4 Exploration

The statistical background of all methods discussed in this chapter can be found *Analysing Ecological Data* by Zuur, Ieno and Smith (2007). Here, we only discuss how to apply the methods in Brodgar. The exploration tools in Brodgar are divided in three groups:

1. Exploration tools using the software package R.
2. General exploration tools.
3. Exploration tools for time series.

In Section 4.4, we show how to obtain some of the graphs from Zuur et al (2007).

4.1 Data exploration using R tools

Using the link to R, Brodgar offers a wide range of exploratory tools (Figure 4.1). If there are no explanatory variables, some of the tools (e.g. lattice plots and coplots) cannot be used. If a particular technique (e.g. a histogram) needs to be applied on an explanatory variable, then temporary change the status of the variable from explanatory variable to response variable in the Data Import menu. Applying the exploration methods in Brodgar is discussed next.
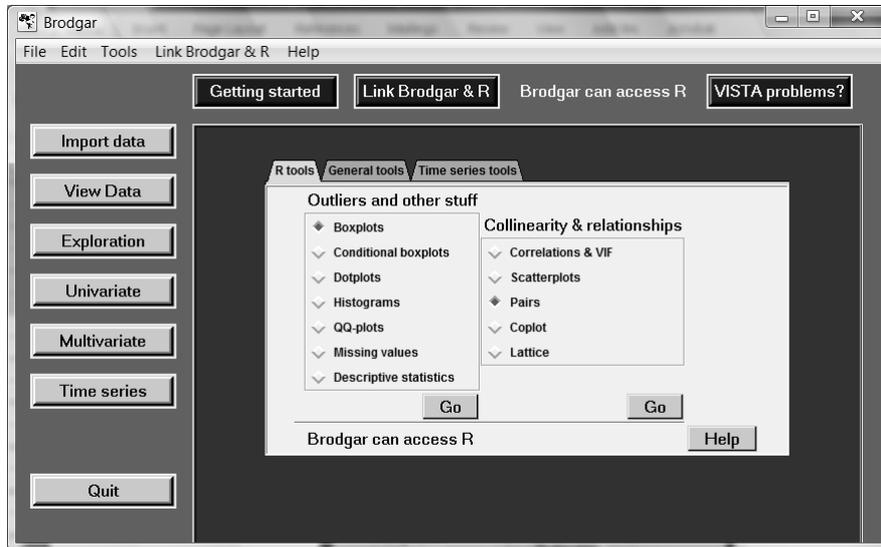
Figure 4.1. Data exploration menu; R tools.

4.1.1 Boxplots and conditional boxplots

These tools can be used to identify potential outliers and visualize heterogeneity. Selecting "Boxplots" and clicking the "Go" button in Figure 4.1 gives the left window in Figure 4.2. Under the variable tab the user selects the response variables that will be used to create the boxplot. For retrieval in later analysis, the selection of variables can be stored.
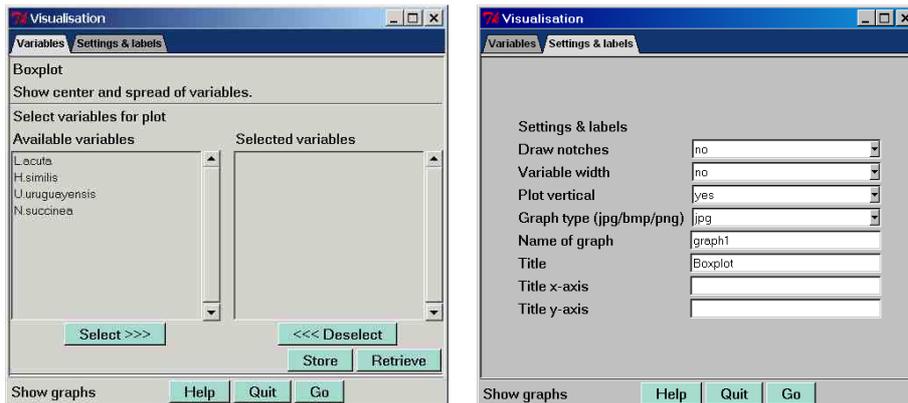
Figure 4.2. Boxplot window.

Under the "Settings & labels" tab (Figure 4.2), the following options are available:

- Draw notches: If `yes' is selected a notch is drawn in each side of the boxes. If the notches of two plots do not overlap then the medians are significantly different at the 5 percent level.
- Variable width: If `yes' is selected, the boxes are drawn with widths proportional to the square roots of the number of observations in the groups.
- Plot vertical. Indicates if the boxplots should be plotted horizontally or vertically.
- Graph type. Allows the user to save graphs in JPG, BMP or PNG format.
- Name of graph. Brodgar will save the graph as: YourName.JPG in the project directory (assuming a JPG file was selected).
- Title. This title will appear at the top of the figure.
- Title x-axis. Allows the user to enter a label for the x-axis.
- Title y-axis. Allows the user to enter a title for the y-axis.
- R-code. For further help, start R, and type: ?boxplot

In a conditional boxplot, the user needs to select one variable and one (or two) nominal variable(s). If one nominal variable (e.g. transect) is selected, a boxplot for each class (each individual transect) is made, and these are plotted next to each other. If two nominal variables are selected, e.g. transect with 3 classes and season with 2 classes, then boxplots of the 3 transect classes are plotted next to each other for each season. On the other hand, if a boxplot conditional on season and transect is made (instead of transect and season), two boxplots are plotted next to each other for each transect.

