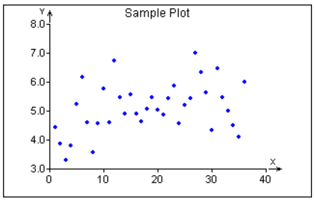
Plotting

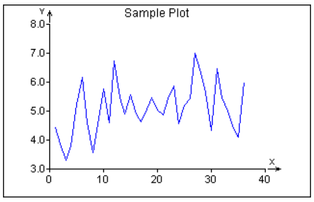
Menu:	QCExpert	Plotting

Module graphs offers various tools for visualization of uni- and multivariate data. Settings and options in different types of graphs allow for modifications and customizations of the graphs. In the following table is a brief graphical overview of the possible graphs. Details of each graph type will be dealt with in the following paragraphs.

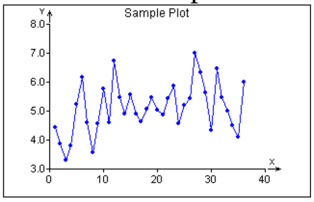
Points



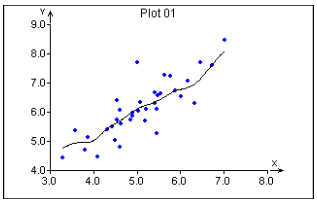
Lines



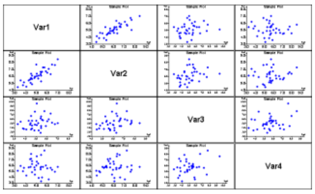
Connected points



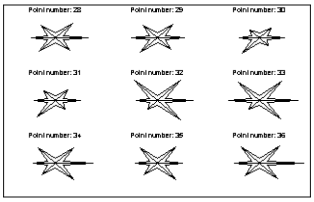
X-Y Scatter



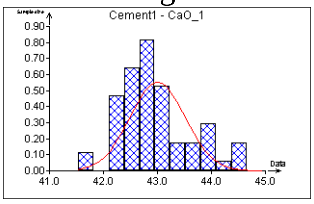
X-Y Matrix



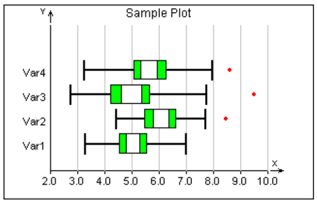
Star Plot



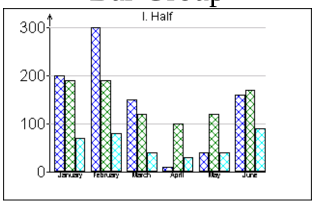
Histogram



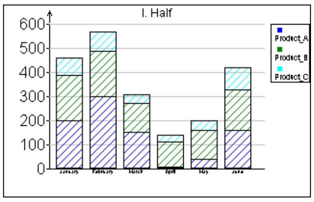
Box Plot



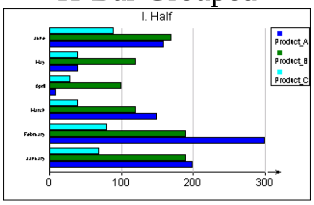
Bar Group



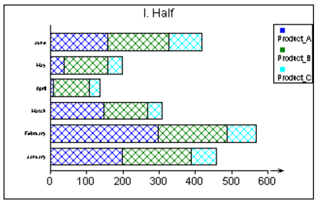
Bar Stacked



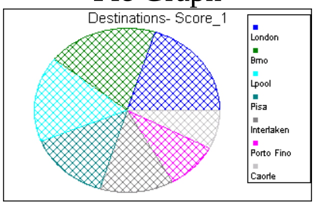
H-Bar Grouped



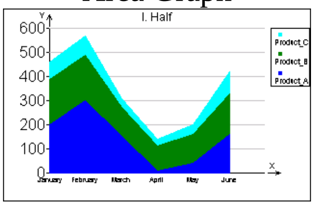
H-Bar Stacked



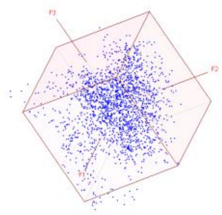
Pie Graph



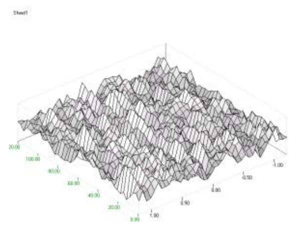
Area Graph



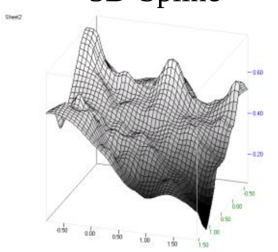
3D-Point Plot



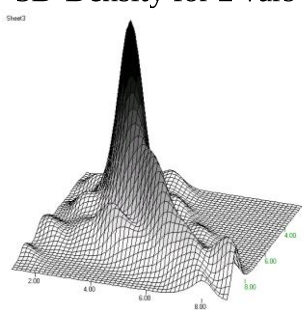
3D-Surface Plot



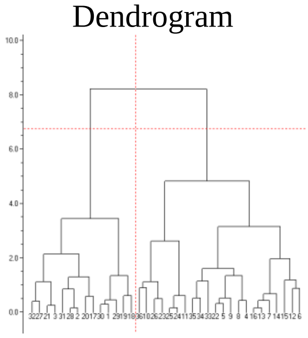
3D-Spline



3D-Density for 2 vars



Dendrogram



Data and Parameters

Input data are expected in one or more columns in current data sheet. By selecting the Graphs menu item, the Graphs dialog window will be opened. The header may be written in the field *Name of graph*. Default header is taken from the name of the respective data sheet. Description of axes may be specified in *X-Label* and *Y-Label* fields. The required type of graph is selected in the drop-down list *Type of graph*. Options and settings are specific to every graph type and will be described in detail for each graph in the respective paragraphs below. The field *Columns* specifies which columns of the data sheet will be used, the group *Data* in the dialog window allows to specify rows to be used for the graph using marking the data. If the checkbox New sheet is checked, every new graph will plot on new graph sheet, otherwise last graph will be overwritten each time a new graph is generated, unless the dialog window Graphs is closed. The pushbutton *Apply* is used to create a graph and leave the dialog box open for further graphs, while the *OK* button will create graph and close the dialog box.

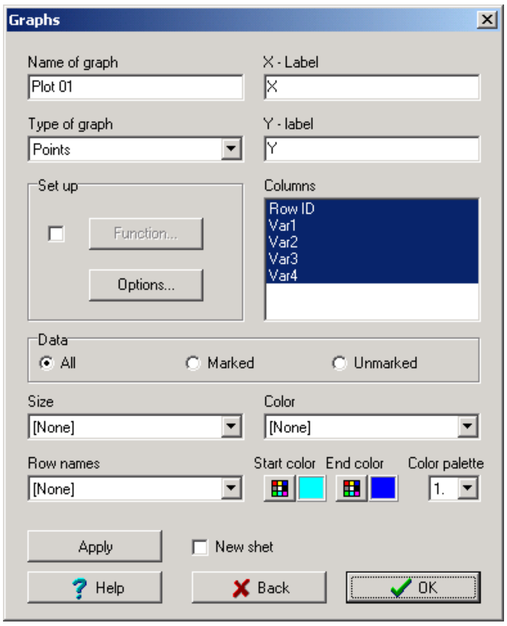


Fig. 1 Dialog window for the module Graphs

The module *Graphs* does not generate any output to *Protocol* with the exception of Dendrogram.

Graph types

Points plot

Creates a plot of data in the form of individual points, on the X-axis is the ordinal index and on the Y-axis is the value, see illustration A. The *Function* button enables to add a curve of a specified function. The function may be specified in the dialog box *Function* (Fig. 2) in field *Y*= in the form *f*(*x*). using common functions and mathematical notation, e.g.: **`3.5+0.05*x`**. Alternatively, a spline may be selected to fit the data. Spline will produce a non-parametric kernel smoothing using specified smoothing parameter. Function will always be plotter only once per plot, while spline will be calculated and plotted for each data column. *No of points* specifies how many intervals will be used to plot the curve of function/spline. The style of plotting the function/spline is affected by selection in the group *Plot*: Points will plot the function value just in the data points, lines will plot the curve. *Color* in the *Function-Setup* dialog window will specify the color of the line. The button *Options* opens *Options* dialog window, Checking the checkbox Legend will add legend in the plot for identifying columns, see illustration C.

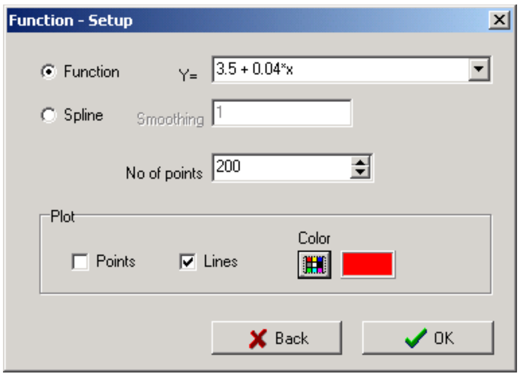
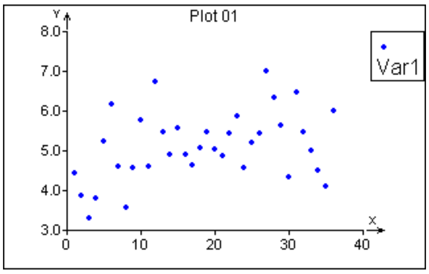


Fig. 2 Function setup in module Graphs

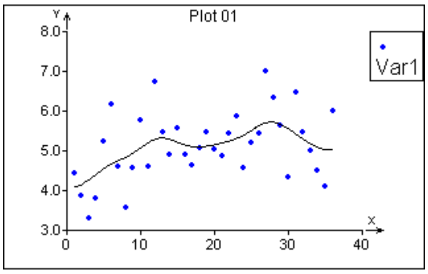
Fig. 3 Legend setup in Points plot

In the Points plot it is possible to select further two columns in *Size* and *Color* drop-down fields to specify size and color of the plotted points. If a column is specified in *Size*, then diameter of each point is determined linearly by the values in the specified column within the range of this column. Similarly, if a column is specified in *Color*, then the color of each point is determined linearly between selected colors in the color space by the values in the specified column within the range of this column. Using color and size allows to visualize up to four dimensions on one 2D plot, specially in the XY-Scatter plot, see below. Number of data in all selected columns must be the same for the plot to work properly. Points plot may be combined with function (illustration D). If a column with row names is selected, these names may be used to describe the points by clicking on them with mouse in interactive plot (after double-clicking on the plot), see illustration E.

