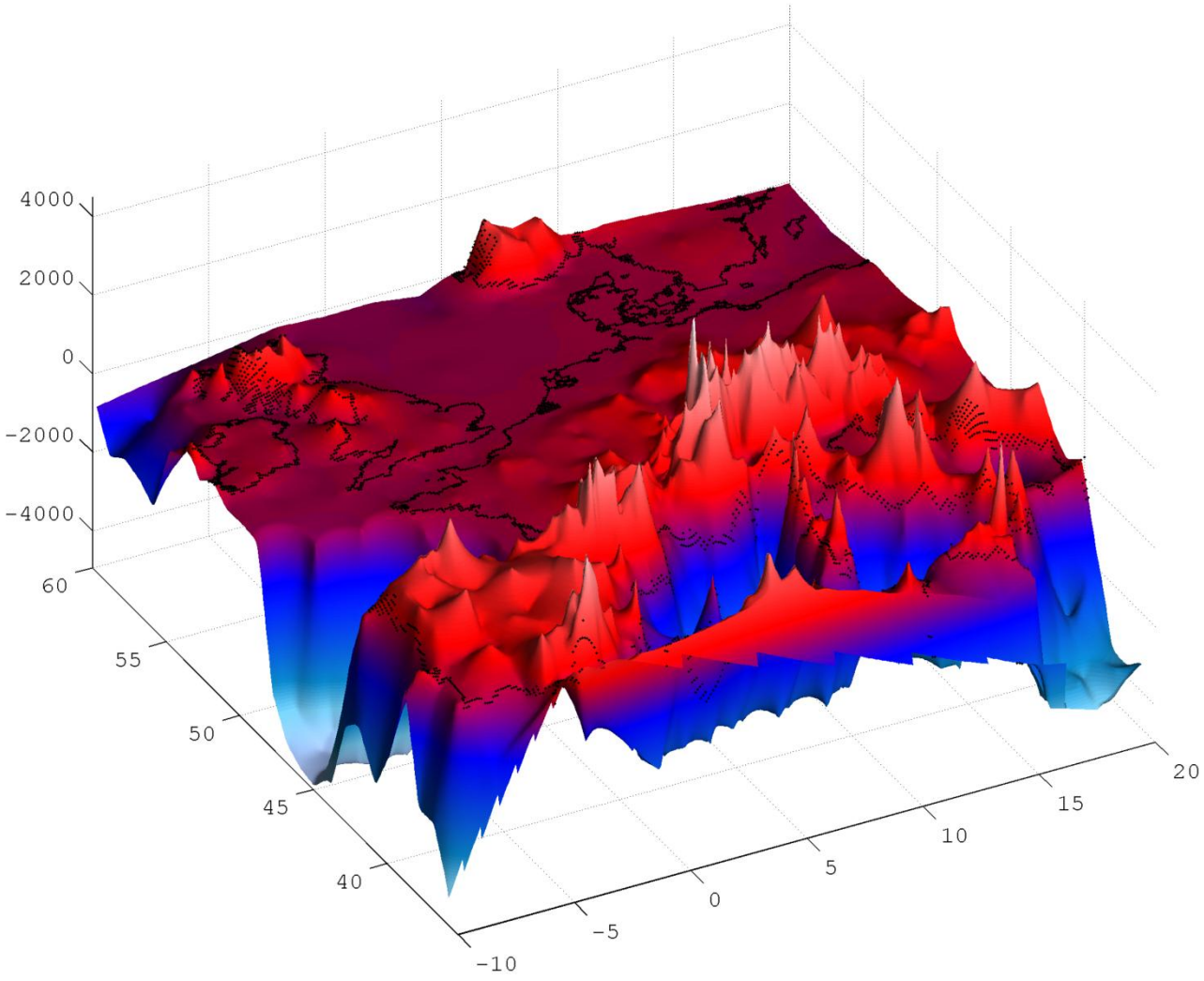
Sybil3 - HELP! Menu



Sybil3: A code for multivariate statistical analysis (method of principal component analysis) of sparse datasets.

Programmer: Leonardo Casini (2013)

Matlab version: r2010a

OS: Windows (win7)

Institution: University of Sassari

HELP! outline

1. Introduction 3

1.1. General information 3
1.2. Software version 3
1.3. OS requirements 3
1.4. Installation 4
1.5. Get started in 3 steps 4

2. Data management 5

2.1. Sybil3 database 5
2.1.1. Database structure 5
2.1.2. Data format 7
2.2. Upload the database 7
2.3. Select samples 8
2.4. Select variables 10
2.5. Append new data to the workspace 10
2.6. Clear the workspace 11

3. Statistical analysis 12

3.1. Cluster analysis 12
3.2. Interpolation of variables 13
3.3. Principal Components Analysis (PCA) 14

4. Results management 16

4.1. Results of cluster analysis 16
4.2. Display the results of spatial analysis (PCA, variables interpolation) 16
4.2.1. Select results type (PCA, variables) 17
4.2.2. Activate results 17
4.2.3. Set the domain of interpolation 18
4.2.4. Set visualization mode (2-3D) 18
4.2.5. Show samples (3D plots only) 19
4.2.6. Set color maps 19
4.2.7. Plot the results of spatial analysis 19
4.3. Export the figures 20
4.4. Export the results of PCA 21

1. Introduction

1.1. General information

Sybil3 is a Matlab® derived code specifically designed for data-mining and multivariate statistical analysis of large datasets of geo-referenced data, although analysis of 1D distribution of samples in single variables datasets is also permitted. Data mining is a set of techniques used to analyze huge datasets. Common data mining approach includes cluster analysis, principal components analysis and other methods that allow to: i) reduction of the dataset dimensionality and simplification of the data structure, making information more readable, ii) filtering of background signal, i.e., elimination of redundant data, iii) automated or semi-automated recognition of significant patterns, and extraction of the most relevant information from complex datasets. These methods are routinely applied by web search engine such as Google® as well as economists, to find out the most relevant websites related to few key words in the first example, and to delineate specific, consumer-oriented marketing strategies (i.e., based on huge databases of consumer preferences) in the latter case. The software is intended as an user-friendly package to apply data mining methods to Earth sciences problems. Although the amount of data usually handled in financial and marketing problems is much larger than that usually analyzed in geosciences, the spatial distribution of data makes the data mining approach potentially useful to reduce the noise, summarizing the information dispersed in several variables in few uncorrelated components that fully describe the total variance of the whole dataset. The maximum code allowance is 2500 samples, each described by up to 50 different variables. The architecture of the code allows to the analysis of sparse datasets; that is, different variables may be characterized by a different number of observations (samples).

(back to main menu)

1.2. Software version

The code has been written using Matlab® r2010a on a PC (O.S. windows 7). The Sybil3 version is the stable release, which would be implemented in future versions. Essential components of Sybil3 includes 11 '.m files' that control the functions used by sybil3 ('sybil3.m', 'get_category.m', 'im2vec.m', 'freqdist1.m', 'data_int.m', 'filt_1.m', 'data_read1.m', 'get_slct_sam.m', 'graph1.m','graph2a.m', 'mva.m', 'stat_report.m') , 9 '.mat files' that stores the default color maps available within Sybil3 ('binary.mat', 'grayscale.mat', 'moon.mat', 'lights.mat', 'magma.mat', 'mountain.mat', 'sharp1.mat', 'sharp2.mat', 'ice.mat'), 3 ' .jpg files' used to place data within a default geographic reference system ('none.jpg', 'world_low.jpg', 'world_prev.jpg'), 1 '.pdf file' ('read_me.pdf'), 1 '.xls file' used to store the user's inputs ('dataset.xls'), and finally 1 '.xls file' to save the outputs of PCA ('statistic_results.xls').

The software is provided with a user-friendly Graphical User Interface (GUI) controlled by the main script sybil3.m. The GUI allow the users to load their own data, perform calculations using the sub-functions ('freqdist1.m', 'mva.m'), and plot the results of statistical analysis simply by clicking on buttons.

(back to main menu)

1.3. OS/software requirements

The present version of Sybil3 requires a PC equipped with Windows 7 (or newer) and Matlab r2010a (or successive versions) . Note that Excel ® ('97-2003 or successive versions) and a PDF viewer are also required.

(back to main menu)

1.4. Installation

As Sybil3 is not compiled, the installation procedure is very simple. To be started, copy and paste the folder SYBIL3 and the files therein within a folder in your computer. After that, open Matlab and add the path to SYBIL3 folder to the current folder menu in the Matlab window (fig. 1). Once completed the installation procedure, Sybile3 is ready to start.

