

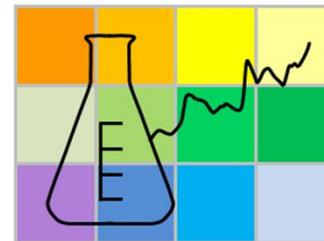


**UNIMORE**

UNIVERSITÀ DEGLI STUDI DI  
MODENA E REGGIO EMILIA

Dipartimento di Scienze della Vita

CHIMSLAB



Chemometrics Imaging  
and Spectroscopy Lab

# COLOURGRAMS GUI V2.2: FAST USER MANUAL

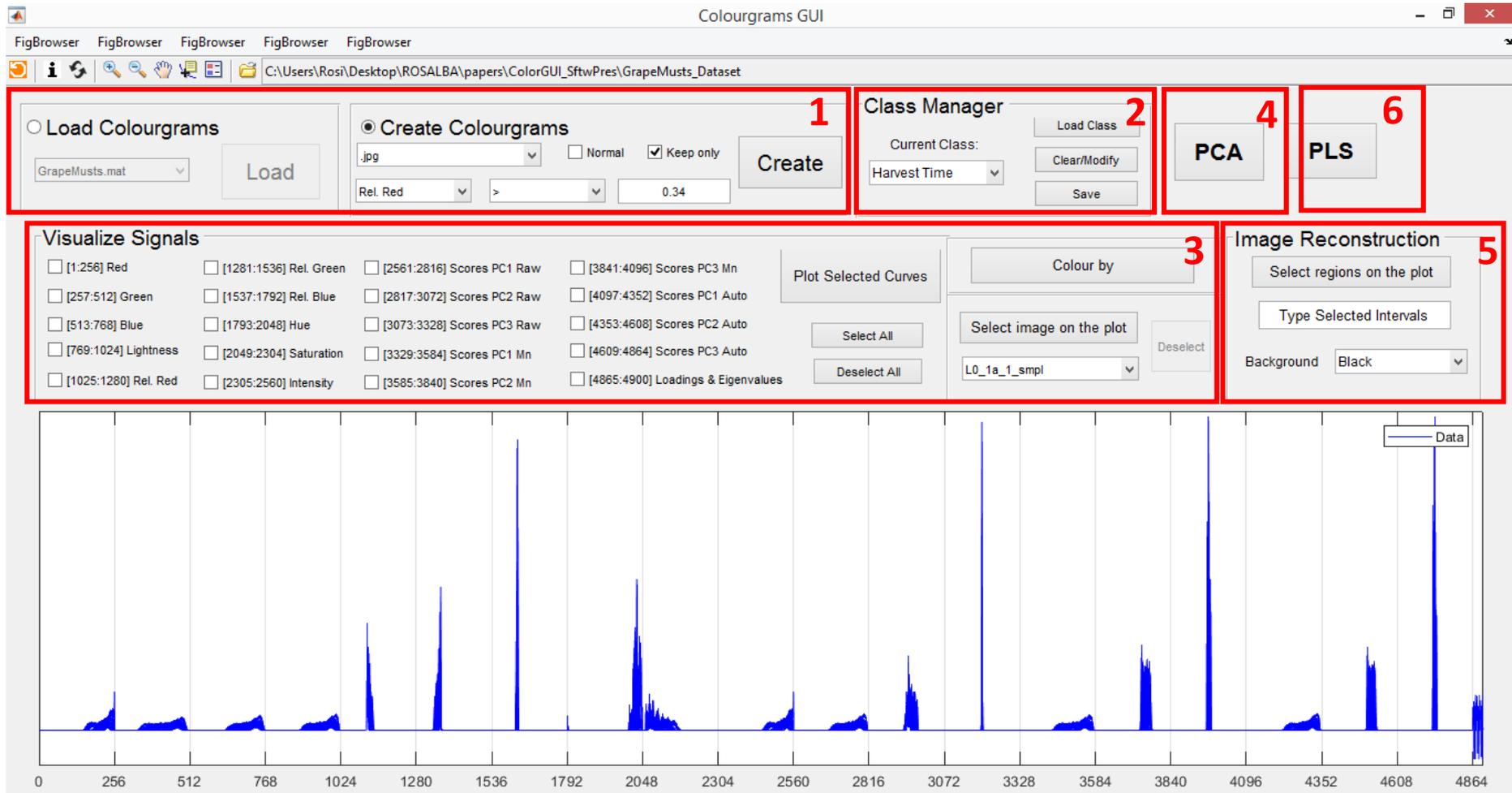
# Getting started

Colourgrams GUI can be easily installed in few steps:

1. Unzip the folder “SoftwareColourgramsGUI\_v2.2”.
2. Download the zip file of the DataSet Object from this link: <http://se.mathworks.com/matlabcentral/fileexchange/39336-dataset-object> .
3. Unzip the folder “dataset” and save it inside Colourgrams GUI main folder “ColourgramsGUI\_2.2”.
4. Open MATLAB, add the folder “ColourgramsGUI\_2.2” and the corresponding subfolders to the MATLAB path (File -> Set path -> Add with subfolders).
5. To get started, type colourgrams\_GUI in the MATLAB command window. The main window of Colourgrams GUI will be displayed and ready to use!!

For any question, problem or technical assistance don't hesitate to send us an e-mail at [chimslab.unimore@gmail.com](mailto:chimslab.unimore@gmail.com).

# The Colourgrams GUI main window

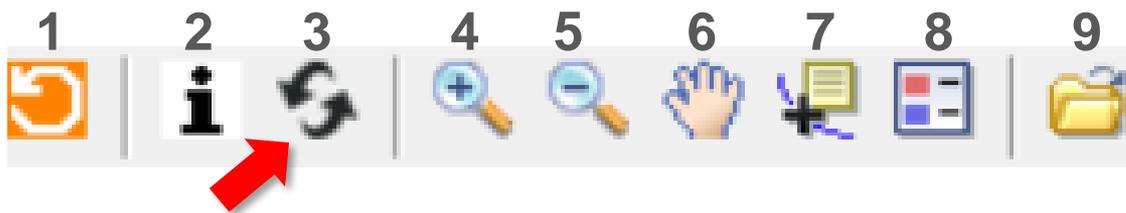


1. Load or create colourgrams dataset.
2. Add supplementary information.
3. Visualize signals.

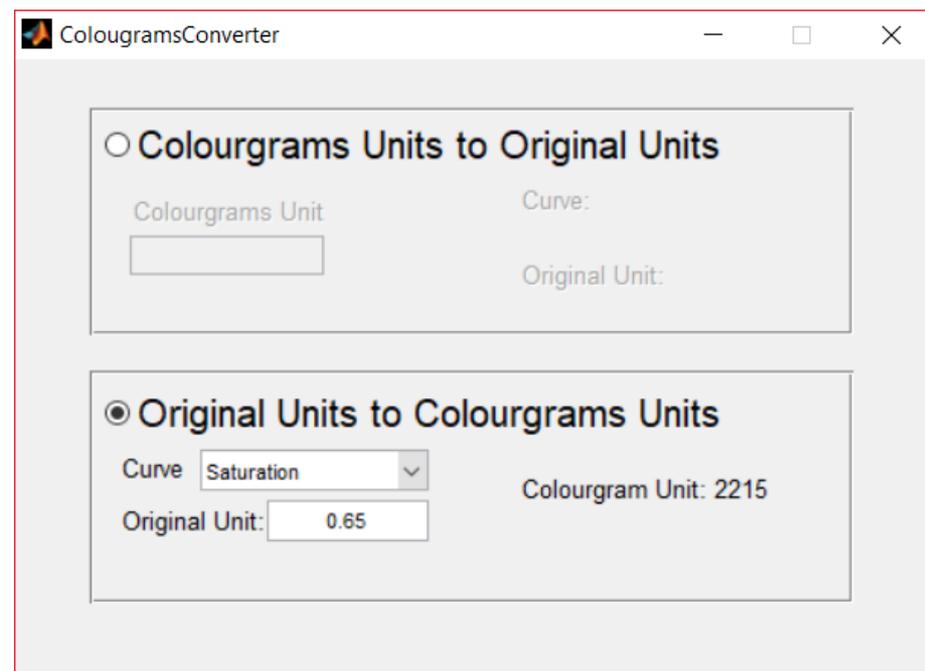
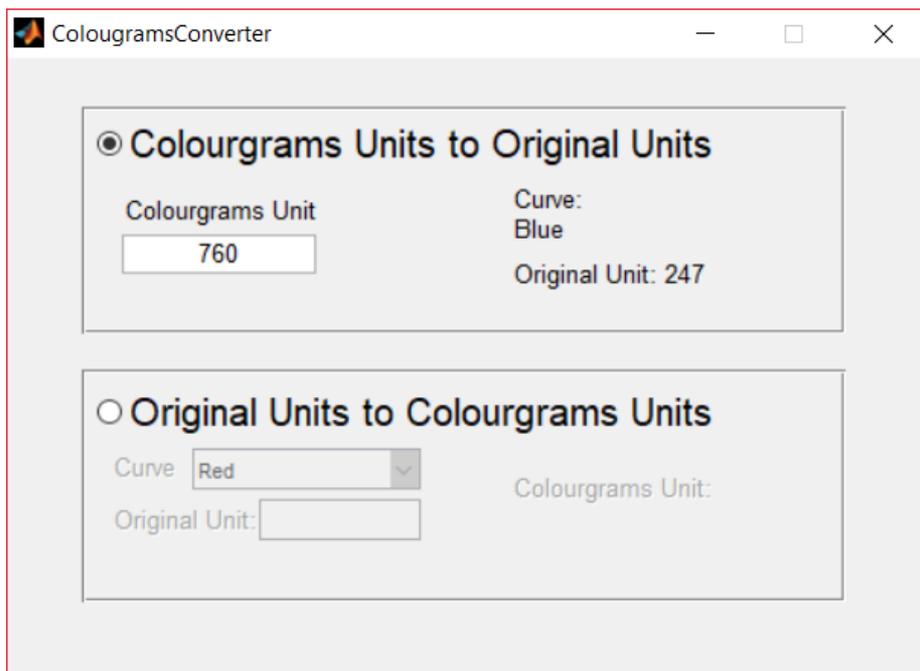
4. Perform PCA.
5. Reconstruct images using features of interest.
6. Perform PLS.

