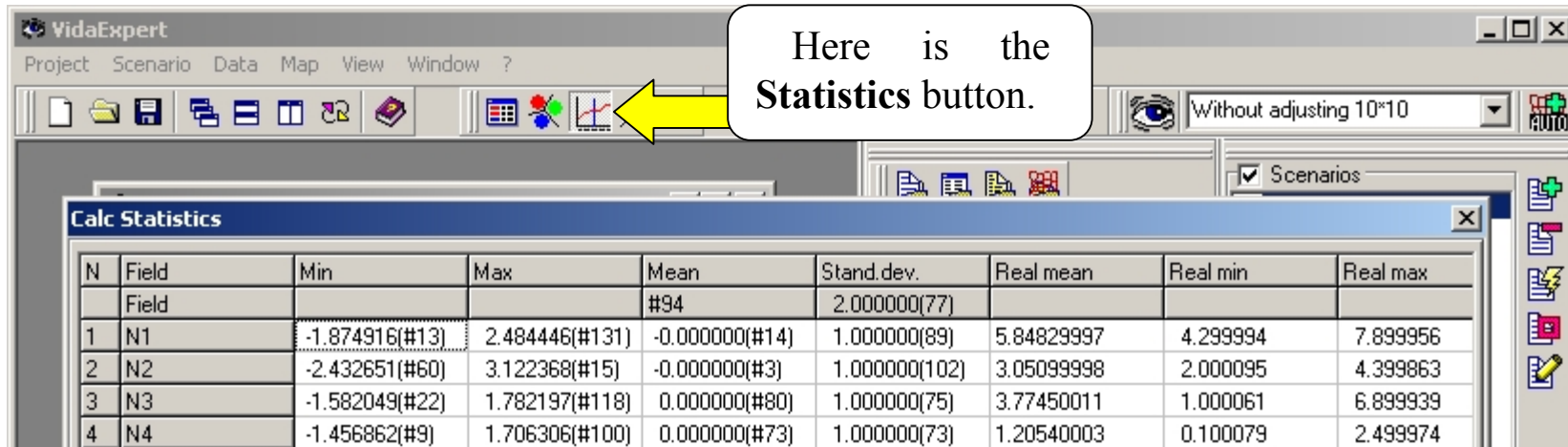


VIDAEXPERT: DATA ANALYSIS



Here is the **Statistics** button.

N	Field	Min	Max	Mean	Stand.dev.	Real mean	Real min	Real max
	Field			#94	2.000000(77)			
1	N1	-1.874916(#13)	2.484446(#131)	-0.000000(#14)	1.000000(89)	5.84829997	4.299994	7.899956
2	N2	-2.432651(#60)	3.122368(#15)	-0.000000(#3)	1.000000(102)	3.05099998	2.000095	4.399863
3	N3	-1.582049(#22)	1.782197(#118)	0.000000(#80)	1.000000(75)	3.77450011	1.000061	6.899939
4	N4	-1.456862(#9)	1.706306(#100)	0.000000(#73)	1.000000(73)	1.20540003	0.100079	2.499974

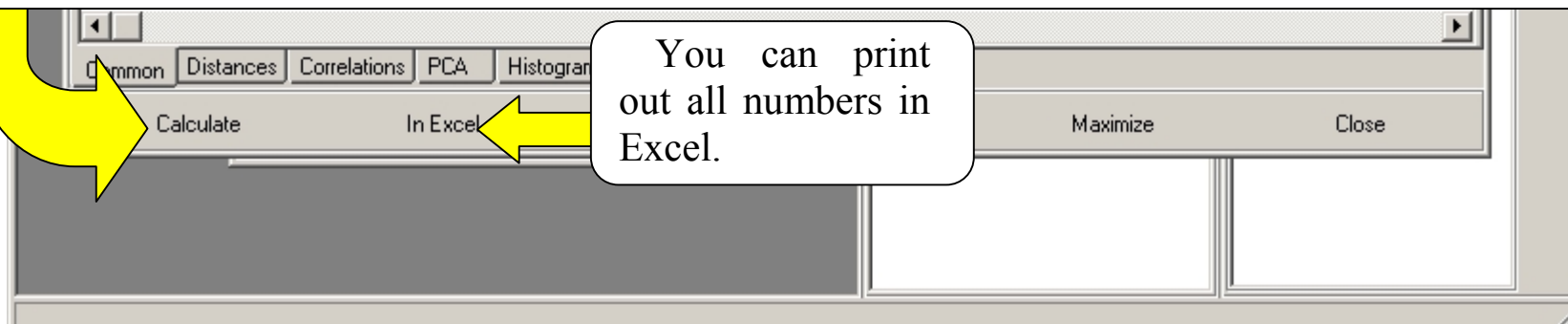
After creating dataset you can analyze it in different ways. First, you can calculate statistics. Open **Statistics** dialog, **Common** tabsheet, click **Calculate**.

Min, Max: minimal and maximal field values after **normalization**. The numbers in parentheses with # symbol mean the numbers of the record where the maximum and minimum values were reached.

The same with **Mean**. The number just below the **Mean** word means the number of object closest to the average point.

Stand.dev.: standard deviations after **normalization**. The numbers in parentheses indicate how many objects are in the interval $(m-\sigma, m+\sigma)$, where m – average, σ – standard deviation. The number just below the **Stand.dev.** word means overall standard deviation.

Real min, Real max, Real mean – Mean, Max, Mean values **before normalization**.



You can print out all numbers in Excel.