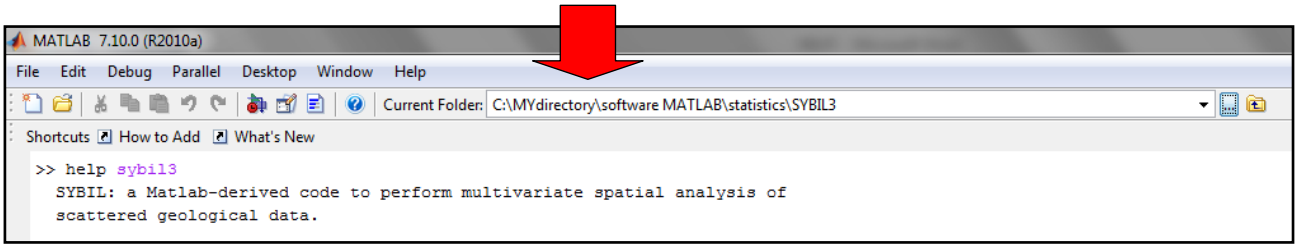


Fig. 1. Add the path to SYBIL3 folder in the current folder menu of Matlab

(back to main menu)

1.5. Get started in 5 steps

To be started follow this simple procedure: i) double click on the 'sybil3.m' file will automatically open the GUI with Matlab; if this does not occurs, open Matlab and set the current folder menu to the path to SYBIL3 folder, then type sybil3 at the command window and click enter, ii) load your own data into the dataset.xls table, following the instruction and guidelines for data formatting, iii) you're now ready to load the data and do calculations with the Sybil3 GUI.

NOTE: type **help sybil3** at the Matlab prompt to visualize the content of sybil3 folder. The list of sybil3 components is provided with a short description of the function behavior.

(back to main menu)

2. Data management

2.1. Sybil3 database

The Sybil3 database ('dataset.xls')is organized in 1 excel sheet consisting of 54 columns and 2500 rows. The first column (fig. 2) is reserved to samples ID. The second column is reserved to sample category, therefore numbers eventually entered in these fields are loaded as strings and stored as sample categories. Basically, you could enter as many samples categories as you need, however the use of unnecessary (redundant) labels would make the analysis less readable; it is therefore recommended to group samples into few relevant categories (i.e., mafic magmatic, felsic, metamorphic, carbonate, sand, and so on…). The third and fourth columns (Fig. 2) are also reserved to sample coordinates, which should be set following the instructions given in the data format section. Columns from 5 to 54 (fig. 2) may be filled with your data; therefore, a maximum of 125000 data values can be uploaded in a single excel file and passed to Sybil3 for analysis.

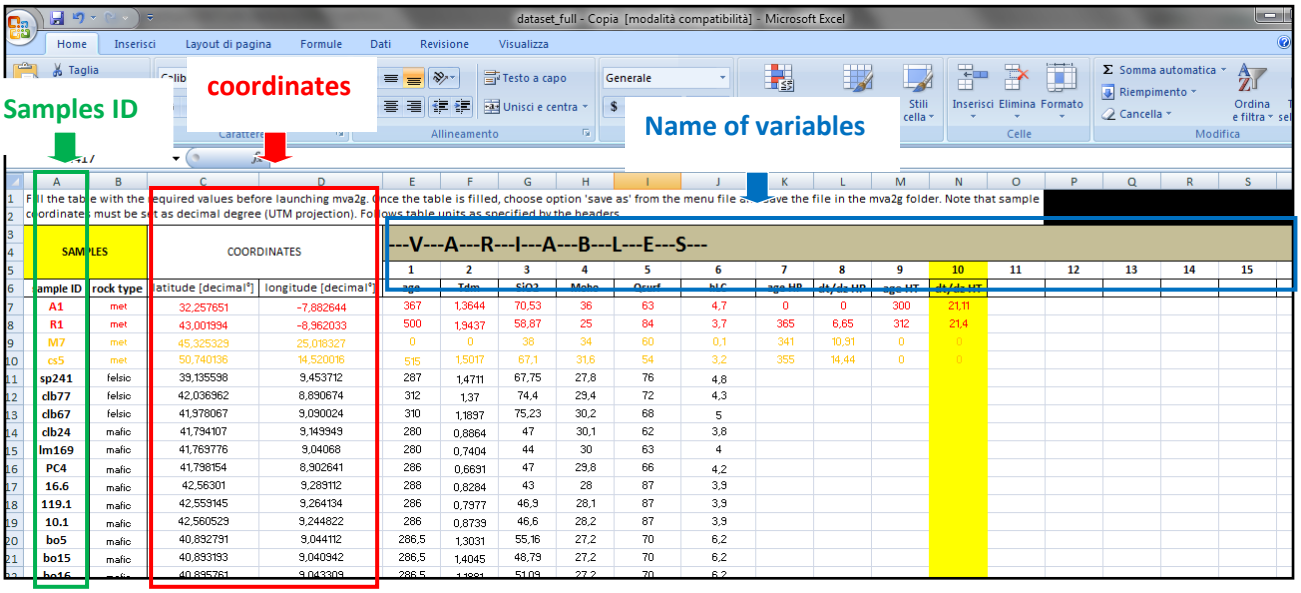


Fig. 2. Lookup of the Sybil3 database

There is an additional row of 50 elements (fig. 2) that might be optionally filled with the name of variables you used. As for samples ID, values eventually entered in this row are loaded as strings and stored as variable names (i.e., entering '1' will appear the first variable to be displayed as '1' in the Sybil3 GUI).

(back to main menu)

2.1.1. Database structure

The data values are loaded from dataset.xls spreadsheet as a 2500-by-50 matrix; two additional vectors of strings are extracted if the user enter any type of character into the appropriate fields for variable names and rock type. The database organization requires that each row represents a samples (i.e., a sample locality whose coordinates may be

common to all observations in the row), whereas each column represent a variable. Therefore, each column represents the value of some variable observed at n localities (maximum allowance is n = 2500 localities).

101	ic4-17	mafic	44,320544	9,883144	306	1,3031	49,6	36	63	5,1			
102	mt10	mafic	44,440809	9,907704	306	1,2143	47,3	36	60	5,12			
103	ic4-10	mafic	44,400987	9,947352	306	1,2879	48,8	36	61	5,1			
104	93202	felsic	40,765772	-4,16175	300	1,925	55	34,6	76	2,8			
105	93208	felsic	40,795828	-4,041471	300	1,9898	53	34,6	73	2,8			
106	90961	felsic	40,671552	-3,72011	300	1,9712	59	34,6	75	2,7			
107	V2	met	45,935808	8,290559			64	58	63	10			299
108	t370	felsic	40,918137	-3,962507	300	1,5622	65	34,6	76	2,8			
109	92193	felsic	40,848577	-3,877974	300	1,3567	64	34,6	78	2,7			
110	95916	felsic	40,652327	-4,007627	300	1,5168	69,67	34,6	75	2,8			
111	56205	felsic	40,865569	-3,627126	310	1,5351	69,09	34,6	79	2,6			
112	cs23	met	50,314143	16,302766	520	0,7341	45,7	32,4	55	4,2			
113	cs2	met	50,535589	14,410275	520	0,7743	61,8	32	55	4,2			

Fig. 3. Example of sparse dataset

Note that the design of Sybil3 is specifically designed for Earth sciences and environmental projects where the spatial distribution of observations is usually heterogeneous while dealing with multivariate problems. Sparse datasets are also allowed (fig. 3), thus there is no need to have always the same number of variables known at each sampling locality (i.e., some samples may have just one variable known while others may be fully described by 15 variables). SPECIAL RESTRICTIONS: running the code will return an error at the Matlab command window unless these special restrictions are satisfied: i) first, you should place the samples with maximum and minimum coordinate values in the rows from 1 to 4 (red-orange fields); basically, you should have a minimum of 2 samples (first 2 rows) if their position coincide with that of two opposite edges of the square domain enclosing the dataset and a maximum of four samples (first four rows) if none of them fit any of the square domain edges (fig. 4), **ii) THESE SAMPLES (row 1 to 4) MUST BE CHARACTERIZED BY THE HIGEST POSSIBLE NUMBER OF VARIABLES, i.e., if the overall number of variables in your dataset is 20, the first 2 to 4 samples must contain values for each of the 20 variables**, iii) both coordinate fields of any sample must contain valid characters (numbers, comma-separated up to six decimal places).

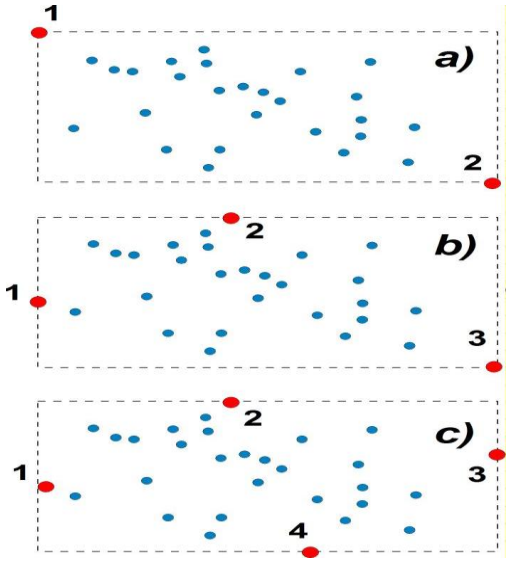


Fig. 4a. Possible distribution of samples within a square domain

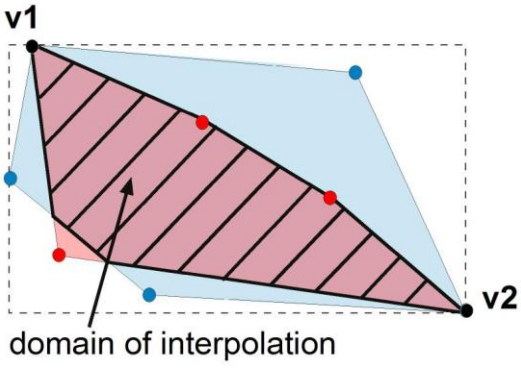


Fig. 4b. Effect of heterogeneous sample distribution on the size of the interpolated domain

Fig. 4b shows an example of the possible shortcomings due to evaluating variables characterized by very different distribution: although both variables are correctly defined at the edges of the domain (in this case, points V1 and V2), using datasets with very different variables distribution such as those indicated in fig. 4b (red and blue shaded regions) may result in a quite small overlapping domain. This effect may be reduced to a minimum by defining the full set of variables at 4 different places, i.e. (case 'c' in fig. 4a).