# Toolbar – Colougrams GUI

1. Restart
2. Dataset Info
3. Converter



4. Zoom in
5. Zoom out
6. Pan
7. Data Cursor
8. Insert Legend
9. Select Folder



The converter allows the fast conversion from colourgram units to original units (and vice versa), facilitating the interpretation of the signals.

# Restart

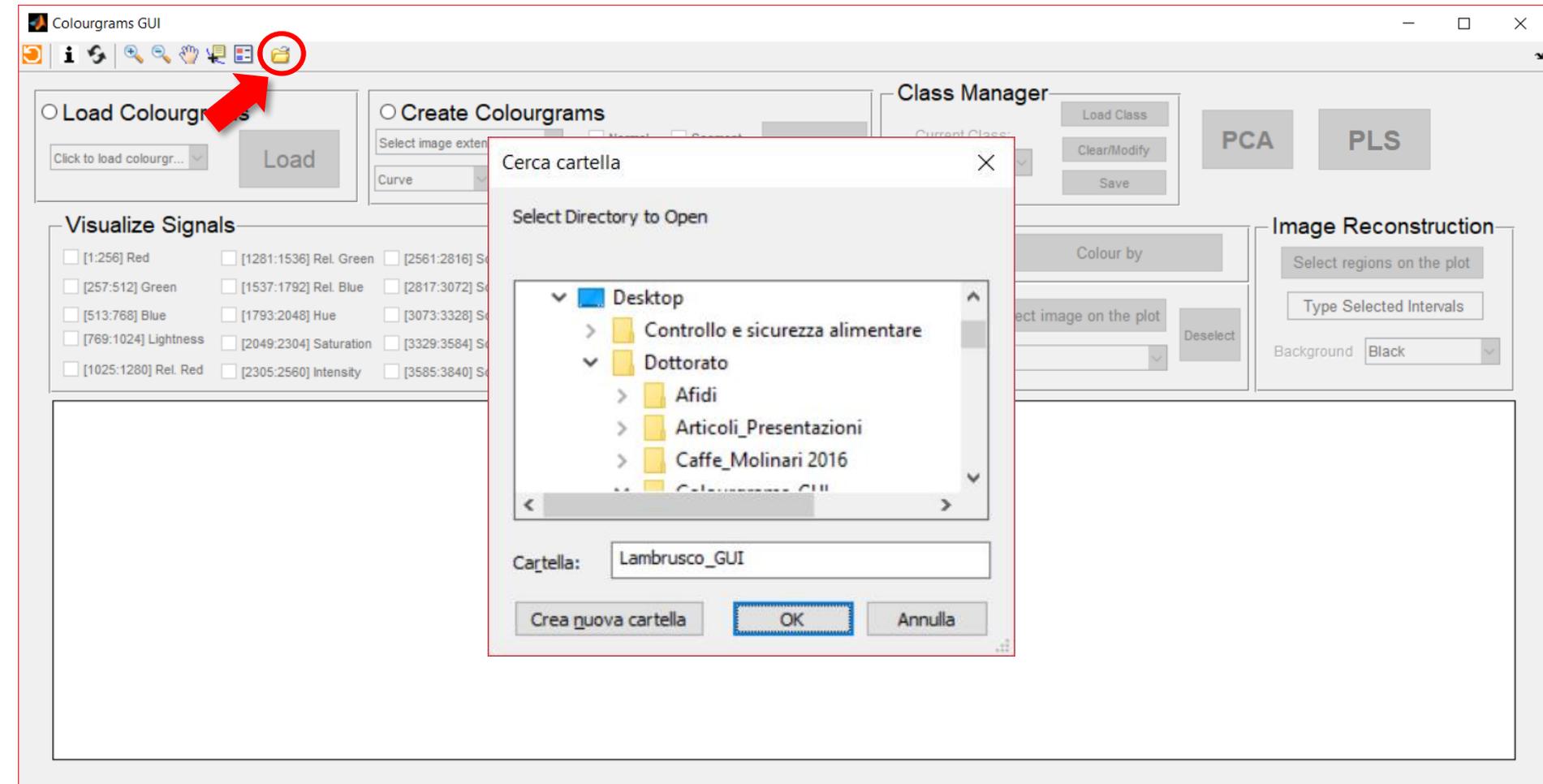


By clicking the “Restart” button, it is possible to start a new session.

Please, pay attention because this button **closes all figures.**

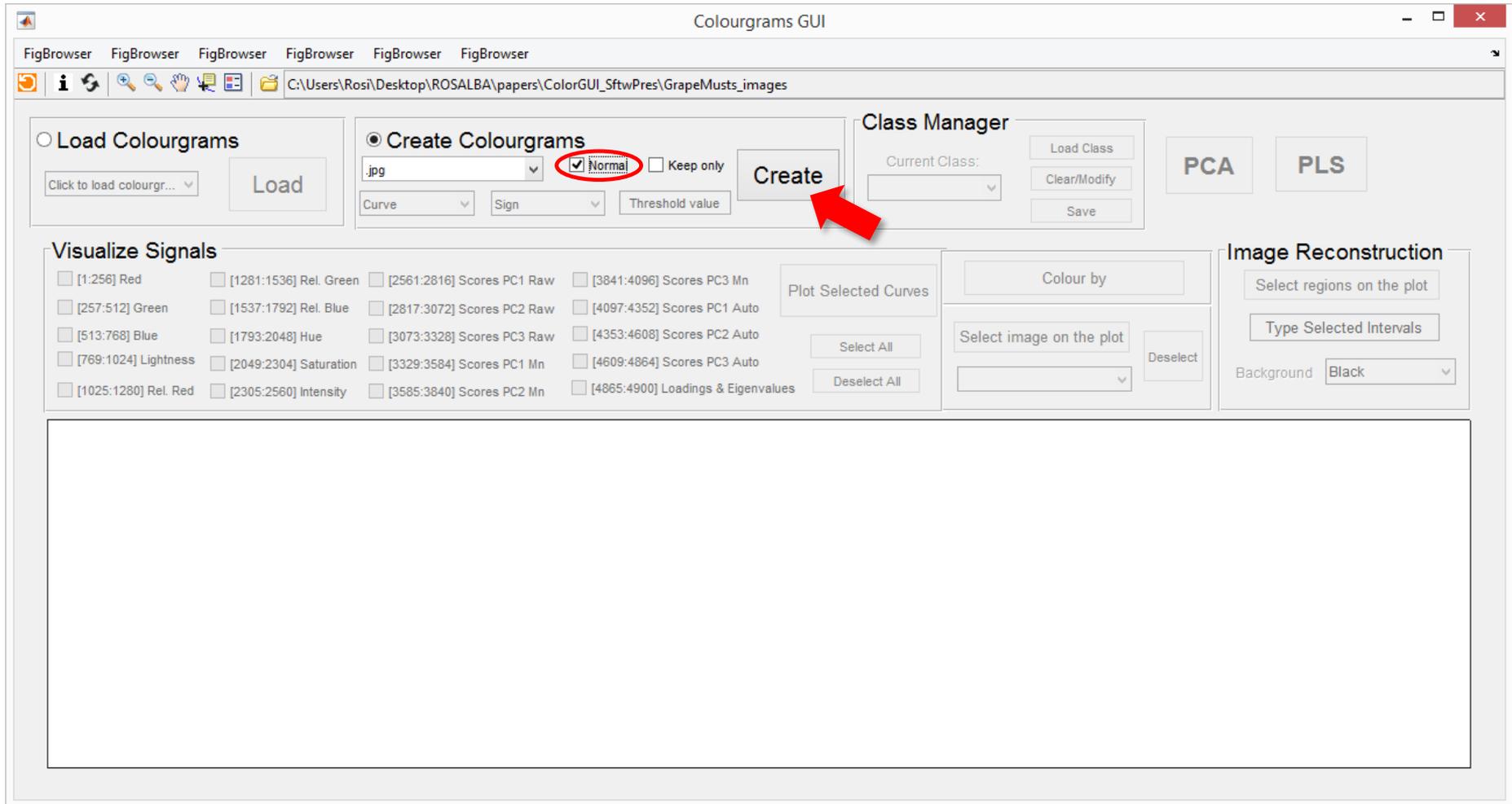
# Selecting folder

By clicking on the folder icon in the toolbar, the user selects the folder containing the images to be converted into colourgrams or a previously saved dataset. Once the current directory is selected, the corresponding path will be displayed in the toolbar.



