking the log transform in the Transform menu. You might want to 'backup' your data column first, using Copy, and then get your original column back using Paste.

t test (one sample)

Typical application	Assumptions	Data needed
Testing whether the mean of a sample is equal to a given value	Normal or almost normal distribution	One column of measured data

The one-sample t test is used to investigate whether the sample is likely to have been taken from a population with a given (theoretical) mean.

Paired t test. Say that a measurement such as length of claw has been taken on the left and right side of a number of crab specimens, and we want to test for directed asymmetry (difference between left and right). A two-sample t test is not appropriate, because the values are not independent. Instead, we can perform a one-sample t test of left minus right against the value zero.

Missing data (?) are supported.

Chi-square (two samples)

Typical application	Assumptions	Data needed
Testing for equal distribution of compartmentalized, counted data	Each compartment containing at least five individuals	Two columns of counted data in different compartments (rows)

The Chi-square test is the one to use if your data consist of the numbers of elements in different bins (compartments). For example, this test can be used to compare two associations (columns) with the number of individuals in each taxon organized in the rows. You should be a little cautious about such comparisons if any of the bins contain less than five individuals.

There are two options that you should select or not for correct results. 'Sample vs. expected' should be ticked if your second column consists of values from a theoretical distribution (expected values) with zero error bars. If your data are from two counted samples each with error bars, leave this box open. This is *not* a small-sample correction.

'One constraint' should be ticked if your expected values have been normalized in order to fit the total observed number of events, or if two counted samples necessarily have the same totals (for example because they are percentages). This will reduce the number of degrees of freedom by one. When "one constraint" is selected, a permutation test is available, with 1000 randomly permutated replicates (row and column sums kept constant).

See Brown & Rothery (1993) or Davis (1986) for details.

Missing data (?) are supported.

Mann-Whitney U (two samples)

Typical application	Assumptions	Data needed
Comparing the medians of two samples	Both samples have $N > 7$, and similar distribution shapes.	Two columns of measured or counted data

Two columns must be selected. The two-tailed (Wilcoxon) Mann-Whitney U test can be used to test whether the medians of two independent distributions are

different. This test is non-parametric, which means that the distributions can be of any shape. PAST uses an approximation based on a z test, which is only valid for $N > 7$. It includes a continuity correction.

See Brown & Rothery (1993) or Davis (1986) for details.

Missing data (?) are supported.

Kolmogorov-Smirnov (two samples)

Typical application	Assumptions	Data needed
Comparing the distributions of two samples	None	Two columns of measured data

Two columns must be selected. The K-S test can be used to test whether two independent distributions of continuous, unbinned numerical data are different. The K-S test is non-parametric, which means that the distributions can be of any shape. If you want to test just the locations of the distribution (medians), you should use instead the Mann-Whitney U test.

See Davis (1986) for details.

Missing data (?) are supported.

Spearman's rho and Kendall's tau (two samples)

Typical application	Assumptions	Data needed
Testing whether two variables are correlated	None	Two columns of measured or counted paired data (such as *x/y* pairs)

These non-parametric rank-order tests are used to test for correlation between two variables.

Missing data (?) are supported.

Correlation matrix

Typical application	Assumptions	Data needed
Quantifying correlation between two or more variables	Normal distribution	Two or more columns of measured or counted variables

A matrix is presented with the correlations between all pairs of columns. Correlation values (Pearson's r) are given in the lower triangle of the matrix, and the probabilities that the columns are uncorrelated are given in the upper.

Variance/covariance matrix

Typical application	Assumptions	Data needed
Quantifying covariance between two or more variables	None	Two or more columns of measured or counted variables

A symmetric matrix is presented with the variances and covariances between all pairs of columns.

Missing data are supported by pairwise deletion.

Contingency table analysis

Typical application	Assumptions	Data needed
Testing for dependence between two variables	None	Matrix of counted data in compartments

A contingency table is input to this routine. Rows represent the different states of one nominal variable, columns represent the states of another nominal variable, and cells contain the counts of occurrences of that specific state (row, column) of the two variables. A measure and probability of association of the two variables (based on Chi-square) is then given.

For example, rows may represent taxa and columns samples as usual (with specimen counts in the cells). The contingency table analysis then gives information on whether the two nominal variables "taxon" and "locality" are associated. If not, the data matrix is not very informative. For details, see Press *et al*. (1992).

One-way ANOVA

Typical application	Assumptions	Data needed
Testing for equality of the means of several univariate samples	Normal distribution and similar variances and sample sizes	Two or more columns of measured or counted data

One-way ANOVA (analysis of variance) is a statistical procedure for testing the null hypothesis that several univariate data sets (in columns) have the same mean. The data sets are required to be close to normally distributed.

See Brown & Rothery (1993) or Davis (1986) for details.

Levene's test for homogeneity of variance (homoskedasticity), that is, whether variances are equal as assumed by ANOVA, is also given.

If the ANOVA shows significant inequality of the means (small p), you can go on to study the given table of "post-hoc" pairwise comparisons, based on Tukey's HSD test. The Studentized Range Statistic Q is given in the lower left triangle of

the array, and the probabilities $p(equal)$ in the upper right. Sample sizes do not have to be equal for the version of Tukey's test used.

Kruskal-Wallis test

Typical application	Assumptions	Data needed
Testing for equality of the medians of several univariate samples	None	Two or more columns of measured or counted data

The Kruskal-Wallis test can be regarded as a non-parametric alternative to ANOVA (Zar 1996). The H statistic and the H statistic corrected for ties (Hc) are given, together with a p value for equality (assuming a chi-squared distribution of Hc).

In the present version, PAST does not include a non-parametric post hoc test.

Similarity/distance indices

Typical application	Assumptions	Data needed
Comparing two or more samples	Equal sampling conditions	Two or more columns of presence/absence (1/0) or abundance data with taxa down the rows

14 similarity and distance measures, as described under Cluster Analysis are available. Note that some of these are similarity indices, while others are distance indices (in cluster analysis, these are all converted to similarities). All pairs of rows are compared, and the results given in a matrix.

Missing data are supported as described under Cluster Analysis.

Mixture analysis

Typical application	Assumptions	Data needed
Fitting a univariate data set to a mixture of two or more Gaussian (normal) distributions	Sampling from a mixture of two or more normally distributed populations	One column of measured data

Mixture analysis is an advanced maximum-likelihood method for estimating the parameters (mean, standard deviation and proportion) of two or more univariate normal distributions, based on a pooled univariate sample. For example, the method can be used to study differences between sexes (two groups), or several species, or size classes, when no independent information about group membership is available.

PAST uses the EM algorithm, which can get stuck on a local optimum. The procedure is therefore automatically run 10 times, each time with new, random starting positions for the means. The starting values for standard deviation are set to s/G, where s is the pooled standard deviation and G is the number of groups. The starting values for proportions are set to $1/G$. The user is still recommended to run the program a few times to check for stability of the solution ("better" solutions have less negative log likelihood values).

7 Multivariate statistics

Principal components analysis

Typical application	Assumptions	Data needed
Reduction and interpretation of large multivariate data sets with some underlying linear structure	Debated	Two or more rows of measured data with three or more variables

Principal components analysis (PCA) is a procedure for finding hypothetical variables (components) which account for as much of the variance in your multidimensional data as possible (Davis 1986, Harper 1999). These new variables are linear combinations of the original variables. PCA has several applications, two of them are:

- Simple reduction of the data set to only two variables (the two most important components), for plotting and clustering purposes.
- More interestingly, you might try to hypothesize that the most important components are correlated with some other underlying variables. For morphometric data, this might be simply age, while for associations it might be a physical or chemical gradient (e.g. latitude or position across the shelf).

The PCA routine finds the eigenvalues and eigenvectors of the variance-covariance matrix or the correlation matrix. Choose var-covar if all your variables are measured in the same unit (e.g. centimetres). Choose correlation (normalized var-covar) if your variables are measured in different units; this implies normalizing all variables using division by their standard deviations. The eigenvalues, giving a measure of the variance accounted for by the corresponding eigenvectors (components) are given for all components. The percentages of variance accounted for by these components are also given. If most of the variance is accounted for by the first one or two components, you have scored a success, but if the variance is spread more or less evenly among the components, the PCA has in a sense not been very successful.

The Jolliffe cut-off value gives an informal indication of how many principal components should be considered significant (Jolliffe, 1986). Components with eigenvalues smaller than the Jolliffe cut-off may be considered insignificant, but too much weight should not be put on this criterion.

The 'Scree plot' (simple plot of eigenvalues) can also be used to informally indicate the number of significant components. After this curve starts to flatten out, the corresponding components may be regarded as insignificant.

The 'View scatter' option allows you to see all your data points (rows) plotted in the coordinate system given by the two most important components. If you have tagged (grouped) rows, the different groups will be shown using different symbols

and colours. You can also plot the Minimal Spanning Tree, which is the shortest possible set of connected lines connecting all points. This may be used as a visual aid in grouping close points. The MST is based on an Euclidean distance measure of the original data points, so it is most meaningful when all your variables use the same unit. The 'Biplot' option will show a projection of the original axes (variables) onto the scattergram. This is another visualisation of the component loadings (coefficients) - see below. Note that the lengths of these axes are arbitrarily scaled, all by the same factor, for giving a clear diagram.

The 'View loadings' option shows to what degree your different original variables (given in the original order along the x axis) enter into the different components (as chosen in the radio button panel). These component loadings are important when you try to interpret the 'meaning' of the components. The 'Coefficients' option gives the PC coefficients, while 'Correlation' gives the correlation between a variable and the PC scores. Do not use the latter if you are doing PCA on the correlation matrix.

The 'SVD' option will enforce use of the supposedly superior Singular Value Decomposition algorithm instead of "classical" eigenanalysis. The two algorithms will normally give almost identical results, except that SVD will center on zero. Also, the eigenvalues will have different absolute values (their relative values remain the same), and axes may be flipped.

For the 'Shape PCA' and 'Shape deform' options, see the section on Geometrical Analysis.

Bruton & Owen (1988) describe a typical morphometrical application of PCA.

Missing data are supported by column average substitution.

Principal coordinates

Typical application	Assumptions	Data needed
Reduction and interpretation of large multivariate data sets with some underlying linear structure	Unknown	Two or more rows of measured, counted or presence/absence data with three or more variables

Principal coordinates analysis (PCO) is another ordination method, somewhat similar to PCA. The algorithm is taken from Davis (1986).

The PCO routine finds the eigenvalues and eigenvectors of a matrix containing the distances between all data points. The Gower measure will normally be used instead of Euclidean distance, which gives results similar to PCA. An additional eleven distance measures are available - these are explained under Cluster Analysis. The eigenvalues, giving a measure of the variance accounted for by the corresponding eigenvectors (coordinates) are given for the first four most important coordinates (or fewer if there are fewer than four data points). The percentages of variance accounted for by these components are also given.

The 'View scatter' option allows you to see all your data points (rows) plotted in the coordinate system given by the PCO. If you have tagged (grouped) rows, the different groups will be shown using different symbols and colours.

Missing data are supported by pairwise deletion (not for the Raup-Crick and rho indices).

Non-metric multidimensional scaling

Typical application	Assumptions	Data needed
Reduction and interpretation of large multivariate ecological data sets	None	Two or more rows of measured, counted or presence/absence data with two or more variables.