(back to main menu)

2.1.2. Data format

Before running experiments you may upload your datasets using the dataset.xls template provided within the SYBIL3 folder. To do this, open the dataset.xls spreadsheet and fill the tables as required. Note that variables are organized in 50 columns, while rows represents observation of the variables at a given location whose coordinates are specified in the first two columns. The values of each variable should be set as numbers, using comma as separator of decimal places. Only numeric values are allowed in the fields of variable values, but letters or a combination of numbers and letters is allowed in the table headings (sample ID and name of variables columns). Once the table(s) is filled, you may save your database as 'your_database_name.xls' using the save as option. Note that the algorithm used to evaluate the matrix of component loadings used in Principal Component Analysis (PCA) is programmed to convert each dataset to standard Z-scores by subtracting to each value in a column the median value and dividing by the standard deviation. This scale the data of each variable so that standardized scores have a median value of 0 and a standard deviation of ± 1. The process of standardization makes possible to compare variables spanning over different orders of magnitude (i.e., the Moho depth [km] and the spacing of metamorphic foliation [mm to μm]). The value of coordinates must be set as decimal degree using up to 6 decimal places; this ensures a 1.1 km precision of interpolants.

(back to main menu)

2.2. Upload the database

Once filled and eventually renamed the dataset.xls file, you may upload your data within Sybil3 automatically. To do this, click the 'load' button in the data management panel of the Sybil3 GUI (fig. 5). Then,

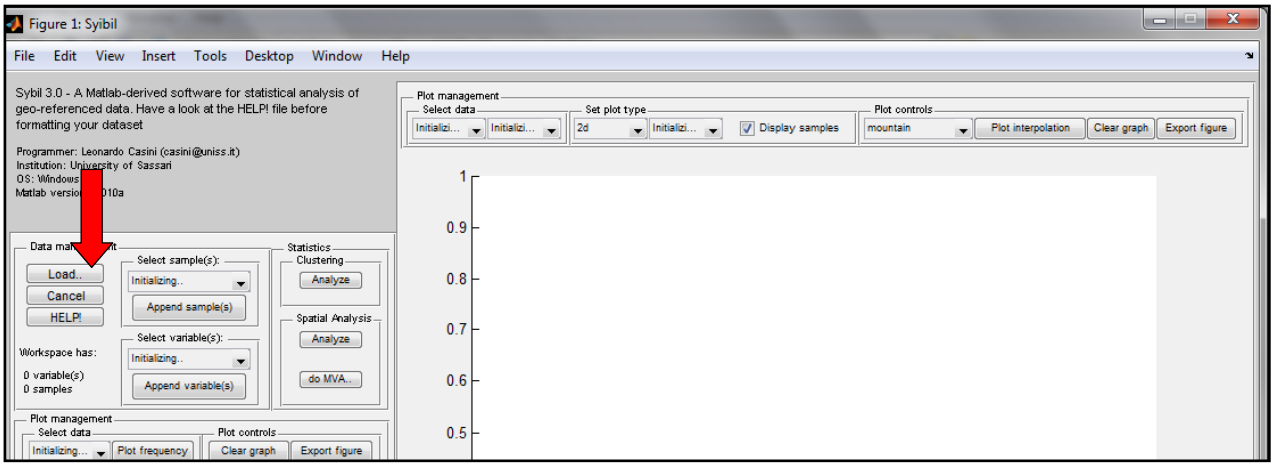


Fig. 5. Load the dataset file

The routine 'data_read1.m' opens a dialog box (fig. 6) which ask for dataset name; type the name of the .xls file where you've saved your dataset and click 'ok'.

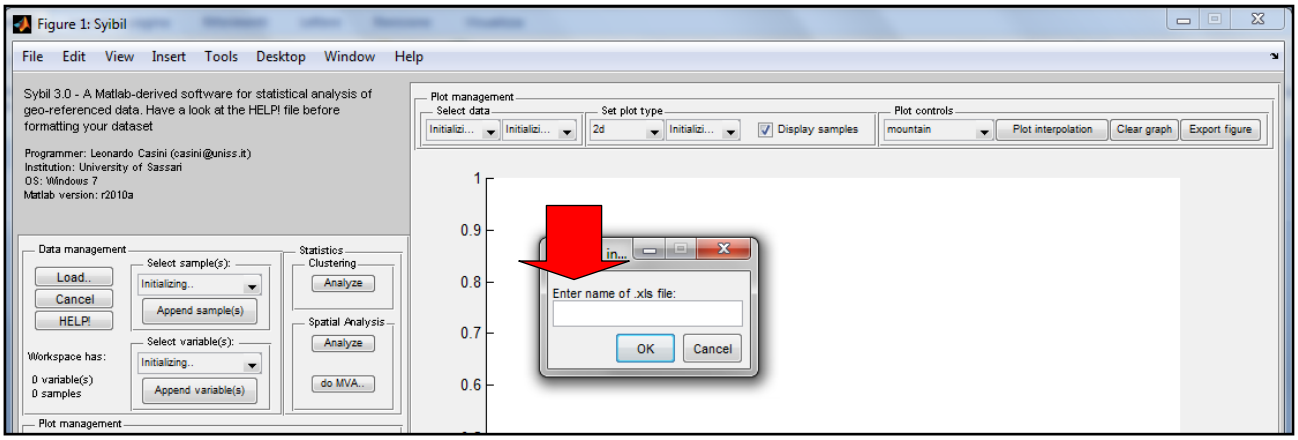


Fig. 6. Enter name of your dataset

Your database, including values and labels of both samples and variables, is now activated and ready to analyze.

(back to main menu)

2.3. Select samples

After data uploading, the pop-up menu in the select sample(s) panel become active. All labels that you've assigned to your samples is now visible in the pop-up menu (fig. 7).

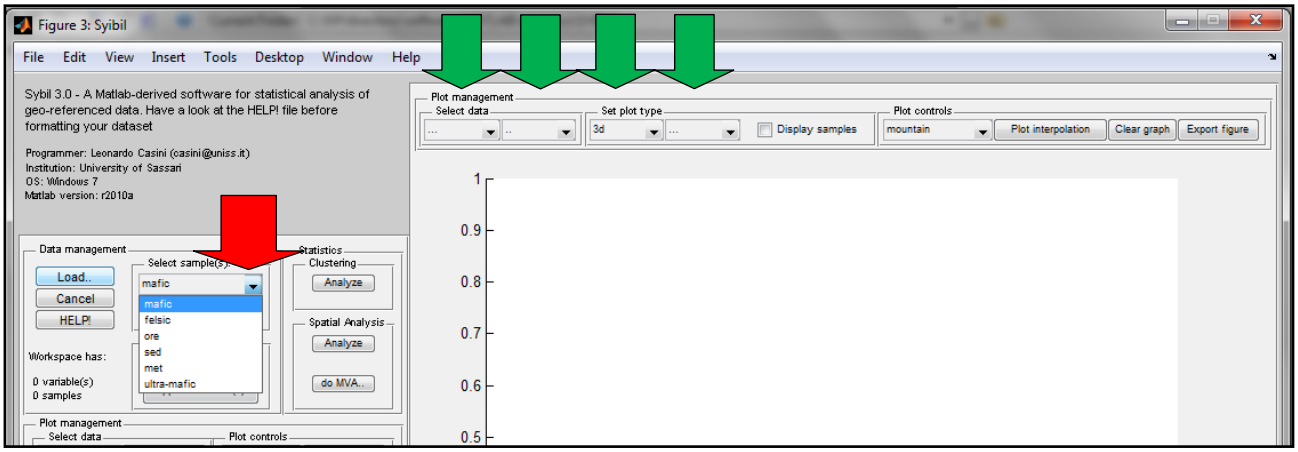


Fig. 7. Select samples menu (green arrows indicate inactive pop-up menus)

To select all samples from a category (i.e., all samples labeled the same way), just click on the category name. This activate the 'get_slct_sam.m' routine that extract the samples you're interested in from the database. Once a selection is made, the name of the selected sample category is removed from the pop-up menu. After the first selection has been made, you may select other categories repeating the procedure as many time as required, until all samples are selected. Once you select all the required sample categories, click on the 'append sample(s)' button to activate your selection (fig. 8).