4.1.2 Histograms

In this tool the user can select one or more response variables. For quick retrieval, selections can be saved. The diversity index uses all the response variables that were selected in the Data Import process. Under the "Settings and labels" tab, the following items can be changed:

- Add density curve. This option allows a density curve to be superimposed over the histogram. The default value is: "no".
- Conditional histograms using nominal variable. The default value is: "no". This feature automatically splits the data into different groups, based on a nominal variable, drawing a separate histogram for each group. The histograms of the same variable are stacked on top of each other. For example, if species abundance data were measured in Summer and Winter, it is useful to draw histograms for the Winter data above the histograms of the Summer data. This makes them easier to compare and might highlight a change in distribution. To do this, create a new nominal variable "Time" in the spreadsheet, import all the data, select "Time" as an explanatory variable (and the species of interest as the

response variables), and in the histogram settings, select "Time" as the nominal variable for splitting the data. Brodgar will generate new titles per histogram. The number above each histogram represents the class of the nominal variable (using integers starting from 1), and the name of the variable is used as a label along the horizontal axis. The maximum number of classes in the nominal variable is approximately 5.

- If conditional histogram, use same horizontal range. This option allows the user to choose whether the range of the histograms along the horizontal axes are equal (default setting) or different per group/class of samples. This method will not work if the nominal variable has more than 5 classes.
- Graph type. This can be JPG, BMP or PNG.
- Name of Graph. Specify the name of the JPG, BMP, PNG or WMF file. The file will be saved in the project directory.
- Title. Names of variables are automatically used as horizontal labels.
- R-code. For further help, start R, and type: ?hist

4.1.3 Dotplot

This is a plot where each observation is presented by a single dot. The value is presented along the horizontal axis. The first observation in the data file is plotted at the bottom and the last observation at the top. For each selected variable, a dotplot is drawn. Under the "Settings and labels" tab, the following options are available:

- Different colours – symbols. Use different colours and symbols conditional on a nominal variable. This option makes a dotplot in which the observations are identified by different colours or symbols based on a nominal variable.
- If yes, on which nominal X. Selects a nominal explanatory variable for different colours and symbols. See previous point.
- Sort by nominal X. If a nominal variable is selected, then the data are grouped. This is a useful tool to detect violation of homogeneity.
- Graph type. This can be JPG, BMP or PNG.
- Name of Graph. Specify the name of the JPG, BMP, PNG or WMF file. The file will be saved in the project directory.
- Title. Allows the user to enter a main title for the graph.
- Title x-axis. Allows the user to enter a label for the x-axis.
- Title y-axis. Allows the user to enter a title for the y-axis.
- R-code. For further help, start R, and type: ?dotchart

4.1.4 QQ-plots

Quantile-Quantile plots are a graphical tool to determine whether the data are normally distributed. Brodgar combines QQ-plots with a power transformation, which is given by:

$$\frac{y^p - 1}{p} \quad \text{if p is not equal to 0}$$

$\log(Y)$ if p is 0

Brodgar offers two options. For each selected (response) variable, either QQ-plots for predetermined values of p are made, or a particular, user selected, value is used. In the first case, Brodgar makes four QQ-plots, one for the untransformed data, the square root transformed data, the cubic root transformed data, and log10 transformed data. All four QQ-plots are drawn in one graph. The alternative is to specify the power p. Under the "Settings and labels" tab, the following options are available:

- Task. Allows the user to accept the default transformations of QQ-plots for four predetermined powers (none, square root, cubic root, log10) or to specify user-defined powers for the transformations (values between 0 and 3 can be chosen).
- Graph type. This can be JPG, BMP or PNG.
- Name of Graph. Allows the user to enter the file name for the saved JPG, BMP, PNG or WMF file. The file will be saved in the project directory.
- Title. Allows the user to enter the main title for the graph.
- R-code. For further help, start R, and type: ?qqnorm

4.1.5 Coplot

Coplots show the relationship between a response and explanatory variable, while changing the values of second (and third) explanatory variable. They are useful for visualizing interactions. Figure 4.3 shows the options for coplots in Brodgar. In the left panel, the user can select one response variable (which will be plotted along the y-axis), and the explanatory variable that will be plotted along the x-axis. The first conditional variable can either be nominal or continuous. The second variable is optional.
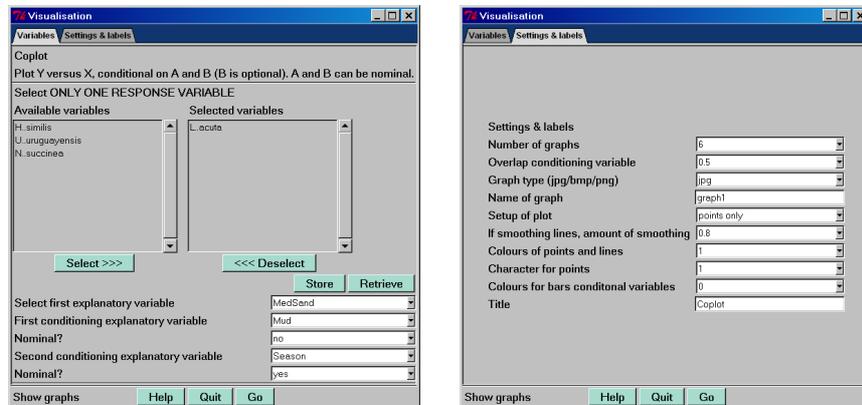
Figure 4.3. Options for coplots.