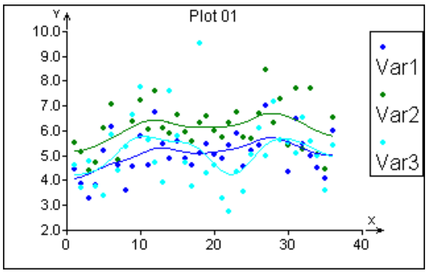
Illustrations



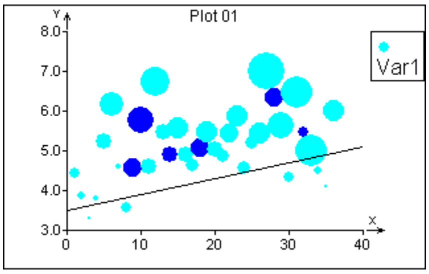
A



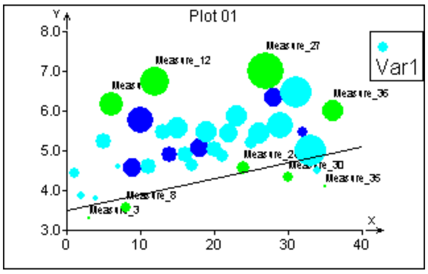
B



C



D

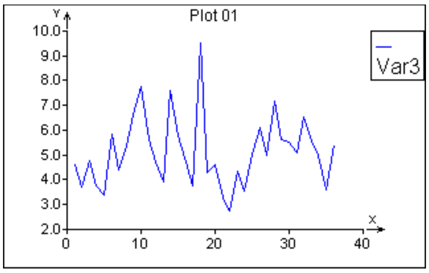


E

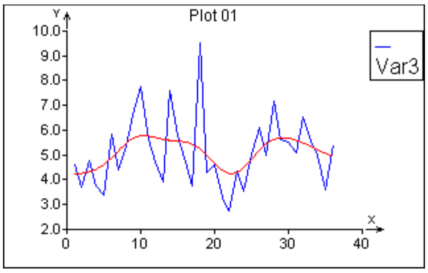
Lines plot

Creates a plot of data in the form of connected line segments, on the X-axis is the ordinal index and on the Y-axis is the value, see illustration A. The *Function* button enables to add a curve of a specified function. The function may be specified in the dialog box *Function* (Fig. 2) in field *Y=* in the form *f*(*x*). using common functions and mathematical notation, e.g.: **`0.5+3*sin(x/6)+0.05*x`**. Alternatively, a spline may be selected to fit the data, illustration B. Spline will produce a non-parametric kernel smoothing using specified smoothing parameter. Function will always be plotter only once per plot, while spline will be calculated and plotted for each data column. *No of points* specifies how many intervals will be used to plot the curve of function/spline. The style of plotting the function/spline is affected by selection in the group *Plot*: Points will plot the function value just in the data points, lines will plot the curve. *Color* in the *Function-Setup* dialog window will specify the color of the line. The button *Options* opens *Options* dialog window, Checking the checkbox *Legend* will add legend in the plot for identifying columns, see illustration C. Columns in Color and Size drop-down menus will affect the points (crosses) of points plotted on selected function or spline, illustration D.

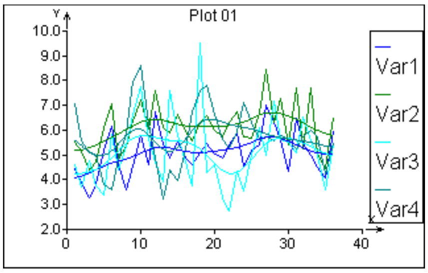
Illustration



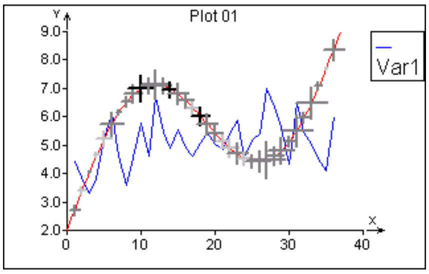
A



B



C



D

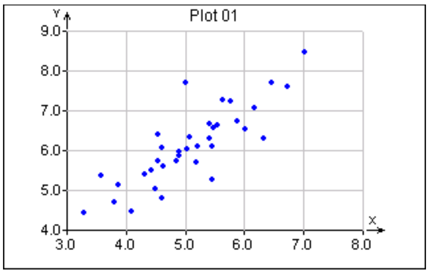
Connected points

Connected points plot is a combination of the two preceding plots and its use is the same.

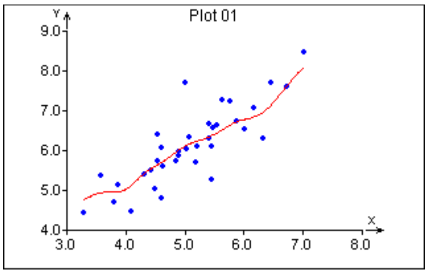
X-Y Scatter plot

X-Y Scatter plot displays two variables in one plot. Two columns must be selected in the field *Columns*. Data in the first column will be plotted on the X axis, second column on the Y axis, see illustration A. Controls of this plot is similar to the Points plot as described above. When combined with other two columns to control color and size of the plotted points, it is possible to visualize four dimensions in one plot, see illustration C, where some correlation between color and size can be spotted. As in the Points plot, the plotted data may be fitted with kernel smoothing curve (illustration B) or a user function can be added. Double clicking on the plot will allow to zoom in or to select and label separate points with row names specified in *Row names* drop-down field.

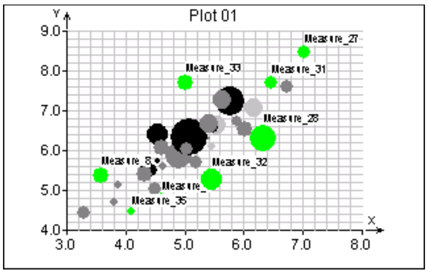
Illustrations



A



B

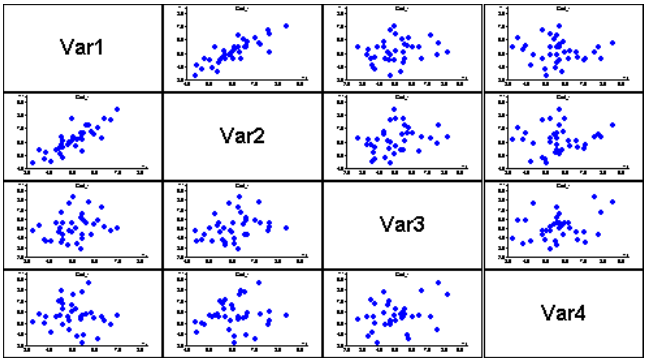


C

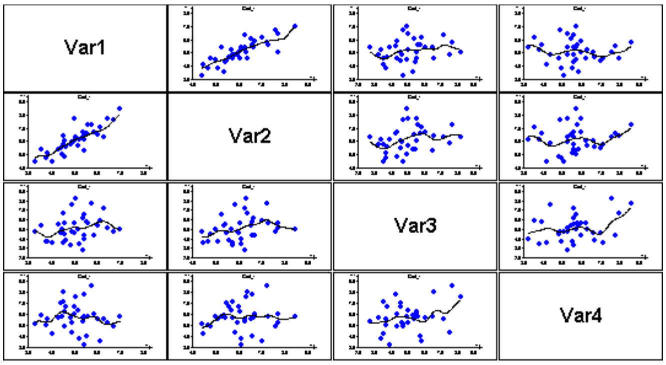
X-Y Matrix Plot

This plot is in fact a generalization of the previous X-Y Scatter plot and an alternative of the graphical output of the Correlation module. It plots scatter plots for all pairs of the selected columns, illustration A. Input data must have at least two columns. Like in the case of X-Y Scatter plot, smoothing curve or a user function may be displayed in each graph using *Function* button, illustration

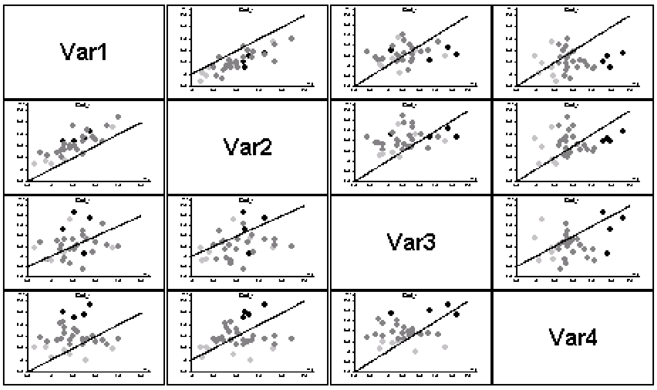
B, C. Other two columns may be used to control size and color on the plots as shown on illustrations C and D.



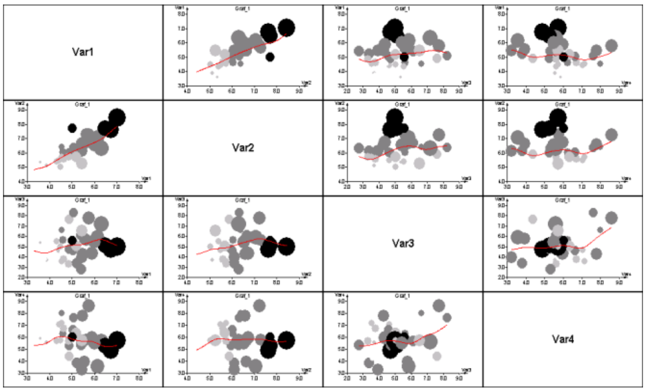
A



B



C



D

Stars plot

Stars plot is a standard tool to visualize multidimensional data. Input data are two or more columns selected in the *Columns* drop-down field. In this plot every data point represented by one row is displayed as one star. Each tip of a star represents one value in the corresponding row. Long tip means big value. This plot is used to review multidimensional data and to find similar or dissimilar values. Different behavior of data can be seen on the last two stars in illustration E. Number of plotted stars corresponds to the number of rows in selected columns. The number of stars to be displayed in one graph (1 to 64) can be set in *Options* window, see illustrations. Checking *Legend* will display the meaning of the pins, illustration B. Checking Grid in *Options* will add a radius in each tip, compare illustrations A and D.