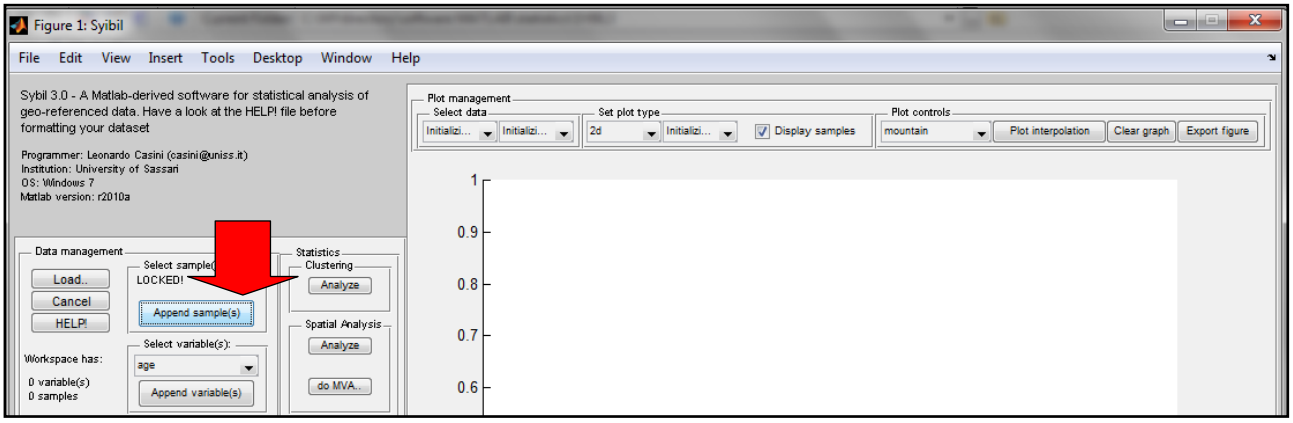


Fig. 8. Activate selected samples

Now the selected sample categories are active and no further selection is possible within the session; the pop-up menu disappear from the GUI and is replaced by the note 'LOCKED!'. Note that this step is required to go on with calculations. In fact, clicking on the 'analyze' and 'do MVA' buttons at this stage will return an error at the Matlab command window:

```
sybil3

??? Index exceeds matrix dimensions.

Error in ==> sybil3>analyze1button_Callback at 490

        htext7 = uicontrol('Style','text','String',...
```

```
??? Error while evaluating uicontrol Callback

??? Index exceeds matrix dimensions.

Error in ==> sybil3>analyze1button_Callback at 490

        htext7 = uicontrol('Style','text','String',...

??? Error while evaluating uicontrol Callback

Warning: popupmenu control requires a non-empty String

Control will not be rendered until all of its parameter values are valid
```

Note also that the pop-up menus on the right 'plot management' panel are still inactive (fig. 7).

Once sample categories have been loaded, the software wait for inspection of 1D samples distribution: clicking any button unless at least one variable is evaluated in the first graph will return a warning (fig. 9):

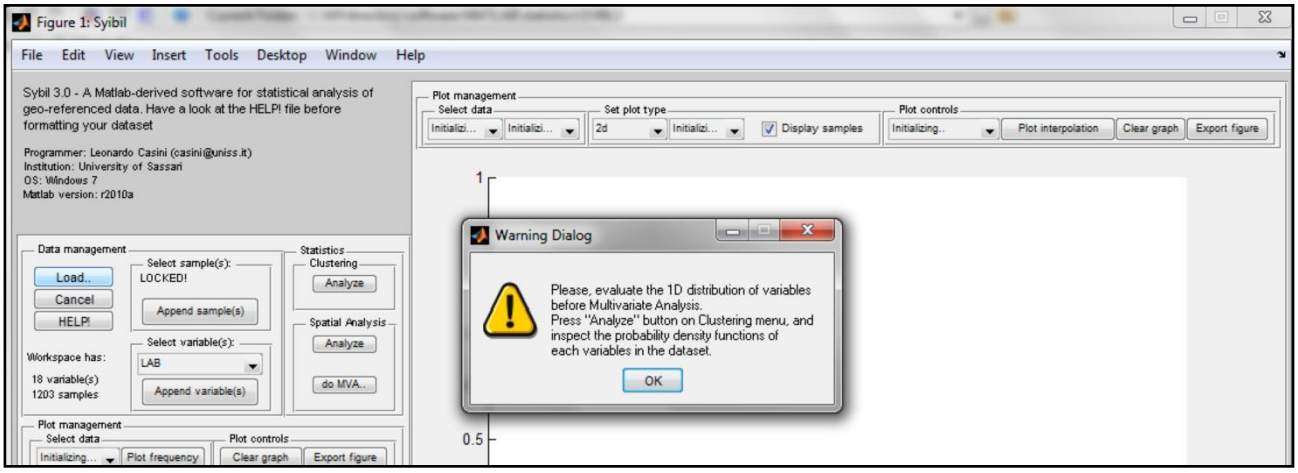


Fig. 9. Warning message

(back to main menu)

2.4. Select variables

Once the selected sample categories have been activated, the pop-up menu in the 'select variable(s)' panel is active. You may select as many variables as you want, simply by clicking on the variable name the same way as for the sample selection step. This call the 'get_category.m' sub-function that extract the selected variables from the database. As for samples, click on the 'append variable(s)' button when the requested variables have been selected (fig. 10).

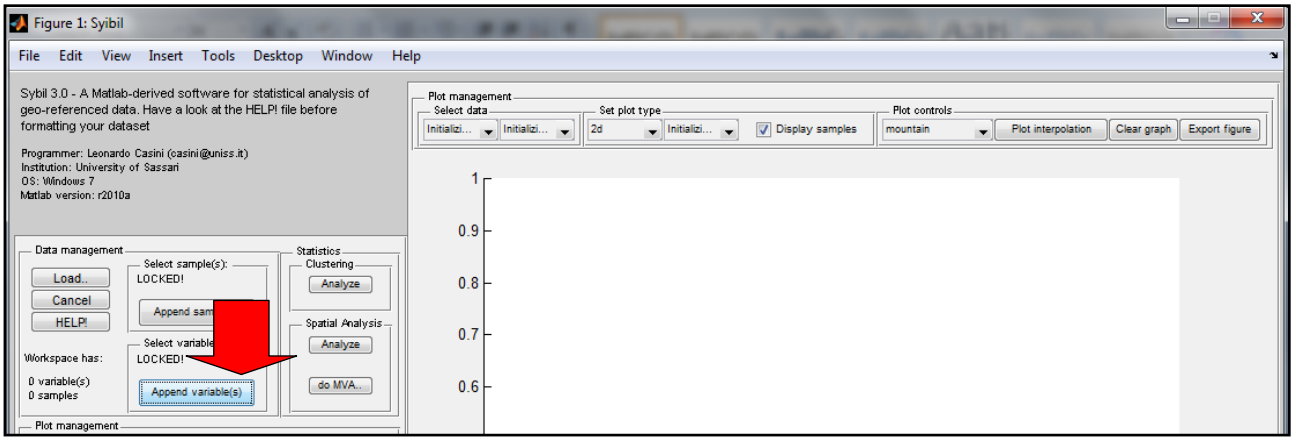


Fig. 10. Activate selected variables

Now the selected variables are active and no further selection is possible unless the 'cancel' button in the data management panel is pressed causing re-initialization of Sybil3.

(back to main menu)

2.5. Append new data to the database

Although Sybil3 is designed to upload a maximum of 125000 variable values (50 variables and 2500 samples) each session, you may append additional data to your database simply loading other 'your_file_name.xls' files during the same Sybil3 session.

(back to main menu)

2.6. Clear the workspace

The 'cancel' button in the data management panel allow the users to clear the current workspace before starting new sessions (fig. 11). This action will remove from the Sybil3 workspace all data previously active. Once the 'cancel' button is pressed, all the pop-up menus and the dynamic text windows are initialized again and a new session can start.

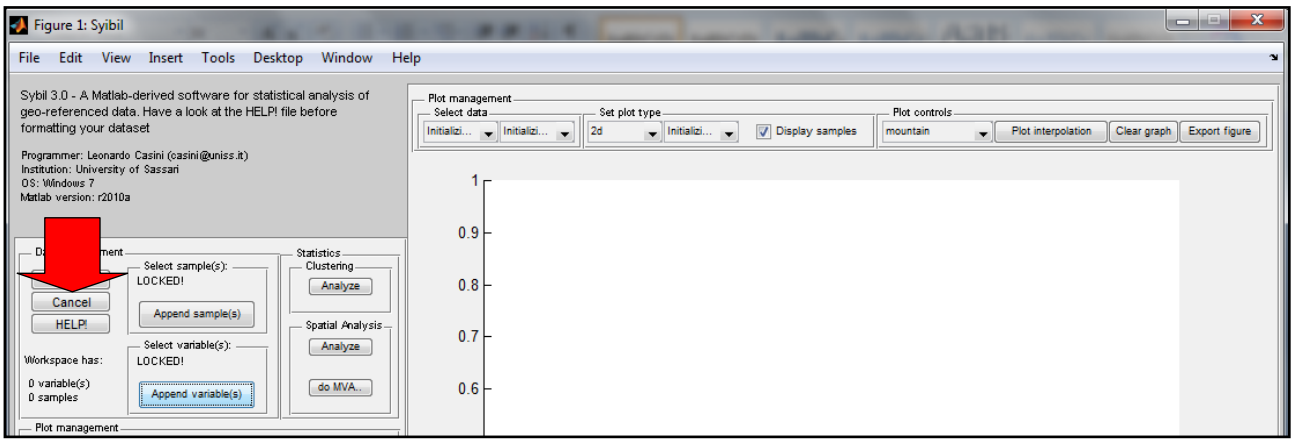


Fig. 11. Clear the workspace before running new session

(back to main menu)

3. Statistical analysis

This chapter deals with some aspects of Sybil3 functioning with emphasis on mathematical and technical background.