In the right panel, various setting and labels can be changed. The following options are available:

- Number of graphs (default: 6). If the conditioning variable is continuous, Brodgar will use 6 graphs. The user can reduce or increase this number. Set this number to 4 if you get an error message related to the layout.
- Overlap conditioning variable. For nominal variables, there is no overlap. For continuous conditioning variables, there will be a certain degree of overlap of the range of this variable (for the different panels). The user can customise the amount of overlap.
- Graph type. This can be JPG, BMP or PNG.
- Name of Graph. Allows the user to enter a name for the saved JPG, BMP, PNG or WMF file. The file will be saved in the project directory.
- Set up of plot. The user can choose from: points only, regression line and points, and smoothing line and points. If smoothing curves are used, please make sure that the explanatory variable along the x-axis has at least 3 different values in each panel.
- If smoothing curve, amount of smoothing. If a smoothing line is used, the user can adjust the amount of smoothing.
- Characters for points. Allows the choice of triangles, circles, etc. instead of points. Each number corresponds to a different type of character.
- Colour of points and lines. Choose from various different options. Each number corresponds to a different colour.
- Colours for bars conditioning variable. Choose from various different options. Each number corresponds to a different colour.
- Title. Allows the user to enter a title for the graph.
- R-code. For further help, start R, and type: ?coplot

4.1.6 Pairs

This tool creates a multi-panel scatterplot. Under the "Settings and labels" tab, the following options are available:

- Graph type. Allows a choice of file type (JPG, BMP or PNG).
- Name of Graph. Specify the name of the JPG, BMP, PNG or WMF file. The file will be saved in the project directory.
- Set up of plot. Allows a choice of: points only, regression line and points, lines only, or smoothing line and points. If a smoothing curve is selected, the amount of smoothing can be changed.
- Alternative pairplot style. Pairplots can either show a scatterplot in all panels (default), scatterplots in all upper and lower diagonal panels and histograms on the diagonal (style 2), or (absolute) correlations in the upper diagonal (the font size is proportional to the correlation) and scatterplots below the diagonal (style 3).
- Title. Allows the user to enter a title for the graph.
- R-code. For further help, start R, and type: ?pairs

4.1.7 Lattice graphs

This is in fact an interface to the xyplot function from the lattice package in R. It creates multi-panel windows. Figure 4.4 shows the various options for lattice plots in Brodgar. In the left panel, the user can select one or more response variables and a continuous explanatory variable. The conditional variables must be nominal.
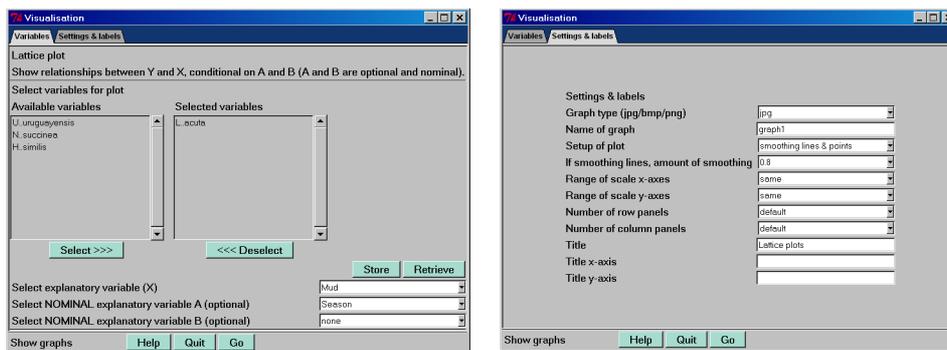
Figure 4.4. Options for Lattice plots. In more recent Brodgar versions, one can also select covariates for the vertical axes.

Under the "Settings and labels" tab, the following options are available:

- Graph type. This can be JPG, BMP or PNG.
- Name of Graph. Specify the name of the graph. The file will be saved in the project directory.
- Set up of plot. Allows a choice of: points only, regression line and points, lines only, or smoothing line and points. If a smoothing curve is selected, the amount of smoothing can be changed.
- Range of scale x-axes. Possible values: same or free. This option allows one to have the same scale along the horizontal axes in each panel of the lattice graph. If the "free" option is selected, the range is allowed to differ.
- Range of scale y-axes. Possible values: same or free. This option allows one to have the same scale along the vertical axes in each panel of the lattice graph. If the "free" option is selected, the range is allowed to differ.
- Number of row panels and number of column panels. The user can choose the number of row and column panels.
- Title. Specify the main title.
- Title x-axis. Specify the label for the horizontal axis.
- Title y-axis. Specify the label for the vertical axis.
- R-code. For further help, start R, load the lattice package from the menu and type: ?xyplot

4.1.8 Scatterplot

This tool makes a scatterplot. Under the "Settings and labels" tab, the following options are available:

- Set up of plot. The options are: points only, regression line and points, and smoothing line and points.
- Thickness of line. Choose between thick and thin lines.
- If smoothing line, add 95% c.i. This option allows one to add 95% pointwise confidence bands around the smoother.
- Different colours and symbols. Select "yes" to plot the samples using different colours and symbols, conditional on a nominal explanatory variable. Samples that have the same value for the nominal variable are plotted with the same symbol and colour.
- Title. Specify the main title.
- Title x-axis. Specify the label for the horizontal axis.
- Title y-axis. Specify the label for the vertical axis.
- Graph type. This can be JPG, BMP or PNG.
- Name of Graph. Specify the name of the graph. The file will be saved in the project directory.
- R-code. For further help, start R, load the lattice package from the menu and type: ?plot

4.1.9 General remarks

To print or copy the graph, click the 'Print/Export' button (or press Control-C, or click on the menu options). Irfanview, a third party graphical software package, will appear and show the graph. In Irfanview, the user can copy the graph to the clipboard (and paste it into Word) or send it to the printer. It is also possible to select another graphical package, see Chapter 2 for further details. Brodgar allows the user to generate WMF or EPS files. These are high quality graphs that can be used for scientific reports and publications. To obtain these graphs, first run the selection exploration tool, and then select the 'Export to WMF' or 'Export to EPS' buttons.