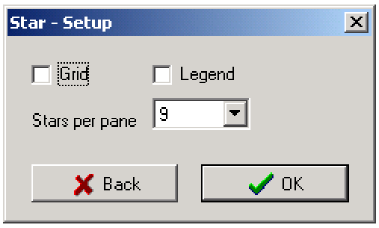
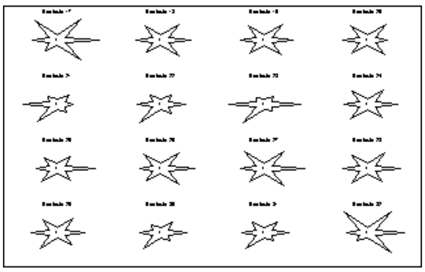
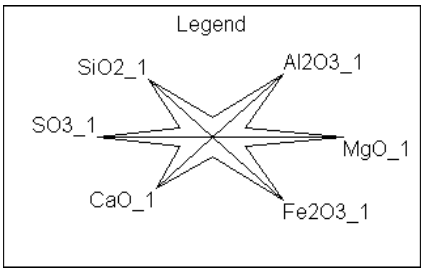


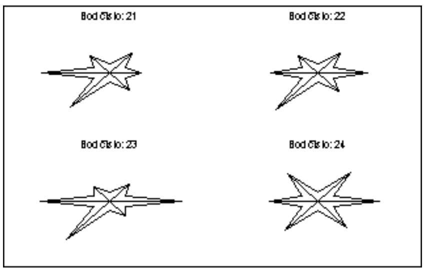
Fig. 4 Options for Star plot



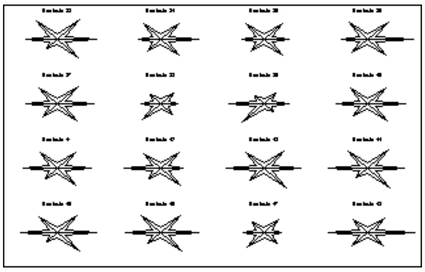
A



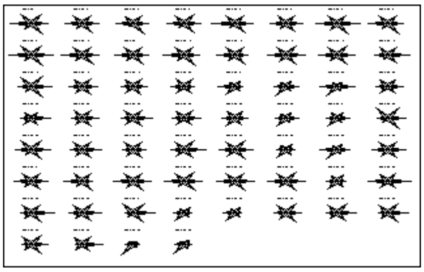
B



C



D



E

Histogram

Draws histograms for all selected columns. By pressing *Options* button it is possible to set properties common to all histograms to be plotted. *Fixed Class width* will plot classical histogram with all bars of the same width, the height of the bars is the count of data in the class, illustrations A, B, C, D. Variable class width will generate variable class histogram, where there is constant number of data in each class (bar). This type of histogram has unit area and has usually higher information value. On illustration A-E histograms are constructed for the same data, only the variable class histogram (E) clearly suggests possible bi-modal distribution. For fixed class, user can specify start and class width, these two values are used for all selected columns and can be used to compare several data samples. Checking *Gauss curve* will generate a normal distribution probability density curve over the histogram, histogram is transformed to the scale of the probability density curve, so that the height of the bars is rather probability density than count. In the Fill field one of three filling grids may be selected.

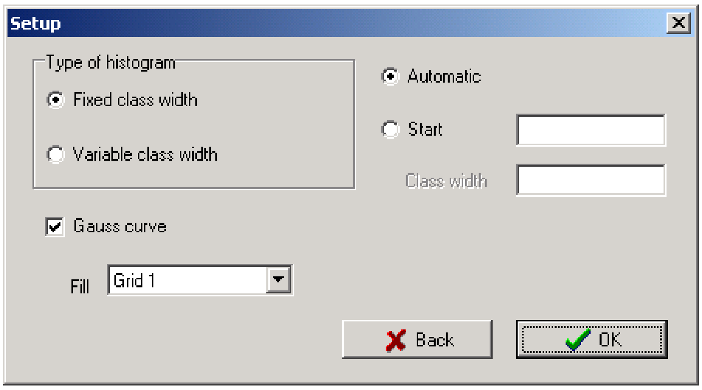
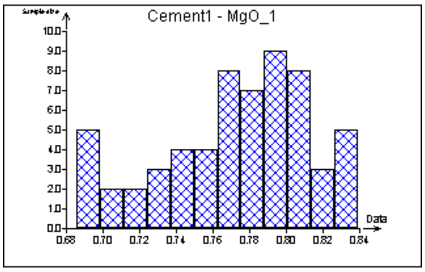
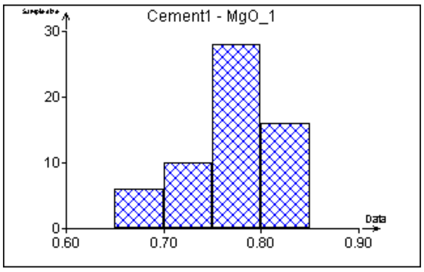


Fig. 5 Setup dialog window for Histogram

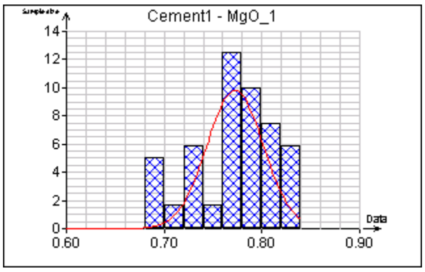
On illustrations A-E there are histograms for the same data with various settings: Automatic class width (A), too wide class (B) and too narrow class (D). Variable class width (E) reveals apparent (yet not statistically confirmed) bimodality of data at 0.77 and 0.79, which is not obvious from any of the previous histograms.



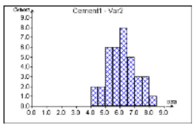
A

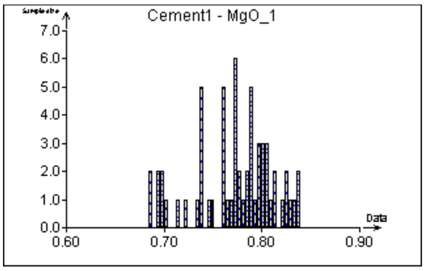


B

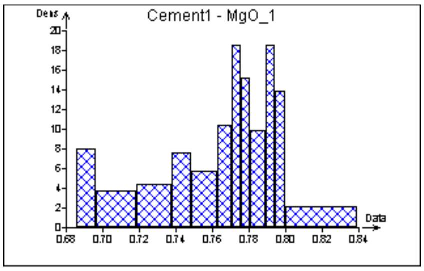


C

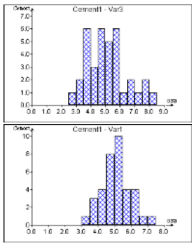




D



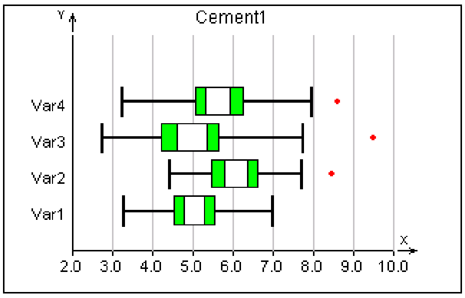
E



F

Box plot

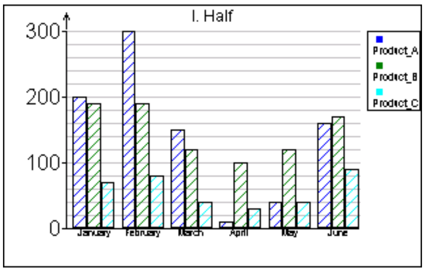
This plot draws box plots for all selected columns, see illustration. Box plot is a standard diagnostic tool used to assess symmetry of data and presence of outliers. The large box contains 50% of the data, its upper edge corresponds to 75th percentile, its lower edge to the 25th percentile. Median is located in the middle of the white rectangle inside the green box. Width of the white rectangle inside the green box corresponds to the width of the confidence interval for the median. Two black lines correspond to the inner fence. The data points outside the inner fence are marked red. They might be considered as outliers. This plot is also generated by ANOVA.



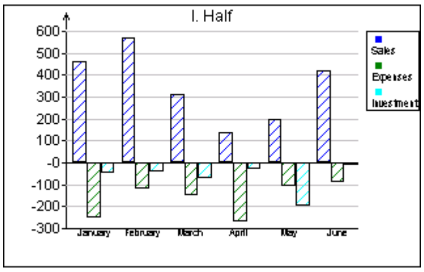
A

Vertical and Horizontal Bar plots

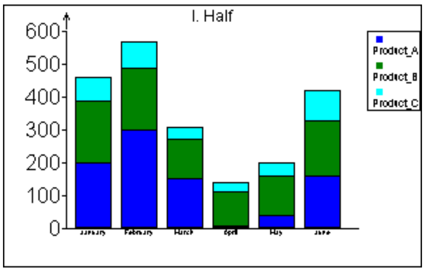
These plots are to visualize data for different classes. Data are expected in one or more columns, one column may contain row names, which should be specified in the *Row names* field. Button *Options* will allow to display a legend on the graph and to select fill type for the bars. Grouped bars will construct a separate bar for each row and each column, illustration A, B, E, F. Bars start from zero and values for bars may contain negative numbers, which plot on the negative part of axis, illustrations B, F, H. Columns are distinguished by color. On Stacked bars, the values form all columns are added into one bar for each row, illustrations C, D, G, H.



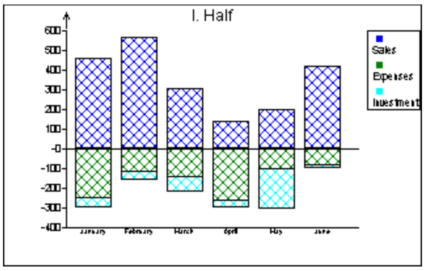
A



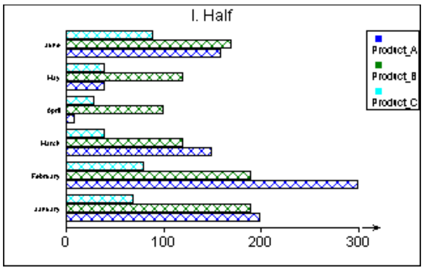
B



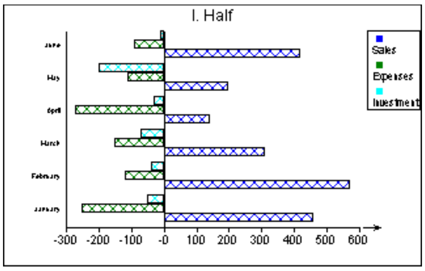
C



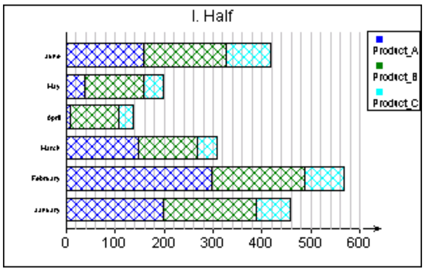
D



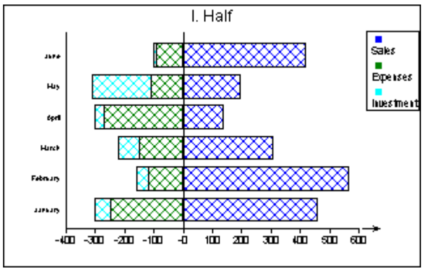
E



F



G



H

Pie chart

Pie chart is used to visualize parts of a defined unity as fractions of a circle. Data may be in one or more columns and must be non-negative (negative data are taken as zeros). One pie chart is constructed for each selected column. Row names if specified are used in legend on the plot, see illustration A, C. If no row names are specified, the legend shows the absolute data values, illustration B. The checkbox *Pool others* in *Options* dialog box will cause the specified fraction of values in columns to form one section on the chart named „Others“, illustration E. If the checkbox *Sort* is checked, the data is sorted in ascending order, illustration D. Fill type for the bars selected.