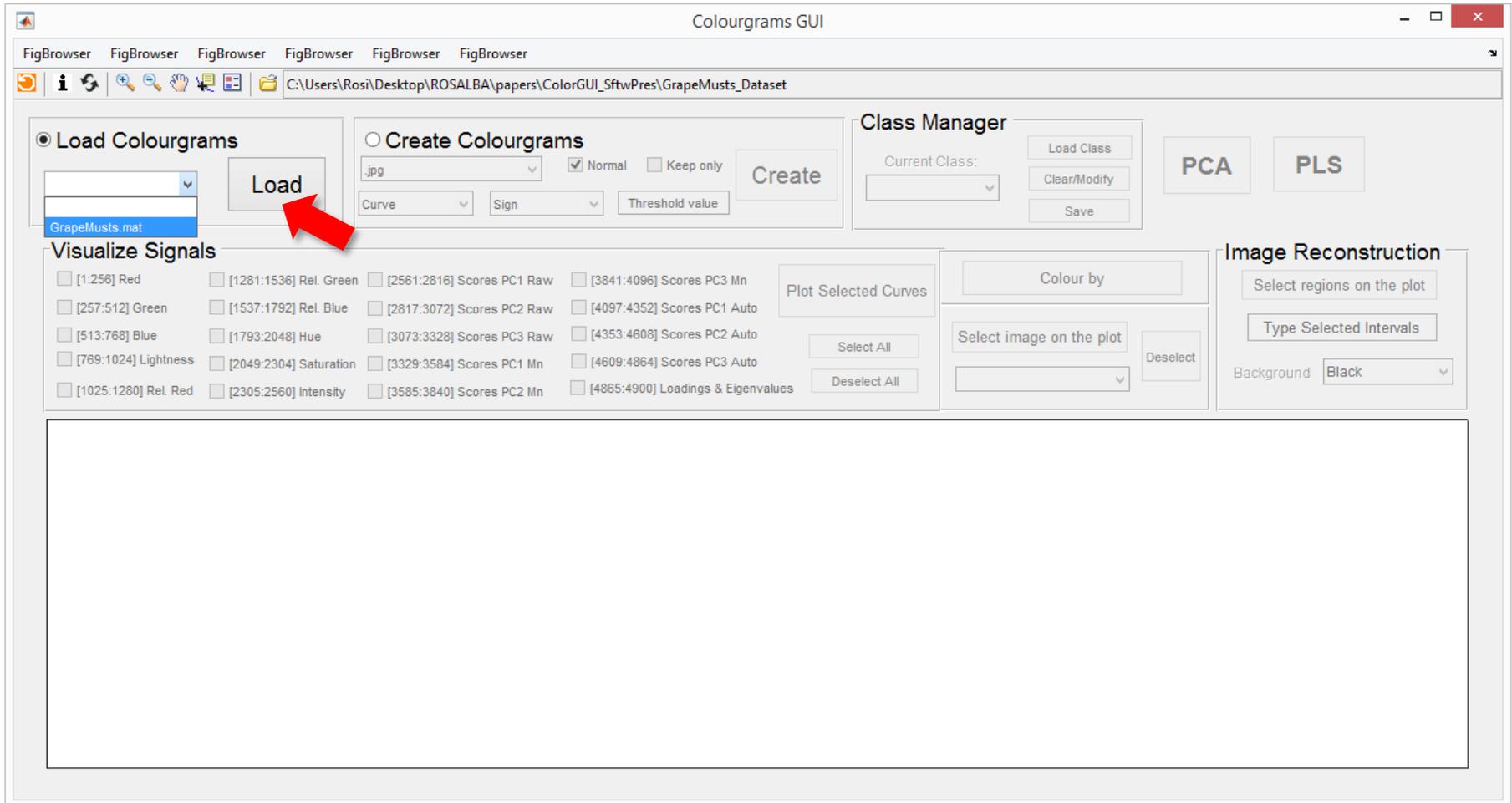
# Create Colourgrams – «Normal» mode

Firstly, the extension of the images has to be defined using the corresponding drop-down menu. Then, the user must define whether calculating the signals from the original images (“Normal” mode) or from the images after background removal (“Keep only” mode).

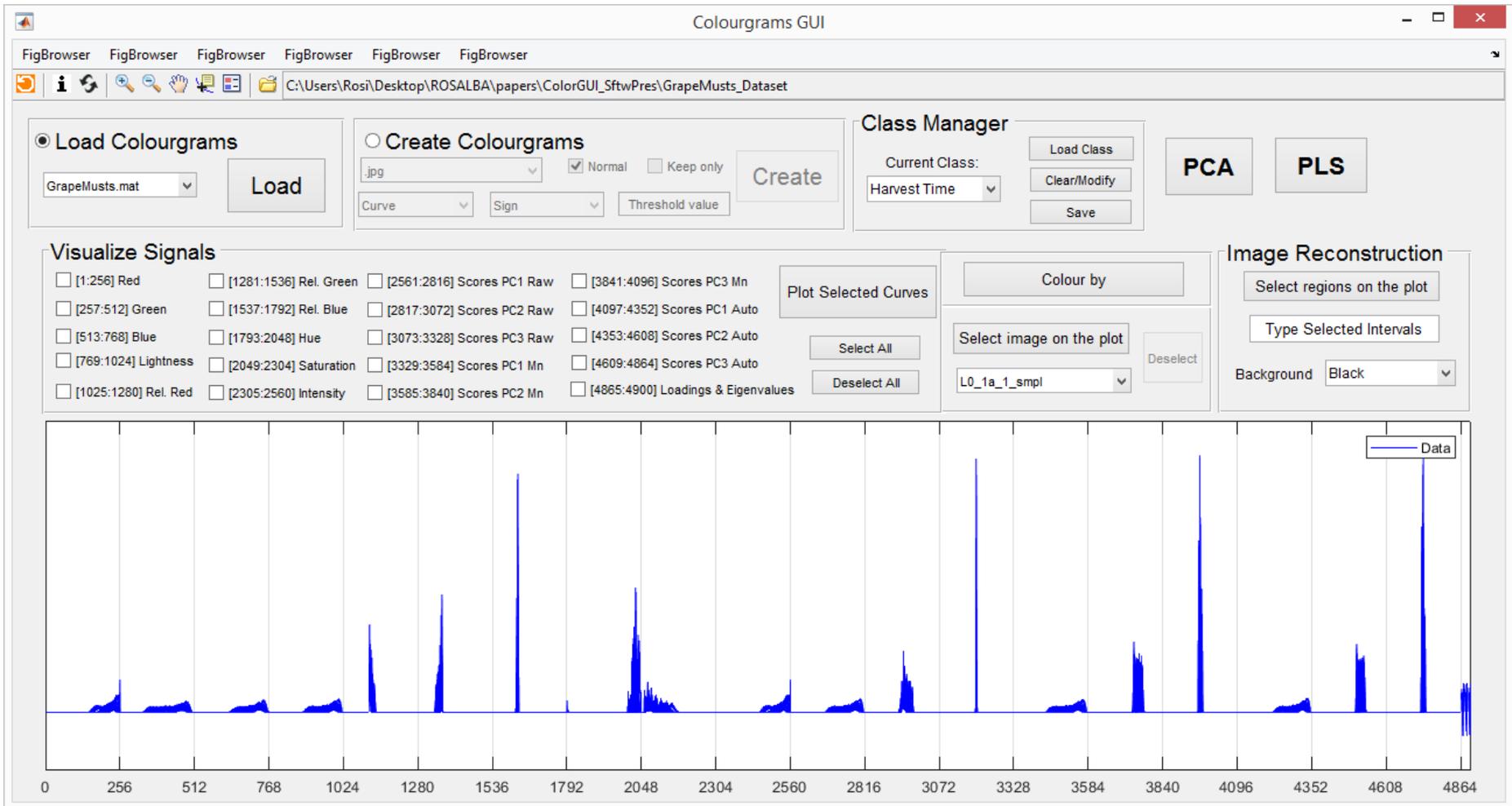


# Load signals

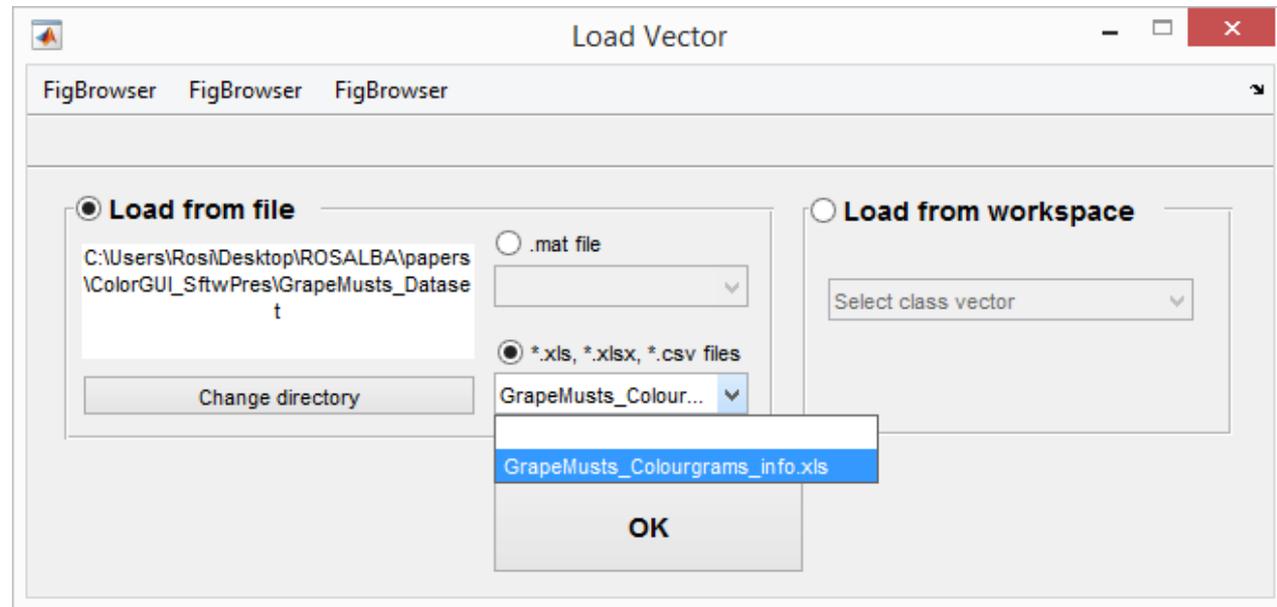
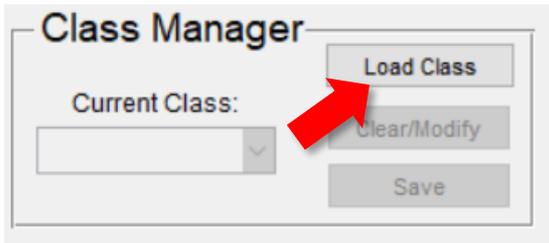
A previously saved DataSet Object of colourgrams can be loaded by clicking on the radio button “Load Colourgrams” and then selecting the corresponding file name from the drop-down menu.



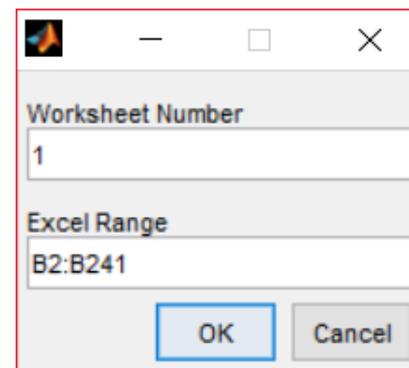
# ...ready to explore signals!!