4.2 General tools

In addition to the specific tools already discussed several general tools are available from the Exploration menu, namely (i) Plot data, (ii) Show correlations larger than…, and (iii) Diversity indices. These options are available from the "General tools" tab in Figure 4.1

4.2.1 Plot data

Depending on the structure of the data, this can be a useful tool, and it is essential for time series data. It was added under the "general tools" tab to make it easily available to all users and not just to those working with time series. The following variables can be plotted against observations (time points in time series analysis):

- Response variables (Y versus samples).
- Explanatory variables (X versus samples).
- Response variables and explanatory variables (Y and X versus samples).
- Coenoclines (s=0.5), coenoclines (s=0.3) or coenoclines (s=0.8). Coenoclines are smoothing curves, and the parameter "s" defines the amount of smoothing; 0.8 results in a very smooth curve, whereas 0.3 is less smooth. Along the x-axis, one explanatory variable is plotted, and along the y-axis smoothing curves of all response variables are drawn. Brodgar makes a graph for each explanatory variable. Smoothing curves cannot be drawn along a nominal explanatory variable, hence one should temporary deselect them in the Import data procedure. Go to the "Info Y &X" tab, choose "Select Variable" and de-select the nominal explanatory variables. Once you have finished de-selecting the appropriate variables, make sure you click on the "Finish Data Import Process" button to make the changes available to Brodgar.

4.2.2 Enumerate large correlations

All correlations larger than a certain threshold value are enumerated. Look for correlations larger than 0.9. Some statistical methods will fail if variables with nearly linear cross-correlations are used.

4.2.3 Diversity indices

The following diversity indices are available in Brodgar:

- Total abundance index
- Shannon weaver index (using either log10, ln or the 2log).
- Species richness.
- Simpson index.
- Macintosh index.
- Berger-Parker index.

Figure 4.5 shows the total abundance index for the Argentinean zoobenthic data used in Chapter 4 in Zuur et al. (2007). The order of the samples along the x-axis is determined by the order of the samples in the original data file. This graph is not generated by R. The File and Edit menus allow the user to print, export and copy (to the clipboard) the file as an EMF (high quality) file. The 'Option' menu in Figure 4.5 can be used to define a title and labels for the axes, and the scale along the y-axis.

The values of the diversity index can be obtained via the 'Index function' menu in Figure 4.5; either access them in a text file, or copy them directly to the clipboard, and paste them into Excel or even better, into the spreadsheet in Brodgar itself. To do this, click on: Index function in Figure 4.5 – Copy diversity index to clipboard, close Figure 4.5, and click on: Import data – Change data to be imported – Column – Add variable from clipboard. Brodgar will then ask for a name (use an appropriate name, **but do not use the names Shannon-Weaver log10, Shannon-Weaver ln, Shannon-Weaver log2, Total abundance, Richness, Simpson, Berger-Parker or Macintosh** as these names are used inside Brodgar when univariate methods like linear regression, GLM, GAM, mixed modelling, etc. are applied. If you do, it may try to calculate the richness of your Richness variable.

To compare different indices, it is useful to calculate various indices, copy each to the spreadsheet, normalise and plot all the indices in one graph using the 'Plot Data' function.

In Brodgar, all univariate modelling methods (regression, generalised linear modelling, generalised additive modelling, regression trees, etc.) can easily be applied to these indices.
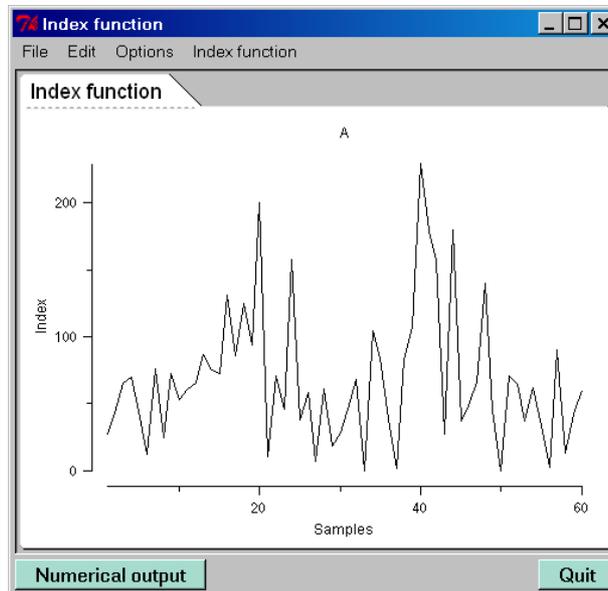
Figure 4.5. Total abundance for Argentinean zoobenthic data.