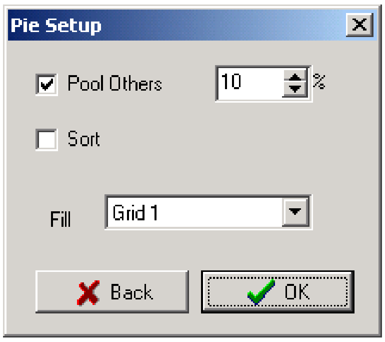
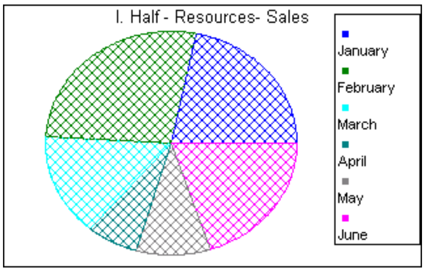
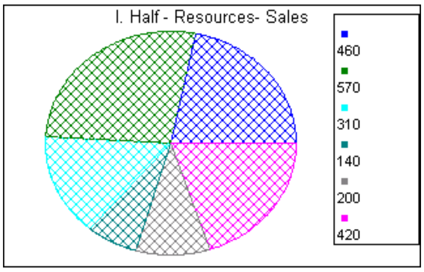


Fig. 6 Options for the Pie chart plot

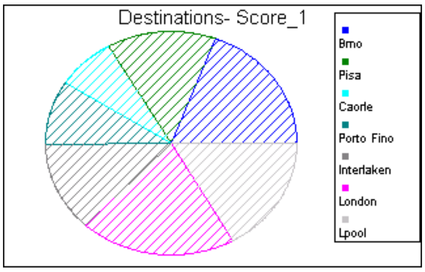
Illustrations



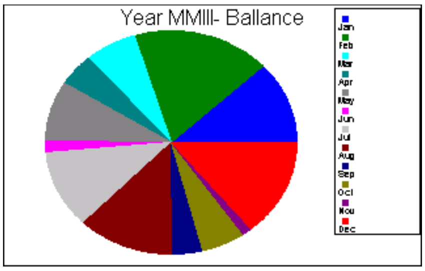
A



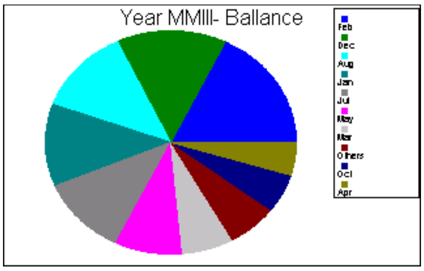
B



C



D

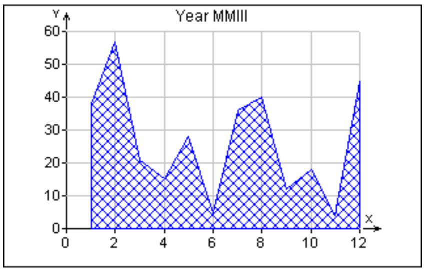


E

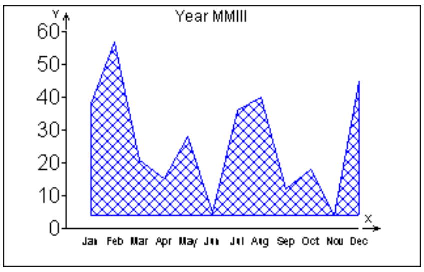
Area plot

Area plot has similar use as the bar plot. Values in the selected columns are plotted on the Y axis. Basic form of the area plot is shown on illustration A. In the case of high values with low variability it may be more suitable to use *From minimum* option instead of the standard *From zero*, see illustration B. If more columns are selected, the areas may be added together by checking *Add* checkbox (illustration C). Negative data can be plotted only on the non-added plot, illustrations E, F. In the added plot, negative values are taken as zeros. Fill type for the bars selected in the options dialog window.

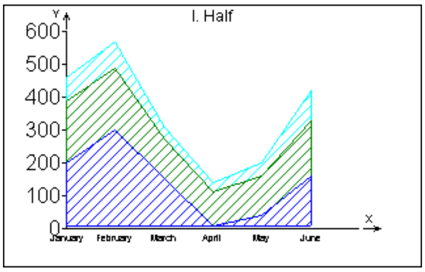
Fig. 7 Options for the Area plot



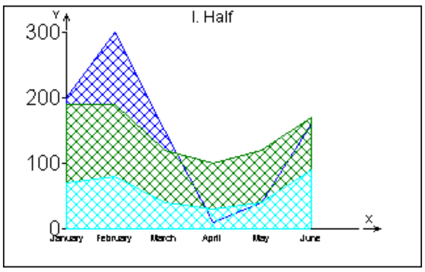
A



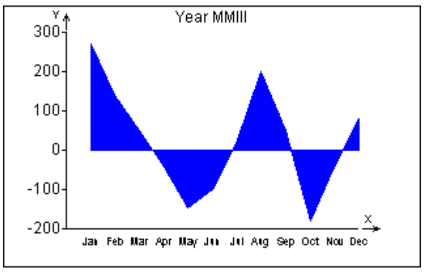
B



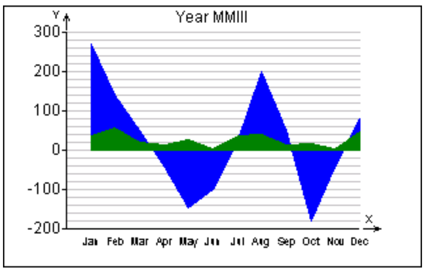
C



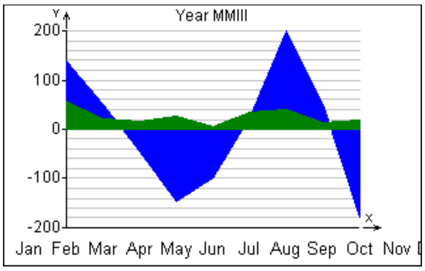
D



E



F



G

3D-point plot

Plots 3 selected columns in a 3d-point plot. This type of a plot is especially suitable for assessing structure, properties relationships or homogeneity of a 3-dimensional data. The 3d-plot can reveal relationships which cannot be observed at 2d scatter plots, pair correlations, etc. 3d-point plot is an extension of the 2-scatter plot. The plot can be rotated, continuously scaled and moved in the graphical window. An individual point can be labeled with a mouse click either by its row number or by a column selected in the *Graphs* dialog window. The scale of the three axes is set so that the values appear normalized. Actual non-normalized scale can be set by the checkbox *Isometric Axes*.

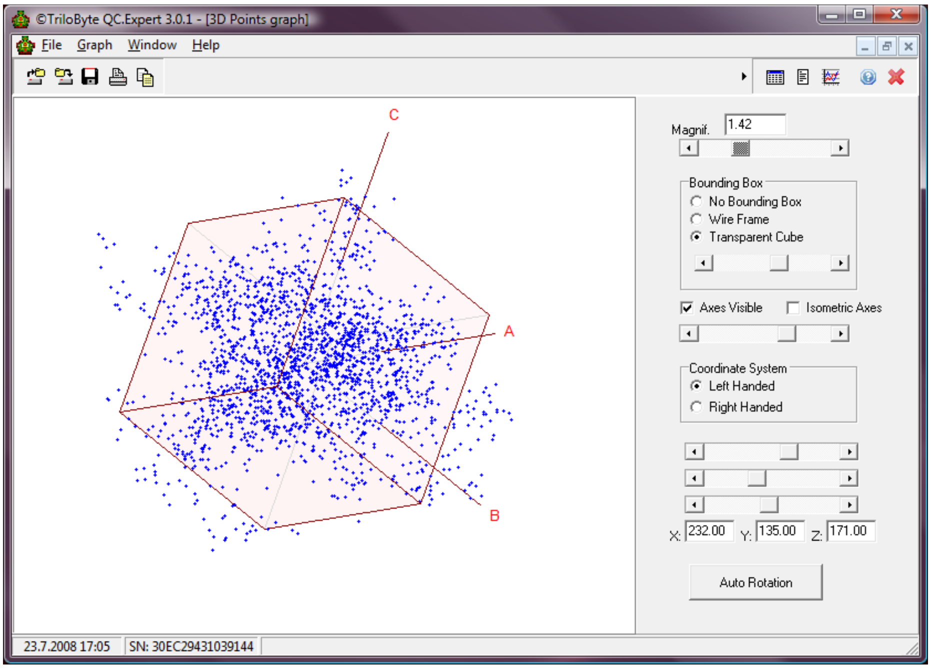
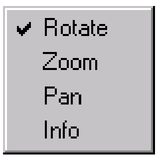


Fig. 8 3D-Point dynamic plot dialog window

To display the 3d-point Plot, first select the graph type and three data columns (see Fig. 1 on page 2), then press OK or Apply. Non-contiguous columns are selected with Ctrl-mouse click. Primarily, the plot is controlled by a mouse. It can be rotated, scaled and moved. Use right mouse-click to select the mouse function:



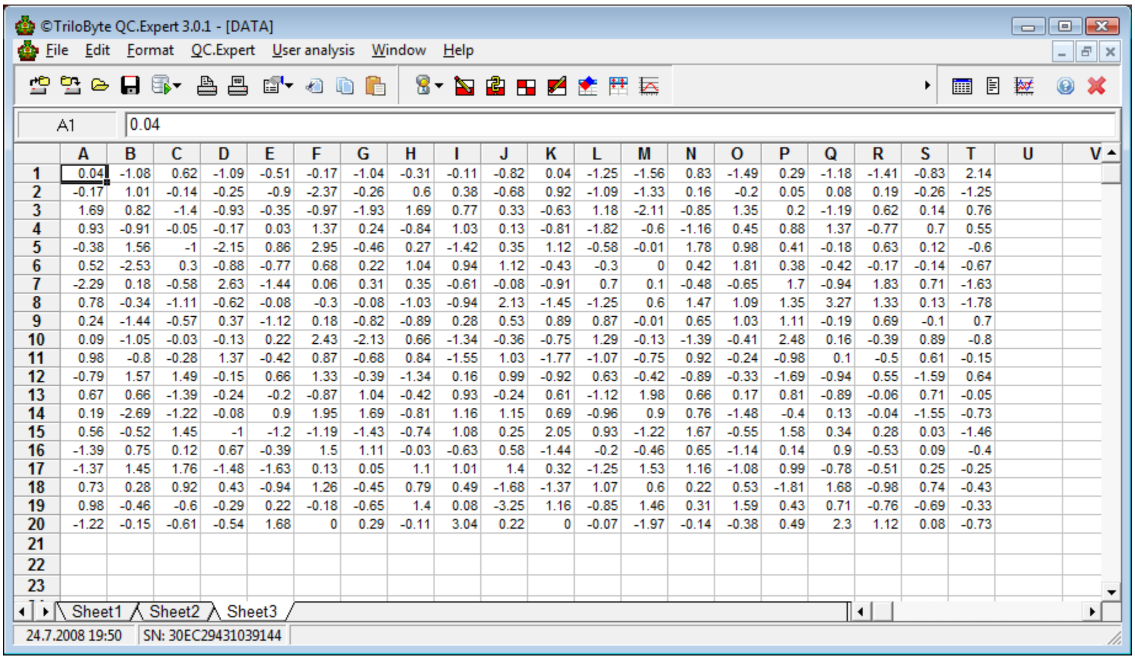
Rotate – rotate the plot with mouse in any direction to explore the data and observe shapes and unusual features in data, like outliers, clusters, non-linearity, correlations.
Zoom – move the mouse up or down to move the plot closer or farther.
Pan – move the plot with a mouse.

Info – Click the mouse near a point in the plot to display its row number or (if selected previously in the *Graphs* dialog window as *Row names*, see Fig. 1 on page 2) its description. This function is applicable only to one point at a time.

The right part of the plot window offers controls for the plot including setting the scale, boxing style, axes visibility, isometry (real scale of the data), left- or right-handed orientation. The Auto Rotation button will rotate the plot automatically in random directions. From the main menu you can copy, save or print the plot.

3D-surface plot

This plot will display an *n* x *m* data matrix as a 3d-plot. Rows and columns are used as X and Y coordinates, the values in the matrix are plotted on the Z axis. Following examples illustrate the use of the plot.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	0.04	-1.08	0.62	-1.09	-0.51	-0.17	-1.04	-0.31	-0.11	-0.82	0.04	-1.25	-1.56	0.83	-1.49	0.29	-1.18	-1.41	-0.83	2.14
2	-0.17	1.01	-0.14	-0.25	-0.9	-2.37	-0.26	0.6	0.38	-0.68	0.92	-1.09	-1.33	0.16	-0.2	0.05	0.08	0.19	-0.26	-1.25
3	1.69	0.82	-1.4	-0.93	-0.35	-0.97	-1.93	1.69	0.77	0.33	-0.63	1.18	-2.11	-0.85	1.35	0.2	-1.19	0.62	0.14	0.76
4	0.93	-0.91	-0.05	-0.17	0.03	1.37	0.24	-0.84	1.03	0.13	-0.81	-1.82	-0.6	-1.16	0.45	0.88	1.37	-0.77	0.7	0.55
5	-0.38	1.56	-1	-2.15	0.86	2.95	-0.46	0.27	-1.42	0.35	1.12	-0.58	-0.01	1.78	0.98	0.41	-0.18	0.63	0.12	-0.6
6	0.52	-2.53	0.3	-0.88	-0.77	0.68	0.22	1.04	0.94	1.12	-0.43	-0.3	0	0.42	1.81	0.38	-0.42	-0.17	-0.14	-0.67
7	-2.29	0.18	-0.58	2.63	-1.44	0.06	0.31	0.35	-0.61	-0.08	-0.91	0.7	0.1	-0.48	-0.65	1.7	-0.94	1.83	0.71	-1.63
8	0.78	-0.34	-1.11	-0.62	-0.08	-0.3	-0.08	-1.03	-0.94	2.13	-1.45	-1.25	0.6	1.47	1.09	1.35	3.27	1.33	0.13	-1.78
9	0.24	-1.44	-0.57	0.37	-1.12	0.18	-0.82	-0.89	0.28	0.53	0.89	0.87	-0.01	0.65	1.03	1.11	-0.19	0.69	-0.1	0.7
10	0.09	-1.05	-0.03	-0.13	0.22	2.43	-2.13	0.66	-1.34	-0.36	-0.75	1.29	-0.13	-1.39	-0.41	2.48	0.16	-0.39	0.89	-0.8
11	0.98	-0.8	-0.28	1.37	-0.42	0.87	-0.68	0.84	-1.55	1.03	-1.77	-1.07	-0.75	0.92	-0.24	-0.98	0.1	-0.5	0.61	-0.15
12	-0.79	1.57	1.49	-0.15	0.66	1.33	-0.39	-1.34	0.16	0.99	-0.92	0.63	-0.42	-0.89	-0.33	-1.69	-0.94	0.55	-1.59	0.64
13	0.67	0.66	-1.39	-0.24	-0.2	-0.87	1.04	-0.42	0.93	-0.24	0.61	-1.12	1.98	0.66	0.17	0.81	-0.89	-0.06	0.71	-0.05
14	0.19	-2.69	-1.22	-0.08	0.9	1.95	1.69	-0.81	1.16	1.15	0.69	-0.96	0.9	0.76	-1.48	-0.4	0.13	-0.04	-1.55	-0.73
15	0.56	-0.52	1.45	-1	-1.2	-1.19	-1.43	-0.74	1.08	0.25	2.05	0.93	-1.22	1.67	-0.55	1.58	0.34	0.28	0.03	-1.46
16	-1.39	0.75	0.12	0.67	-0.39	1.5	1.11	-0.03	-0.63	0.58	-1.44	-0.2	-0.46	0.65	-1.14	0.14	0.9	-0.53	0.09	-0.4
17	-1.37	1.45	1.76	-1.48	-1.63	0.13	0.05	1.1	1.01	1.4	0.32	-1.25	1.53	1.16	-1.08	0.99	-0.78	-0.51	0.25	-0.25
18	0.73	0.28	0.92	0.43	-0.94	1.26	-0.45	0.79	0.49	-1.68	-1.37	1.07	0.6	0.22	0.53	-1.81	1.68	-0.98	0.74	-0.43
19	0.98	-0.46	-0.6	-0.29	0.22	-0.18	-0.65	1.4	0.08	-3.25	1.16	-0.85	1.46	0.31	1.59	0.43	0.71	-0.76	-0.69	-0.33
20	-1.22	-0.15	-0.61	-0.54	1.68	0	0.29	-0.11	3.04	0.22	0	-0.07	-1.97	-0.14	-0.38	0.49	2.3	1.12	0.08	-0.73

Fig. 9 20 x 20 data matrix for the 3D surface plot

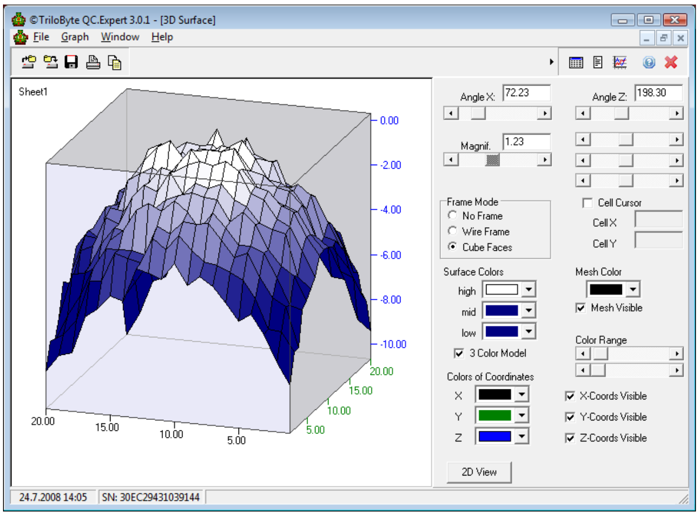
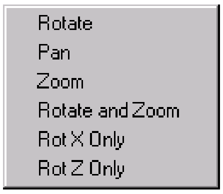


Fig. 10 3D-surface dynamic plot for spatial data, 20 x 20 data matrix

The plot menu is activated by the right mouse click. The menu offers the following control functions. The 3D-plot can be rotated, zoomed and moved interactively with mouse. The Z-level of the surface can be colored with 2 or 3 colors. There are more controls on the 3D-spline plot control panel which are used to modify the look of the plot. The button *2D View* is used to display an orthogonal projection into XY plane. For this view it is recommended to uncheck the *Mesh Visible* checkbox and select contrast colors. Right mouse click will display the control menu:



Rotate – Mouse movement will rotate the plot.
Pan –Move the plot with mouse.
Zoom – Mouse movement will zoom the plot.
Rotate and Zoom – Rotate with left mouse button, zoom with right mouse button.
Rotate X – Rotate only around X-axis
Rotate Z – Rotate only around Z-axis

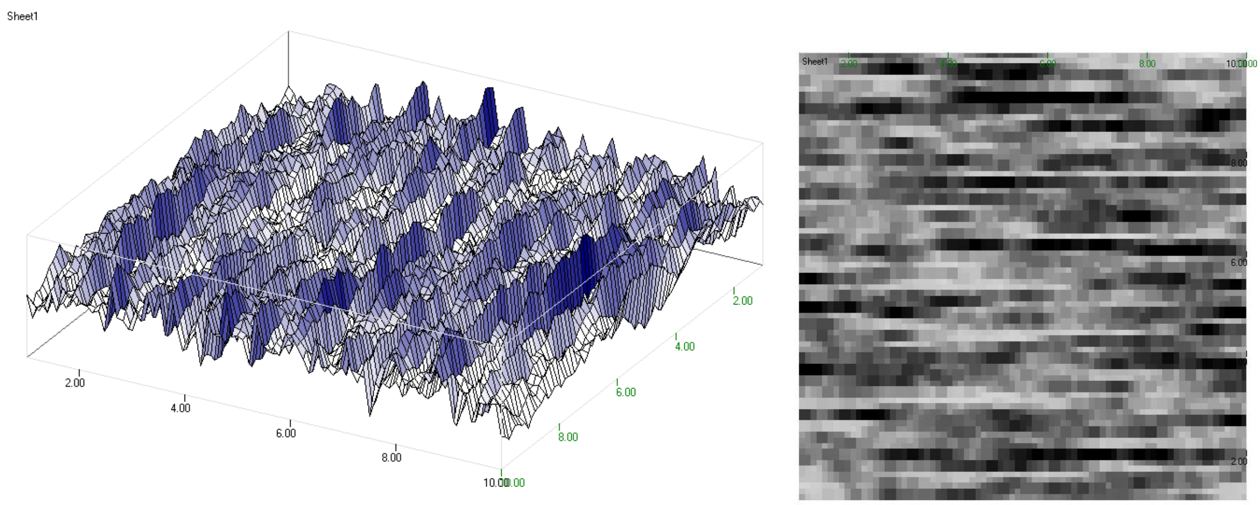


Fig. 11 a, b Multivariate time series view in general angle and in its orthogonal projection.