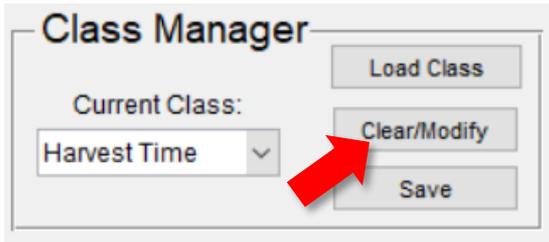
# Class manager



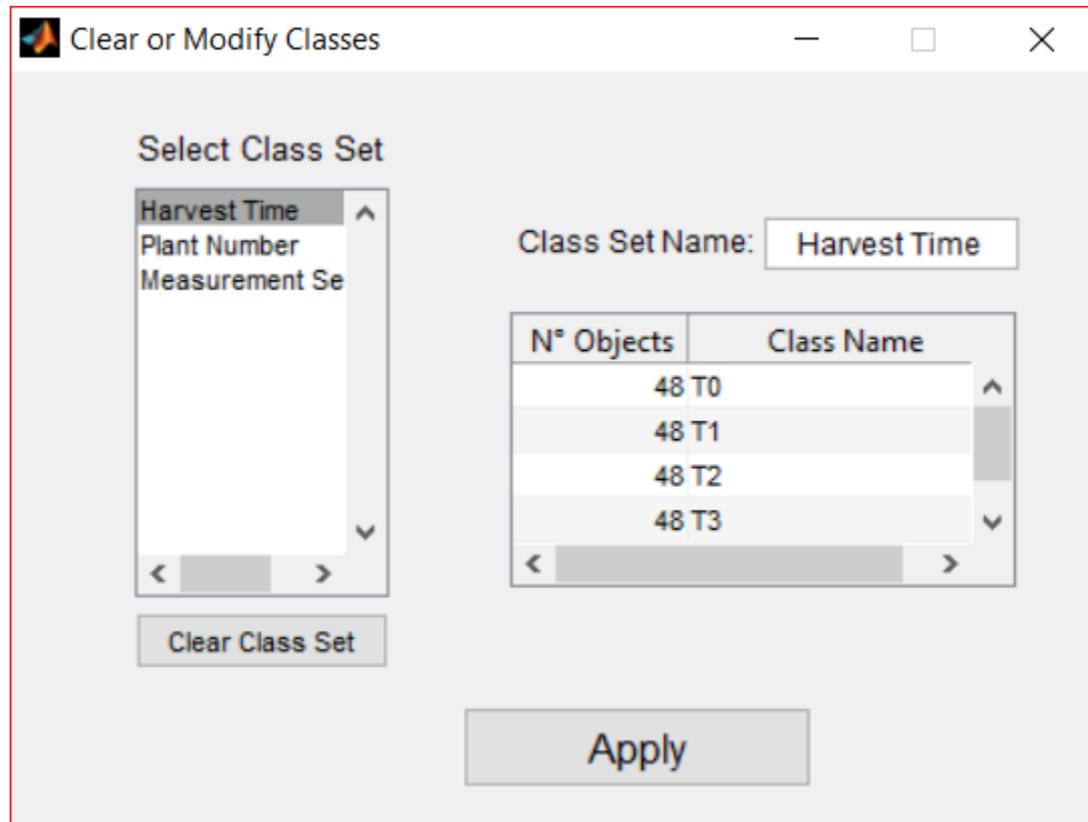
If the class vector is loaded from an .xls or .xlsx file, the user is required to specify the number of the Excel worksheet and the Excel range of the class vector



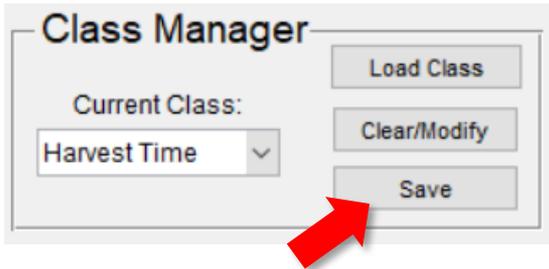
# Class manager



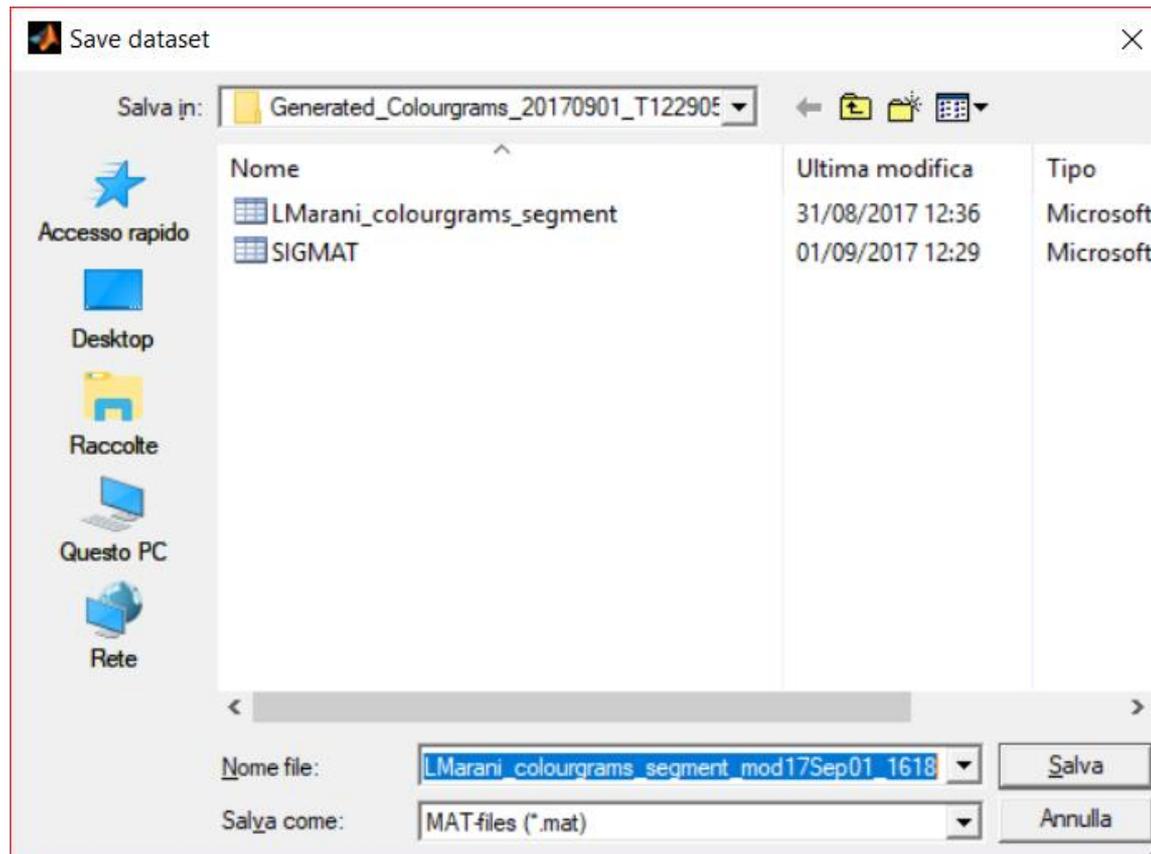
The push button “Clear/Modify” allows to clear or modify an existing class.



# Class manager

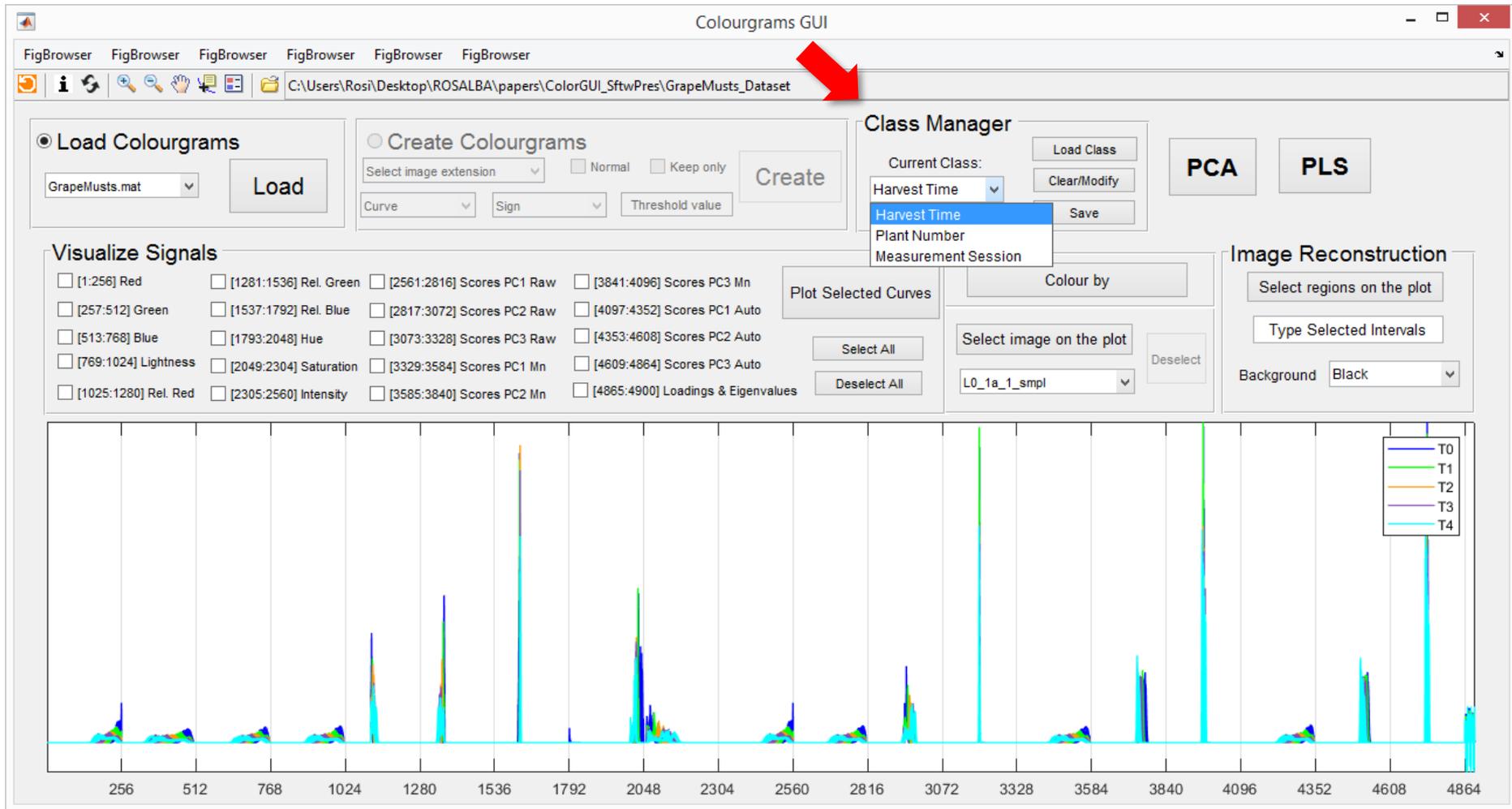


If the user has loaded new class sets, modified the names of existing classes or deleted a class set, the push button “Save” allows to save the updated DataSet Object to a .mat file.