4.3 Time series tools

One of the demo data sets in Brodgar is a lobster data set, which is used to illustrate exploratory time series tools. This data set consists of catch per unit effort (CPUE) time series of lobsters in 11 areas in the Atlantic Ocean between 1960 and 1999. The data were available on an annual basis. There is one explanatory variable, the North Atlantic Oscillation index (NAO), which is an environmental index function. The time series plot in Figure 4.6 (not produced by R) was created with the following mouse clicks:

1. Import data – Demo data – Load data (fish.brd) – Finish data import process.
2. Exploration – General tools – Go (button for plotting Y versus samples).

There are various settings that can be changed. Close the time series graph window (Figure 4.6), and click on the "Options" button in the "Exploration – General tools" window, or in the "Exploration – Time series tools". The window in Figure 4.7 will appear. In this window, changes can be made to the size of the time series plot (change: Length of x-axis and/or Length of y-axis), and a legend can be added or removed. If a legend is added (Select "yes" for Plot legend in Figure 4.7), clicking on any name in the legend in Figure 4.6, will change the colour of the corresponding time series. It is also possible to use different colours. Other options available in this dialog box are discussed later in this section.
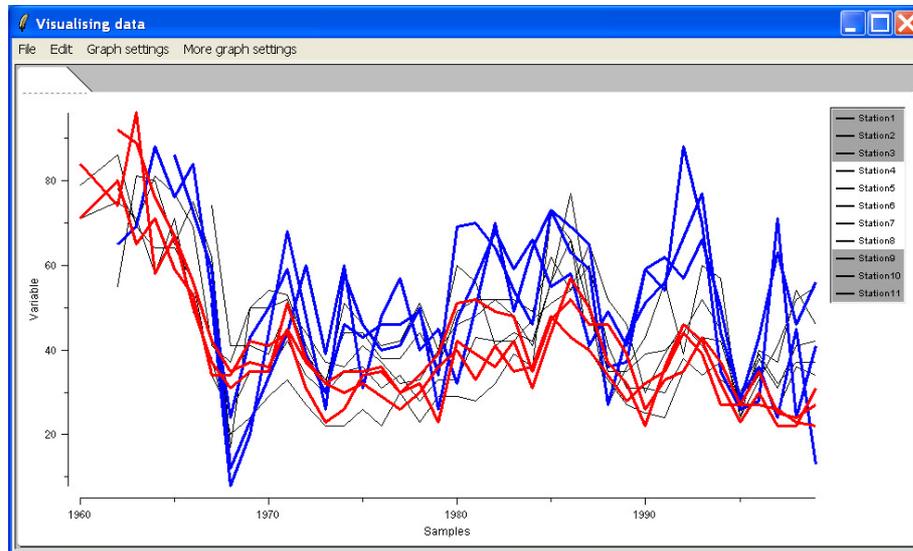
Figure 4.6. Plot of time series. Newer versions of Brodgar allow the user the change the colour of the lines via the menu. To do this, make sure the legend is switched on (Graph settings in Figure 4.6 or Plot legend: yes in Exploration - General tools -Options).

Further tools for time series analysis are available by selecting the Exploration – Time series tools, (Figure 4.8). Under this tab a range of specialised exploratory time series analysis tools are available including auto-correlations and cross-correlations, combined with metric multi-dimensional scaling (to visualise the correlation matrix).
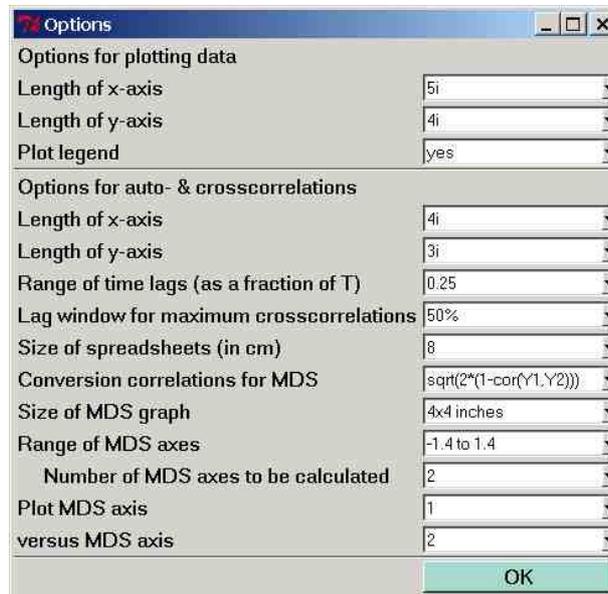
Figure 4.7. Options.
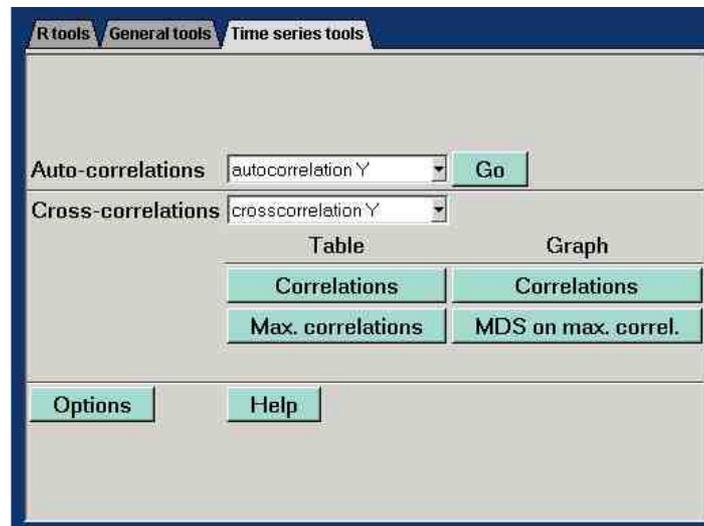
Figure 4.8. Tools available under Exploration –Time series tools.