3D-spline

This plot smoothens data in three columns with a Gaussian local kernel smoother

$$z_s(x,y)=\frac{\sum_{i=1}^{n} z_i \exp-\left(\frac{r_{xy}}{u}\sqrt{(x_i-x)^2+(y_i-y)^2}\right)}{\sum_{i=1}^{n} \exp-\left(\frac{r_{xy}}{u}\sqrt{(x_i-x)^2+(y_i-y)^2}\right)},$$

where u is the smoothing parameter, the coefficient r_{xy} is calculated from range of x and y. Thus, the smoothing parameter is independent of the ranges of variables.

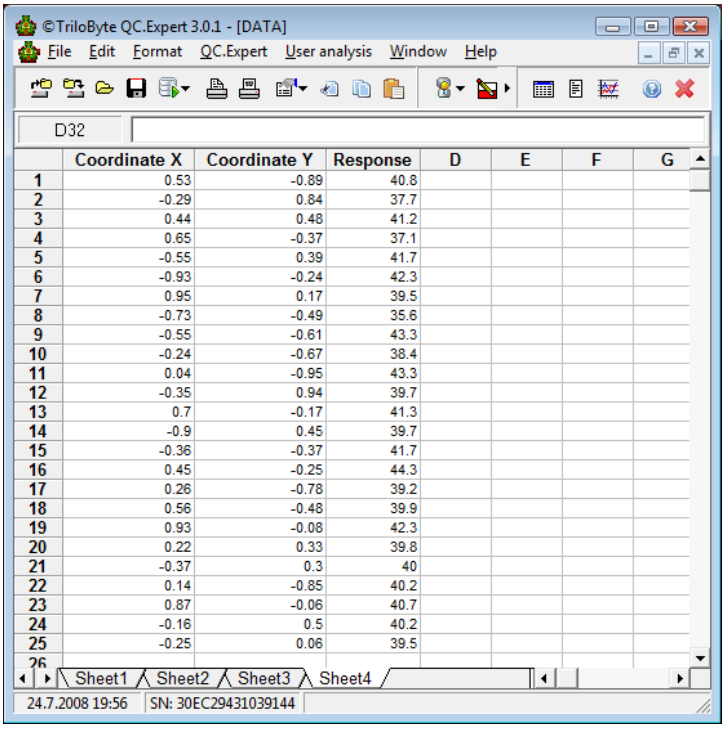


Fig. 12 Three columns for construction of the 3D spline

Data are expected in three columns, In the dialog box, two columns x, y must be selected in the field *Columns X,Y* and one response column must be selected in the *Z-Label* field. In Options form, the user may choose the density of the spline framework, or grid resolution and the value of the smoothing parameters. For most cases the default values can be left unchanged.

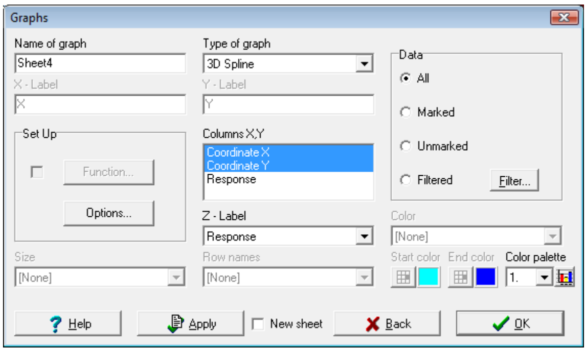
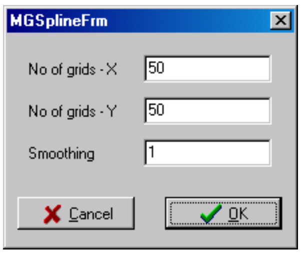


Fig. 13 3D-spline dialog box



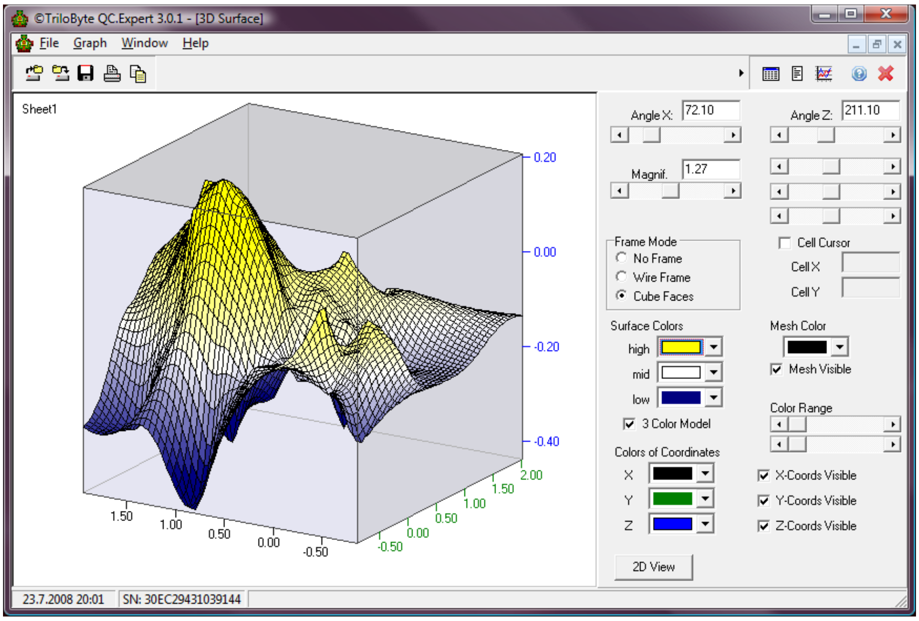
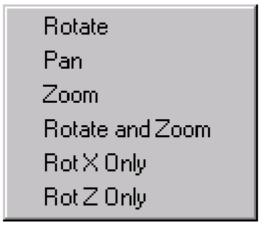


Fig. 14 3D-spline for data in 3 columns (X, Y, Z) can characterize response surfaces

Similarly like the previous plot, the 3D-spline can be rotated, zoomed and moved interactively with mouse. The Z-level of the surface can be colored with 2 or 3 colors. There are more controls on the 3D-spline plot control panel which are used to modify the look of the plot. The button *2D View* is used to display an orthogonal projection into XY plane. For this view it is recommended to uncheck the *Mesh Visible* checkbox and select contrast colors. Right mouse click will display the control menu:



Rotate – Mouse movement will rotate the plot.
Pan – Move the plot with mouse.
Zoom – Mouse movement will zoom the plot.
Rotate and Zoom – Rotate with left mouse button, zoom with right mouse button.
Rotate X – Rotate only around X-axis
Rotate Z – Rotate only around Z-axis

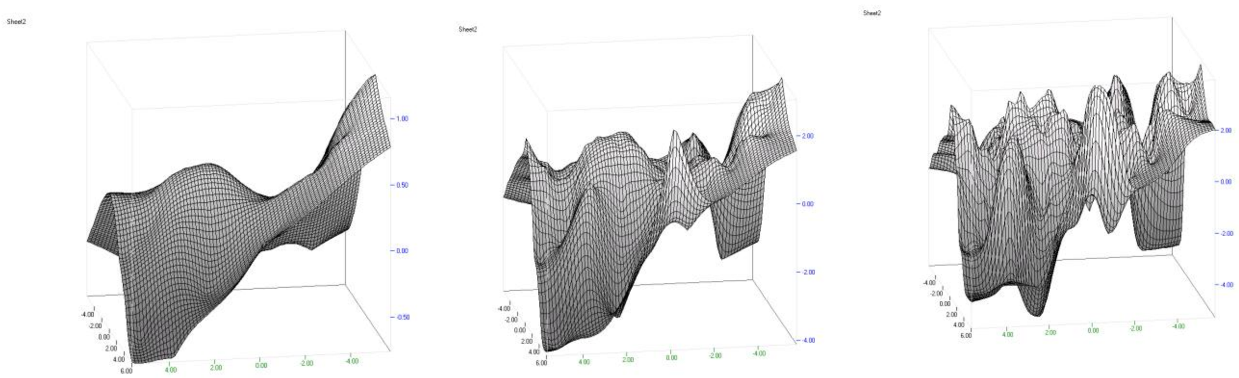


Fig. 15 a, b, c Non-parametric kernel smoothing for the same measured data with smoothing parameter 3, 1 a 0.5