Non-metric multidimensional scaling is based on a distance matrix computed with any of 14 supported distance measures, as explained under Cluster Analysis below. The algorithm then attempts to place the data points in a two-dimensional coordinate system such that the *ranked differences* are preserved. For example, if the original distance between points 4 and 7 is the ninth largest of all distances between any two points, points 4 and 7 will ideally be placed such that their euclidean distance in the plane is still the ninth largest. Non-metric multidimensional scaling intentionally does not take absolute distances into account.

The program will converge on a different solution in each run, depending upon the random initial conditions.

The algorithm implemented in PAST, which seems to work very well, is based on a new approach developed by Taguchi & Oono (in press).

Shepard plot: This plot of obtained versus observed (target) ranks indicates the quality of the result. Ideally, all points should be placed on a straight ascending line ($x = y$).

Missing data are supported by pairwise deletion (not for the Raup-Crick and rho indices).

Correspondence analysis

Typical application	Assumptions	Data needed
Reduction and interpretation of large multivariate ecological data sets with environmental or other gradients	Unknown	Two or more rows of counted data in three or more compartments

Correspondence analysis (CA) is yet another ordination method, somewhat similar to PCA but for counted data. For comparing associations (columns) containing counts of taxa, or counted taxa (rows) across associations, CA is the more appropriate algorithm. The algorithm is taken from Davis (1986).

The CA routine finds the eigenvalues and eigenvectors for a matrix containing the Chi-squared distances between all data points. The eigenvalues, giving a measure of the similarity accounted for by the corresponding eigenvectors, are given for the first four most important eigenvectors (or fewer if there are less than four variables). The percentages of similarity accounted for by these components are also given. Note that the very first, so-called 'trivial' eigenvector is not included in the output.

The 'View scatter' option allows you to see all your data points (rows) plotted in the coordinate system given by the CA. If you have tagged (grouped) rows, the different groups will be shown using different symbols and colours.

In addition, the variables (columns, associations) can be plotted in the same coordinate system (Q mode), optionally including the column labels. If your data are 'well behaved', taxa typical for an association should plot in the vicinity of that association.

If you have more than two columns in your data set, you can choose to view a scatter plot on the second and third axes.

Relay plot: This is a composite diagram with one plot per column. The plots are ordered according to CA column scores. Each data point is plotted with CA first-axis row scores on the vertical axis, and the original data point value (abundance) in the given column on the horizontal axis. This may be most useful when samples are in rows and taxa in columns. The relay plot will then show the taxa ordered according to their positions along the gradients, and for each taxon the corresponding plot should ideally show a unimodal peak, partly overlapping with the peak of the next taxon along the gradient (see Hennebert & Lees 1991 for an example from sedimentology).

Missing data are supported by column average substitution.

Detrended correspondence analysis

Typical application	Assumptions	Data needed
Reduction and interpretation of large multivariate ecological data sets with environmental or other gradients	Unknown	Two or more rows of counted data in three or more compartments

The Detrended Correspondence (DCA) module uses the same algorithm as Decorana (Hill & Gauch 1980), with modifications according to Oxanen & Minchin (1997). It is specialized for use on 'ecological' data sets with abundance data (taxa in rows, localities in columns). When the 'Detrending' option is switched off, a basic Reciprocal Averaging will be carried out. The result should be similar to Correspondence Analysis (see above) plotted on the first and second axes.

Eigenvalues for the first three ordination axes are given as in CA, indicating their relative importance in explaining the spread in the data.

Detrending is a sort of normalization procedure in two steps. The first step involves an attempt to 'straighten out' points lying in an arch, which is a common occurrence. The second step involves 'spreading out' the points to avoid clustering of the points at the edges of the plot. Detrending may seem an arbitrary procedure, but can be a useful aid in interpretation.

Missing data are supported by column average substitution.

Cluster analysis

Typical application	Assumptions	Data needed
Finding hierarchical groupings in multivariate data sets	None	Two or more rows of counted, measured or presence/absence data in one or more variables or categories

The hierarchical clustering routine produces a 'dendrogram' showing how data points (rows) can be clustered. For 'R' mode clustering, putting weight on groupings of taxa, taxa should be in rows. It is also possible to find groupings of variables or associations (Q mode), by entering taxa in columns. Switching between the two is done by transposing the matrix (in the Edit menu).

Three different algorithms are available:

- Unweighted pair-group average (UPGMA). Clusters are joined based on the average distance between all members in the two groups.
- Single linkage (nearest neighbour). Clusters are joined based on the smallest distance between the two groups.
- Ward's method. Clusters are joined such that increase in within-group variance is minimized.

One method is not necessarily better than the other, though single linkage is not recommended by some. It can be useful to compare the dendrograms given by the different algorithms in order to informally assess the robustness of the groupings. If a grouping is changed when trying another algorithm, that grouping should perhaps not be trusted.

For Ward's method, a Euclidean distance measure is inherent to the algorithm. For UPGMA and single linkage, the distance matrix can be computed using 13 different measures:

- The Euclidean distance (between rows) is a robust and widely applicable measure. Distance is converted to similarity by changing the sign.

$$Euclidean_{jk} = \sqrt{\sum_{i=1}^{s}(x_{ij} - x_{ik})^2}$$

- Correlation (of the variables along rows) using Pearson's r. A little meaningless if you have only two variables.
- Correlation using Spearman's rho (basically the r value of the ranks). Will often give the same result as correlation using r.
- Dice (Sorensen) coefficient for absence-presence (coded as 0 or positive numbers). Puts more weight on joint occurences than on mismatches.

  When comparing two columns (associations), a match is counted for all taxa with presences in both columns. Using 'M' for the number of matches and 'N' for the the total number of taxa with presences in just one column, we have

  Dice similarity = 2M / (2M+N)
- Jaccard similarity = M / (M+N)
- The Simpson index is defined as M/N_{min}, where N_{min} is the smaller of the numbers of presences in the two associations. This index treats two associations as identical if one is a subset of the other, making it useful for fragmentary data.
- Bray-Curtis measure for abundance data.

$$Bray - Curtis_{jk} = \frac{\sum_{i=1}^{s} |x_{ij} - x_{ik}|}{\sum_{i=1}^{s} (x_{ij} + x_{ik})}$$

- Cosine distance for abundance data - one minus the inner product of abundances each normalised to unit norm.
- Chord distance for abundance data (converted to similarity by changing the sign). Recommended!

$$Chord_{jk} = \sqrt{2 - 2\frac{\sum_{i=1}^{s}(x_{ij}x_{ik})}{\sqrt{\sum_{i=1}^{s} x_{ij}^2 \sum_{i=1}^{s} x_{ik}^2}}}$$

- Morisita's index for abundance data. Recommended!

$$\begin{aligned} \lambda_1 &= \frac{\sum_{i=1}^{s}(x_{ij}(x_{ij}-1))}{\sum_{i=1}^{s} x_{ij} \left(\sum_{i=1}^{s} x_{ij} - 1\right)} \\ \lambda_2 &= \frac{\sum_{i=1}^{s}(x_{ik}(x_{ik}-1))}{\sum_{i=1}^{s} x_{ik} \left(\sum_{i=1}^{s} x_{ik} - 1\right)} \\ Morisita_{jk} &= \frac{2\sum_{i=1}^{s}(x_{ij}x_{ik})}{(\lambda_1 + \lambda_2)\sum_{i=1}^{s} x_{ij} \sum_{i=1}^{s} x_{ik}} \end{aligned} \quad (1)$$

- Raup-Crick index for absence-presence data. Recommended! This index (Raup & Crick 1979) uses a randomization ("Monte Carlo") procedure, comparing the observed number of species ocurring in both associations with the distribution of co-occurrences from 200 random replicates.

- Horn's overlap index for abundance data (Horn 1966).

$$
\begin{aligned}
N_j &= \sum_{i=1}^{s} x_{ij} \\
N_k &= \sum_{i=1}^{s} x_{ik} \\
Ro_{jk} &= \frac{\sum_{i=1}^{s}\left[(x_{ij}+x_{ik})\ln(x_{ij}+x_{ik})\right] - \sum_{i=1}^{s}\left[x_{ij}\ln x_{ij}\right] - \sum_{i=1}^{s}\left[x_{ik}\ln x_{ik}\right]}{(N_j+N_k)\ln(N_j+N_k) - N_j\ln N_j - N_k\ln N_k}
\end{aligned}
\tag{2}
$$

- Hamming distance for categorical data as coded with integers. The Hamming distance is the number of differences (mismatches), so that the distance between (3,5,1,2) and (3,7,0,2) equals 2. In PAST, this is normalised to the range (0,1).

- Manhattan distance: The sum of differences in each variable (converted to similarity by changing the sign).

See Harper (1999) or Davis (1986) for details.

Missing data: The cluster analysis algorithm can handle missing data, coded as -1 or question mark (?). This is done using pairwise deletion, meaning that when distance is calculated between two points, any variables that are missing are ignored in the calculation. Missing data are not supported for Ward's method, nor for the Rho or the Raup-Crick similarity measures.

Two-way clustering: The two-way option allows simultaneous clustering in R-mode and Q-mode. The graphics only support relatively small data sets.

Stratigraphically constrained clustering: This option will allow only adjacent rows (or groups of rows) to be joined during the agglomerative clustering procedure. May produce strange-looking (but correct) dendrograms.

K-means clustering

Typical application	Assumptions	Data needed
Non-hierarchical clustering of multivariate data into a specifi ed number of groups	None	Two or more rows of counted or measured data in one or more variables

K-means clustering (e.g. Bow 1984) is a non-hierarchical clustering method. The number of clusters to use is specified by the user, usually according to some hypothesis such as there being two sexes, four geographical regions or three species in the data set

The cluster assignments are initially random. In an iterative procedure, items are then moved to the cluster which has the closest cluster mean, and the cluster means are updated accordingly. This continues until items are no longer "jumping" to other clusters. The result of the clustering is to some extent dependent upon the initial, random ordering, and cluster assignments may therefore differ from run to run. This is not a bug, but normal behaviour in k-means clustering.

The cluster assignments may be copied and pasted back into the main spreadsheet, and corresponding colors (symbols) assigned to the items using the 'Numbers to colors' option in the Edit menu.

Missing data are supported by column average substitution.

Seriation

Typical application	Assumptions	Data needed
Stratigraphical or environmental ordering of taxa and localities	None	Presence/absence (1/0) matrix with taxa in rows

Seriation of an absence-presence matrix using the algorithm described by Brower and Kyle (1988). This method is typically applied to an association matrix with taxa (species) in the rows and populations in the columns. For constrained seriation (see below), columns should be ordered according to some criterion, normally stratigraphic level or position along a presumed faunal gradient.

The seriation routines attempt to reorganize the data matrix such that the presences are concentrated along the diagonal. There are two algorithms: Constrained and unconstrained optimization. In constrained optimization, only the rows (taxa) are free to move. Given an ordering of the columns, this procedure finds the 'optimal' biozonation, that is, the ordering of taxa which gives the prettiest range plot. Also, in the constrained mode, the program runs a 'Monte Carlo' simulation, generating and seriating 30 random matrices with the same number of occurences within each taxon, and compares these to the original matrix to see if it is more informative than a random one (this procedure is time-consuming for large data sets).

In the unconstrained mode, both rows and columns are free to move.

Discriminant analysis and Hotelling's T^2

Typical application	Assumptions	Data needed
Testing for separation and equal means of two multivariate data sets	Multivariate normality. Hotelling's test assumes equal covariances.	Two multivariate data sets of measured data, marked with different colors

Given two sets of multivariate data, an axis is constructed which maximizes the difference between the sets. The two sets are then plotted along this axis using a histogram.

This module expects the rows in the two data sets to be tagged with dots (black) and crosses (red), respectively.