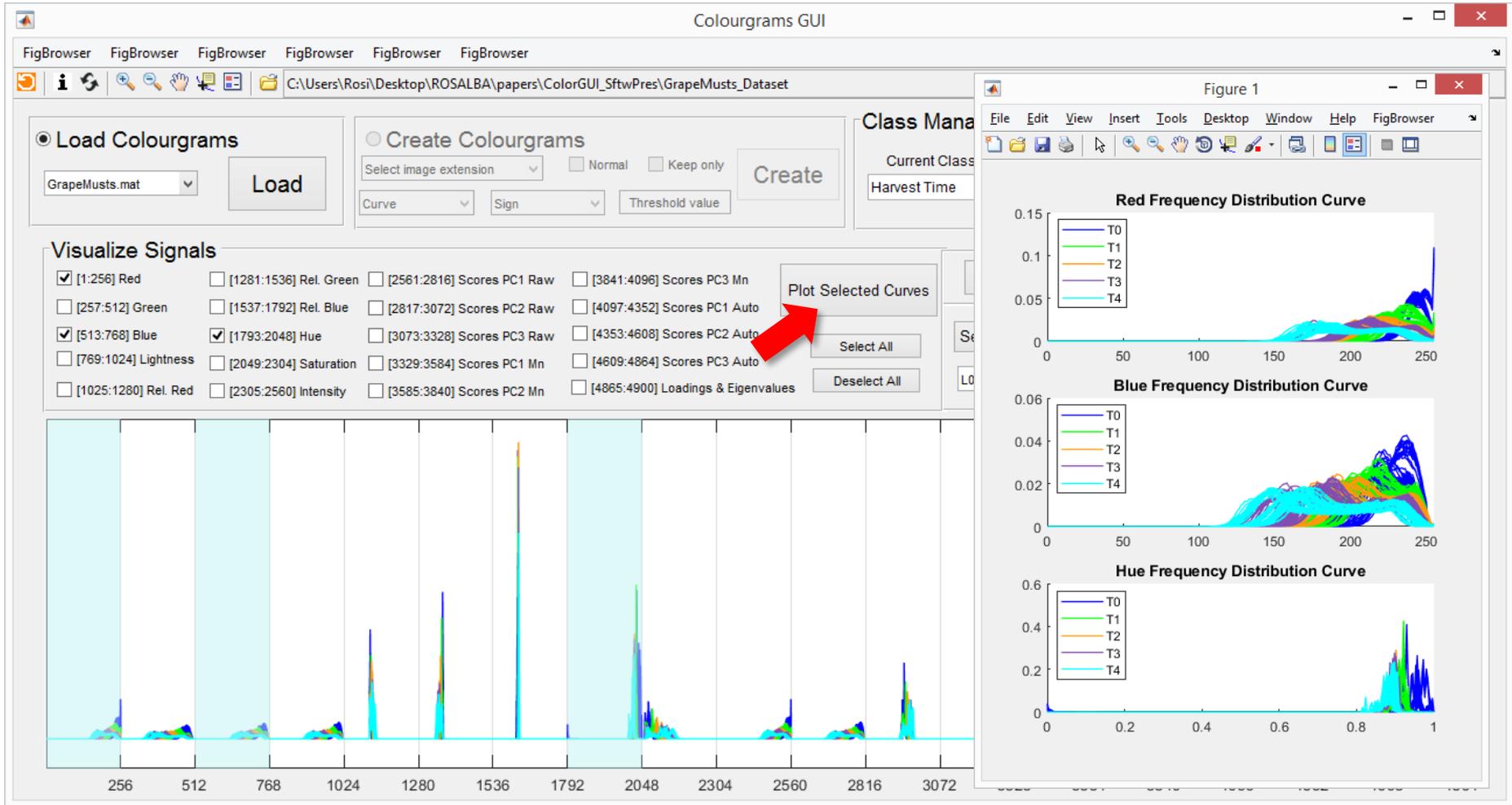
# Class manager

By using the “Current Class” drop-down menu of the “Class Manager” section, it is possible to change the class identifiers set used to visualize the signals.



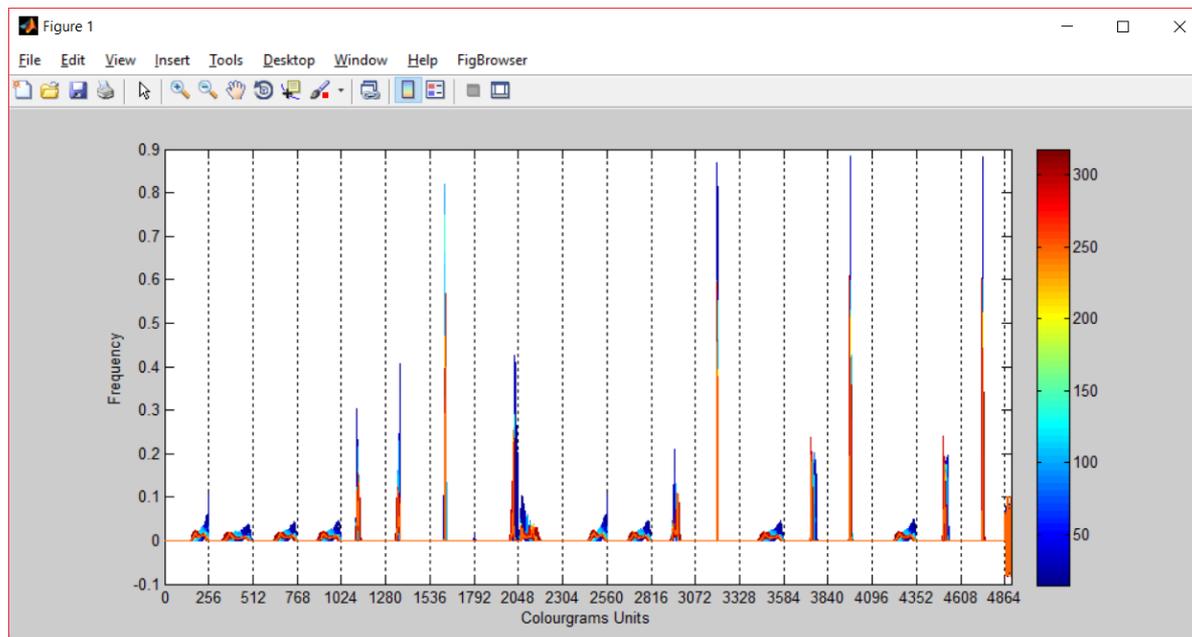
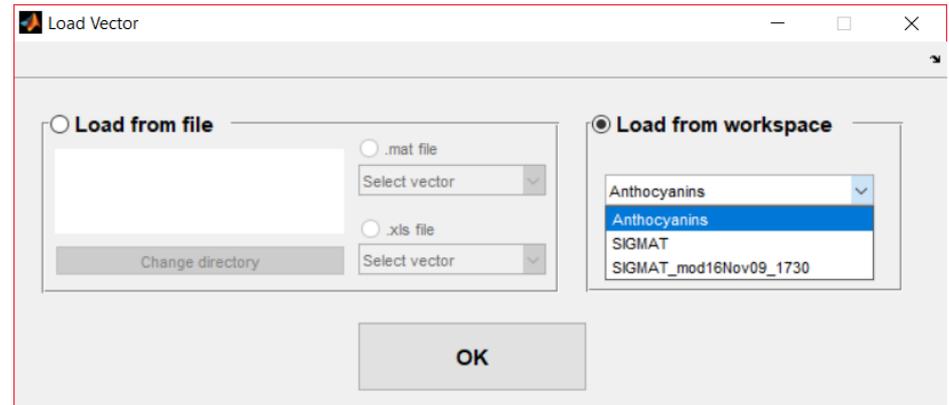
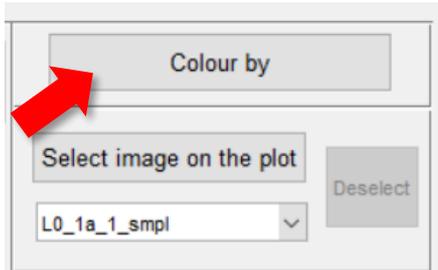
# Visualize Signals

The “Plot Selected Curves” push button allows to plot the frequency distribution curves of the colour-related parameters selected with the check boxes in a separate figure.