(back to main menu)

3.1. Cluster analysis

Cluster analysis is performed simultaneously on the currently active variables by clicking on the 'analyze' button (fig. 12) in the clustering panel (right in the data management panel). Once the analysis is completed, a message 'done!' appears just below the 'analyze' button.

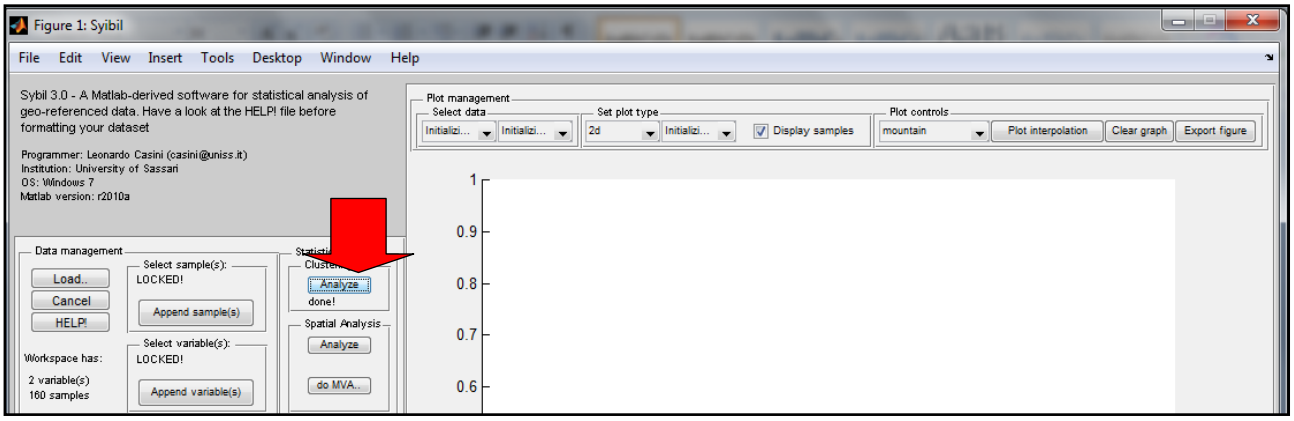


Fig. 12. Do cluster analysis

Note that now the status of text boxes in the left side of the data management panel has changed showing the current number of active variables and samples.

Cluster analysis is performed by the sub-function 'freqdist1.m'. That function uses a simple search algorithm that allow to find group of samples that are distributed around a common median. In a first step, the algorithm evaluate the distribution density (p_ρ^i) of a given samples population, i.e., N observations of some variable, as:

$$p_\rho^i = \frac{100 * n^i}{N} \quad (1)$$

where N is the total number of samples within the population, and the term n^i represents the bin density of samples, that is number of samples within 100 bins of 0.01*N size defined between the minimum and maximum values of population N. The expression for n^i may be written as:

$$n^i = \sum_{i=1}^{100} s < bin^i \quad (2)$$

where s is the number of samples whose values are lower than a given bin bin^i in the interval. Once the distribution density of a given sample population N is defined, the search algorithm is automatically activated by the sub-function 'filt_1.m'. That routine uses the Matlab built-in 'clusterdata.m function', which calculates the Euclidean distance *dist* between pairs of data in a given 2D space with coordinates (x,y) and (a,b) as:

$$dist_{\{(x,y),(a,b)\}} = \sqrt{(x-a)^2 + (y-b)^2} \quad (3)$$

Then, 'clusterdata.m' creates a hierarchical tree structure from the Euclidean distance vector, based on the shortest distance method. Finally, 'clusterdata.m' extract the three most significant clusters from the agglomerative hierarchical cluster tree. The median and standard deviation of each three clusters is finally calculated and outputted by the 'filt_1.m' routine.

(back to main menu)

3.2. Interpolation of variables

Interpolation of selected variables is performed simultaneously on the currently active variables by clicking on the 'analyze' button (fig. 13) in the spatial analysis panel (just below the statistics panel). As before with cluster analysis, once calculation has done a message 'done!' appears just below the 'analyze' button.

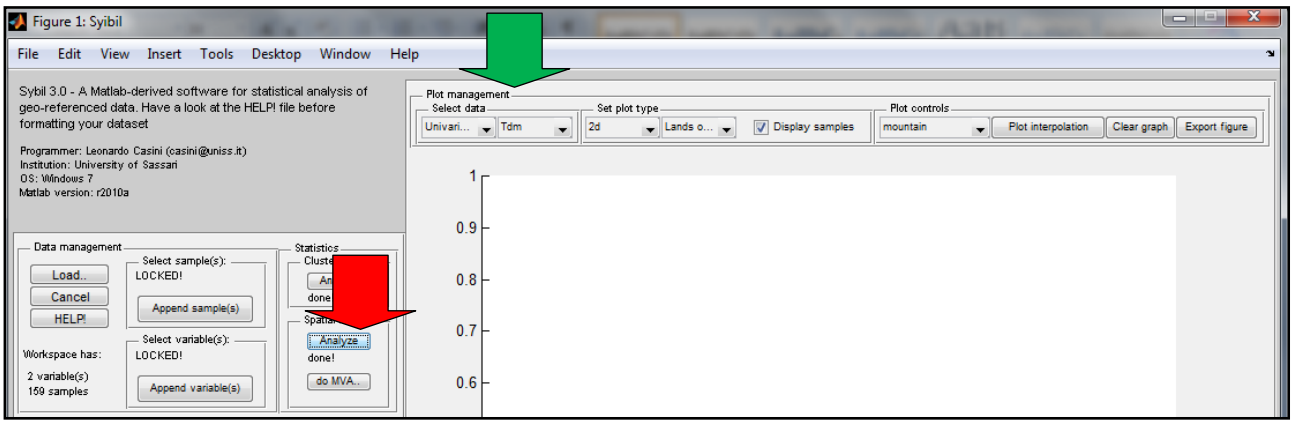


Fig. 13. Interpolate the variables over a common domain

Note that now the pop-up menus in the plot management panel (above the main plot of Sybil3, green arrow in fig. 112)have been initialized. At this stage, the first pop-up menu in the select data panel shows only one option, 'univariate' as PCA has not been done yet. The second pop-up menu shows instead the name of selected variables, for which plotting commands are enabled.

Interpolation of variables over a given spatial domain is performed by the sub-functions 'data_int.m' and 'im2vec.m'. The 'data_int.m' script first construct K m-by-n matrices which defines the spatial domains of each of the selected variables. The size of matrices is sensitive to sample distribution; that is, the number of elements within each matrix are defined as:

$$\begin{cases} m^i = 10 * \left[\max\left(var^i_{(x)}\right) - \min\left(var^i_{(x)}\right)\right] \\ n^i = 10 * \left[\max\left(var^i_{(y)}\right) - \min\left(var^i_{(y)}\right)\right] \end{cases} \quad (4)$$

where $var^i_{(x)}$ and $var^i_{(y)}$ are vectors of longitude and latitude, expressed as decimal degree, of the i*th* variable. Note that a given amount of samples distributed within a 10*5° region corresponds to a 100-by-50 matrix and so on. Once domains of interpolation have been defined, the function 'data_int.m' interpolates the datasets separately, creating Delaunay triangulation (i.e., a triangulation in which no points of the data set is contained in the circum-spherical volume for any triangle of the 3D surface). At this point, the 'im2vec.m' function automatically extract from the reference images 'none.jpg', 'world_low.jpg', or 'world_prev.jpg' a grayscale image as p-by-q matrix. Note that p = m,q = n, therefore the reference image is automatically sized as the domain of variables interpolation.

(back to main menu)

3.3. Principal Component Analysis (PCA)

The analysis of principal components is the main advantage Sybil3 has respect to other commercial software routinely used to interpolates variables over spatial domain, either in a geo-referenced framework or not. PCA analysis is performed by the function 'mva.m'. The routine may be activated simply clicking on the 'do MVA' button in the spatial analysis panel (fig. 14).