Equality of the means of the two groups is tested by a multivariate analogue to the t test, called *Hotelling's T-squared*, and a p value for this test is given. Normal distribution of the variables is required, and also that the number of cases is at least two more than the number of variables.

Number of constraints: For correct calculation of the Hotelling's p value, the number of dependent variables(constraints) must be specified. It should normally be left at 0, but for Procrustes fitted landmark data use 4 (for 2D) or 6 (for 3D).

Discriminant analysis is a standard method for visually confirming or rejecting the hypothesis that two species are morphologically distinct. Using a cutoff point at zero (the midpoint between the means of the discriminant scores of the two groups), a classification into two groups is shown in the "view numbers" option. The percentage of correctly classified items is also given.

Discriminant function: New specimens can be classified according to the discriminant function. Take the inner product between the measurements on the new specimen and the given discriminant function factors, and then subtract the given offset value.

Beware: The combination of discriminant analysis and Hotelling's T^2 test is sometimes misused. One should not be surprised to find a statistically significant difference between two samples which have been chosen with the objective of maximizing distance in the first place! The division into two groups should ideally be based on independent evidence.

See Davis (1986) for details.

Missing data are supported by column average substitution.

Paired Hotelling's T^2

Typical application	Assumptions	Data needed
Testing for equal means of a paired multivariate data set	Multivariate normality.	A multivariate data set of paired measured data, marked with different colors

The paired Hotelling's test expects two groups of multivariate data, marked with different colours. Rows within each group must be consecutive. The first row of the first group is paired with the first row of the second group, the second row is paired with the second, etc.

Missing data are supported by column average substitution.

Permutation test for two multivariate groups

Typical application	Assumptions	Data needed
Testing for qual means of two multivariate data sets	The two groups have similar distributions (variances)	Two multivariate data sets of measured data, marked with different colors

This module expects the rows in the two data sets to be grouped into two sets by colouring the rows, e.g. with black (dots) and red (crosses). Rows within each group must be consecutive.

Equality of the means of the two groups is tested using permutation with 2000 replicates, and the Mahalanobis squared distance measure. The permutation test is an alternative to Hotelling's test when the assumptions of multivariate normal distributions and equal covariance matrices do not hold.

Missing data are supported by column average substitution.

Box's M test

Typical application	Assumptions	Data needed
Testing for equivalence of the covariance matrices for two data sets	Multivariate normality	Two multivariate data sets of measured data, or two (square) variance-covariance matrices, marked with different colors

This test is rather specialized, testing for the equivalence of the covariance matrices for two multivariate data sets. You can use either two original multivariate data sets from which the covariance matrices are automatically computed, or two specified variance-covariance matrices. In the latter case, you must also specify the sizes (number of individuals) of the two samples.

The Box's M statistic is given, together with a significance value based on a chi-square approximation. Note that this test is supposedly very sensitive. This means that a high p value will be a good, although informal, indicator of equality, while a highly significant result (low p value) may in practical terms be a somewhat too sensitive indicator of inequality.

One-way MANOVA and Canonical Variates Analysis

Typical application	Assumptions	Data needed
Testing for equality of the means of several multivariate samples, and ordination based on maximal separation (multigroup discriminant analysis)	Multivariate normal distribution, similar variances-covariances	Two or more samples of multivariate measured data, marked with different colors. The number of cases must exceed the number of variables.

One-way MANOVA (Multivariate ANalysis Of VAriance) is the multivariate version of the univariate ANOVA, testing whether several samples have the same mean. If you have only two samples, you would perhaps rather use the two-sample Hotelling's T^2 test.

Two statistics are provided: Wilk's lambda with it's associated Rao's F and the Pillai trace with it's approximated F. Wilk's lambda is probably more commonly used, but the Pillai trace may be more robust.

Number of constraints: For correct calculation of the p values, the number of dependent variables(constraints) must be specified. It should normally be left at 0, but for Procrustes fitted landmark data use 4 (for 2D) or 6 (for 3D).

Canonical Variates Analysis

An option under MANOVA, CVA produces a scatter plot of specimens along the two first canonical axes, producing maximal and second to maximal separation between all groups (multigroup discriminant analysis). The axes are linear combinations of the original variables as in PCA, and eigenvalues indicate amount of variation explained by these axes.

Missing data are supported by column average substitution.

One-way ANOSIM

Typical application	Assumptions	Data needed
Testing for difference between two or more multivariate groups, based on any distance measure	Ranked dissimilarities within groups have similar median and range.	Two or more groups of multivariate data, marked with different colours.

ANOSIM (ANalysis Of Similarities) is a non-parametric test of significant difference between two or more groups, based on any distance measure (Clarke 1993). The distances are converted to ranks. ANOSIM is normally used for ecological taxa-in-samples data, where groups of samples are to be compared.

In a rough analogy with ANOVA, the test is based on comparing distances between groups with distances within groups. Let r_b be the mean rank of all distances between groups, and r_w the mean rank of all distances within groups. The test statistic R is then defined as

$$R = \frac{r_b - r_w}{N(N-1)/4}.$$

Large positive R (up to 1) signifies dissimilarity between groups. The significance is computed by permutation of group membership, with 5000 replicates.

Missing data are supported by pairwise deletion (not for the Raup-Crick and Rho indices).

One-way NPMANOVA

Typical application	Assumptions	Data needed
Testing for difference between two or more multivariate groups, based on any distance measure	The groups have similar distributions (similar variances)	Two or more groups of multivariate data, marked with different colors.

NPMANOVA (Non-Parametric MANOVA) is a non-parametric test of significant difference between two or more groups, based on any distance measure (Anderson 2001). NPMANOVA is normally used for ecological taxa-in-samples data, where groups of samples are to be compared, but may also be used as a general non-parametric MANOVA

NPMANOVA calculates an F value in analogy with ANOVA. In fact, for univariate data sets and the Euclidean distance measure, NPMANOVA is equivalent to ANOVA and gives the same F value.

The significance is computed by permutation of group membership, with 5000 replicates.

8 Fitting data to functions

Linear

Typical application	Assumptions	Data needed
Fitting data to a straight line, or exponential or power function	None	One or two columns of counted or measured data

If two columns are selected, they represent x and y values, respectively. If one column is selected, it represents y values, and x values are taken to be the sequence of positive integers (1,2,...). A straight line $y = ax+b$ is fitted to the data. There are two different algorithms available: Standard regression and Reduced Major Axis (the latter is selected by ticking the box). Standard regression keeps the x values fixed, and finds the line which minimizes the squared errors in the y values. Use this if your x values have very small errors associated with them. Reduced Major Axis tries to minimize both the x and the y errors. RMA fitting and standard error estimation is according to Miller & Kahn (1962), *not* Davis (1986)!

Also, both x and y values can be log-transformed (base 10), in effect fitting your data to the 'allometric' function $y = 10^b x^a$. An a value around 1 indicates that a straight-line ('isometric') fit may be more applicable.

The values for a and b, their errors, a Chi-square correlation value (not for RMA), Pearson's r correlation, and the probability that the columns are not correlated are given.

The calculation of standard errors for slope and intercept assumes normal distribution of residuals and independence between the variables and the variance of residuals. If these assumptions are strongly broken, it is preferable to use the bootstrapped 95 percent confidence intervals (2000 replicates). The number of random points selected for each replicate should normally be kept as N, but may be reduced for special applications.

In Standard regression (not RMA), a 95 percent "Working-Hotelling" confidence band for the fitted line (not for the data points!) is available.

Residuals

The Residuals window reports the distances from each data point to the regression line, in the x and y directions. Only the latter is of interest when using ordinary linear regression rather than RMA. The residuals can be copied back to the spreadsheet and inspected for normal distribution and independence between independent variable and residual variance (homoskedasticity).

Exponential functions

Your data can be fitted to an exponential function $y = e^b e^{ax}$ by first log-transforming just your y column (in the Transform menu) and then performing a straight-line fit.

Sinusoidal

Typical application	Assumptions	Data needed
Fitting data to a set of periodic, sinusoidal functions	None	Two columns of counted or measured data

Two columns must be selected (x and y values). A sum of up to eight sinusoids with periods specified by the user, but with unknown amplitudes and phases, is fitted to the data. This can be useful for modelling periodicities in time series, such as annual growth cycles or climatic cycles, usually in combination with spectral analysis. The algorithm is based on a least-squares criterion and singular value decomposition (Press *et al.* 1992). By default, the periods are set to the range of the x values, and harmonics (1/2, 1/3, 1/4, 1/5, 1/6, 1/7 and 1/8 of the fundamental period). These values can be changed, and need not be in harmonic proportion.

With a little effort, frequencies can also be estimated by trial and error, by adjusting the frequency so that amplitude is maximized (this procedure is difficult with more than a single sinusoidal).

It is not meaningful to specify periodicities that are smaller than two times the typical spacing of data points.

Each sinusoid is given by $y = a\,\cos(2\pi x/T - \phi)$, where a is the amplitude, T is the period and ϕ is the phase.

Logistic

Typical application	Assumptions	Data needed
Fitting data to a logistic or von Bertalanffy growth model	None	Two columns of counted or measured data

Attempts to fit the data to the logistic equation $y = a/(1 + b * e^{-cx})$. For numerical reasons, the x axis is normalized. The algorithm is a little complicated. The value of a is first estimated to be the maximal value of y. The values of b and c are then estimated using a straight-line fit to a linearized model.

Though acceptable, this estimate can optionally be improved by using the estimated values as an initial guess for a Levenberg-Marquardt nonlinear optimization (tick the box). This procedure can sometimes improve the fit, but due to the numerical instability of the logistic model it often fails with an error message.

The logistic equation can model growth with saturation, and was used by Sepkoski (1984) to describe the proposed stabilization of marine diversity in the late Palaeozoic.

The 95 percent confidence intervals are based on 2000 bootstrap replicates, not using the Levenberg-Marquardt optimization step.

Von Bertalanffy

An option in the 'Logistic fit' window. Uses the same algorithm as above, but fits to the von Bertalanffy equation $y = a * (1 - b * e^{-cx})$. This equation is used for modelling growth of multi-celled animals (in units of length or width, not volume).

B-splines

Typical application	Assumptions	Data needed
Smoothing noisy data	None	Two columns of counted or measured data

Two columns must be selected (x and y values). The data are fitted with a least-squares criterion to a B-spline, which is a sequence of third-order polynomials, continuous up to the second derivative. A typical application of this is the construction of a smooth curve going through a noisy data set.

A decimation factor is set by the user, and controls how many data points contribute to each polynomial section. Larger decimation gives a smoother curve.

Note that sharp jumps in your data can give rise to oscillations in the curve, and that you can also get large excursions in regions with few data points.

Abundance models

Typical application	Assumptions	Data needed
Fitting taxon abundance distribution to one of three models	None	One column of abundance counts for a number of taxa in a sample

This module can be used for plotting logarithms of taxon abundances in descending rank order (Whittaker plot), or number of species in abundance octave classes (as shown when fitting to log-normal distribution). It can also fit the data to one of three different standard abundance models:

- Geometric, where the 2nd most abundant species should have a taxon count of k<1 times the most abundant, the 3rd most abundant a taxon count of k times the 2nd most abundant etc. for a constant k. This will give a straight descending line in the Whittaker plot. Fitting is by simple linear regression of the log abundances.

- Log-series, with two parameters α and x. The fitting algorithm is from Krebs (1989).