VIDAEXPERT: DATA ANALYSIS

The screenshot displays the VidaExpert software interface. A 'Calc Statistics' dialog box is open, showing a table of Euclidean Distances matrix (after normalization). The table has 20 rows and 10 columns, with the first column representing the object index (0-19) and the subsequent columns representing distances to objects 0-9. The cell at row 7, column 4 (value 2.631052) is highlighted. A yellow arrow points to the 'Distances' tab in the dialog box. A text box provides instructions: 'Euclidean Distances matrix (after normalization). Choose Distances tabsheet, click Calculate. (Not very useful if you have thousands of objects. ;))'. The dialog box also features 'Calculate', 'Maximize', and 'Close' buttons.

	0	1	2	3	4	5	6	7	8	9
0										
1	0.525547									
2	0.434508	0.285032								
3	1.394024	0.996191	1.255868							
4	2.191980	1.861481	2.109420	0.903063						
5	1.003253	0.499579	0.709072	0.682849	1.524610					
6	0.935462	0.599392	0.846628	0.466417	1.287027	0.505213				
7	0.648199	0.785738	0.525546	1.775664	2.631052	1.189689	1.367653			
8	0.272389	0.377630	0.386450	1.172442	1.991137	0.828775	0.717074	0.775567		
9	1.730588	1.439046	1.693255	0.539854	0.544777	1.200522	0.846627	2.213211	1.520724	
10	0.940661	0.508153	0.737609	0.534746	1.392823	0.298381	0.248809	1.259745	0.719338	1.006628
11	0.178982	0.499580	0.364486	1.415862	2.247872	0.992629	0.967704	0.552776	0.267373	1.781462
12	0.757973	0.692307	0.504915	1.641277	2.527276	1.043021	1.282609	0.338926	0.795918	2.113950
13	2.560878	2.281719	2.545637	1.344861	0.662541	2.017567	1.701874	3.060955	2.360896	0.863737
14	3.392872	3.043584	3.301163	2.054250	1.218324	2.674369	2.478978	3.821634	3.185703	1.681235
15	2.186038	1.847455	2.109420	0.888545	0.228085	1.516056	1.287027	2.626105	1.991137	0.544777
16	1.189689	0.858777	1.189689	1.189689	1.189689	1.189689	1.189689	1.632338	0.994266	0.605713
17	2.100904	1.861481	2.100904	2.100904	2.100904	2.100904	2.100904	2.619966	1.909465	0.464675
18	1.872958	1.481137	1.872958	1.872958	1.872958	1.872958	1.872958	2.253570	1.659287	0.450463
19	1.119387	0.992307	1.119387	1.119387	1.119387	1.119387	1.119387	1.683733	0.937618	0.703680

Euclidean Distances matrix (after normalization). Choose Distances tabsheet, click Calculate. (Not very useful if you have thousands of objects. ;))

VIDAEXPERT: DATA ANALYSIS

The screenshot shows the VidaExpert software interface. The main window has a menu bar (Project, Scenario, Data, Map, View, Window, ?) and a toolbar with various icons. A 'Calc Statistics' dialog box is open, displaying a correlation matrix for four variables (N1, N2, N3, N4). The matrix values are: N1-N2: -0.103784; N1-N3: 0.871283; N1-N4: 0.816971; N2-N3: -0.415218; N2-N4: -0.350733; N3-N4: 0.962314. The values 0.871283, 0.816971, and 0.962314 are highlighted in red. The value -0.415218 is highlighted in green. The 'Correlations' tab is selected, and a yellow arrow points to it from a text box. The text box contains the following text: 'Correlations matrix. Red color indicate values greater than 0.8, blue – greater then 0.6, green – greater than 0.4.'

	N1	N2	N3	N4
N1		-0.103784	0.871283	0.816971
N2			-0.415218	-0.350733
N3				0.962314
N4				

Correlations matrix. Red color indicate values greater than 0.8, blue – greater then 0.6, green – greater than 0.4.

VIDAEXPERT: DATA ANALYSIS

Calc Statistics

N	No	Field	1	2	3	4
		Dispers	0.726151	0.231524	0.037115	0.005210
		Overall	0.726151	0.957674	0.994790	1.000000
1	3	N1	0.526813	0.363793	0.721485	0.263784
2	4	N2	-0.250294	0.929538	-0.240373	-0.124630
3	1	N3	0.581995	0.013402	-0.138809	-0.801146
4	2	N4	0.566659	0.058602	-0.634360	0.522542

Common Distances Correlations PCA Histograms Obj

Calculate In Excel

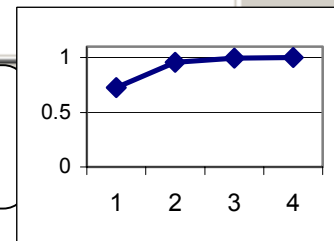
First line - eigen values of the corresponding vector.

Second line – overall contribution in summary dispersion. From this you can determine **effective linear dimension** of the dataset.