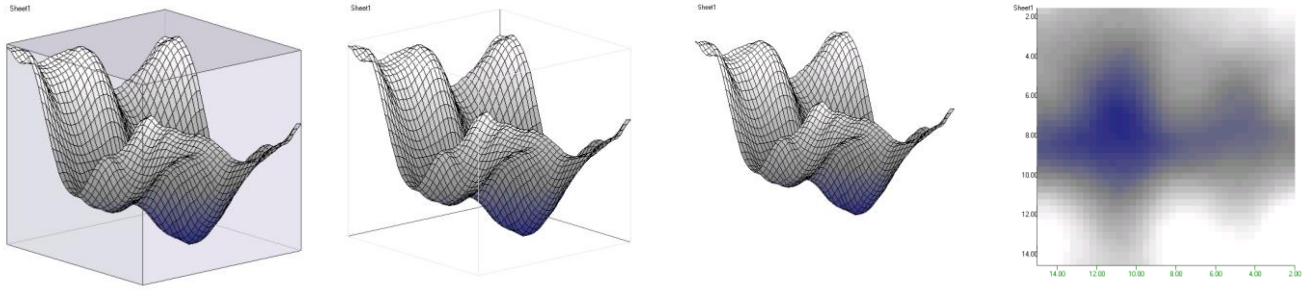


Fig. 16 a, b, c, d Various framing for 3D-plot and a 2D-view

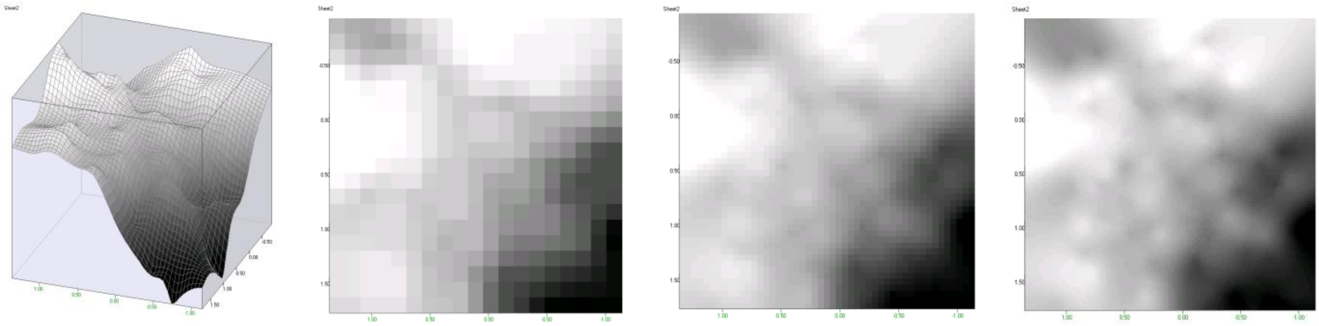


Fig. 17 a, b, c, d 2D-Same data, different grid resolution

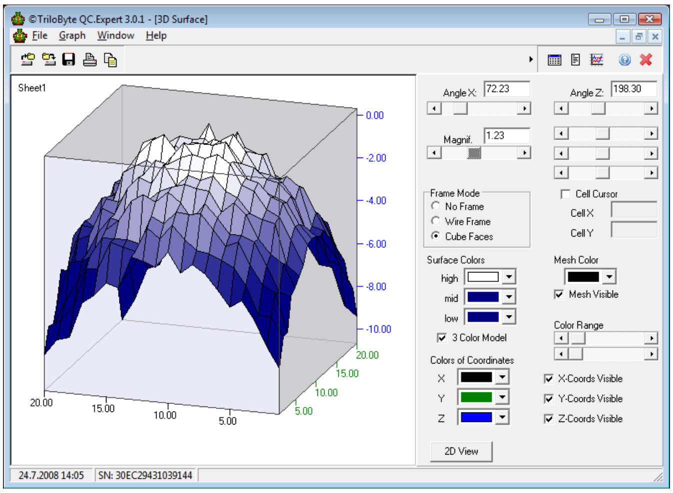


Fig. 18 Spline with smoothing parameter *u* = 0.01, spline with small *u* is an interpolation alternative of the 3D-Surface plot.

3D-kernel probability density for 2 variables

This plot is used to estimate and visualize a non-parametric estimate of probability density for 2 variables. The density function is constructed as a Gaussian kernel density estimate with a user-defined kernel width (or smoothing parameter) and grid resolution.

$$f_s(x,y)=\frac{1}{K}\sum_{i=1}^{n}\exp-\left(\frac{1}{u}\sqrt{\frac{(x_i-x)^2}{s_x}+\frac{(y_i-y)^2}{s_y}}\right),$$

where u is the relative kernel width and s_x and s_y are standard deviation estimates for x and y respectively. Smoothing parameter and grid resolution are defined in the *Options* form. Plot controls are the same as in the previous plot. Examples of use are given below.

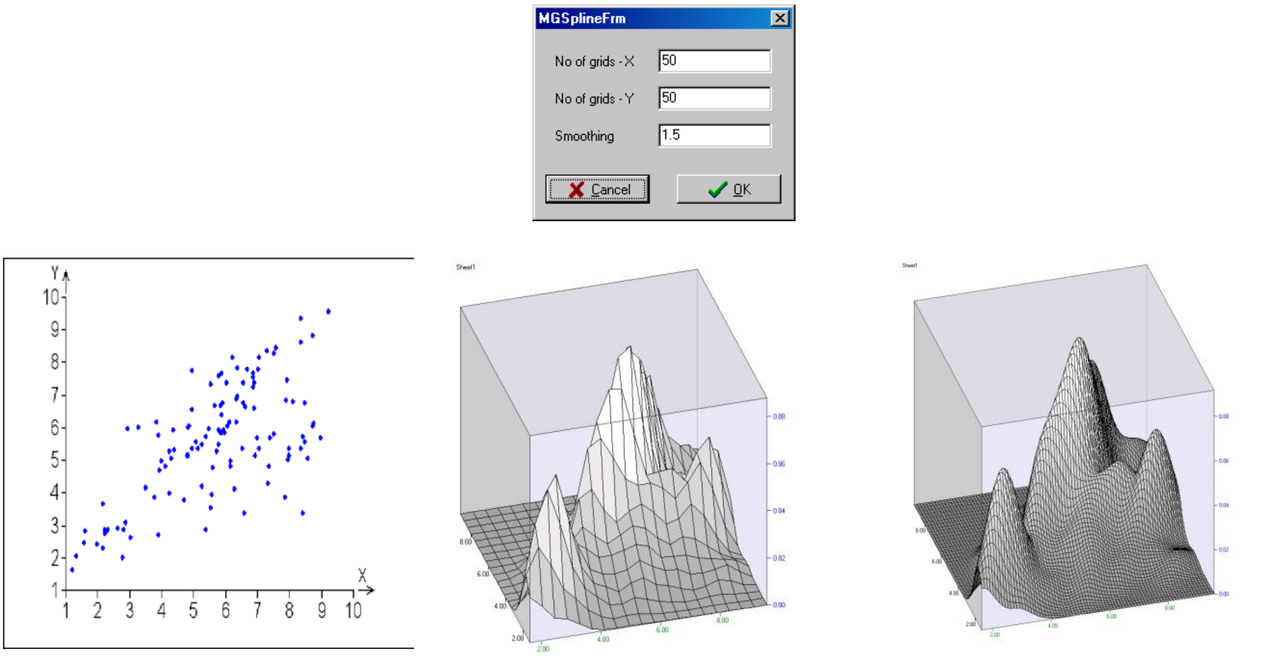


Fig. 19 Data plot and corresponding probability density estimates with 2 different grid resolutions

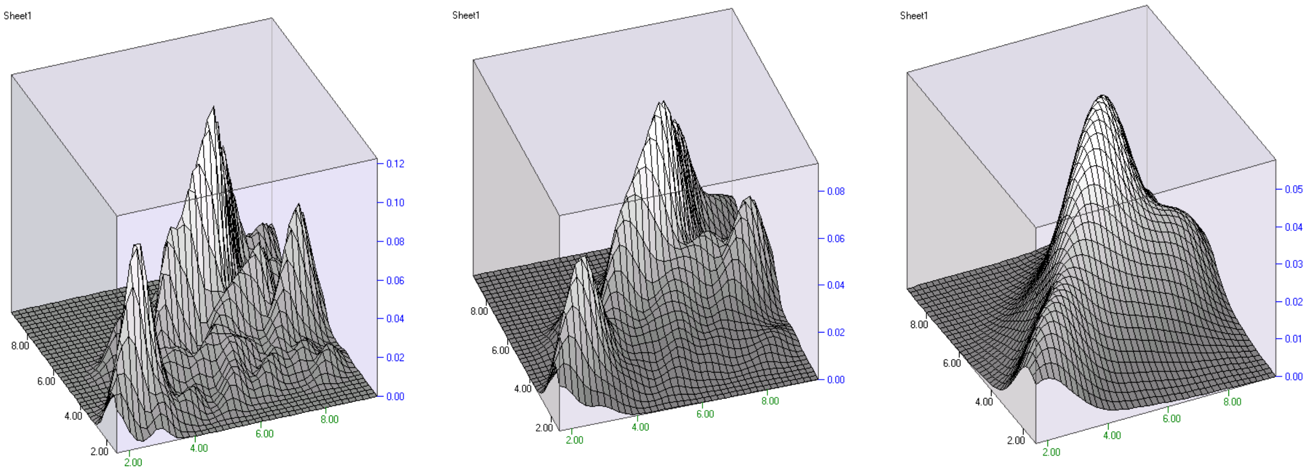


Fig. 20 a, b, c Kernel probability density estimates for the same data with 3 different smoothing parameters

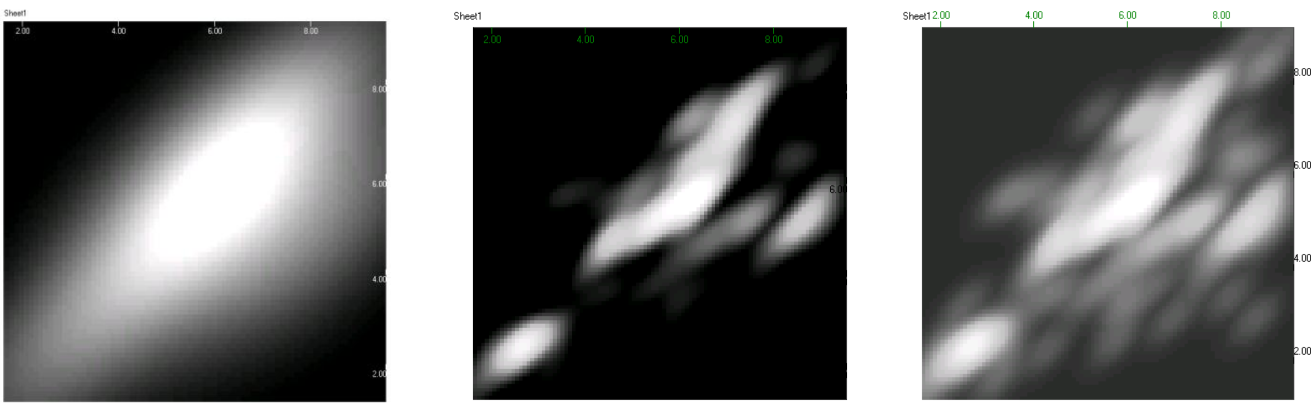


Fig. 21 a, b, c Projection of the previous plots into 2D "contour" plots with smoothing parameter 2 (a) and 0.7 (b, c)