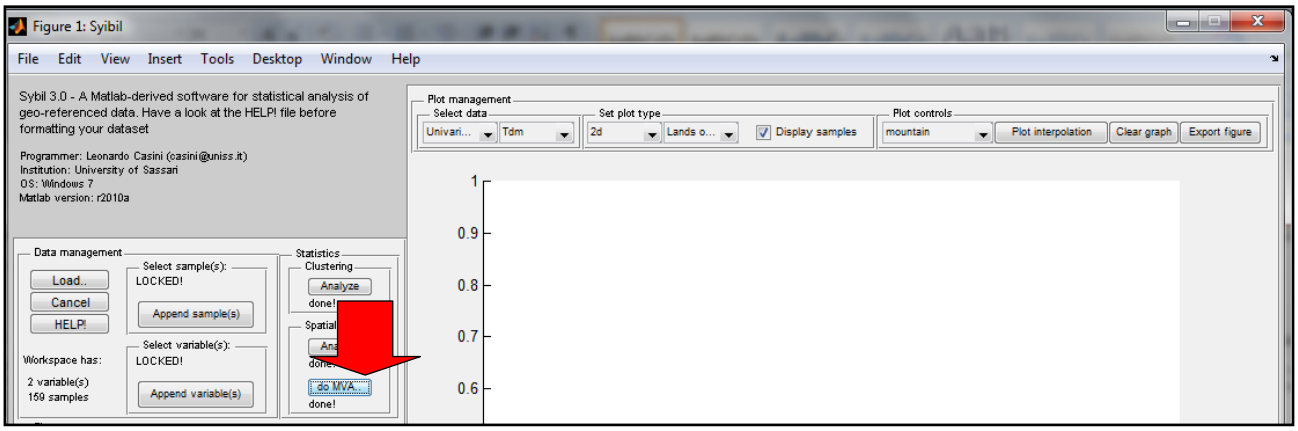


Fig. 14. Perform Principal Component Analysis

Once the 'do MVA' button is pressed, the sub-function 'mva.m' first transform the values of selected variables into standard scores (v_j^i) as:

$$v_j^i = (val_j^i - V^i)/stdV^i \quad (5)$$

where (val_j^i) are j observations of the ith variable, V^i is the median of and ($stdV^i$) is the standard deviation of the ith variable, respectively. The 'mva.m' function then calculates the i-by-i coefficient matrix and extract the significant eigenvalues, that is, eigenvalues greater than 1. Two cases are possible: i) there is only one eigenvalue greater than 1, or ii) there are multiple results. In the first case, a message box (fig. 15a) informs the user that Principal Component Analysis is not significant; the results will not be shown and the pop-up menus in the

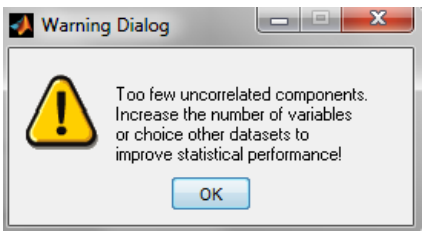


Fig. 15a. Warning message; it is shown when PCA analysis returns less than 2 Principal Components

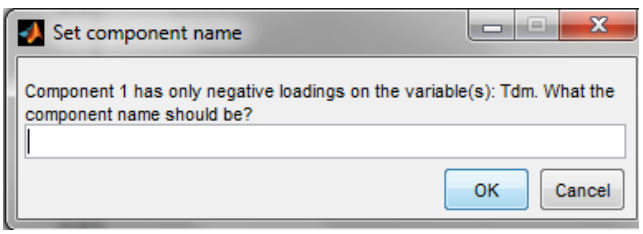


Fig. 15b. input dialog box; Sybil3 informs about the loadings each principal component have, and ask for name of that principal component.

plot management panel are not updated. If the latter option holds true more than 1 principal component can be defined; in this case, a set of message boxes (fig. 15b)inform the user about the loadings that each of the principal components has on the relevant variables, and ask for that principal component name. Once typed a name, click on the 'ok' button; a new message box will be displayed until all principal components have been labeled. At this moment, the pop-up menus on the plot management panel are updated (fig. 16) and the user can plot the results of both single variable interpolation and PCA.

(back to main menu)

4. Results management

This chapter explain how to plot the results of cluster analysis, interpolation of single variables, and Principal Component Analysis. The different settings of each plot, and the way Sybil3 allow to export the graphic objects are also illustrated.

4.1. Results of cluster analysis

To show the results of cluster analysis, first select the variable to be displayed from the pop-up menu in the plot management panel on the right (fig. 16a), just beneath the data management panel, then click on the plot frequency button (fig. 16b).

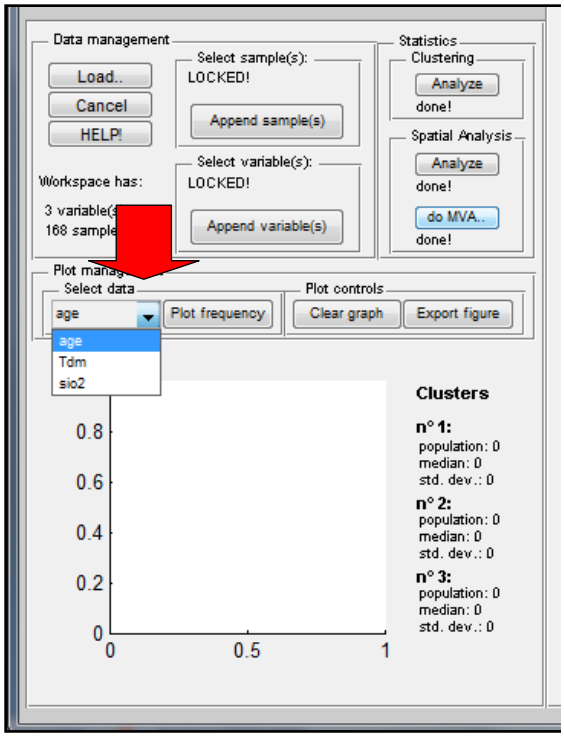


Fig. 16a. Selecting results of cluster analysis to be displayed in plot 1 (choose one option from the pop-up menu)

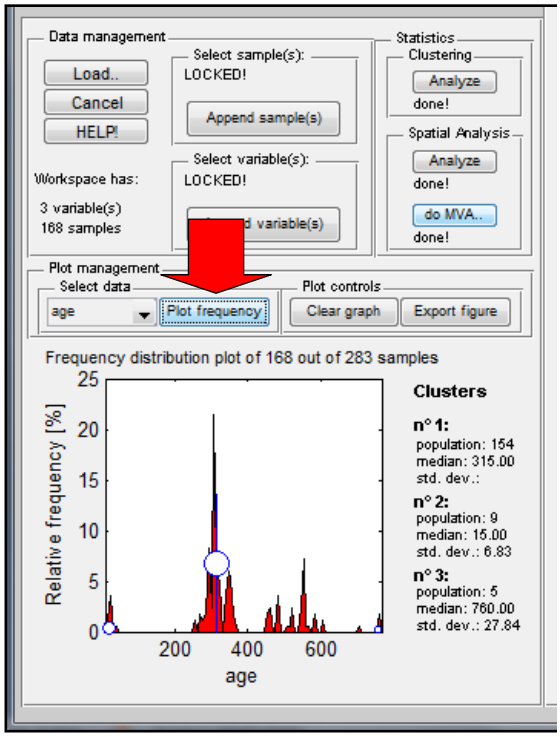


Fig. 16b. Plot the results for selected data (cluster analysis)

Note that the sample population, median value and standard deviation of each cluster are displayed right of the plot; these values are automatically updated as new variables are selected from the pop-up menu. To clear the plot, click on the clear graph button in the plot controls panel.

(back to main menu)

4.2. Display the results of spatial analysis (PCA, variables interpolation)

This section explain how to select the results of spatial analysis to be plotted in the main Sybil3 window, either interpolation of single variables or Principal Component Analysis.

(back to main menu)

4.2.1. Select results type (PCA, variables)

Before plot the interpolants, you may select the type of results to be displayed in the main plot of Sybil3, on the left side of the GUI. If principal component analysis returned more than one principal component, the first pop-up menu shows two options: i) 'univariate', and ii) 'PCA' (fig. 17). If PCA analysis returned only one principal component, the only option in the pop-up menu is 'univariate'.