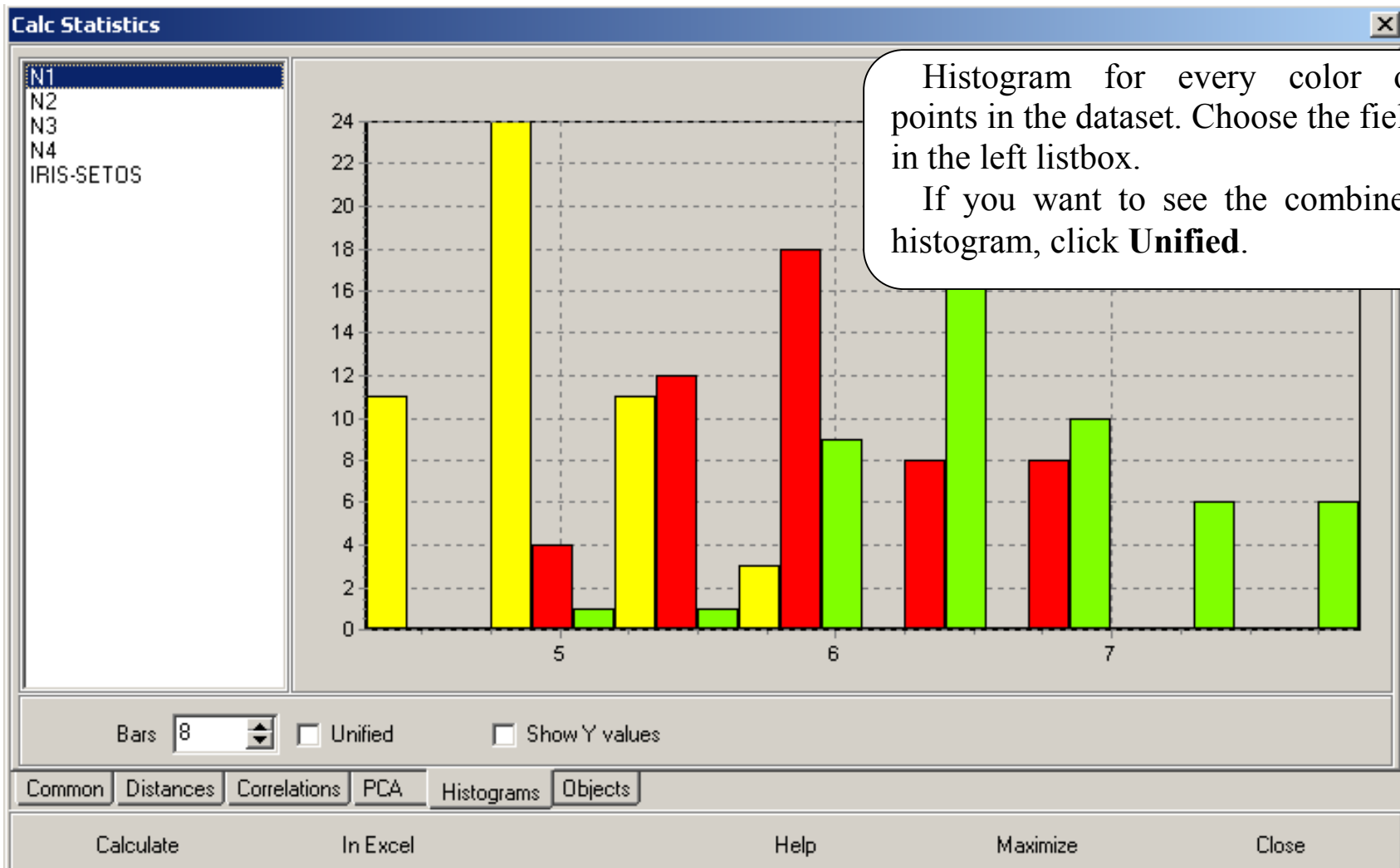
In this example **Iris** dataset is 2D dimensional with accuracy 0.05 and 3D dimensional with accuracy 0.01.

Principal components analysis. Coordinates of principal vectors are shown. Red color indicate the maximum value. Green – several big values.

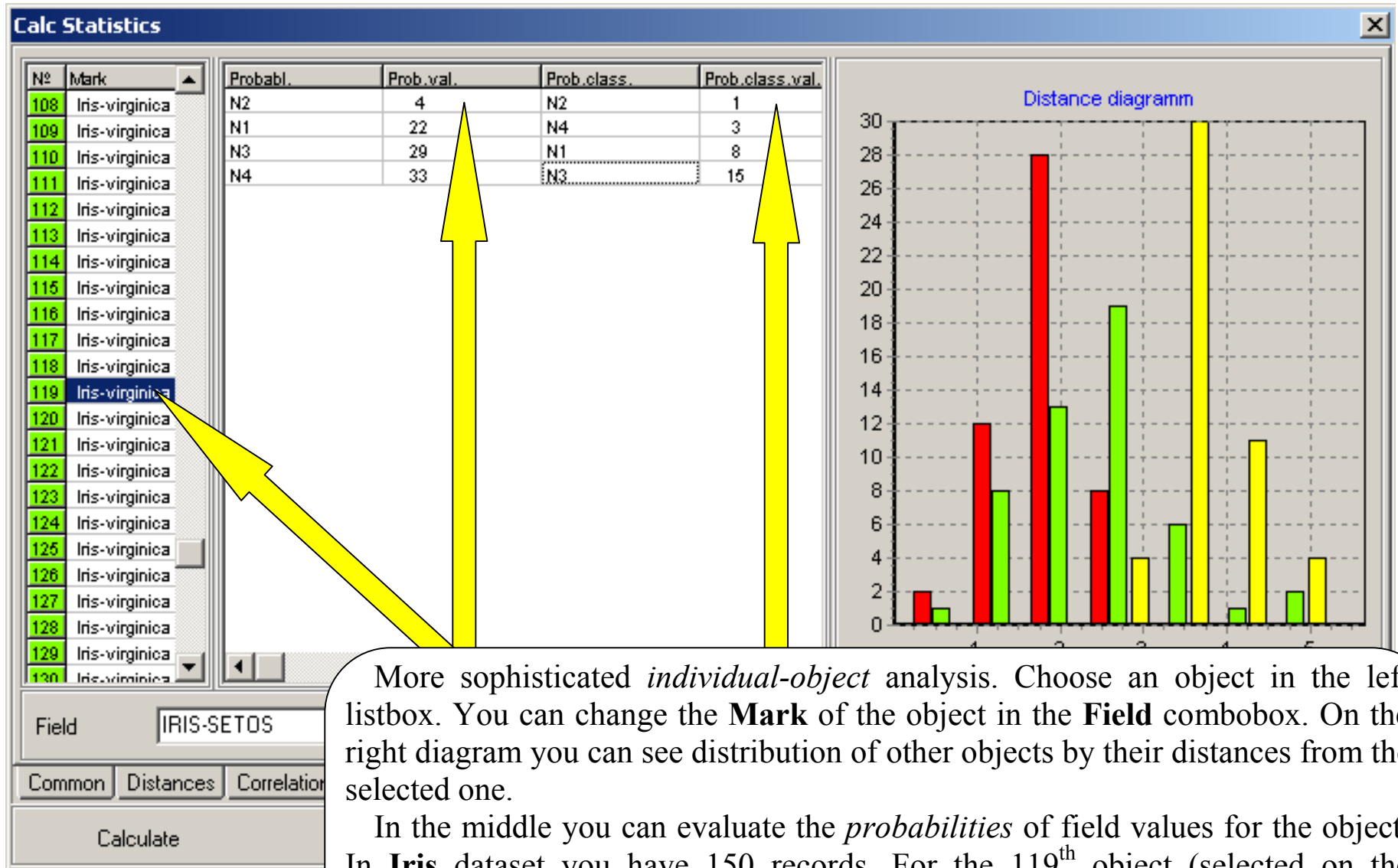
Tip: print the table in Excel to plot different diagrams.