- Log-normal. The fitting algorithm is from Krebs (1989). The logarithm (base 10) of the fitted mean and variance are given. The octaves refer to power-of-2 abundance classes:

Octave	Abundance
1	1
2	2-3
3	4-7
4	8-15
5	16-31
6	32-63
7	64-127
...	...

A significance value based on chi-squared is given for each of these models, but the power of the test is not the same for the tree models and the significance values should therefore not be compared. It is important, as always, to remember that a high p value can not be taken to imply a good fit. A low value does however imply a bad fit.

9 Diversity

Diversity statistics

Typical application	Assumptions	Data needed
Quantifying taxonomical diversity in samples	Representative samples	One or more columns, each containing counts of individuals of different taxa down the rows

These statistics apply to association data, where numbers of individuals are tabulated in rows (taxa) and possibly several columns (associations). The available statistics are as follows, for each association:

- Number of taxa (S)
- Total number of individuals (n)
- Dominance=1-Simpson index. Ranges from 0 (all taxa are equally present) to 1 (one taxon dominates the community completely). $D = \sum \left(\frac{n_i}{n}\right)^2$ where n_i is number of individuals of taxon i.
- Simpson index=1-dominance. Measures 'evenness' of the community from 0 to 1. Note the confusion in the literature: Dominance and Simpson indices are often interchanged!
- Shannon index (entropy). A diversity index, taking into account the number of individuals as well as number of taxa. Varies from 0 for communities with only a single taxon to high values for communities with many taxa, each with few individuals. $H = -\sum \frac{n_i}{n} \ln\left(\frac{n_i}{n}\right)$
- Buzas and Gibson's evenness: e^H/S
- Menhinick's richness index - the ratio of the number of taxa to the square root of sample size.
- Margalef's richness index: $(S-1)/\ln(n)$, where S is the number of taxa, and n is the number of individuals.
- Equitability. Shannon diversity divided by the logarithm of number of taxa. This measures the evenness with which individuals are divided among the taxa present.
- Fisher's alpha - a diversity index, defined implicitly by the formula $S = \alpha \ln(1 + n/\alpha)$ where S is number of taxa, n is number of individuals and α is the Fisher's alpha.
- Berger-Parker dominance: simply the number of individuals in the dominant taxon divided by n.

Most of these indices are explained in Harper (1999).

Approximate confidence intervals for all the indices can be computed with a bootstrap procedure. 1000 random samples are produced (200 prior to version 0.87b), each with the same total number of individuals as in the original sample. The random samples are taken from the total, pooled data set (all columns). For each individual in the random sample, the taxon is chosen with probabilities according to the original abundances. A 95 percent confidence interval is then calculated. Note that the diversity in the replicates will often be less than, and never larger than, the pooled diversity in the total data set.

Since these confidence intervals are all computed with respect to the pooled data set, they do not represent confidence intervals for the individual samples. They are mainly useful for identifying samples where the given diversity index falls outside the confidence interval. Bootstrapped comparison of diversity indices in two samples is provided in the "Compare diversities" module.

Quadrat richness

Typical application	Assumptions	Data needed
Estimating species richness from several quadrat samples	Representative, random quadrats of equal size	Two or more columns, each containing presence/absence (1/0) of different taxa down the rows

Four non-parametric species richness estimators are included in PAST: Chao 2, first- and second-order jackknife, and bootstrap. All of these require presence-absence data in two or more sampled quadrats of equal size. Colwell & Coddington (1994) reviewed these estimators, and found that the Chao2 and the second-order jackknife performed best.

Taxonomic distinctness

Typical application	Assumptions	Data needed
Quantifying taxonomical distinctness in samples	Representative samples	One or more columns, each containing counts of individuals of different taxa down the rows. In addition, the leftmost row(s) must contain names of genera/families etc. (see below).

Taxonomic diversity and taxonomic distinctness as defined by Clarke & Warwick (1998), including confidence intervals computed from 200 random replicates taken from the pooled data set (all columns). Note that the "global list" of Clarke &

Warwick is not entered directly, but is calculated internally by pooling (summing) the given samples.

These indices depend on taxonomic information also above the species level, which has to be entered for each species as follows. Species names go in the name column (leftmost, fixed column), genus names in column 1, family in column 2 etc. Species counts follow in the columns thereafter. The program will ask for the number of columns containing taxonomic information above the species level.

For presence-absence data, taxonomic diversity and distinctness will be valid but equal to each other.

Compare diversities

Typical application	Assumptions	Data needed
Comparing diversities in two samples of abundance data	Equal sampling conditions	Two columns of abundance data with taxa down the rows

This module computes a number of diversity indices for two samples, and then compares the diversities using two different randomization procedures as follows.

Bootstrapping

The two samples A and B are pooled. 1000 random pairs of samples (A_i, B_i) are then taken from this pool (200 prior to version 0.87b), with the same numbers of individuals as in the original two samples. For each replicate pair, the diversity indices $div(A_i)$ and $div(B_i)$ are computed. The number of times $|div(A_i) - div(B_i)|$ exceeds or equals $|div(A) - div(B)|$ indicates the probability that the observed difference could have occurred by random sampling from one parent population as estimated by the pooled sample.

A small probability value $p(equal)$ then indicates a significant difference in diversity index between the two samples.

Permutation

1000 random matrices with two columns (samples) are generated, each with the same row and column totals as in the original data matrix. The p value is computed as for the boostrap test.

Diversity t test

Typical application	Assumptions	Data needed
Comparing Shannon diversities in two samples of abundance data	Equal sampling conditions	Two columns of abundance data with taxa down the rows

Comparison of the Shannon diversities (entropies) in two samples, using a t test described by Poole (1974). This is an alternative to the randomization test available in the Compare diversities module.

Note that the Shannon indices here include a bias correction term (Poole 1974), and may diverge slightly from the uncorected estimates calculated elsewhere in PAST, at least for small samples.

Diversity profiles

Typical application	Assumptions	Data needed
Comparing diversities in two samples of abundance data	Equal sampling conditions	Two columns of abundance data with taxa down the rows

The validity of comparing diversities in two samples can be criticized because of arbitrary choice of diversity index. One sample may for example contain a larger number of taxa, while the other has a larger Shannon index. It may therefore be a good idea to try a number of diversity indices in order to make sure that the diversity ordering is robust. A formal way of doing this is to define a family of diversity indices, dependent upon a single continuous parameter (Tothmeresz 1995).

PAST uses the exponential of the so-called Renyi index, which depends upon a parameter alpha. For alpha=0, this function gives the total species number; alpha=1 gives an index proportional to the Shannon index, while alpha=2 gives an index which behaves like the Simpson index.

$$H = \frac{\ln \sum_{i=1}^{s} p_i^{\alpha}}{1-\alpha},$$

where p_i are proportional abundances of individual taxa and s is the number of species.

The program plots two such diversity profiles together. If the profiles cross, the diversities are non-comparable.

Rarefaction

Typical application	Assumptions	Data needed
Comparing taxonomical diversity in samples of different sizes	When comparing samples: Samples should be taxonomically similar, obtained using standardised sampling and taken from a similar 'habitat'.	Single column of counts of individuals of different taxa

Given a column of abundance data for a number of taxa, this module estimates how many taxa you would expect to find in a sample with a smaller total number of

individuals. With this method, you can compare the number of taxa in samples of different size. Using rarefaction analysis on your largest sample, you can read out the number of expected taxa for any smaller sample size. The algorithm is from Krebs (1989). An example application in palaeontology can be found in Adrain *et al.* (2000).

Let N be the total number of individuals in the sample, s the total number of species, and N_i the number of individuals of species number i. The expected number of species $E(S_n)$ in a sample of size n and the variance $V(S_n)$ are then given by

$$E(S_n) = \sum_{i=1}^{s} \left[1 - \frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right]$$

$$\begin{aligned} V(S_n) &= \sum_{i=1}^{s} \left[\frac{\binom{N-N_i}{n}}{\binom{N}{n}} \left(1 - \frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right) \right] \\ &+ 2 \sum_{j=2}^{s} \sum_{i=1}^{j-1} \left[\frac{\binom{N-N_i-N_j}{n}}{\binom{N}{n}} - \frac{\binom{N-N_i}{n}\binom{N-N_j}{n}}{\binom{N}{n}\binom{N}{n}} \right] \end{aligned} \quad (3)$$

Standard errors (square roots of variances) are given by the program. In the graphical plot, these standard errors are converted to 95 percent confidence intervals.

Diversity curves

Typical application	Assumptions	Data needed
Plotting diversity curves from occurrence data	None	Abundance or presence/absence matrix with samples in rows (lowest sample at bottom) and taxa in columns

Found in the 'Strat' menu, this simple tool allows plotting of diversity curves from occurrence data in a stratigraphical column. Note that samples should be in stratigraphical order, with the uppermost (youngest) sample in the uppermost row. Data are subjected to the range-through assumption (absences between first and last appearance are treated as presences). Originations and extinctions are in absolute numbers, not percentages.

The 'Endpoint correction' option counts a FAD or LAD in a sample as 0.5 instead of 1 in that sample. Both FAD and LAD in the sample counts as 0.33.

10 Time series analysis

Spectral analysis

Typical application	Assumptions	Data needed
Finding periodicities in counted or measured data	Time series long enough to contain at least four cycles	One or two columns of counted or measured data

Two columns must be selected (x and y values). Since palaeontological data are often unevenly sampled, the FFT algorithm can be difficult to use. PAST therefore includes the Lomb periodogram algorithm for unevenly sampled data, with time values given in the first column.

The frequency axis is in units of 1/(x unit). If for example, your x values are given in millions of years, a frequency of 0.1 corresponds to a period of 10 million years. The power axis is in units proportional to the square of the amplitudes of the sinusoids present in the data.

Also note that the frequency axis extends to very high values. If your data are evenly sampled, the upper half of the spectrum is a mirror image of the lower half, and is of little use. If some of your regions are closely sampled, the algorithm may be able to find useful information even above the half-point (Nyquist frequency).

The highest peak in the spectrum is presented with its frequency and power value, together with a probability that the peak could occur from random data. The 0.01 and 0.05 significance levels ('white noise lines') are shown as red dashed lines.

You may want to remove any linear trend in the data (Edit menu) before applying the Lomb periodogram. Failing to do so can produce annoying peaks at low frequencies.

Autocorrelation

Typical application	Assumptions	Data needed
Finding periodicities in counted or measured data	Time series long enough to contain at least two cycles. Even spacing of data points.	One column of counted or measured data

Autocorrelation (Davis 1986) is carried out on separate column(s) of evenly sampled temporal/stratigraphic data. Lag times up to $N/2$, where N is the number of values in the vector, are shown along the x axis (positive lag times only - the autocorrelation function is symmetrical around zero). A predominantly zero autocorrelation signifies random data - periodicities turn up as peaks.

The "95 percent confidence interval" option will draw lines at plus/minus $1.76\sqrt{\frac{1}{n-\tau+3}}$, after Davis (1986). This is the confidence interval for random, independent points.

This module handles missing data, coded with question marks ('?').

Cross-correlation

Typical application	Assumptions	Data needed
Finding an optimal alignment of two time series	Even spacing of data points.	Two columns of counted or measured data

Cross-correlation (Davis 1986) is carried out on two column(s) of evenly sampled temporal/stratigraphic data. The x axis shows the displacement of the second column with respect to the first, the y axis the correlation between the two time series for a given displacement. The "p values" option will draw the significance of the correlation, after Davis (1986).