Auto-and cross-correlations

Clicking the auto-correlation's "Go" button (Figure 4.8), creates eleven graphs for the CPUE Nephrops data, one for each sampling station (Figure 4.9). Each window shows the auto-correlations of a time series. Tab 1 corresponds to the auto-correlation function of the first time series, tab 2 to the second time series, etc. (use the scrollbar at the bottom of the graph to see other tabs). Dotted lines in each figure represent 95% confidence intervals (Zuur et al. 2007, Diggle 1990).

Brodgar can also calculate cross-correlations of Y, X or Y, and X. Click the button labelled "Correlations" under Table in Figure 4.8 to estimated the cross-correlations. Results are presented in a spreadsheet, see Figure 4.10. Estimated cross-correlations in a grey (non-white) cell are significantly different from 0 at the 5% significance level (Diggle 1990). Estimated auto- and cross-correlations can be accessed in various ascii files in the project directory (mycorrel.txt, correlmtx.txt, mdsmaxcorrel.txt).

Cross-correlations can be plotted versus time lags, see Figure 4.11. This figure was obtained by clicking the "Correlations" button under Graph in Figure 4.8.
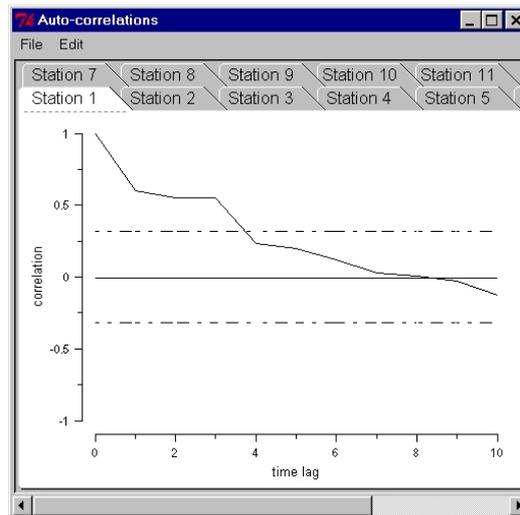
Figure 4.9. Auto-correlations for the CPUE series.
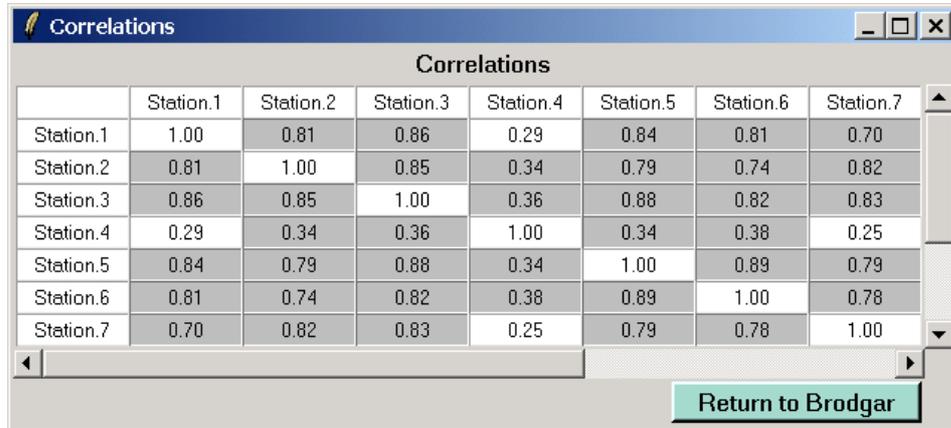
Figure 4.10. Cross-correlations for the CPUE series. Numbers in a grey cell are significantly different at the 5% level.
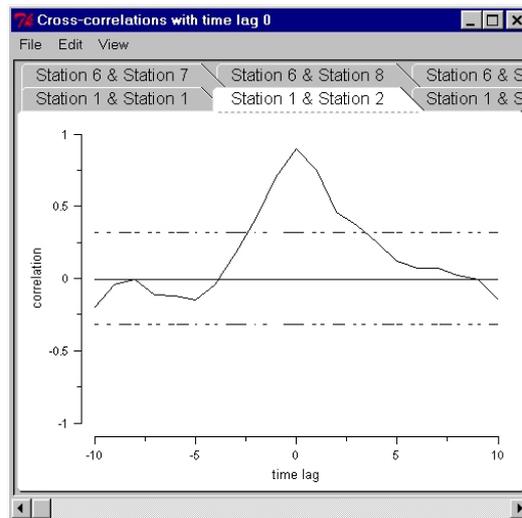
Figure 4.11. Cross-correlations plotted versus time lags for the CPUE series.

Brodgar also estimates the maximum cross-correlation between $Z1(t+k)$ and $Z2(t)$; where $Z1$ and $Z2$ are two variables (time series) and t represents time. The maximum is taken with respect to the time lag k, where k takes values between -L and L. For each combination of variables, the maximum value and the corresponding lag are estimated. These can be viewed in a table by clicking the button labelled "Max. correlation" under the Table column (Figure 4.8), see Figure 4.12.