VIDAEXPERT: DATA ANALYSIS



VIDAEXPERT: DATA ANALYSIS



VIDAEXPERT: DATA ANALYSIS

Data analysis dialog

On Clustering tabsheet select method of clustering, number of clusters (not for all methods) and click **Analyze**. You will see the results on the **Map** panel.

Cluster	Color	Objects	Interclass deviation
1	Blue	52/0.91303	
2	Green	49/0.99429	
3	Red	48/1.01319	

Class-class deviation = 0.973007

Here you see the colors of the clusters, number of objects in every cluster and interclass deviation value (compactness).

Cancel will cancel the clustering.
Remember colors will assign for every point it's current color (result of clustering).
Numbers in table will put information about cluster's numbers in the datatable.
Distances in table will put information about distances from the **centroids** in the datatable.

VIDAEXPERT: DATA ANALYSIS

The screenshot shows the VidaExpert software interface. The main window displays a data table with columns N1, N2, N3, N4, IRIS-SETOS, and Class. The table is color-coded by cluster number. A yellow arrow points to the 'Numbers in table' button in the toolbar. An 'Object analysis' dialog box is open, showing the 'Clustering' tab with 'K-Means' selected and 'Number of clusters' set to 3. The dialog also displays 'Clusters colors' and 'Class-class deviation = 0.973007'. A text box provides instructions on how to use the 'Numbers in table' button.

Click this to switch between **Map** and **Table** panels.

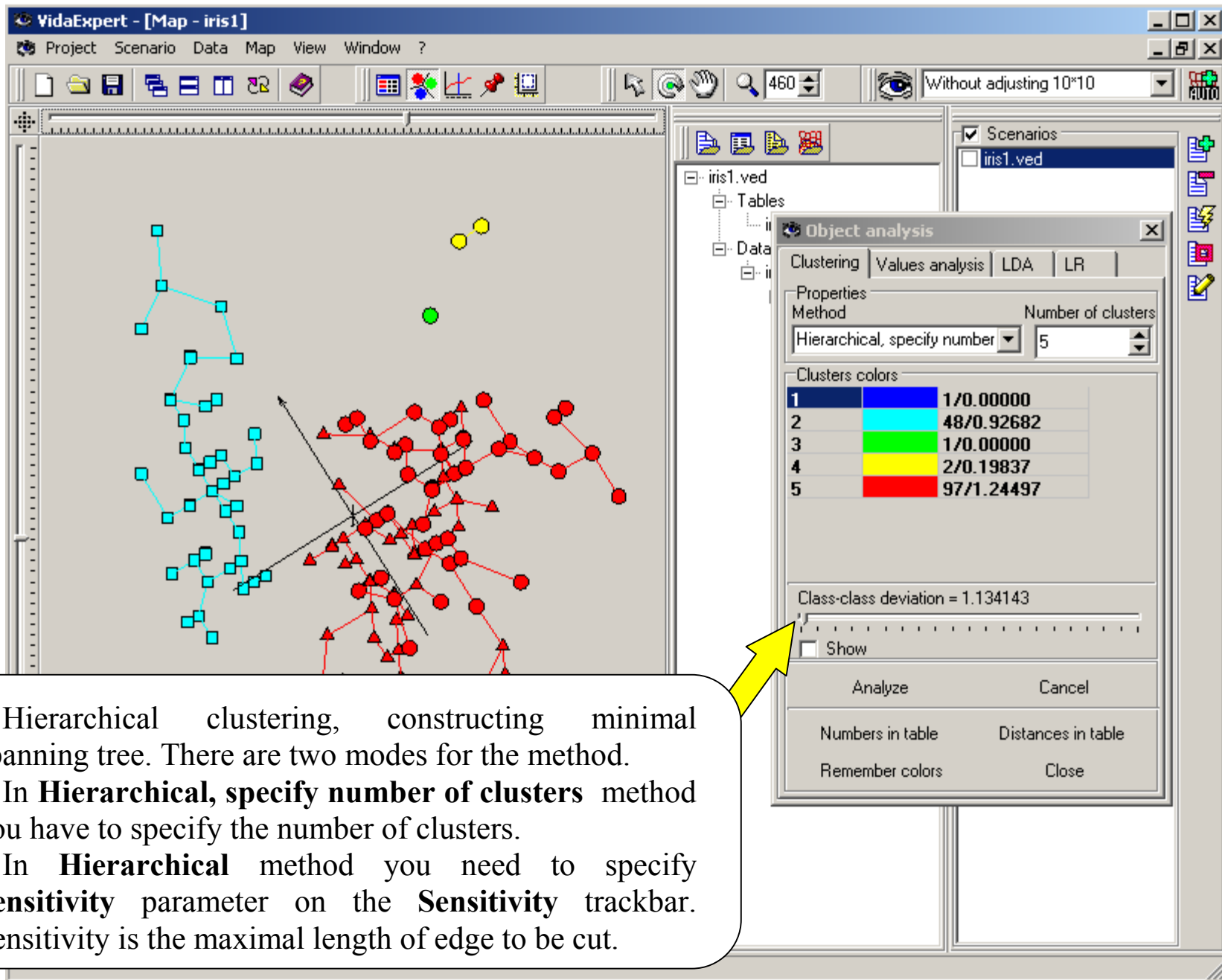
You can see records colored accordingly to the cluster number.
If you have clicked the **Numbers in table** button, you will see the numbers of clusters.
Tip: right-click on the table and print the table in Excel worksheet or CSV file for the further analysis.