Wavelet transform

Typical application	Assumptions	Data needed
Inspection of time series at different scales	Even spacing of data points	One column of counted or measured data

The continuous wavelet transform (CWT) is an analysis method where a data set can be inspected at small, intermediate and large scales simultaneously. It can be useful for detecting periodicities at different wavelengths, self-similarity and other features. The vertical axis in the plot is a logarithmic size scale, with the signal observed at a scale of only two consecutive data points at the bottom, and at a scale of one fourth of the whole sequence at the top. One unit on this axis corresponds to a doubling of the size scale. The bottom of the figure thus represents a detailed, fine-grained view, while the top represents a smoothed overview of longer trends. Signal energy (or more correctly correlation strength with the scaled mother wavelet) is shown with a grayscale.

The shape of the mother wavelet can be set to Morlet, Gauss or Sombrero. The Morlet wavelet usually performs best.

The algorithm is based on fast convolution of the signal with the wavelet at different scales, using the FFT.

The wavelet transform was used by Prokoph et al. (2000) for illustrating cycles in diversity curves for planktic foraminiferans.

Walsh transform

Typical application	Assumptions	Data needed
Spectral analysis (finding periodicities) of binary or ordinal data	Even spacing of data points	One column of binary (0/1) or ordinal (integer) data

The normal methods for spectral analysis are perhaps not optimal for binary data, because they decompose the time series into sinusoids rather than "square

waves". The Walsh transform may then be a better choice, using basis functions that flip between -1 and +1. These basis functions have different "frequencies" (number of transitions divided by two), known as *sequencies*. In PAST, each pair of even ("cal") and odd ("sal") basis functions (one pair for each integer-valued sequency) is combined into a power value using $cal^2 + sal^2$, producing a "power spectrum" that is comparable to the Lomb periodogram.

Note that the Walsh transform is slightly "exotic" compared with the Fourier transform, and its interpretation must be done cautiously. For example, the effects of the duty cycle (percentage of ones versus zeros) are somewhat difficult to understand.

In PAST, the data values are pre-processed by multiplying with two and subtracting one, bringing 0/1 binary values into the -1/+1 range optimal for the Walsh transform.

Runs test

Typical application	Assumptions	Data needed
Testing for randomness in a time series	None	One column containing a time series. The values are converted to 0 ($x \leq 0$) or 1 ($x > 0$).

The runs test is a non-parametric test for randomness in a sequence of values. Non-randomness may include such effects as autocorrelation, trend and periodicity.

The test is based on a dichotomy between two values ($x \leq 0$ or $x > 0$). It counts the number of runs (groups of consecutive equal values) and compares this to a theoretical value. The runs test can therefore be used directly for sequences of binary data. Continuous data can be converted to an appropriate form by subtracting the mean (Transform menu), or taking the difference from one value to the next (use "x-u" in the Evaluate Expression function).

11 Geometrical analysis

Directional analysis

Typical application	Assumptions	Data needed
Displaying and testing for random distribution of directional data	See below	One column of directional data in degrees (0-360)

Plots a rose diagram (polar histogram) of directions given in a column of degree values (0 to 360). Used for plotting current-oriented specimens, orientations of trackways, orientations of morphological features (e.g. terrace lines), etc.

By default, the 'mathematical' angle convention of anticlockwise from east is chosen. If you use the 'geographical' convention of clockwise from north, tick the box.

You can also choose whether to have the abundances proportional to radius in the rose diagram, or proportional to area (equal area).

The mean angle, together with the $\bar{R}$ value (Rayleigh's spread), are given:

$$\bar{R} = \sqrt{\left(\sum_{i=1}^{n} \cos\theta_i\right)^2 + \left(\sum_{i=1}^{n} \sin\theta_i\right)^2} \tag{4}$$

$\bar{R}$ is further tested against a random distribution using Rayleigh's test for directional data (Davis 1986). Note that this procedure assumes evenly or unimodally distributed data - the test is not appropriate for bidirectional data. Also, the test is not accurate for N>200; it will then report a too high p value.

A four-bin Chi-square test is also available, giving the probability that the directions are randomly and evenly distributed.

Point distribution

Typical application	Assumptions	Data needed
Testing for clustering or overdispersion of two-dimensional position values	Elements small compared to their distances, mainly convex domain, N>50.	Two columns of *x/y* positions

Point distribution statistics using nearest neighbour analysis (modified from Davis 1986). The area is estimated using the convex hull, which is the smallest convex polygon enclosing the points. This is inappropriate for points in very concave domains. Also, there is no correction for boundary effects, meaning that the statistics are reasonably valid only for large N (N>50).

The probability that the distribution is random (Poisson process, giving an exponential nearest neighbour distribution) is presented, together with the R value:

$$R = \frac{2\bar{d}}{\sqrt{A/N}},$$

where $\bar{d}$ is the observed mean distance between nearest neighbours, A is the area of the convex hull, and N is the number of points. Clustered points give R<1, Poisson patterns give R 1, while overdispersed points give R>1.

Applications of this module include spatial ecology (are in-situ brachiopods clustered) and morphology (are trilobite tubercles overdispersed).

Multivariate allometry

Typical application	Assumptions	Data needed
Finding and testing for allometry in a multivariate morphometric data set	None	A multivariate data set with variables (distance measurements) in columns, specimens in rows.

This advanced method for investigating allometry in a multivariate data set is based on Jolicoeur (1963) with extensions by Kowalewski et al. (1997). The data are (automatically) log-transformed and subjected to PCA. The first principal component (PC1) is then regarded as a size axis (this is only valid if the variation accounted for by PC1 is large, say more than 80 percent). The allometric coefficient for each original variable is estimated by dividing the PC1 loading for that variable by the mean PC1 loading over all variables.

95 percent confidence intervals for the allometric coefficients are estimated by bootstrapping specimens. 2000 bootstrap replicates are made.

Missing data are supported by column average substitution.

Fourier shape analysis

Typical application	Assumptions	Data needed
Analysis of fossil outline shape (2D)	Shape expressible in polar coordinates, sufficient number of digitized points to capture features.	Digitized x/y coordinates around an outline. Specimens in rows, coordinates of alternating x and y values in columns (see Procrustes fitting below).

Accepts $X - Y$ coordinates digitized around an outline. More than one shape (row) can be simultaneously analyzed. Points do not need to be totally evenly spaced. The shape must be expressible as a unique function in polar co-ordinates, that is, any straight line radiating from the centre of the shape must cross the outline only once.

The origin for the polar coordinate system is found by numerical approximation to the centroid. 128 points are then produced at equal angular increments around the outline, through linear interpolation. The centroid is then re-computed, and the radii normalized (size is thus removed from the analysis). The cosine and sine components are given for the first ten harmonics, but note that only $N/2$ harmonics are 'valid', where N is the number of digitized points. The coefficients can be copied to the main spreadsheet for further analysis (e.g. by PCA).

The 'Shape view' window allows graphical viewing of the Fourier shape approximation(s).

Elliptic Fourier shape analysis

Typical application	Assumptions	Data needed
Analysis of fossil outline shape	Sufficient number of digitized points to capture features.	Digitized x/y coordinates around an outline. Specimens in rows, coordinates of alternating x and y values in columns (see Procrustes fitting below).

More than one shape (row) can be simultaneously analyzed.

Elliptic Fourier shape analysis is in some respects superior to simple Fourier shape analysis. One advantage is that the algorithm can handle complicated shapes which may not be expressible as a unique function in polar co-ordinates. Elliptic Fourier shapes is now a standard method of outline analysis. The algorithm used in PAST is described in Ferson et al. (1985).

Cosine and sine components of x and y increments along the outline for the first 10 harmonics are given, but only the first $N/2$ harmonics should be used, where N is the number of digitized points. Size and positional translation are normalized away, and do not enter in the coefficients. However, no attempt is made to standardize rotation or starting point, so all specimens should be measured in a standard orientation. The coefficients can be copied to the main spreadsheet for further analysis (e.g. by PCA).

The 'Shape view' window allows graphical viewing of the elliptic Fourier shape approximation(s).

Eigenshape analysis

Typical application	Assumptions	Data needed
Analysis of fossil outline shape	Sufficient number of digitized points to capture features.	Digitized x/y coordinates around several outlines. Specimens in rows, coordinates of alternating x and y values in columns (see Procrustes fitting below).

Eigenshapes are principal components of outlines. The scatter plot of outlines in principal component space can be shown, and linear combinations of the eigenshapes themselves can be visualized.

The implementation in PAST is partly based on MacLeod (1999). It finds the optimal number of equally spaced points around the outline using an iterative search, so the original points need not be equally spaced. The eigenanalysis is based on the covariance matrix of the non-normalized turning angle increments around the outlines. The algorithm does not assume a closed curve, and the endpoints are therefore not constrained to coincide in the reconstructed shapes. Landmark-registered eigenshape analysis is not included. All outlines must start at the 'same' point.

Procrustes fitting (2D or 3D)

Typical application	Assumptions	Data needed
Standardization of morphometrical landmark coordinates	None	Digitized x/y or $x/y/z$ landmark coordinates. Specimens in rows, coordinates of alternating x and y (or $x/y/z$) values in columns.

The Procrustes option in the Transform menu will transform your measured coordinates to Procrustes coordinates. Specimens go in different rows and landmarks along each row. If you have three specimens with four landmarks, your data should look as follows:

x1 y1 x2 y2 x3 y3 x4 y4
x1 y1 x2 y2 x3 y3 x4 y4
x1 y1 x2 y2 x3 y3 x4 y4

For 3D the data will be similar, but with additional columns for z.

Landmark data in this format could be analyzed directly with the multivariate methods in PAST, but it is recommended to standardize to so-called Procrustes coordinates by removing position, size and rotation. A further transformation to Procrustes residuals (approximate tangent space coordinates) is achieved by selecting 'Subtract mean' in the Edit menu. Note: You must always convert to Procrustes coordinates first, then to Procrustes residuals.

Here is a typical sequence of operations for landmark analysis:

- Conversion of measured coordinates to Procrustes coordinates
- Conversion of Procrustes coordinates to Procrustes residuals (this must not be done before Thin-plate Spline Transformation or Shape PCA analysis, see below).
- Multivariate analysis of tangent space coordinates, with e.g. PCA or cluster analysis.

A thorough description of Procrustes and tangent space coordinates is given by Dryden & Mardia (1998). Algorithms for Procrustes fitting are as given in this reference (a closed-form algorithm for 2D, an iterative algorithm for 3D).

Missing data are supported by column average substitution.

Shape PCA

This is an option in the Principal Components module (Multivar menu). PCA on landmark data can be carried out as normal PCA analysis on Procrustes residuals for 2D or 3D (see above), but for 2D landmark data some extra functionality is available in the PCA module by choosing Shape PCA. The conversion to Procrustes residuals is then done automatically, so your data must be Procrustes fitted, but not with subtracted mean. The var-covar option is enforced, and the 'Shape deform (2D)' button enabled. This allows you to view the displacement of landmarks from the mean shape (plotted as points or symbols) in the direction of the different principal components, allowing interpretation of the components. The displacements are plotted as lines (vectors).

Another implementation of Shape PCA is available under Relative Warps (see below), by setting the parameter *alpha* to zero.

Thin-plate spline transformation grids

Typical application	**Assumptions**	**Data needed**
Visualization of shape change	None	Digitized x/y landmark coordinates. Specimens in rows, coordinates of alternating x and y values in columns. Procrustes standardization recommended.

The first specimen (first row) is taken as a reference, with an associated square grid. The warps from this to all other specimens can be viewed. You can also choose the mean shape as the reference.