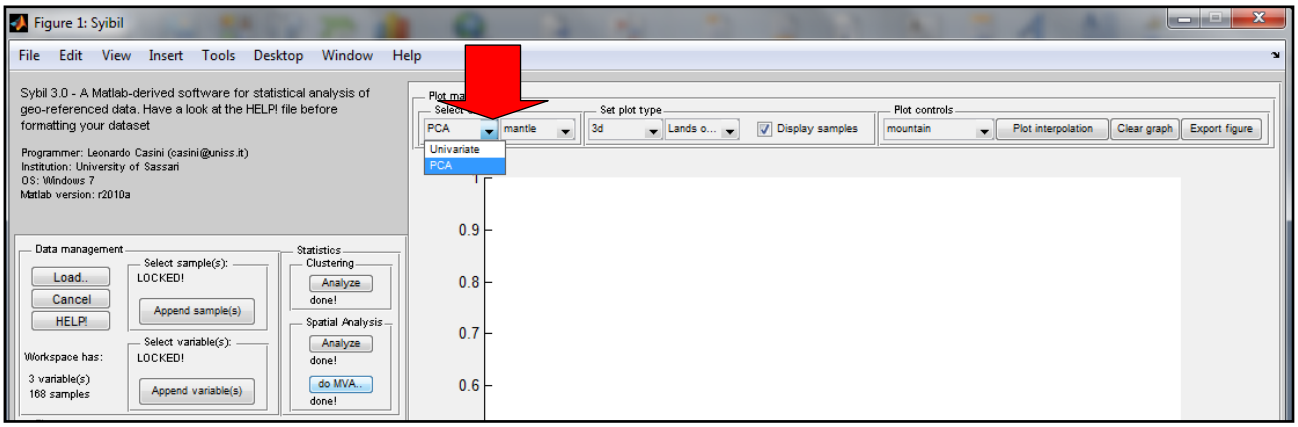


Fig. 17. Select type of results to be displayed

(back to main menu)

4.2.2. Activate results

Once selected one of these type of results, the second pop-up menu in the select data panel is updated showing the available datasets, either variable interpolants or Principal Component scores. To activate a specific dataset, select the required category from the pop-up menu(fig. 18).

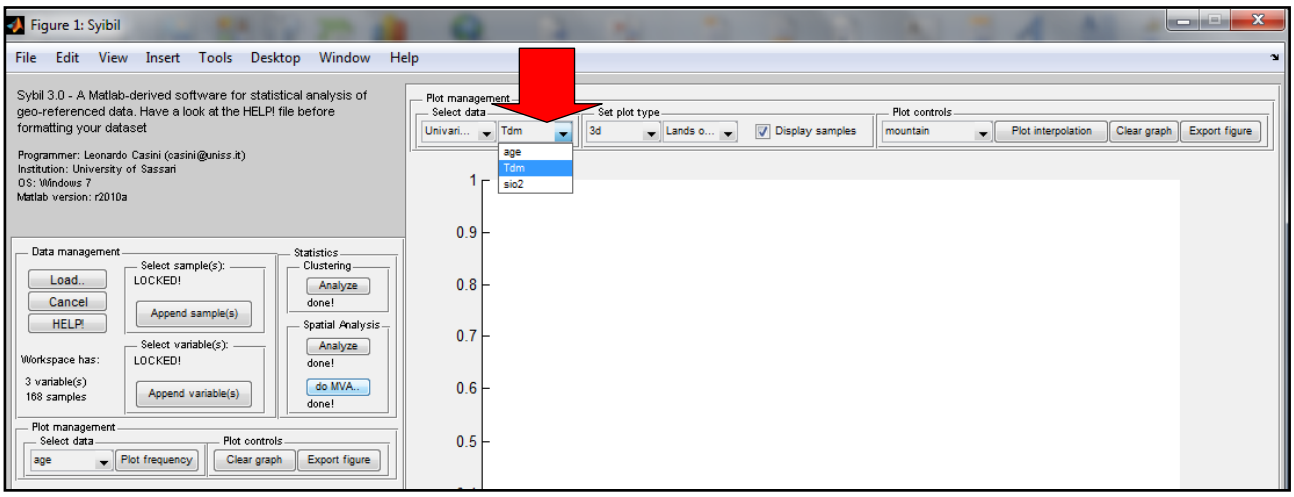


Fig. 18. Select category to be displayed in the main plot

(back to main menu)

4.2.3. Set the domain of interpolation

After a selection has been made, you may decide to plot the results of spatial analysis either as interpolation over the lands, or as interpolation over the spatial domain of interest including both lands and seas. To choose between one of these options, click on the second pop-up menu in the set plot type panel (fig. 19), and select either 'lands only (option 1), or 'all' (option 2). Note that interpolation over the whole domain is allowed only for the results of single variable interpolation; the reason for this is that Principal Component Analysis is performed only onto emerged areas to avoid statistical artifacts due to extrapolating the value of variables (used to calculate the coefficient matrix) over the seas where no samples are usually available.

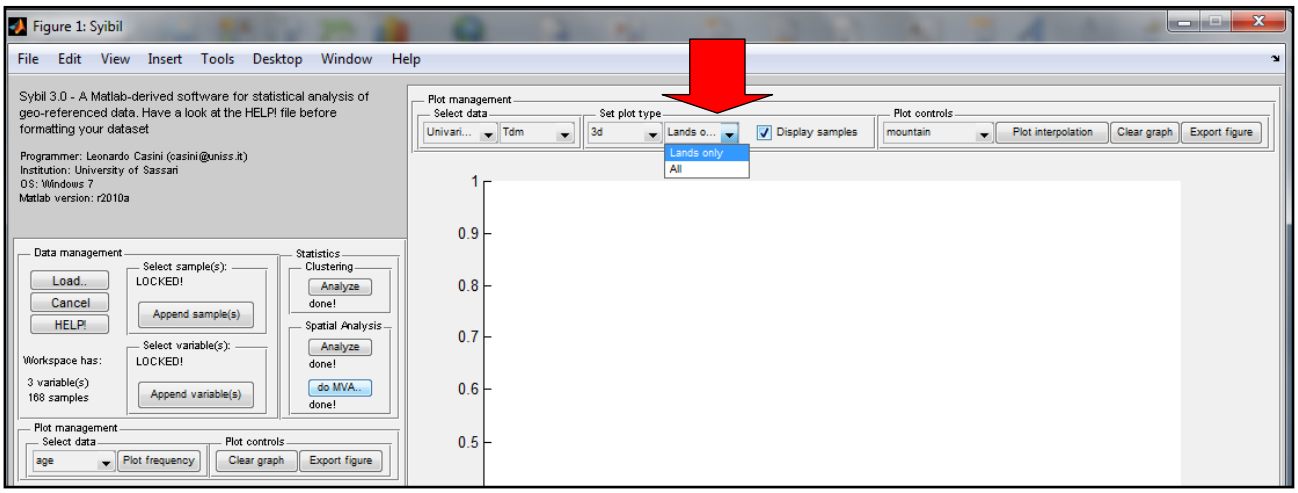


Fig. 19. Select the domain of interpolation

(back to main menu)

4.2.4. Set visualization mode (2-3D)

You may set the visualization mode, either 2d or 3d, selecting one option from the first pop-up menu in the select plot type panel(fig. 20)

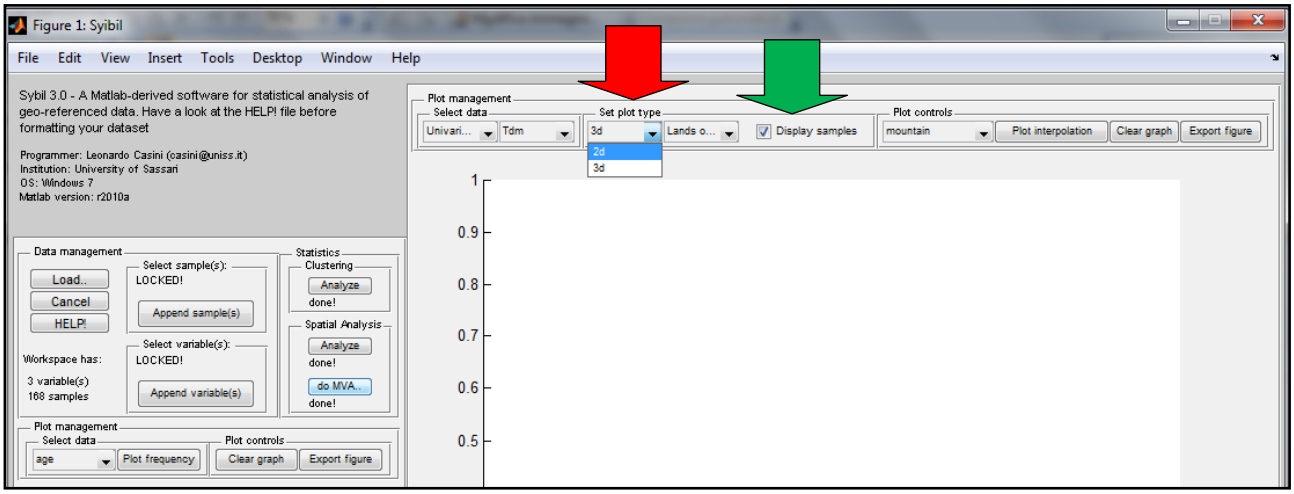


Fig. 20. Set the visualization mode, either 2d or 3d (red arrow); option to plot the position of selected samples (green arrow).

If a 2d option is activated, the results are displayed as filled contour maps; on the other hand, if a 3d option is on, the results are shown as shaded surfaces in a 3d graph. Note that the z values are proportional to the value the selected variable has at any given point in the interpolated surface.

(back to main menu)

4.2.5. Show samples (3D plots only)

If a 3d option is on, you may decide to show both the interpolated surface and the position of the currently active samples by selecting the check box in the set plot type panel (fig. 20).