Results in this window show that the correlation between station 4 and station 7 have the highest value in absolute sense if a time lag of k=-2 is chosen. To enhance visual inspection of this table, metric multidimensional scaling can be applied by clicking the "MDS on max. correl." button, see Figure 4.13. Due to edge effects there might be small differences between the cross-correlations of Z(t+k) and Z2(t) versus Z2(t+k) and Z1(t). As a result the cross-correlation matrix might not be symmetric. The matrix containing the correlations is made symmetric via: A+A' / 2, and the correlations are converted to dissimilarity coefficients. The default conversion method is taken from Krzanowski (1988):

$$d_{1,2} = \sqrt{2(1 - cor(Z_1, Z_2))}$$

This conversion method classifies two variables having a high positive correlation as "very similar" and two variables having a high negative correlation as "not similar". The later conversion might not be appropriate for all data sets. For this reason, we added the following conversion method:

$$d_{1,2} = \sqrt{2(1 - |cor(Z_1, Z_2)|)}$$

where | | is the absolute value. Now, positive and negative correlations are handled in the same way. Two alternative conversion methods available in Brodgar are:

$$d_{1,2} = 1 - cor(Z_1, Z_2) \quad \text{and} \quad d_{1,2} = 1 - |cor(Z_1, Z_2)|$$

The conversion method can be changed via the "Options" button in Figure 4.8, see Figure 4.7.

Maximum cross-correlations over h between Y1(t) and Y2(t+h)

Maximum cross-correlations

	Station.1	Station.2	Station.3	Station.4	Station.5	Station.6	Station.7
Station.1	0	0.81	0.86	0.32	0.84	0.81	0.70
Station.2	0	0	0.85	0.40	0.79	0.74	0.82
Station.3	0	0	0	0.38	0.88	0.82	0.83
Station.4	-1	-4	-1	0	0.42	0.46	0.47
Station.5	0	0	0	1	0	0.89	0.79
Station.6	0	0	0	2	0	0	0.78
Station.7	0	0	0	2	0	0	0

Info   Return to Brodgar

Figure 4.12. Maximum cross-correlations over time lags. Correlations in a grey cell are significantly different at the 5% level.
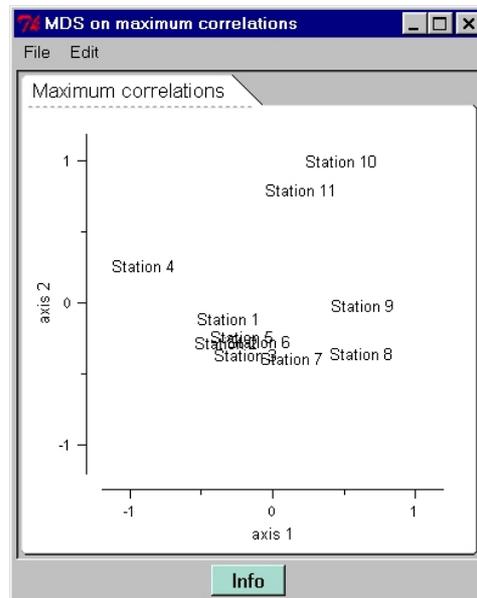
Figure 4.13. MDS applied on maximum cross-correlations. The 'Info' button gives the stress.

Obviously, the exact interpretation of the ordination diagram in Figure 4.13 depends on the chosen conversion method. For the default conversion method, stations close to each other correspond to stations that are positively correlated with each other.

The graphs can be printed, exported and imported into Microsoft Word in the same way as for time series plots. Please note that the mathematical calculations are not repeated unless you leave the Data exploration menu. Various general settings can be changed via the "Options" button, e.g.:

- Size of graphs. Changing the length of the x-axis and the length of the y-axis increases the size of the window in Figure 4.6. The MDS graphs can be enlarged by changing 4x4 inches to 5x5 or 6x6. The same holds for the auto- and cross-correlation graphs.
- The value of L. It increases or decreases the range of k (time lag) for which cross-correlations are calculated. By default L is set to 40% of the length of the time series.
- Lag window for maximum cross-correlations. Set this value to a relative low percentage (e.g. 20%) if you only want to look at maximum cross-correlations with time lags from –k, ..,-1, 0, 1, ..k, for small values of k, say 2.
- The way correlation coefficients are transformed to measures of dissimilarity.

- Size of the spreadsheet containing the correlations (change Size of spreadsheets to 9 or 10 cm).

4.4 Examples from Zuur et al. (2007)

In this section, we discuss how to reproduce some of the graphs in Zuur et al. (2007).

4.4.1 Figure 21.2 in Zuur et al. (2007)

Download the *Solea solea* data from www.highstat.com/Soleasolea.xls. Import the data (the first row and column contain labels) and select all variables as explanatory variables, except for the column Soleasolea; this is the response variable.

Click Exploration and select "Pairs" in Figure 4.1; click on the Go button. Select the variables gravel, large sand, medium fine sand and mud content. To get the Pearson correlation coefficients in the lower diagonal panels, click "Settings & labels", and select "Style 3" from "Alternative pairplot style". You should now get Figure 21.2 in Zuur et al. (2007). Figure 21.3 can be created in the same way.

4.4.2 Figure 21.4 in Zuur et al. (2007)

Download the *Solea solea* data from www.highstat.com/Soleasolea.xls. Import the data (the first row and column contain labels) and select all variables as explanatory variables, except for the column Soleasolea; this is the response variable.

Click Univariate, and select the tab "General methods". Select "Design plot" and click on the Go button. Select the categorical variables season, month and area, and clicking the Go button gives Figure 21.4. You should only use categorical variables in this tool!