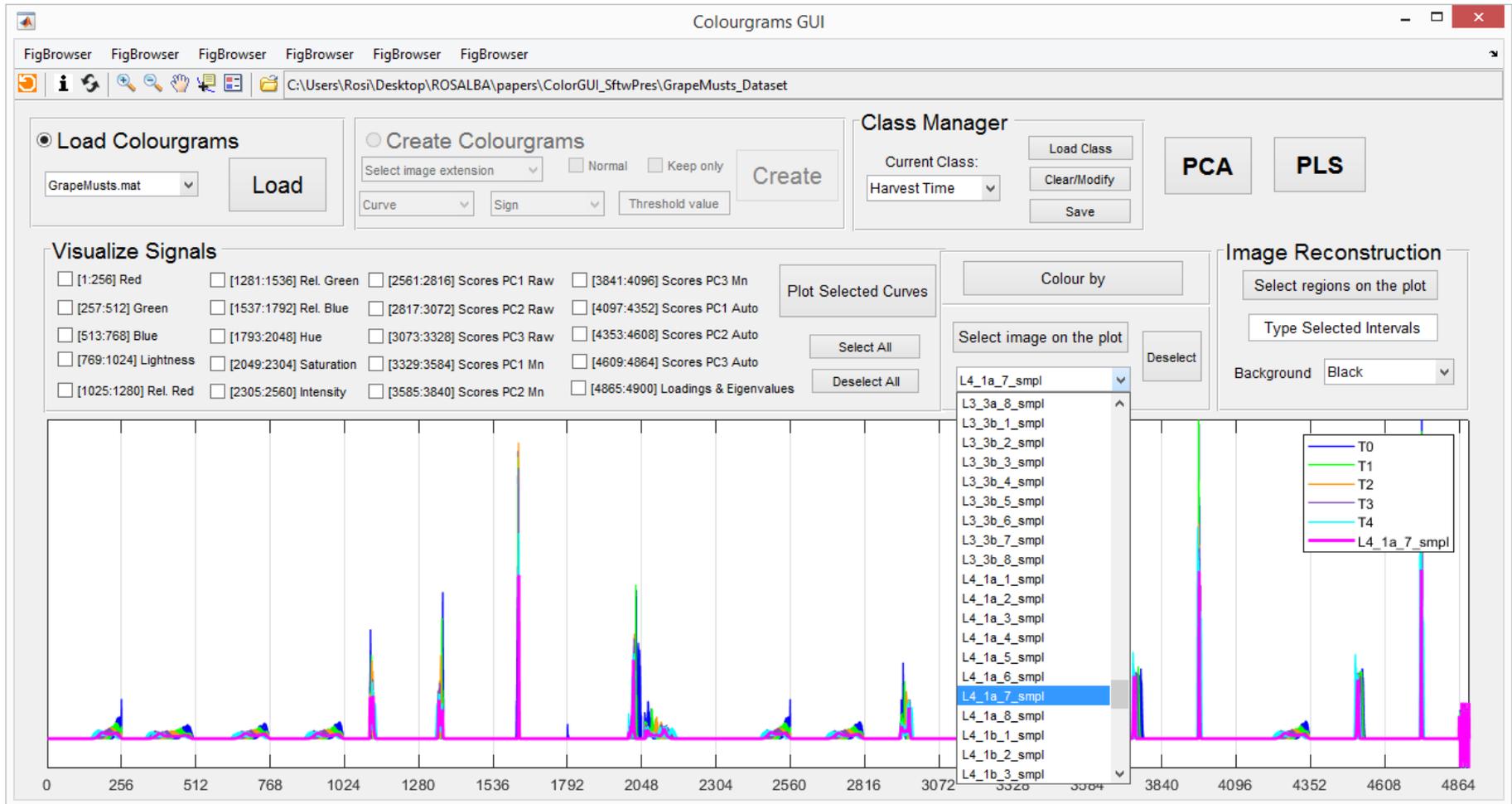
# Visualize Signals

The “Colour By” push button allows to highlight the colourgrams according to a continuously varying quantitative property of the samples (i.e., Total anthocyanin content mg/L).



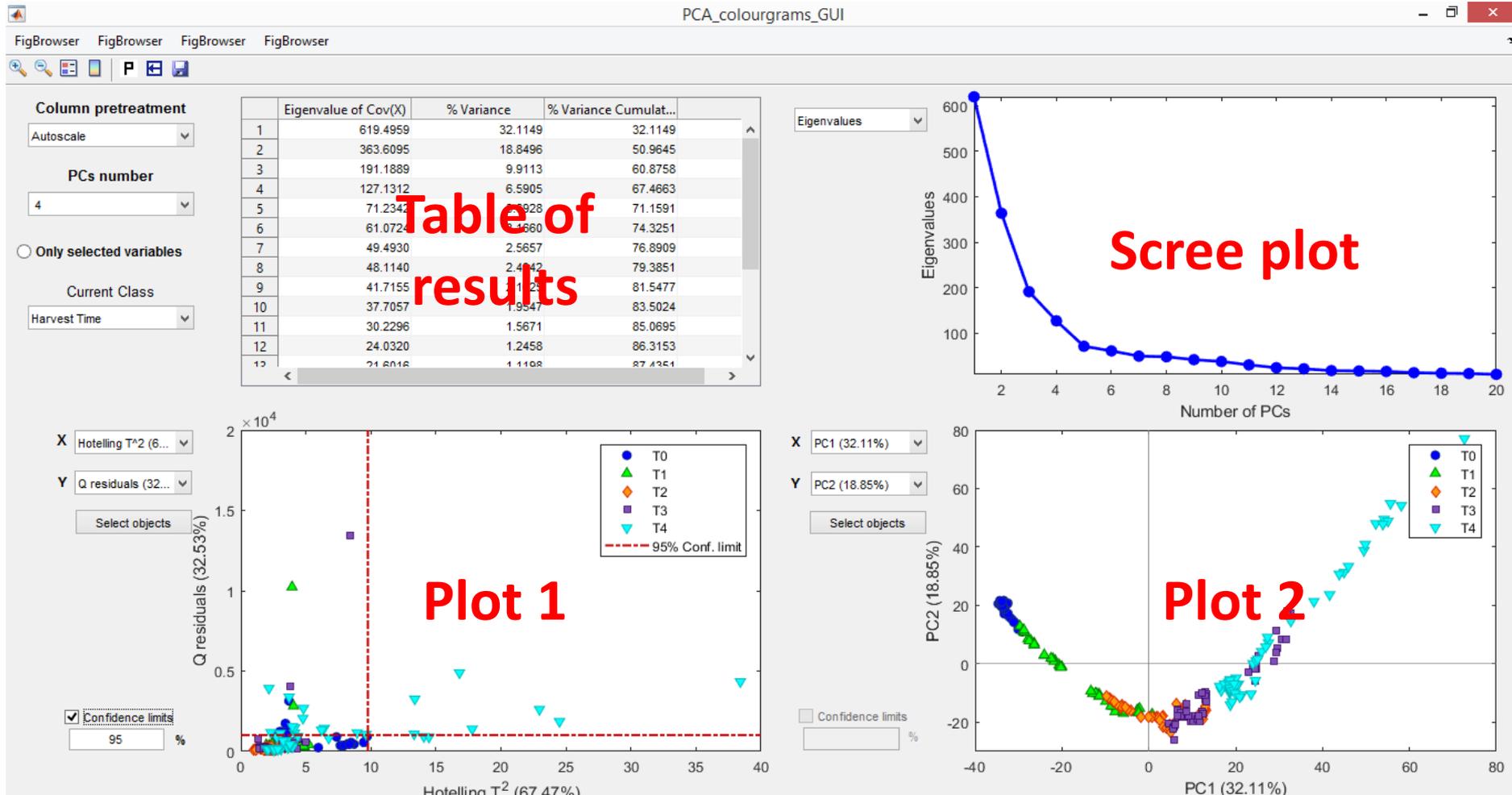
# Visualize Signals

Identification of a specific signal (i.e., L4\_1a\_7\_smpl).

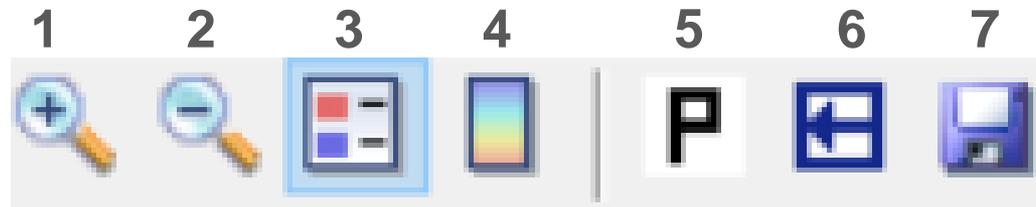


# PCA of the colourgrams matrix

By clicking the “PCA” button of the Colourgrams GUI main window, the PCA window will be automatically displayed.



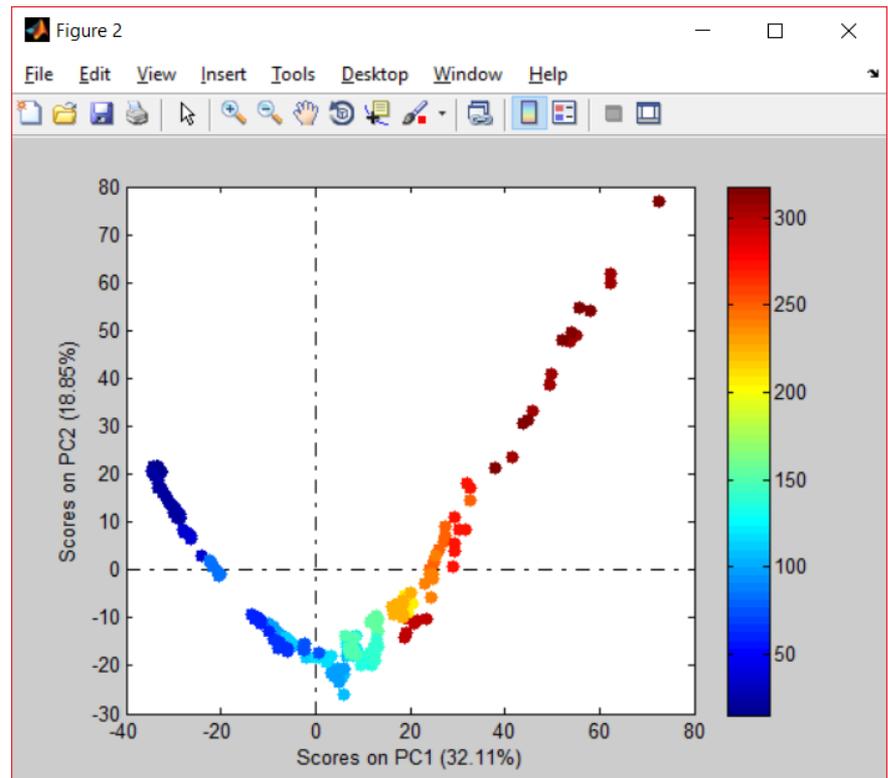
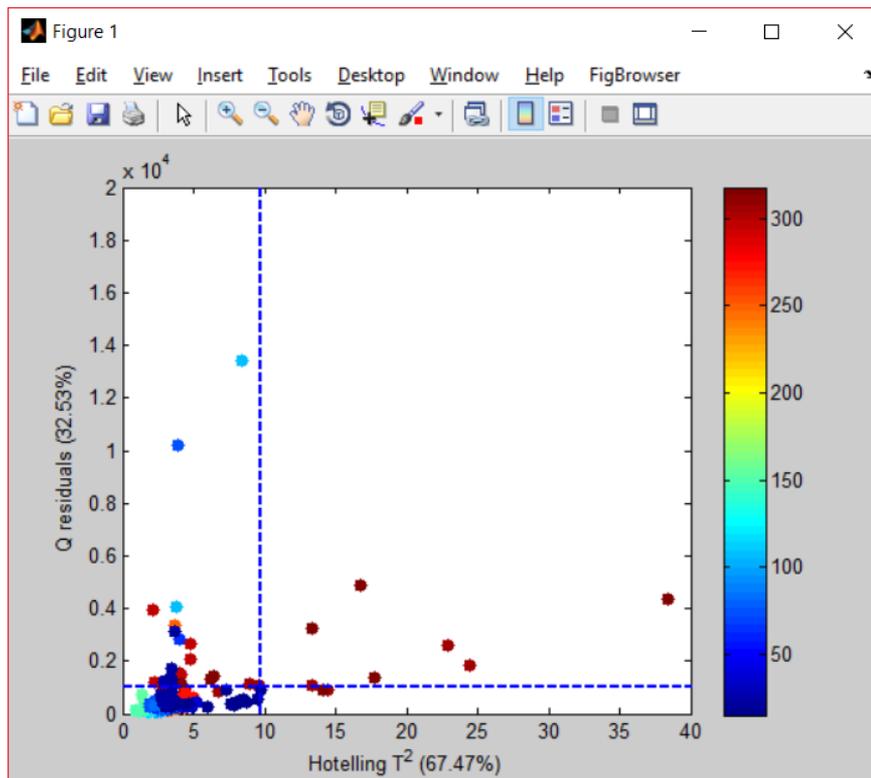
# Toolbar – PCA Colourgrams GUI