	N1	N2	N3	N4	IRIS-SETOS	Class
41	4.5	2.3	1.3	0.3	Iris-setosa	2
42	4.4	3.2	1.3	0.2	Iris-setosa	2
43	5	3.5	1.6	0.6	Iris-setosa	2
44	5.1	3.8	1.9	0.4	Iris-setosa	2
45	4.8	3	1.4	0.3	Iris-setosa	2
46	5.1	3.8	1.6	0.2	Iris-setosa	2
47	4.6	3.2	1.4	0.2	Iris-setosa	2
48	5.3	3.7	1.5	0.2	Iris-setosa	2
49	5	3.3	1.4	0.2	Iris-setosa	2
50	7	3.2	4.7	1.4	Iris-versicolor	3
51	6.4	3.2	4.5	1.5	Iris-versicolor	3
52	6.9	3.1	4.9	1.5	Iris-versicolor	3
53	5.5	2.3	4	1.3	Iris-versicolor	1
54	6.5	2.8	4.6	1.5	Iris-versicolor	1
55	5.7	2.8	4.1	1.5	Iris-versicolor	1
56	6.3	3.3	5.8	1.6	Iris-versicolor	1
57	4.9	2.4	3.6	1	Iris-versicolor	1
58	6.6	2.9	5.6	1.4	Iris-versicolor	1
59	5.2	2.7	4.3	1.5	Iris-versicolor	1
60	5	2	3.5	1	Iris-versicolor	1
61	5.9	3	5.1	1.4	Iris-versicolor	1
62	6	2.2	3.5	1	Iris-versicolor	1
63	6.1	2.9	5.6	1.4	Iris-versicolor	1
64	5.6	2.9	5.1	1.4	Iris-versicolor	1
65	6.7	3.1	5.4	1.5	Iris-versicolor	1
66	5.6	3	4.5	1.5	Iris-versicolor	1
67	5.8	2.7	4.1	1	Iris-versicolor	1
68	6.2	2.2	4.5	1.5	Iris-versicolor	1
69	5.6	2.5	3.9	1.1	Iris-versicolor	1

iris.vet

Object analysis dialog box details:
Clustering: Values analysis, LDA, LR
Properties: Method: K-Means, Number of clusters: 3
Clusters colors:
1: 52/0.91303
2: 49/0.99429
3: 48/1.01319
Class-class deviation = 0.973007
Buttons: Analyze, Cancel, Numbers in table, Distances in table, Remember colors, Close
Checkboxes: Colorings, Selections, Axis X, Axis Y, Axis Z, Not draw all