The 'Expansion factors' option will display the area expansion (or contraction) factor around each landmark in yellow numbers, indicating the degree of local growth. This is computed using the Jacobian of the warp. Also, the expansions are colour-coded for all grid elements, with green for expansion and purple for contraction.

At each landmark, the principal strains can also be shown, with the major strain in black and minor strain in brown. These vectors indicate directional stretching.

A description of thin-plate spline transformation grids is given by Dryden & Mardia (1998).

Partial warps

From the thin-plate spline window, you can choose to see the partial warps for a particular spline deformation. The first partial warp will represent some long-range (large scale) deformation of the grid, while higher-order warps will normally be connected with more local deformations. The affine component of the warp (also known as zeroth warp) represents linear translation, scaling, rotation and shearing. In the present version of PAST you can not view the principal warps.

When you increase the magnification factor from zero, the original landmark configuration and a grid will be progressively deformed according to the selected partial warp.

Partial warp scores

From the thin-plate spline window, you can also choose to see the partial warp scores of all the specimens. Each partial warp score has two components (x and y), and the scores are therefore presented in scatter plots.

Relative warps

Typical application	Assumptions	Data needed
Ordination of a set of shapes	None	Digitized x/y landmark coordinates. Specimens in rows, coordinates of alternating x and y values in columns. Procrustes standardization recommended.

The relative warps can be viewed as the principal components of the set of thin-plate transformations from the mean shape to each of the shapes under study. It provides an alternative to direct PCA of the landmarks (see Shape PCA above).

The parameter alpha can be set to one of three values:

- *alpha=-1* emphasizes small-scale variation.
- *alpha=0* is PCA of the landmarks directly, and is equivalent to Shape PCA (see above) of the non-affine part of shape variation.
- *alpha=1* emphasizes large-scale variation.

The relative warps are ordered according to importance, and the first and second warps are usually the most informative. Note that the percentage values of the eigenvalues are relative to the total non-affine part of the transformation - the affine part is not included.

The relative warps are visualized with thin-plate spline transformation grids. When you increase or decrease the amplitude factor away from zero, the original

landmark configuration and grid will be progressively deformed according to the selected relative warp.

The relative warp scores of pairs of consecutive relative warps can shown in scatter plots, and all scores can be shown in a numerical matrix.

The algorithm for computing the relative warps is taken from Dryden & Mardia (1998).

Size from landmarks (2D or 3D)

Typical application	Assumptions	Data needed
Size estimation from landmarks	None	Digitized x/y or $x/y/z$ landmark coordinates. Specimens in rows, coordinates with alternating x and y (and z for 3D) values in columns. Must not be Procrustes fitted or normalized for size!

Calculates the centroid size for each specimen (Euclidean norm of the distances from all landmarks to the centroid).

The values in the 'Normalized' column are centroid sizes divided by the square root of the number of landmarks - this might be useful for comparing specimens with different numbers of landmarks.

Normalize size

The 'Normalize size' option in the Transform menu allows you to remove size by dividing all coordinate values by the centroid size for each specimen. For 2D data you may instead use Procrustes coordinates, which are also normalized with respect to size.

See Dryden & Mardia (1998), p. 23-26.

Distance from landmarks (2D or 3D)

Typical application	Assumptions	Data needed
Calculating distances between two landmarks	None	Digitized x/y or $x/y/z$ landmark coordinates. Specimens in rows, coordinates with alternating x and y (and z for 3D) values in columns. May or may not be Procrustes fitted or normalized for size.

Calculates the Euclidean distances between two fixed landmarks for one or many specimens. You must choose two landmarks - these are named according to the name of the first column for the landmark (x value).

All distances from landmarks (EDMA)

Typical application	Assumptions	Data needed
Calculating distances between all pairs of landmarks	None	Digitized x/y or $x/y/z$ landmark coordinates. Specimens in rows, coordinates with alternating x and y (and z for 3D) values in columns. May or may not be Procrustes fitted or normalized for size.

This function will replace the landmark data in the data matrix with a data set consisting of distances between all pairs of landmarks, with one specimen per row. The number of pairs is N(N-1)/2 for N landmarks. This transformation will allow multivariate analysis of distance data, which are not sensitive to rotation or translation of the original specimens, so a Procrustes fitting is not mandatory before such analysis. Using distance data also allows log-transformation, and analysis of fit to the allometric equation for pairs of distances.

Missing data are supported by column average substitution.

Landmark linking

This function in the Geomet menu allows the selection of any pairs of landmarks to be linked with lines in the morphometric plots (thin-plate splines, partial and relative warps, etc.), to improve readability. The landmarks must be present in the main spreadsheet before links can be defined.

Pairs of landmarks are selected or deselected by clicking in the symmetric matrix. The set of links can also be saved in a text file. Note that there is little error checking in this module.

Burnaby size removal

This function in the Transform menu will log-transform your multivariate distance measurement data set, and project it onto a space orthogonal to the first principal component. Burnaby's method may (or may not!) remove size but not shape from the data, for further "size-free" data analysis. Note that the implementation in PAST does not center the data within groups - it assumes that all specimens (rows) belong to one group.

Gridding (spatial interpolation)

Typical application	Assumptions	Data needed
Spatial interpolation of scattered data points onto a regular grid	Some degree of smoothness	Three columns with position (x,y) and corresponding data values

Gridding (spatial interpolation) allows the production of a map showing a continuous spatial estimate of some variate such as fossil abundance or thickness of a rock unit, based on scattered data points. The user can specify the size of the grid (number of rows and columns), but in the present version the spatial coverage of the map is generated automatically based on the positions of data points (the map will always be square).

A least-squares linear surface (trend) is automatically fitted to the data, removed prior to gridding and finally added back in. This is primarily useful for the semivariogram modelling and the kriging method.

Three algorithms are available:

Moving average

The value at a grid node is simply the average of the N closest data points, as specified by the user (the default is to use all data points). The points are given weight in inverse proportion to distance. This algorithm is simple and will not always give good (smooth) results. One advantage is that the interpolated values will never go outside the range of the data points.

Thin-plate spline

Maximally smooth interpolator. Can overshoot in the presence of sharp bends in the surface.

Kriging

This advanced method is implemented in a simple version in PAST. The user is required to specify a model for the semivariogram, by choosing one of three models (spherical, exponential or Gaussian) and corresponding parameters to fit the empirical semivariances as well as possible. See e.g. Davis (1986) for more information. The kriging procedure also provides an estimate of standard errors across the map (this depends on the semivariogram model being accurate). Kriging in PAST does not provide for anisotropic semivariance.

12 Cladistics

Typical application	Assumptions	Data needed
Semi-objective analysis of relationships between taxa from morphological or genetic evidence	Many! See Kitchin *et al.* (1998)	Character matrix with taxa in rows, outgroup in first row. For calculation of stratigraphic congruence indices, first and last appearance datums must be given in the first two columns.

Warning: the cladistics package in PAST is fully operational, but lacking in comprehensive functionality. The heuristic algorithms seem not to perform quite as well as in some other programs (this is being looked into). The PAST cladistics package is adequate for education and initial data exploration, but for more 'serious' work we recommend a specialized program such as PAUP. Algorithms are from Kitchin et al. (1998).

Parsimony analysis

Character states should be coded using integers in the range 0 to 255. The first taxon is treated as the outgroup, and will be placed at the root of the tree.

Missing values are coded with a question mark (?) or the value -1. Please note that PAST does not collapse zero-length branches. Because of this, missing values can lead to a proliferation of equally shortest trees ad nauseam, many of which are in fact equivalent.

There are four algorithms available for finding short trees:

Branch-and-bound

The branch-and-bound algorithm is guaranteed to find all shortest trees. The total number of shortest trees is reported, but a maximum of 1000 trees are saved. You should not use the branch-and-bound algorithm for data sets with more than 12 taxa.

Exhaustive

The exhaustive algorithm evaluates all possible trees. Like the branch-and-bound algorithm it will necessarily find all shortest trees, but it is very slow. For 12 taxa, more than 600 million trees are evaluated! The only advantage over branch-and-bound is the plotting of tree length distribution. This histogram may indicate the 'quality' of your matrix, in the sense that there should be a tail to the left such that few short trees are 'isolated' from the greater mass of longer trees (but see Kitchin et al. 1998 for critical comments on this). For more than 8 taxa, the histogram is based on a subset of tree lengths and may not be accurate.

Heuristic, nearest neighbour interchange

This heuristic algorithm adds taxa sequentially in the order they are given in the matrix, to the branch where they will give least increase in tree length. After each taxon is added, all nearest neighbour trees are swapped to try to find an even shorter tree.

Like all heuristic searches, this one is much faster than the algorithms above and can be used for large numbers of taxa, but is not guaranteed to find all or any of the most parsimonious trees. To decrease the likelihood of ending up on a suboptimal local minimum, a number of reorderings can be specified. For each reordering, the order of input taxa will be randomly permutated and another heuristic search attempted.

Please note: Because of the random reordering, the trees found by the heuristic searches will normally be different each time. To reproduce a search exactly, you will have to start the parsimony module again from the menu, using the same value for "Random seed". This will reset the random number generator to the seed value.

Heuristic, subtree pruning and regrafting

This algorithm (SPR) is similar to the one above (NNI), but with a more elaborate branch swapping scheme: A subtree is cut off the tree, and regrafting onto all other branches in the tree is attempted in order to find a shorter tree. This is done after each taxon has been added, and for all possible subtrees. While slower than NNI, SPR will often find shorter trees.

Heuristic, tree bisection and reconnection

This algorithm (TBR) is similar to the one above (SPR), but with an even more complete branch swapping scheme. The tree is divided into two parts, and these are reconnected through every possible pair of branches in order to find a shorter tree. This is done after each taxon is added, and for all possible divisions of the tree. TBR will often find shorter trees than SPR and NNI, at the cost of longer computation time.

Character optimization criteria

Three different optimization criteria are availiable:

Wagner

Characters are reversible and ordered, meaning that 0->2 costs more than 0->1, but has the same cost as 2->0.

Fitch

Characters are reversible and unordered, meaning that all changes have equal cost. This is the criterion with fewest assumptions, and is therefore generally preferable.

Dollo

Characters are ordered, but acquistition of a character state (from lower to higher value) can happen only once. All homoplasy is accounted for by secondary reversals. Hence, 0->1 can only happen once, normally relatively close to the root of the tree, but 1->0 can happen any number of times further up in the tree. (This definition has been debated on the PAST mailing list, especially whether Dollo characters need to be ordered).

Bootstrap

Bootstrapping is performed when the 'Bootstrap replicates' value is set to non-zero. The specified number of replicates (typically 100 or even 1000) of your character matrix are made, each with randomly weighted characters. The bootstrap value for a group is the percentage of replicates supporting that group. A replicate supports the group if the group exists in the majority rule consensus tree of the shortest trees made from the replicate.

Warning: Specifying 1000 bootstrap replicates will clearly give a thousand times longer computation time than no bootstrap! Exhaustive search with bootstrapping is unrealistic and is not allowed.

Cladogram plotting

All shortest (most parsimonious) trees can be viewed, up to a maximum of 1000 trees. If bootstrapping has been performed, a bootstrap value in percents is given at the root of the subtree specifying each group.

Character states can also be plotted onto the tree, as selected by the 'Character' buttons. This character reconstruction is unique only in the absence of homoplasy. In case of homoplasy, character changes are placed as close to the root as possible, favouring one-time acquisition and later reversal of a character state over several independent gains (so-called *accelerated transformation*).