4.4.3 Figure 22.3 in Zuur et al. (2007)

Download the bee data from www.highstat.com/Bees.xls. Import the data (the first row contains labels and the first column contains data), and select all variables as explanatory variables, except for the column Bees; this is the response variable. Go to Exploration, select coplot in Figure 4.1 and click Go.

Figure 4.14 shows how we selected the variables. Note the option "yes" for "Nominal?" To get a smoother, click "Settings & labels", and select "smoothing lines & points" under "Setup of plot". We used a span width of 0.5 (amount of smoothing).
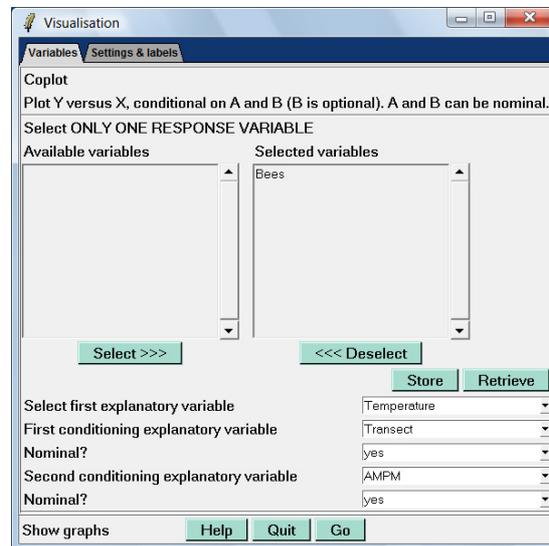
Figure 4.14. Settings for creating the coplot in Figure 22.3 in Zuur et al. (2007).

4.4.4 Figure 33.1 in Zuur et al. (2007)

Download and import the Portuguese fisheries landings data from www.highstat.com/Algarve1982-1999.xls. Do not highlight the first two rows! The first row and column contain labels, and select the second column Year as explanatory variable. Note that the first column labelled Year is used as a label. To select the standardisation, go to "Info Y & X" (under Import), and select "normalisation" for "Standardisation for Y". This will normalise (standardise) all variables. If you only want to standardise a few particular series, then do not select a standardisation at this stage, but do it under "Change Data – variables and Transformation".

Go to Exploration, select Lattice in Figure 4.1 and click Go. Select all species. Under "Settings & labels", select the option "lines only" for "Setup of plot". Else it will plot points with a smoother.

4.4.5 Figure 16.1 in Zuur et al. (2007)

Download and import the Squid data from www.highstat.com/Squid.xls. The first row and column contain labels. Select GSI as response variable and all other variables as explanatory variables. Ensure that no standardisation is selected (just in case you do just imported the Portuguese fisheries landings data).

Go to Exploration, and select "Conditional boxplots" in Figure 4.1, and click the Go button. Double click on GSI and use Month as "First nominal explanatory variable". Clicking the Go button gives Figure 16.1 in Zuur et al. (2007).

4.4.6 Figure 4.7 in Zuur et al. (2007)

Download the wedge clam data from www.highstat.com/WedgeclamII.xls. The first row and column contain labels. Select AFD as response variable and month and length as explanatory variables. Click "Change data to be imported", and click "Variables and Transformations". Apply a log transformation on length and AFD, see Figure 4.15. Click on "Continue" and then on "Save changes and Finish Import Data Process". Go to Exploration, and select "Scatterplot" in Figure 4.1. Select AFD and length, and click the Go button. To add a line, go to "Settings and labels" in the previous step, and select "regression line and points" for "Setup of plot". To remove the outlier, go to Import and then to "Change data to be imported". Scroll to row 108, put the mouse on the label 108, right-mouse click, click "Delete row(s)" – "Yes" – "Continue" and do not forget to click "Save Changes and Finish Data Import Process".

Variables	Response (Y)	Explanatory (X)	Y transformation	Y standardization	X transformation	X standardization
MONTH	No	Yes				
LENGTH	No	Yes			log(X)	
AFD	Yes	No	log(Y)			
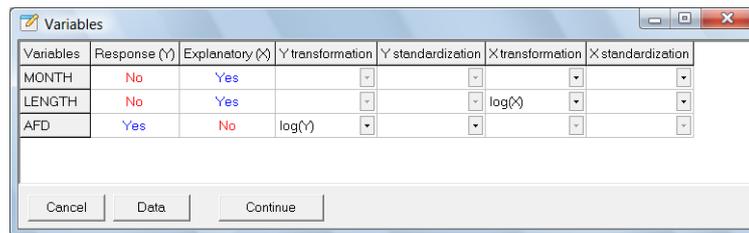
Figure 4.15. Logarithmic transformation applied on the AFD and length data for the wedge clam data.

4.4.7 Figure 4.8 in Zuur et al. (2007)

Download the decapod data from www.highstat.com/Decapod.xls. The first row and column contain labels. Select the columns B to M as response variables, and N – T as explanatory variables. Click on Data exploration – General Tools, and in the diversity index box select Richness, and press on the Go button. The diversity function will be displayed in a graph. From the menu of this window, click on Index Function – Copy diversity index to clipboard. Now close the figure, and go to: Import data – Change Data to be Imported – Column – Add variable from clipboard. Brodgar will ask for a name (**do not use Richness as name!**). Type in a name, and click on Continue – Save Changes and Finish Data Import Process. Go to Exploration and in Figure 4.1, select "Pairs". Select the diversity index, and also click T1m, S1m, S45_35m and Ch0_10m. Clicking the Go button gives Figure 4.8 in Zuur et al. (2007).