VIDAEXPERT: DATA ANALYSIS



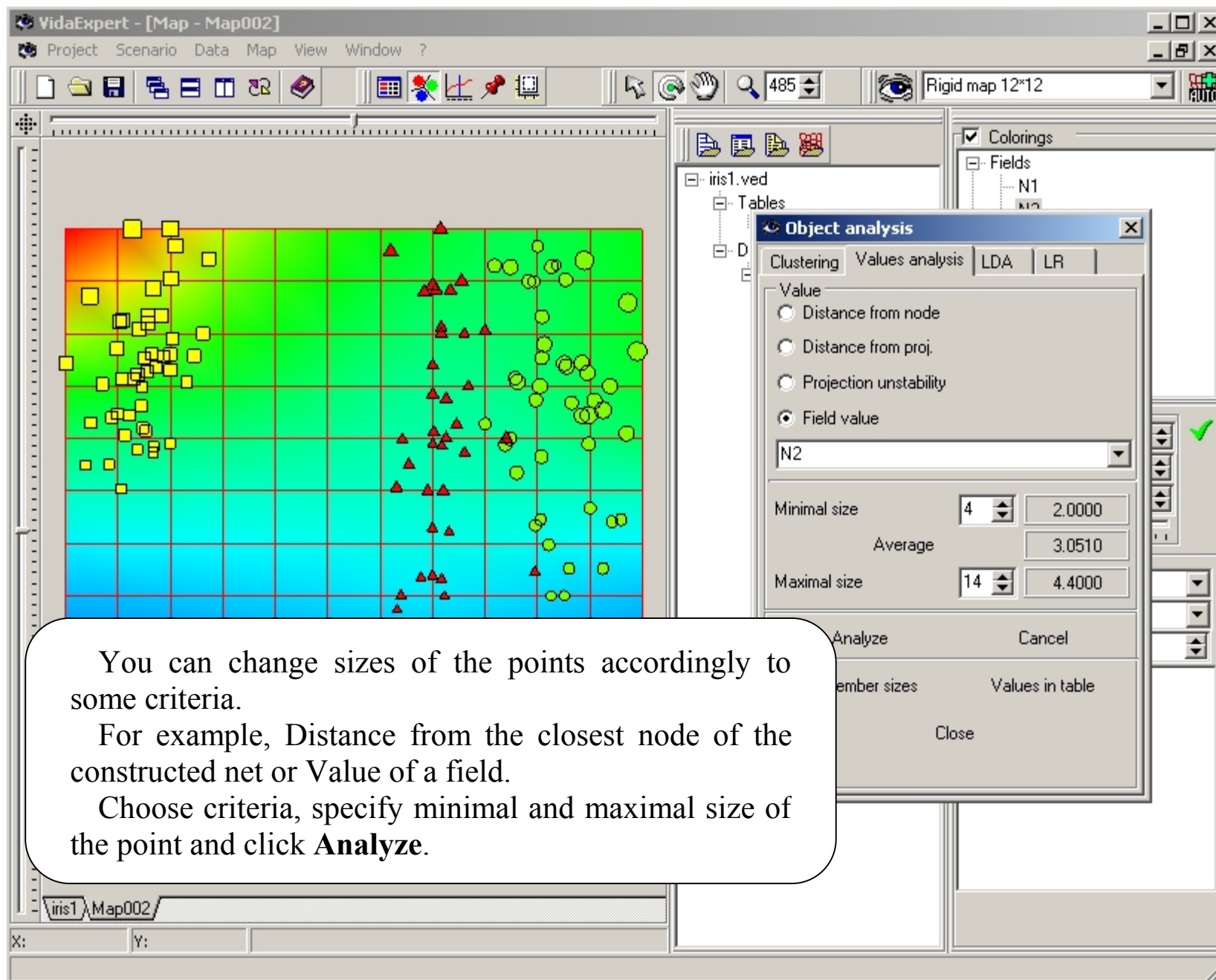
Hierarchical clustering, constructing minimal spanning tree. There are two modes for the method.

In **Hierarchical, specify number of clusters** method you have to specify the number of clusters.

In **Hierarchical** method you need to specify **Sensitivity** parameter on the **Sensitivity** trackbar. Sensitivity is the maximal length of edge to be cut.

Cluster	Color	Count/Deviation
1	Blue	1/0.00000
2	Cyan	48/0.92682
3	Green	1/0.00000
4	Yellow	2/0.19837
5	Red	97/1.24497

VIDAEXPERT: DATA ANALYSIS



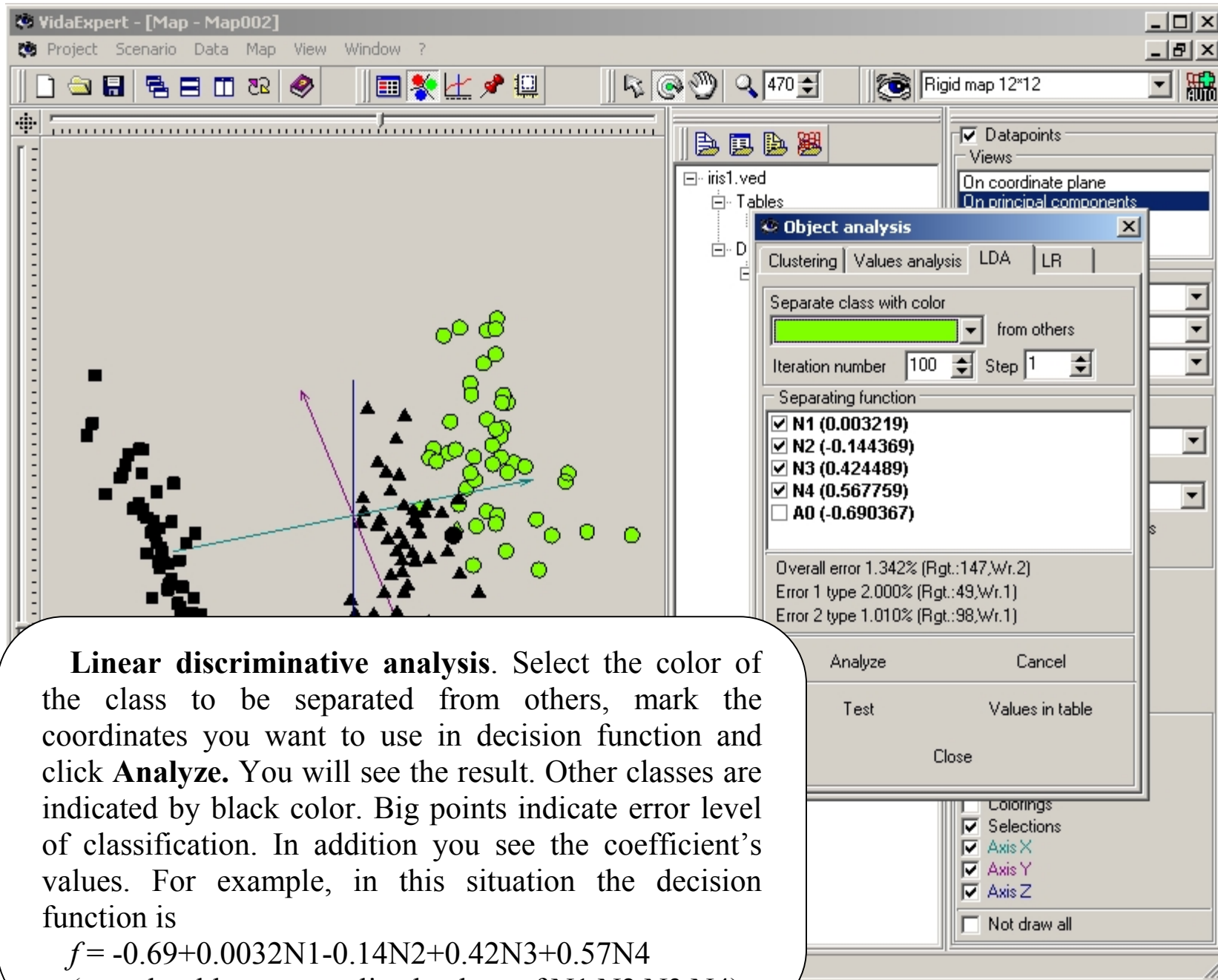
The screenshot displays the VidaExpert software interface. The main window shows a map with a grid overlay. The map is color-coded, with a gradient from red at the top to blue at the bottom. Data points are plotted on the map, including yellow squares, red triangles, and green circles. The size of these points varies, indicating they are being sized based on a specific criterion. An 'Object analysis' dialog box is open, showing options for 'Clustering', 'Values analysis', 'LDA', and 'LR'. The 'Values analysis' tab is selected, and the 'Field value' option is chosen. The dialog also shows a dropdown menu set to 'N2', and fields for 'Minimal size' (4) and 'Maximal size' (14). The 'Analyze' button is visible.