Consistency index

The per-character consistency index (ci) is defined as m/s, where m is the minimum possible number of character changes (steps) on any tree, and s is the actual number of steps on the current tree. This index hence varies from one (no homoplasy) and down towards zero (a lot of homoplasy). The ensemble consistency index CI is a similar index summed over all characters.

Retention index

The per-character retention index (ri) is defined as $(g - s)/(g - m)$, where m and s are as for the consistency index, while g is the maximal number of steps for the character on any cladogram (Farris 1989). The retention index measures the amount of synapomorphy on the tree, and varies from 0 to 1.

Consensus tree

The consensus tree of all shortest (most parsimonious) trees can also be viewed. Two consensus rules are implemented: Strict (groups must be supported by all trees) and majority (groups must be supported by more than 50 percent of the trees).

Bremer support (decay index)

The Bremer support for a clade is the number of extra steps you need to construct a tree (consistent with the characters) where that clade is no longer present. There are reasons to prefer this index rather than the bootstrap value. PAST does not compute Bremer supports directly, but for smaller data sets it can be done 'manually' as follows:

- Perform parsimony analysis with exhaustive search or branch-and-bound. Take note of the clades and the length N of the shortest tree(s) (for example 42). If there are more than one shortest tree, look at the strict consensus tree. Clades which are no longer found in the consensus tree have a Bremer support value of 0.
- In the box for 'Longest tree kept', enter the number $N+1$ (43 in our example) and perform a new search.
- Additional clades which are no longer found in the strict consensus tree have a Bremer support value of 1.
- For 'Longest tree kept', enter the number $N + 2$ (44) and perform a new search. Clades which now disappear in the consensus tree have a Bremer support value of 2.
- Continue until all clades have disappeared.

Stratigraphic congruence indices

For calculation of stratigraphic congruence indices, the first two columns in the data matrix must contain the first and last appearance datums, respectively, for each taxon. These datums must be given such that the youngest age (or highest stratigraphic level) has the highest numerical value. You may need to use negative

values to achieve this (e.g. 400 million years before present is coded as -400.0). The box "FADs/LADs in first columns" in the Parsimony dialogue must be ticked.

The Stratigraphic Congruence Index (SCI) of Huelsenbeck (1994) is defined as the proportion of stratigraphically consistent nodes on the cladogram, and varies from 0 to 1. A node is stratigraphically consistent when the oldest first occurrence above the node is the same age or younger than the first occurrence of its sister taxon (node).

The Relative Completeness Index (RCI) of Benton & Storrs (1994) is defined as $(1-MIG/SRL)x100$ percent, where MIG (Minimum Implied Gap) is the sum of the durations of ghost ranges and SRL is the sum of the durations of observed ranges. The RCI can become negative, but will normally vary from 0 to 100.

The Gap Excess Ratio (GER) of Wills (1999) is defined as $1-(MIG-G_{min})/(G_{max}-G_{min})$ where G_{min} is the minimum possible sum of ghost ranges on any tree (that is, the sum of distances between successive FADs), and G_{max} is the maximum (that is, the sum of distances from first FAD to all other FADs).

These indices are further subjected to a permutation test, where all dates are randomly redistributed across the different taxa 1000 times. The proportion of permutations where the recalculated index exceeds the original index is given. If small (e.g. p<0.05), this indicates a statistically significant departure from the null hypothesis of no congruency between cladogram and stratigraphy (in other words, you have significant congruency). The permutation probabilities of RCI and GER are equal for any given set of permutations, because they are based on the same value for MIG.

13 Biostratigraphy

Unitary associations

Typical application	Assumptions	Data needed
Quantitative biostratigraphical correlation	None	Presence/absence (1/0) matrix with horizons in rows, taxa in columns

Unitary Associations analysis (Guex 1991) is a method for biostratigraphical correlation (see Angiolini & Bucher 1999 for a typical application). The data input consists of a presence/absence matrix with samples in rows and taxa in columns. Samples belonging to the same section (locality) are tagged with the same color, and ordered stratigraphically within each section such that the lowermost sample is entered in the lowest row. Colours can be re-used in data sets with large numbers of sections (see alveolinid.dat for an example).

Overview of the method

The method of Unitary Associations is logical, but rather complicated, consisting of a number of steps. For details, see Guex 1991. The implementation in PAST includes most of the features found in the standard program, called BioGraph (Savary & Guex 1999), and thanks to a fruitful co-operation with Jean Guex it also includes a number of options and improvements not found in the present version of that program.

The basic idea is to generate a number of assemblage zones (similar to 'Oppel zones') which are optimal in the sense that they give maximal stratigraphic resolution with a minimum of superpositional contradictions. One example of such a contradiction would be a section containing a species A above a species B, while assemblage 1 (containing species A) is placed below assemblage 2 (containing species B). PAST (and BioGraph) carries out the following steps:

1. Residual maximal horizons

The method makes the range-through assumption, meaning that taxa are considered to have been present at all levels between the first and last appearance in any section. Then, any samples with a set of taxa that is contained in another sample are discarded. The remaining samples are called residual maximal horizons. The idea behind this throwing away of data is that the absent taxa in the discarded samples may simply not have been found even though they originally existed. Absences are therefore not as informative as presences.

2. Superposition and co-occurrence of taxa

Next, all pairs (A,B) of taxa are inspected for their superpositional relationships: A below B, B below A, A together with B, or unknown. If A occurs below B in one locality and B below A in another, they are considered to be co-occurring although they have never actually been found together.

The superpositions and co-occurrences of taxa can be viewed in the *biostratigraphic graph*. In this graph, taxa are coded as numbers. Co-occurrences between pairs of taxa are shown as solid blue lines. Superpositions are shown as dashed red lines, with long dashes from the above-occurring taxon and short dashes from the below-occurring taxon.

Some taxa may occur in so-called forbidden sub-graphs, which indicate inconsistencies in their superpositional relationships. Two of the several types of such sub-graphs can be plotted in PAST: C_n cycles, which are superpositional cycles (A->B->C->A), and S_3 circuits, which are inconsistencies of the type 'A co-occurring with B, C above A, and C below B'. Interpretation of such forbidden sub-graphs is described by Guex (1991).

3. Maximal cliques

Maximal cliques are groups of co-occurring taxa not contained in any larger group of co-occurring taxa. The maximal cliques are candidates for the status of unitary associations, but will be further processed below. In PAST, maximal cliques receive a number and are also named after a maximal horizon in the original data set which is identical to, or contained in (marked with asterisk), the maximal clique.

4. Superposition of maximal cliques

The superpositional relationships between maximal cliques are decided by inspecting the superpositional relationships between their constituent taxa, as computed in step 2. Contradictions (some taxa in clique A occur below some taxa in clique B, and vice versa) are resolved by a 'majority vote'. The contradictions between cliques can be viewed in PAST.

The superpositions and co-occurrences of cliques can be viewed in the maximal clique graph. In this graph, cliques are coded as numbers. Co-occurrences between pairs of cliques are shown as solid blue lines. Superpositions are shown as dashed red lines, with long dashes from the above-occurring clique and short dashes from the below-occurring clique. Also, cycles between maximal cliques (see below) can be viewed as green lines.

5. Resolving cycles

It will sometimes be the case that maximal cliques are now ordered in cycles: A is below B, which is below C, which is below A again. This is clearly contradictory. The 'weakest link' (superpositional relationship supported by fewest taxa) in such cycles is destroyed.

6. Reduction to unique path

At this stage, we should ideally have a single path (chain) of superpositional relationships between maximal cliques, from bottom to top. This is however often not the case, for example if A and B are below C, which is below D, or if we have isolated paths without any relationships (A below B and C below D). To produce a single path, it is necessary to merge cliques according to special rules.

7. Post-processing of maximal cliques

Finally, a number of minor manipulations are carried out to 'polish' the result: Generation of the 'consecutive ones' property, reinsertion of residual virtual co-occurrences and superpositions, and compaction to remove any generated non-

maximal cliques. For details on these procedures, see Guex 1991. At last, we now have the Unitary Associations, which can be viewed in PAST.

The unitary associations have associated with them an index of similarity from one UA to the next, called D:

$$D_i = |UA_i - UA_{i-1}|/|UA_i| + |UA_{i-1} - UA_i|/|UA_{i-1}|$$

8. Correlation using the Unitary Associations

The original samples are now correlated using the unitary associations. A sample may contain taxa which uniquely places it in a unitary association, or it may lack key taxa which could differentiate between two or more unitary associations, in which case only a range can be given. These correlations can be viewed in PAST.

9. Reproducibility matrix

Some unitary associations may be identified in only one or a few sections, in which case one may consider to merge unitary associations to improve the geographical reproducibility (see below). The reproducibility matrix should be inspected to identify such unitary associations. A UA which is uniquely identified in a section is shown as a black square, while ranges of UAs (as given in the correlation list) are shown in gray.

10. Reproducibility graph and suggested UA merges (biozonation)

The reproducibility graph (Gk' in Guex 1991) shows superpositions of unitary associations that are actually observed in the sections. PAST will internally reduce this graph to a unique maximal path (Guex 1991, section 5.6.3), and in the process of doing so it may merge some UAs. These mergers are shown as red lines in the reproducibility graph. The sequence of single and merged UAs can be viewed as a suggested biozonation.

Special functionality

The implementation of the Unitary Associations method in PAST includes a number of options and functions which have not yet been described in the literature. For questions about these, please contact us.

Ranking and Scaling

Typical application	Assumptions	Data needed
Quantitative biostratigraphical correlation	None	Table of depths, with wells in rows and events in columns

Ranking-Scaling (Agterberg & Gradstein 1999) is a method for quantitative biostratigraphy based on *events* in a number of wells or sections. The data input consists of wells in rows with one well per row, and events (e.g. FADs and/or LADs) in columns. The values in the matrix are depths of each event in each

well, increasing upwards (you may want to use negative values to achieve this). Absences are coded as zero. If only the order of events is known, this can be coded as increasing whole numbers (ranks, with possible ties for co-occurring events) within each well.

The implementation of ranking-scaling in PAST is not comprehensive, and advanced users are referred to the RASC and CASC programs of Agterberg and Gradstein.

Overview of the method

The method of Ranking-Scaling proceeds in two steps:

1. Ranking

The first step of Ranking-Scaling is to produce a single, comprehensive stratigraphic ordering of events, even if the data contains contradictions (event A over B in one well, but B over A in another), or longer cycles (A over B over C over A). This is done by 'majority vote', counting the number of times each event occurs above, below or together with all others. Technically, this is achieved by "presorting" followed by the Modified Hay Method (Agterberg & Gradstein 1999).

2. Scaling

The biostratigraphic analysis may end with ranking, but additional insight may be gained by estimating stratigraphic distances between the consecutive events. This is done by counting the number of observed superpositional relationships (A above or below B) between each pair (A,B) of consecutive events. A low number of contradictions implies long distance.

Some computed distances may turn out to be negative, indicating that the ordering given by the ranking step was not optimal. If this happens, the events are re-ordered and the distances re-computed in order to ensure only positive inter-event distances.

RASC in PAST

Parameters

Well threshold: The minimum number of wells in which an event must occur in order to be included in the analysis

Pair threshold: The minimum number of times a relationship between events A and B must be observed in order for the pair (A,B) to be included in the ranking step

Scaling threshold: Pair threshold for the scaling step

Tolerance: Used in the ranking step (see Agterberg & Gradstein)

Ranking

The ordering of events after the ranking step is given, with the first event at the bottom of the list. The "Range" column indicates uncertainty in the position.