1. Zoom in
2. Zoom out
3. Insert Legend
4. Colour by
5. View Loadings
6. Reininclude samples
7. Save Data

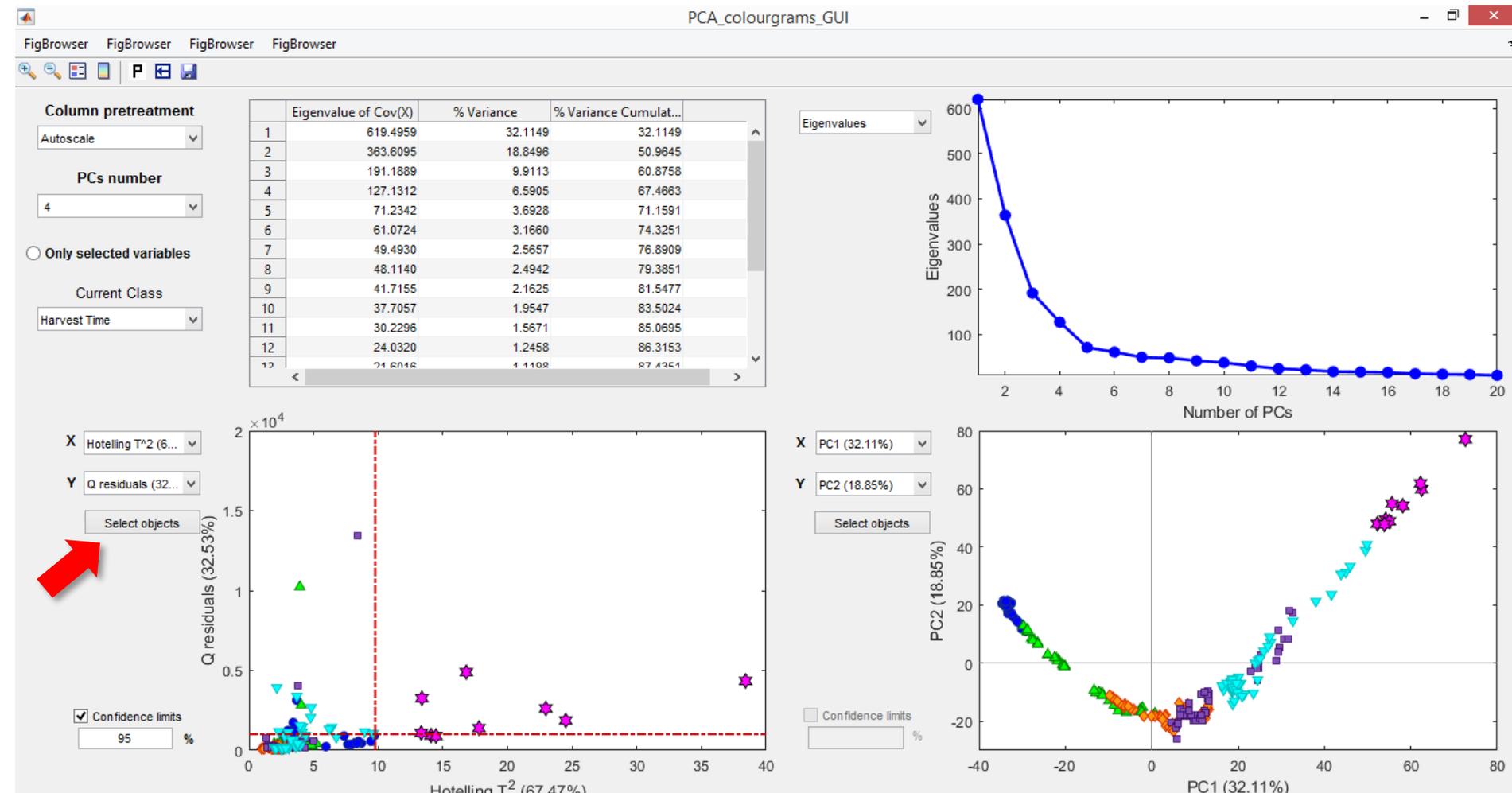
# Colour by

The “Colour by” button in the toolbar allows to colour the samples reported in Plot 1 and Plot 2 according to a defined property. This button opens a window to load a vector of numeric values from a MATLAB workspace variable, from a .mat file or from a .xls / .xlsx / .csv file. Once the vector has been loaded, two new figures are displayed, one for Plot 1 and one for Plot 2, with the samples coloured according to the chosen property values.



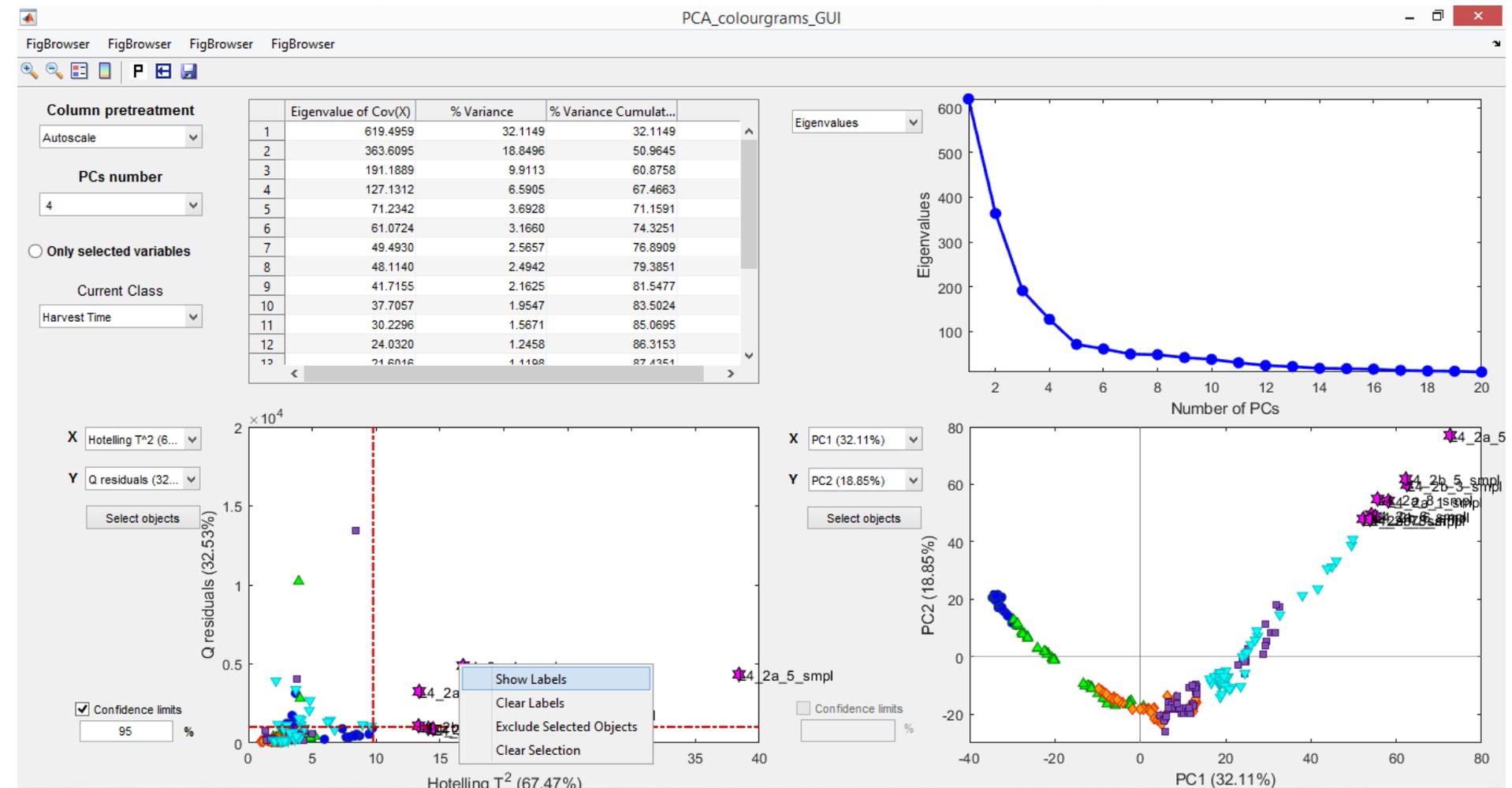
# Select objects

The “Select objects” push buttons associated with Plot 1 and Plot 2 allow to select some objects of interest directly on the plots, for example objects identified as outliers.