You can change sizes of the points accordingly to some criteria.

For example, Distance from the closest node of the constructed net or Value of a field.

Choose criteria, specify minimal and maximal size of the point and click **Analyze**.

VIDAEXPERT: DATA ANALYSIS



VIDAEXPERT: DATA ANALYSIS

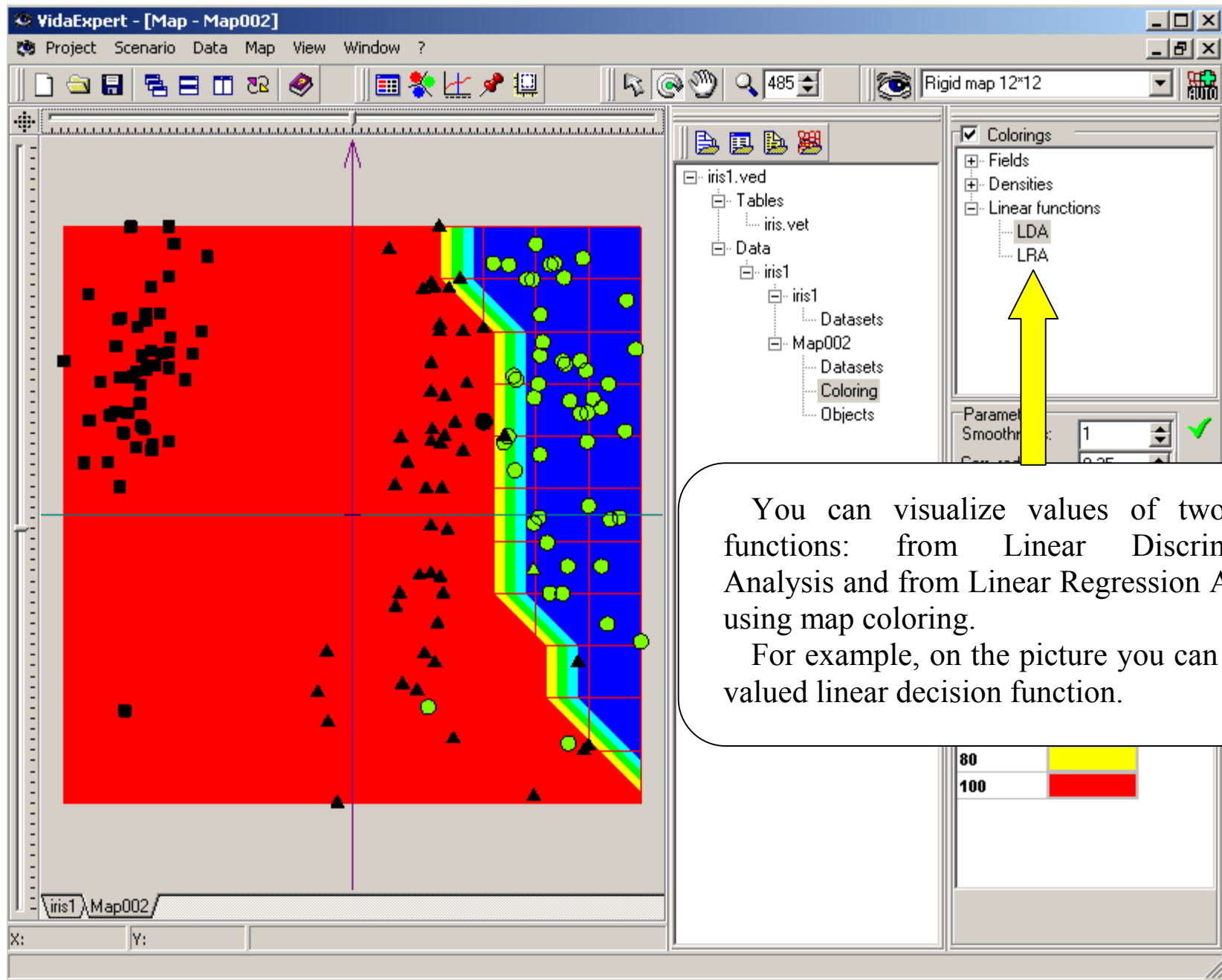
Linear regression analysis. Select the field to be calculated, mark the coordinates you want to use in regression function and click **Analyze**. Big points indicate error level of calculation. For example, in this situation the function is

$$N1 = 5.84 + 0.28N2 + 1.24N3 - 0.42N4$$

(you should use normalized values of N2, N3, N4, but the resulting value is not normalized).