Scaling

The ordering of the events after the scaling step is given, with the first event at the bottom of the list. For an explanation of all the columns, see Agterberg & Gradstein (1999).

Event distribution

A plot showing the number of events in each well, with the wells ordered according to number of events.

Scattergrams

For each well, the depth of each event in the well is plotted against the optimum sequence (after scaling). Ideally, the events should plot in an ascending sequence.

Dendrogram

Plot of the distances between events in the scaled sequence, including a dendrogram which may aid in zonation.

Constrained Optimization (CONOP)

Typical application	Assumptions	Data needed
Quantitative biostratigraphical correlation	None	Table of depths/levels, with wells/sections in rows and event pairs in columns: FADs in odd columns and LADs in even columns. Missing events are coded with zeros.

PAST includes a simple version of Constrained Optimization (Kemple et al. 1989). Both FAD and LAD of each taxon must be specified in alternate columns. Using so-called Simulated Annealing, the program searches for a global (composite) sequence of events that implies a minimal total amount of range extension (penalty) in the individual wells/sections. The parameters for the optimization procedure include an initial annealing temperature, the number of cooling steps, the cooling ratio (percentage lower than 100), and the number of trials per step. For explanation and recommendations, see Kemple et al. 1989.

Output windows include the optimization history with the temperature and penalty as function of cooling step, the global composite solution and the implied ranges in each individual section.

The implementation of CONOP in PAST is based on a FORTRAN optimization core provided by Kemple and Sadler.

Unitary Associations, Ranking-Scaling or CONOP?

(The below is a personal opinion of O. Hammer only!)

There are now three main paradigms in the field of quantitative stratigraphy (in addition to the semi-quantitative approach of graphical correlation): Unitary Associations, Ranking-Scaling and Constrained Optimization. These methods have

different aims, use different types of data, and are based on different philosophies. The discussion continues about which method is 'best', but to some extent the choice of method will depend on the purpose of the investigation. As a gross generalization, it might be expected that the probabilistic approach of ranking-scaling will produce high resolution, but at the cost of basing some of the correlations and zonal boundaries on facies-controlled or geographically constrained events rather than global (evolutionary) originations and extinctions. Unitary Associations is a more conservative approach that will probably be more robust to low lateral reproducibility, but at the cost of lower resolution. Hence, it could perhaps be argued that Unitary Associations might be preferred for 'academic' use, while Ranking-Scaling is preferable in e.g. hydrocarbon exploration where resolution is important and diachronous units are to some extent acceptable. CONOP combines the potentially high resolution of RASC with the preservation of co-occurrences of UA, at the cost of non-uniqueness of the solution and long computation times.

A major difference between the methods is that the UA method is based on association data (presence/absence in samples), while RASC and CONOP use so-called events such as FADs or LADs. The choice of method may therefore to some extent be dictated by the type of data available.

Range confidence intervals

Typical application	Assumptions	Data needed
Estimation of confidence intervals for first or last appearances and total range, for one taxon.	Random distribution of fossiliferous horizons through the stratigraphic column or through time. Section should be continuously sampled.	The number of horizons containing the taxon, and levels or dates of first and last occurrences of the taxon.

Assuming a random (Poisson) distribution of fossiliferous horizons, confidence intervals for the stratigraphic range of one taxon can be calculated given the first occurrence datum (level), last occurrence datum, and total number of horizons where the taxon is found (Strauss & Sadler 1989, Marshall 1990).

No data are needed in the spreadsheet. The program will ask for the number of horizons where the taxon is found, and levels or dates for the first and last appearances. If necessary, use negative values to ensure that the last appearance datum has a higher numerical value than the first appearance datum. 80, 95 and 99 percent confidence intervals are calculated for the FAD considered in isolation, the LAD considered in isolation, and the total range. The value *alpha* is the length of the confidence interval divided by the length of the observed range.

Be aware that the assumption of random distribution will not hold in many real situations.

Distribution free range confidence intervals

Typical application	Assumptions	Data needed
Estimation of confidence intervals for first or last appearances.	No correlation between stratigraphic position and gap size. Section should be continuously sampled.	One column per taxon, with levels or dates of all horizons where the taxon is found.

This method (Marshall 1994) does not assume random distribution of fossiliferous horizons. It requires that the levels or dates of all horizons containing the taxon are given.

The program outputs upper and lower bounds on the lengths of the confidence intervals, using a 95 percent confidence probability, for confidence levels of 50, 80 and 95 percent. Values which can not be calculated are marked with an asterisk (see Marshall 1994).

14 Acknowledgments

PAST was inspired by and includes many functions found in PALSTAT, which was programmed by P.D. Ryan with assistance from J.S. Whalley. Harper thanks the Danish Natural Science Research Council (SNF) for support. Frits Agterberg and Felix Gradstein allowed OH access to source code for RASC, and Peter Sadler provided source code for CONOP. Jean Guex provided a series of ideas for improvement and extension of the Unitary Associations module, and tested it intensively.

Many users of PAST have given us ideas for improvement and reported bugs. Among these are Charles Galea Bonavia, Hans Arne Nakrem, Mikael Fortelius, Knut Rognes, Julian Overnell, Kirsty Brown, Paolo Tomassetti, Jose Luis Navarrete-Heredia, Wally Woolfenden, Erik Telie, Fernando Archuby, Ian J. Slipper, James Gallagher, Marcio Pie, Hugo Bucher, Alexey Tesakov, Craig Macfarlane, José Camilo Hurtado Guerrero, Wolfgang Kiessling and Bastien Wauthoz.

15 References

Adrain, J.M., S.R. Westrop & D.E. Chatterton 2000. Silurian trilobite alpha diversity and the end-Ordovician mass extinction. *Paleobiology* 26:625-646.

Anderson, M.J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26:32-46.

Angiolini, L. & H. Bucher 1999. Taxonomy and quantitative biochronology of Guadalupian brachiopods from the Khuff Formation, Southeastern Oman. *Geobios* 32:665-699.

Benton, M.J. & G.W. Storrs. 1994. Testing the quality of the fossil record: paleontological knowledge is improving. *Geology* 22:111-114.

Bow, S.-T. 1984. Pattern recognition. Marcel Dekker, New York.

Brower, J.C. & K.M. Kyle 1988. Seriation of an original data matrix as applied to palaeoecology. *Lethaia* 21:79-93.

Brown, D. & P. Rothery 1993. Models in biology: mathematics, statistics and computing. John Wiley & Sons, New York.

Bruton, D.L. & A.W. Owen 1988. The Norwegian Upper Ordovician illaenid trilobites. *Norsk Geologisk Tidsskrift* 68:241-258.

Clarke, K.R. 1993. Non-parametric multivariate analysis of changes in community structure. *Australian Journal of Ecology* 18:117-143.

Clarke, K.R. & Warwick, R.M. 1998. A taxonomic distinctness index and its statistical properties. *Journal of Applied Ecology* 35:523-531.

Colwell, R.K. & J.A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society (Series B) 345:101-118.

Davis, J.C. 1986. Statistics and Data Analysis in Geology. John Wiley & Sons, New York.

Dryden, I.L. & K.V. Mardia 1998. Statistical Shape Analysis. Wiley.

Farris, J.S. 1989. The retention index and the rescaled consistency index. *Cladistics* 5:417-419.
Ferson, S.F., F.J. Rohlf & R.K. Koehn 1985. Measuring shape variation of two-dimensional outlines. *Systematic Zoology* 34:59-68.
Guex, J. 1991. Biochronological Correlations. Springer Verlag, Berlin.
Harper, D.A.T. (ed.). 1999. Numerical Palaeobiology. John Wiley & Sons, Chichester.
Hennebert, M. & A. Lees. 1991. Environmental gradients in carbonate sediments and rocks detected by correspondence analysis: examples from the Recent of Norway and the Dinantian of southwest England. *Sedimentology* 38:623-642.
Hill, M.O. & H.G. Gauch Jr. 1980. Detrended Correspondence analysis: an improved ordination technique. *Vegetatio* 42:47-58.
Horn, H.S. 1966. Measurement of overlap in comparative ecological studies. *American Naturalist* 100:419-424.
Huelsenbeck, J.P. Comparing the stratigraphic record to estimates of phylogeny. *Paleobiology* 20:470-483.
Jolicoeur, P. 1963. The multivariate generalization of the allometry equation. *Biometrics* 19:497-499.
Jolliffe, I.T. 1986. Principal Component Analysis. Springer-Verlag, Berlin.
Kemple, W.G., P.M. Sadler & D.J. Strauss. 1989. A prototype constrained optimization solution to the time correlation problem. In Agterberg, F.P. & G.F. Bonham-Carter (eds), Statistical Applications in the Earth Sciences. Geological Survey of Canada Paper 89-9:417-425.
Kitchin, I.J., P.L. Forey, C.J. Humphries & D.M. Williams 1998. Cladistics. Oxford University Press, Oxford.
Kowalewski, M., E. Dyreson, J.D. Marcot, J.A. Vargas, K.W. Flessa & D.P. Hallmann. 1997. Phenetic discrimination of biometric simpletons: paleobiological implications of morphospecies in the lingulide brachiopod *Glottidia*. *Paleobiology* 23:444-469.
Krebs, C.J. 1989. Ecological Methodology. Harper & Row, New York.
MacLeod, N. 1999. Generalizing and extending the eigenshape method of shape space visualization and analysis. *Paleobiology* 25:107-138.
Marshall, C.R. 1990. Confidence intervals on stratigraphic ranges. *Paleobiology* 16:1-10.
Marshall, C.R. 1994. Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of randomly distributed fossil horizons. *Paleobiology* 20:459-469.
Miller, R.L. & Kahn, J.S. 1962. Statistical Analysis in the Geological Sciences. John Wiley & Sons, New York.
Oxanen, J. & P.R. Minchin. 1997. Instability of ordination results under changes in input data order: explanations and remedies. *Journal of Vegetation Science* 8:447-454.
Poole, R.W. 1974. An introduction to quantitative ecology. McGraw-Hill, New York.

Press, W.H., S.A. Teukolsky, W.T. Vetterling & B.P. Flannery 1992. Numerical Recipes in C. Cambridge University Press, Cambridge.
Prokoph, A., A.D. Fowler & R.T. Patterson. 2000. Evidence for periodicity and nonlinearity in a high-resolution fossil record of long-term evolution. *Geology* 28:867-870.
Raup, D. & R.E. Crick. 1979. Measurement of faunal similarity in paleontology. *Journal of Paleontology* 53:1213-1227.
Ryan, P.D., Harper, D.A.T. & Whalley, J.S. 1995. PALSTAT, Statistics for palaeontologists. Chapman & Hall (now Kluwer Academic Publishers).
Savary, J. & J. Guex. 1999. Discrete Biochronological Scales and Unitary Associations: Description of the BioGraph Computer Program. *Memoires de Geologie (Lausanne)* 34.
Sepkoski, J.J. 1984. A kinetic model of Phanerozoic taxonomic diversity. *Paleobiology* 10:246-267.
Strauss, D. & P.M. Sadler. 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. *Mathematical Geology* 21:411-427.
Taguchi, Y-H. & Oono, Y. In press. Novel non-metric MDS algorithm with confidence level test.
Tothmeresz, B. 1995. Comparison of different methods for diversity ordering. *Journal of Vegetation Science* 6:283-290.
Wills, M.A. 1999. The gap excess ratio, randomization tests, and the goodness of fit of trees to stratigraphy. *Systematic Biology* 48:559-580.
Zar, J.H. 1996. Biostatistical Analysis. 3rd ed. Prentice Hall, New York.