(back to main menu)

4.2.6. Set color maps

Default for maps and interpolated surfaces is gray scale; however, after a plot has appeared in the window you may change its color scale by selecting one of the nine option from the first pop-up menu in the plot controls panel (fig. 21).

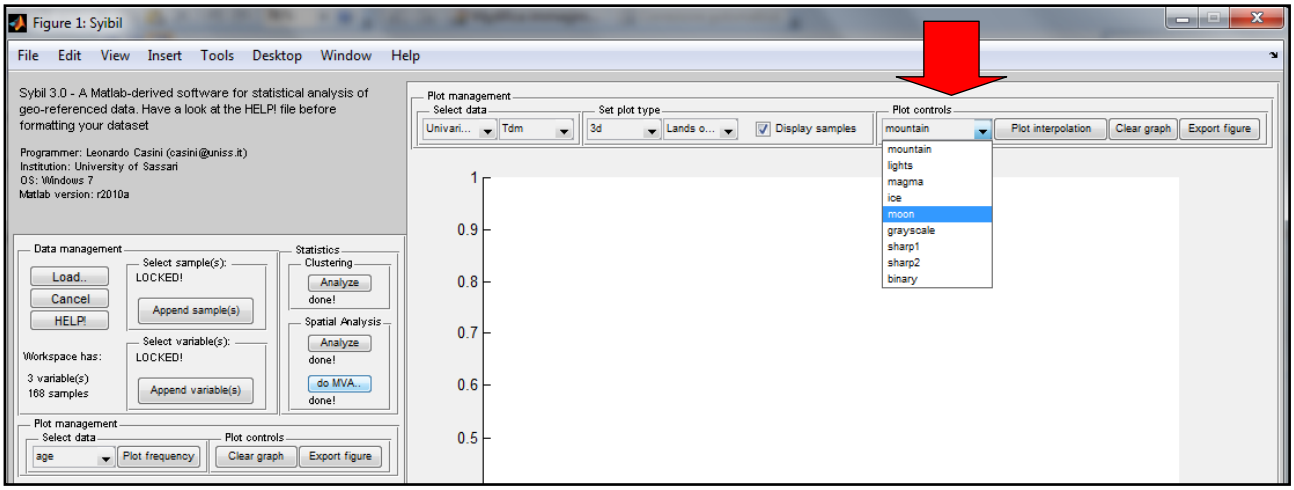


Fig. 21. Set the color scale.

(back to main menu)

4.2.7. Plot the results of spatial analysis

Now that all settings have been defined, you're about to display the results of spatial analysis! To show the results, just click on the second pop-up menu in the plot controls panel (i.e., 'plot interpolation', fig. 22) and see how your results looks like.

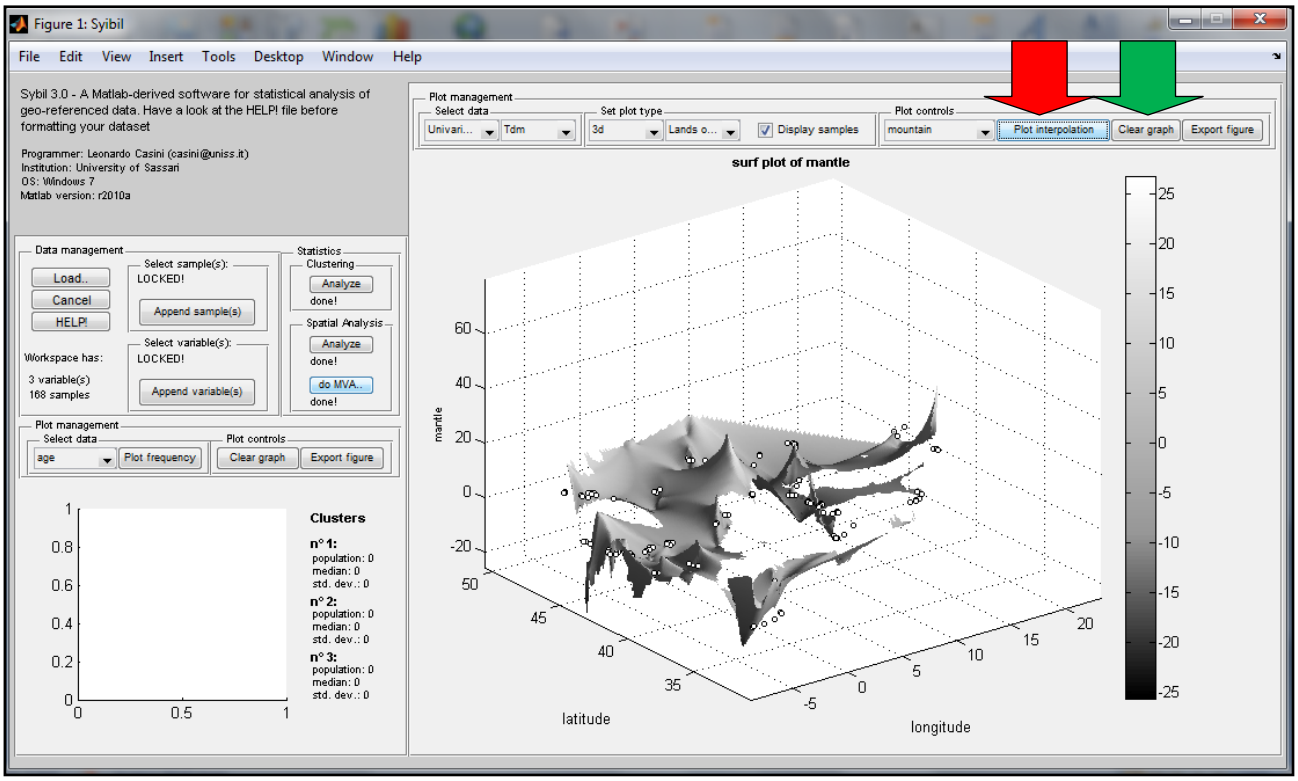


Fig. 22. Display the results of spatial analysis (red arrow) and clear the graph as required before next plot (green arrow). Note that default color is gray scale; you may change the color scale choosing one of the nine options from the pop-up menu on the right of the 'plot interpolation' button.

You may cancel the plot simply by clicking on the 'clear graph' button in the plot controls panel.

(back to main menu)

4.3. Export the figures

You may save high-resolution graphics of the results (both spatial analysis and cluster analysis) using the predefined figure GUI of Matlab. To do this, click on the 'export figure' buttons. The plot is now open in a new window; select the option 'save as' from the GUI 'file' menu to save the image in several graphic formats, i.e., pdf, jpg, bmp, tiff, wmf and so on (fig. 23). Note that the new window allow the user to zoom on/off the figure and, for 3d plots only, rotate the point of view and 3d navigation of the image using the commands of the menu bar (fig. 24).

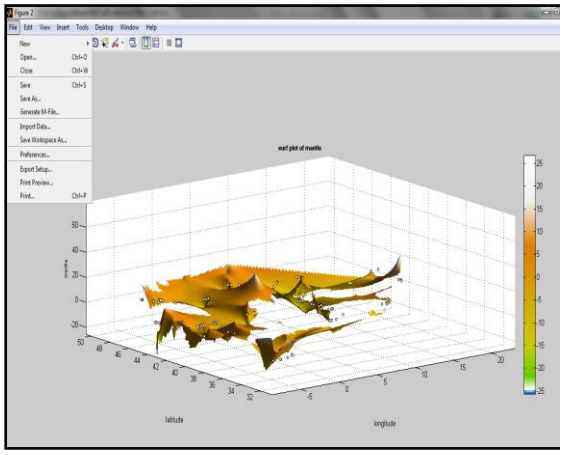


Fig. 23. Save the plots.

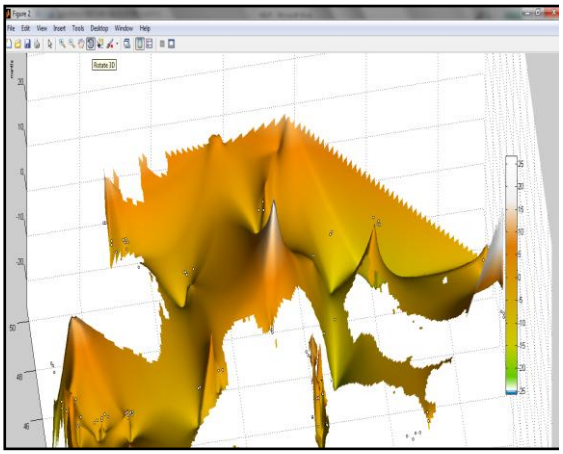


Fig. 24. Navigation of 3d plot

[illegible] an excel table (spreadsheet #1 in excel file **statistic_results.xls**) by the sub-routine 'stat_report.m'. The table shows, from left to right: i) the non-rotated matrix of principal component loadings, ii) the eigenvalues vector sorted from largest to smallest eigenvalue, iii) the percent of total variance explained by each non-rotated component, and iv) the rotated component loadings matrix.

The software is programmed to overwrite data, therefore it is recommended to save each file statistic_results.xls with a different name before starting a new sybil3 session.

(back to main menu)