Try to change value of **Quality** to test accuracy level.

VIDAEXPERT: DATA ANALYSIS



You can visualize values of two linear functions: from Linear Discriminative Analysis and from Linear Regression Analysis using map coloring.

For example, on the picture you can see 0-1 valued linear decision function.

VIDAEXPERT: DATA ANALYSIS

The screenshot displays the VidaExpert software interface. The main window is titled "VidaExpert - [Map - Map002]". The menu bar includes "Project", "Scenario", "Data", "Map", "View", and "Window". The toolbar contains various icons for file operations, map navigation, and analysis. The main map area shows a 12x12 grid with data points represented by yellow squares, red triangles, green circles, and blue triangles. A dashed black rectangle highlights a selection of blue triangles. A yellow arrow points from the "Select" button in the toolbar to the map. Another yellow arrow points from the "Selection..." button in the right-hand panel to the "Marks" list. The "Selection..." dialog is open, showing a "Marks" list with the following items: N1, N2 (checked), N3, N4, and IRIS-SETOS. A callout box with a white background and a black border contains the following text:

Select points by clicking on the **Select** button and dragging the mouse over the map.
You can change properties of displaying of the selected points in the **Selection** dialog.

VIDAEXPERT: DATA ANALYSIS

Click this to switch between **Map** and **Table** panels.

	N1	N2	N3	N4	IRIS-SETOS	ED
46	5.1	3.8	1.6	0.2	Iris-setosa	5.356821
47	4.6	3.2	1.4	0.2	Iris-setosa	4.822156
48	5.3	3.7	1.5	0.2	Iris-setosa	5.220302
49	5	3.3	1.4	0.2	Iris-setosa	4.887563
50	7	3.2	4.7	1.4	Iris-versicolor	6.494551
51	6.4	3.2	4.5	1.5	Iris-versicolor	6.296136
52	6.9	3.1	4.9	1.5	Iris-versicolor	6.515177
53	5.5	2.3	4	1.3	Iris-versicolor	5.464298
54	6.5	2.8	4.6	1.5	Iris-versicolor	6.105620
55	5.7	2.8	4.5	1.3	Iris-versicolor	6.146891
56	6.3	3.3	4.7	1.6	Iris-versicolor	6.447575
57	4.9	2.4	3.3	1	Iris-versicolor	5.200496
58	6.6	2.9	4.6	1.3	Iris-versicolor	6.283410
59	5.2	2.7	3.9	1.4	Iris-versicolor	5.598621
60	5	2	3.5	1	Iris-versicolor	5.081092
61	5.9	3	4.2	1.5	Iris-versicolor	5.951986
62	6	2.2	4	1	Iris-versicolor	5.567465
63	6.1	2.9	4.7	1.4	Iris-versicolor	6.298330
64	5.6	2.9	3.6	1.3	Iris-versicolor	5.572291
65	6.7	3.1	4.4	1.4	Iris-versicolor	6.215808
66	5.6	3	4.5	1.5	Iris-versicolor	6.165322
67	5.8	2.7	4.1	1	Iris-versicolor	5.965612
68	6.2	2.2	4.5	1.5	Iris-versicolor	5.642067
69	5.6	2.5	3.9	1.1	Iris-versicolor	5.636382
70	5.9	3.2	4.8	1.8	Iris-versicolor	6.340897
71	6.1	2.8	4	1.3	Iris-versicolor	5.791332
72	6.3	2.5	4.9	1.5	Iris-versicolor	6.122736
73	6.1	2.8	4.7	1.2	Iris-versicolor	6.345306
74	6.4	2.9	4.3	1.3	Iris-versicolor	6.070074

Here you can see the selected points as records in the datatable.

VIDAEXPERT: DATA ANALYSIS

The screenshot displays the VidaExpert software interface. The main window is titled "VidaExpert - [Map - Map002]". The menu bar includes "Project", "Scenario", "Data", "Map", "View", and "Window". The toolbar contains various icons, with a yellow arrow pointing to the "Annotation" icon. The main map area shows a grid of data points, with several points labeled with text such as "1.1/Iris-versicolor", "1.3/Iris-versicolor", "1.7/Iris-versicolor", "1.7/Iris-virginica", and "1.3/Iris-versicolor". The "Annotation" dialog box is open on the right side of the screen. It has a title bar "Annotation" and a close button. The "Points" section has three radio buttons: "Randomly selected" (with a value of 50 and a percentage sign), "By color", and "Selected" (which is selected). The "Parameters" section has a "Field names" list with checkboxes for "N1", "N2", "N3", "N4" (checked), and "IRIS-SETOS". Below this, there are radio buttons for "All fields" (selected), "Max values", "Min values", "Unexpected features", and "Unexpected in class". The "Depth" is set to 1 and "Font size" is 9. The "Max length" is 50. At the bottom of the dialog, there are "Apply", "Clear", "To table", and "Close" buttons.

Annotation dialog

You can assign for some points labels with their short description.
In **Annotation** dialog choose which points do you need to label, content of the labels and click **Apply**.