Cluster analysis - Dendrogram

Dendrogram is a useful tool in cluster analysis. It is a tree-like plot where the multidimensional data represented by rows are connected by lines. The length of path connecting two points roughly defines the dissimilarity or distance between the two points. Thus, we can see if there are any distinct groups of data, we can recognize data that are close to each other. The shape of the dendrogram may be strongly dependent on the selected distance measure and clustering method. This module offers 5 distance measures and 6 clustering methods, one of which (Flexible linkage) is parameterized and can be adjusted continuously. Variables are expected in *m* columns and *n* rows, the data matrix may have more columns than rows. Individual rows are taken as points in *m*-dimensional space. After defining columns and constructing the dendrogram we can use controls in the right part of the plot window to modify the dendrogram. On the right mouse click we get interactive menu:

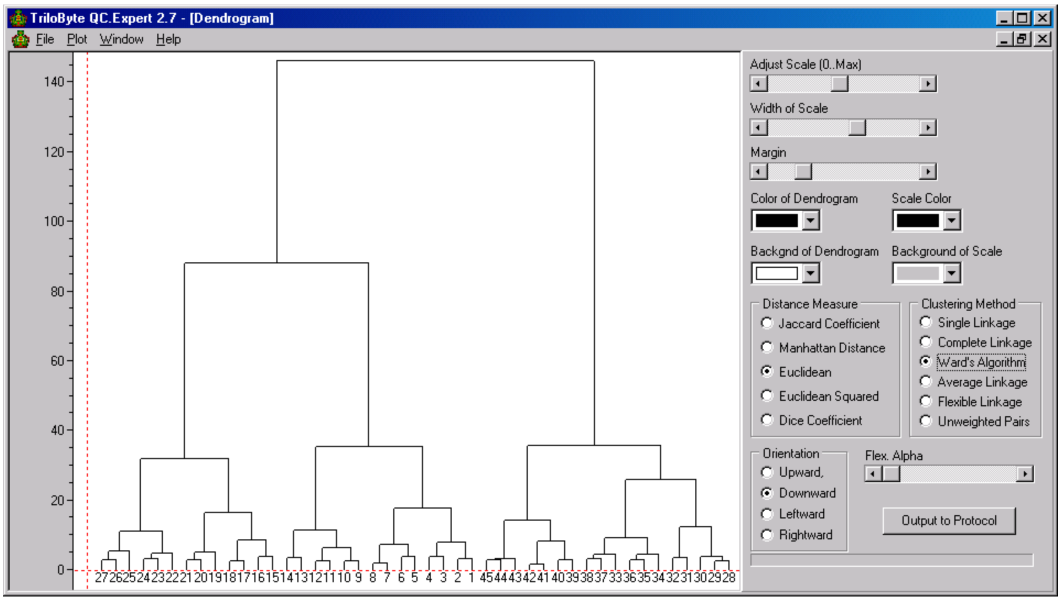
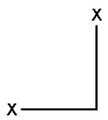


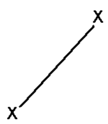
Fig. 22 Dendrogram control window with parameters and dendrogram tree

Distance measures definition:



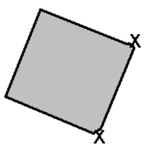
Manhattan distance

$$d_{ij} = \sum_{k=1}^{m} \left| x_{ik} - x_{jk} \right|$$



Euclidean distance

$$d_{ij} = \sqrt{\sum_{k=1}^{m} \left(x_{ik} - x_{jk} \right)^2}$$



Squared Euclidean distance

$$d_{ij} = \sum_{k=1}^{m} \left(x_{ik} - x_{jk} \right)^2$$

Continuous Jaccard coefficient $d_{ij} = \dfrac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\|x_i\| \cdot \|x_j\| - \sum_{k=1}^{m} x_{ik} x_{jk}}$,

where $\|x\|$ is the quadratic norm of x.

For binary data (e.g. ones and zeroes) four similarity parameters *a*, *b*, *c*, *d* are computed: *a* = number of cases where in both rows to compare are zeroes in the same column, *b* = Number of cases with 1 in the first row and 0 in the second row, *c* = Number of cases with 0 in the first row and 1 in the second row, *d* = number of cases where in both rows to compare are ones. Binary data example is given below.

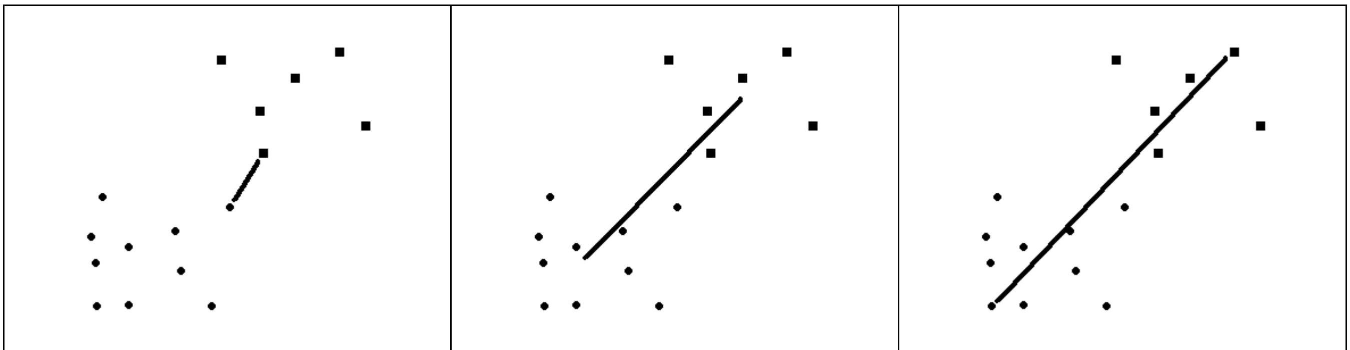
		Row 1	
Row 2		0	1
	0	a	b
	1	c	d

0	0	0	1	0	1	1
0	**0**	**1**	**1**	**0**	**1**	**1**
0	1	1	1	1	1	1
0	0	0	1	0	1	0
0	**1**	**1**	**0**	**0**	**1**	**0**
0	1	0	0	0	0	1
1	1	0	0	1	0	0

For the given example, when we compare the second and fifth row, we get: *a*=2, *b*=1, *c*=2, *d*=2. Dissimilarity coefficient d_{25} is then computed using Jaccard or dice coefficient:

		Row 1	
Row 2		0	1
	0	2	1
	1	2	2

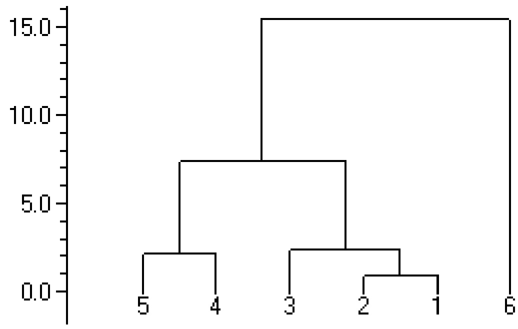
Jaccard coefficient $d_{ij} = \dfrac{a}{a+b+c}$ and dice-coefficient $d_{ij} = \dfrac{2a}{2a+b+c}$.



a - single linkage **b -** average linkage **c -** complete linkage

Fig. 23 a, b, c Three methods to build clusters

3-dimensional data example				
Object no.	p1	p2	p3	
1	2	2	2	
2	2	3	2	
3	4	1	3	
4	7	7	7	
5	8	7	5	
6	12	16	-7	



With the button *Export to Protocol* we get the following table in the Protocol window which is read as follows: Objects 1 and 2 with distance 1 form cluster 7; objects 4 and 5 have distance 2.23 and form cluster 8. Cluster 7 and object 3 form cluster 9, distance between 3 and 7 is 2.45, etc. The distances are plotted on the vertical axis.

Obj.1	Obj.2	New cluster	Distance
1	2	7	1
4	5	8	2.2361
7	3	9	2.4495
9	8	10	7.4833
10	6	11	15.5242

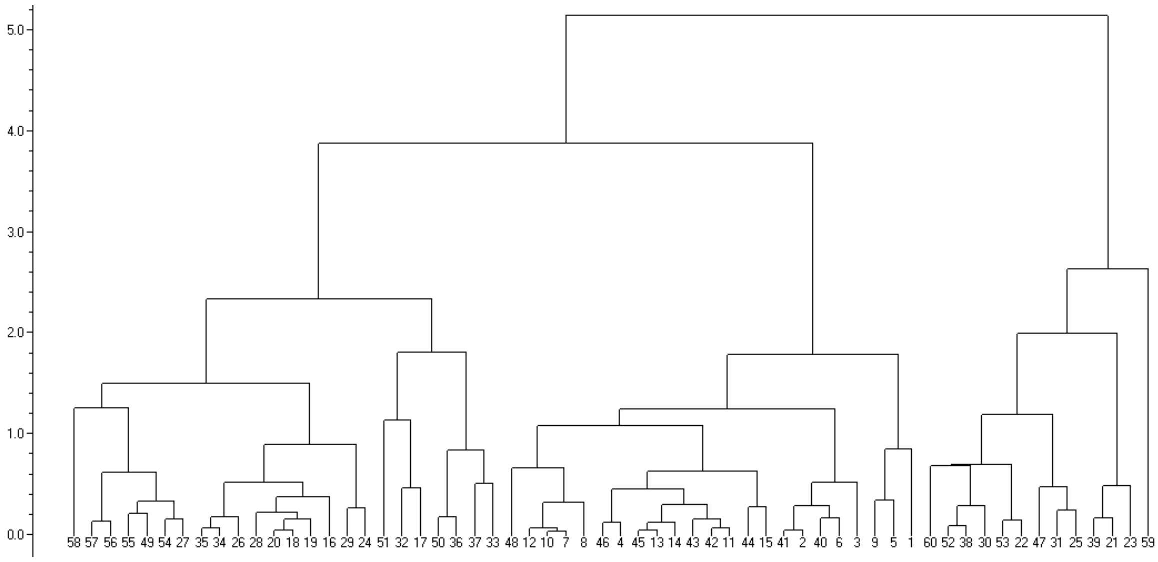


Fig. 24 Example dendrogram for 6 variables on 60 samples of cement, 3 dominant clusters are detected, containing 24, 23 a 13 points(or cement samples)