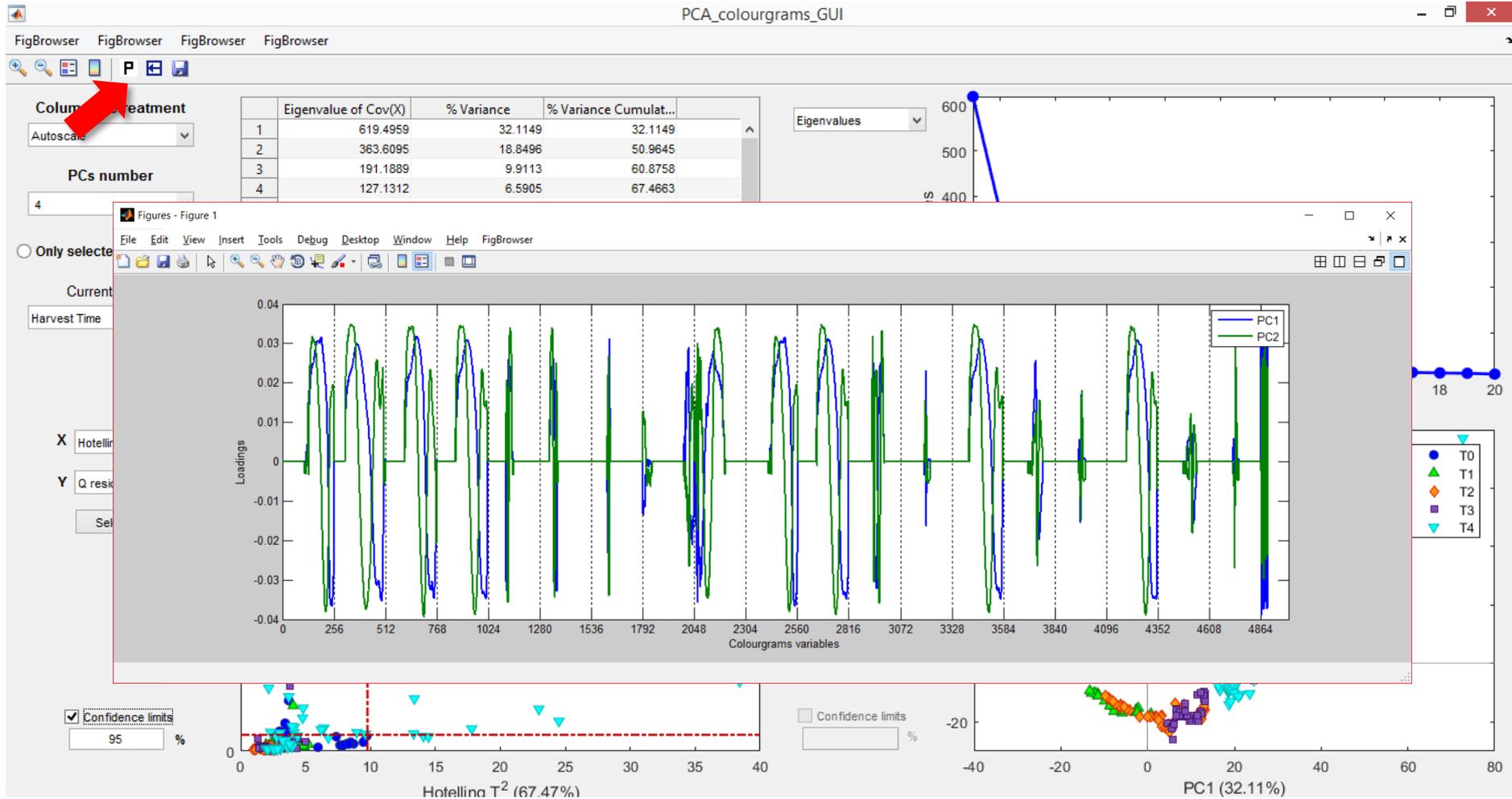


# Select objects

By right-clicking on the magenta hexagrams and using the resulting context menu, the user can decide whether visualizing the labels of the selected objects, deselecting the objects or eliminating them from the dataset. If the selected objects are eliminated, the PCA model is automatically updated

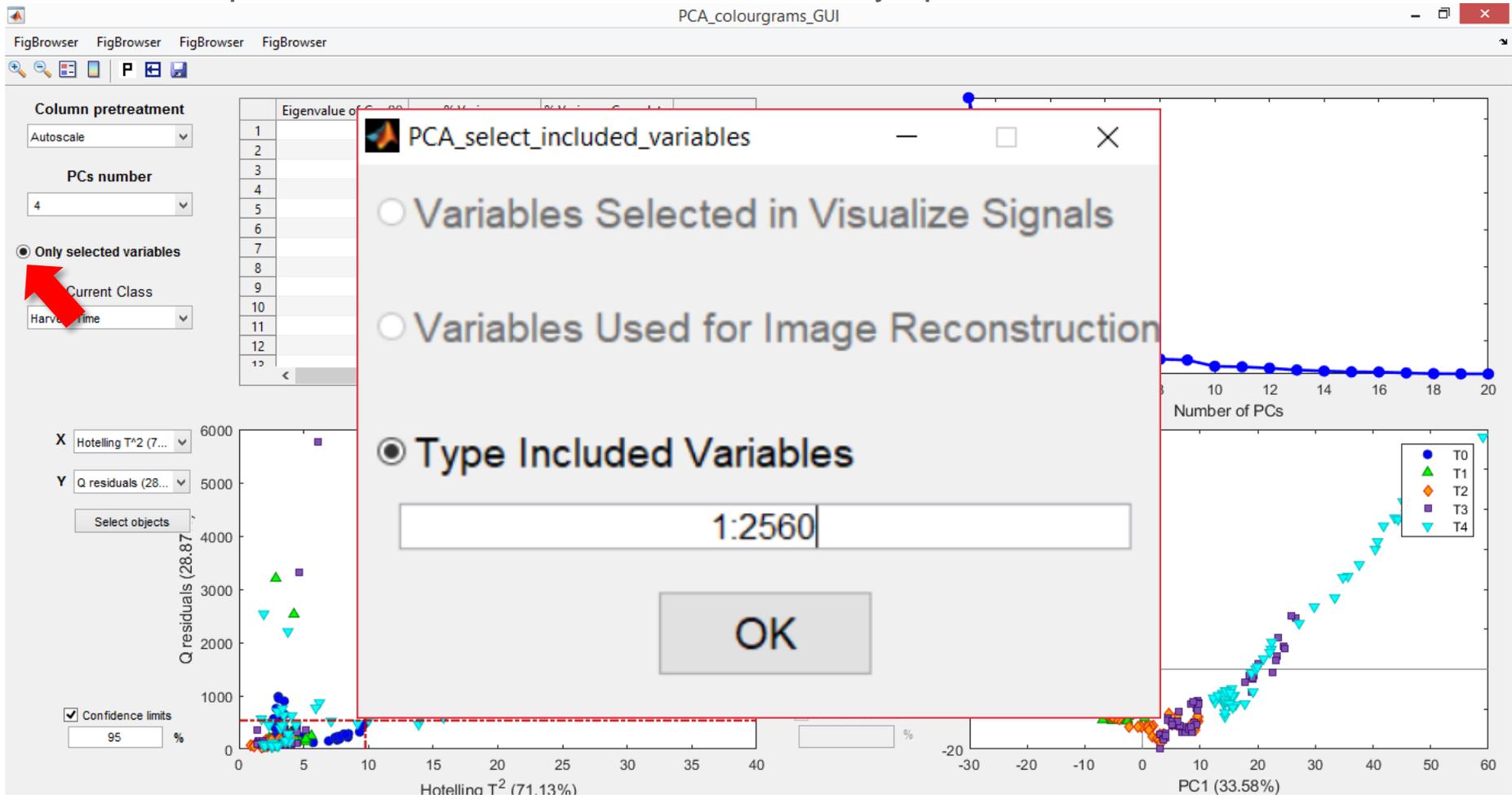


# Loading vectors



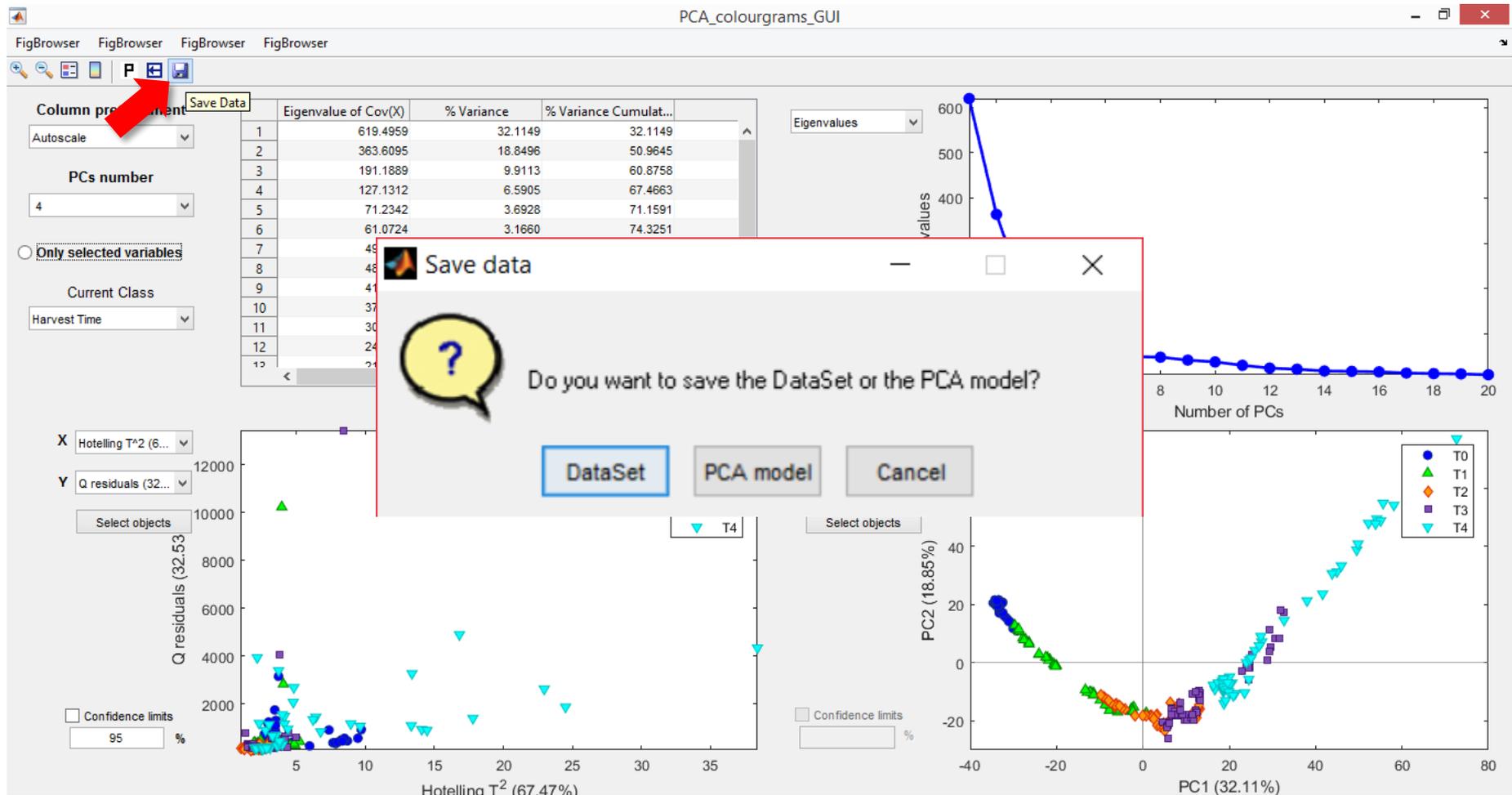
# Variable Selection

By clicking on the “Only selected variables” radio button a specific window appears, allowing to define the variables to be used for the calculation of the PCA model. If the included variables are specified, the PCA model is automatically updated



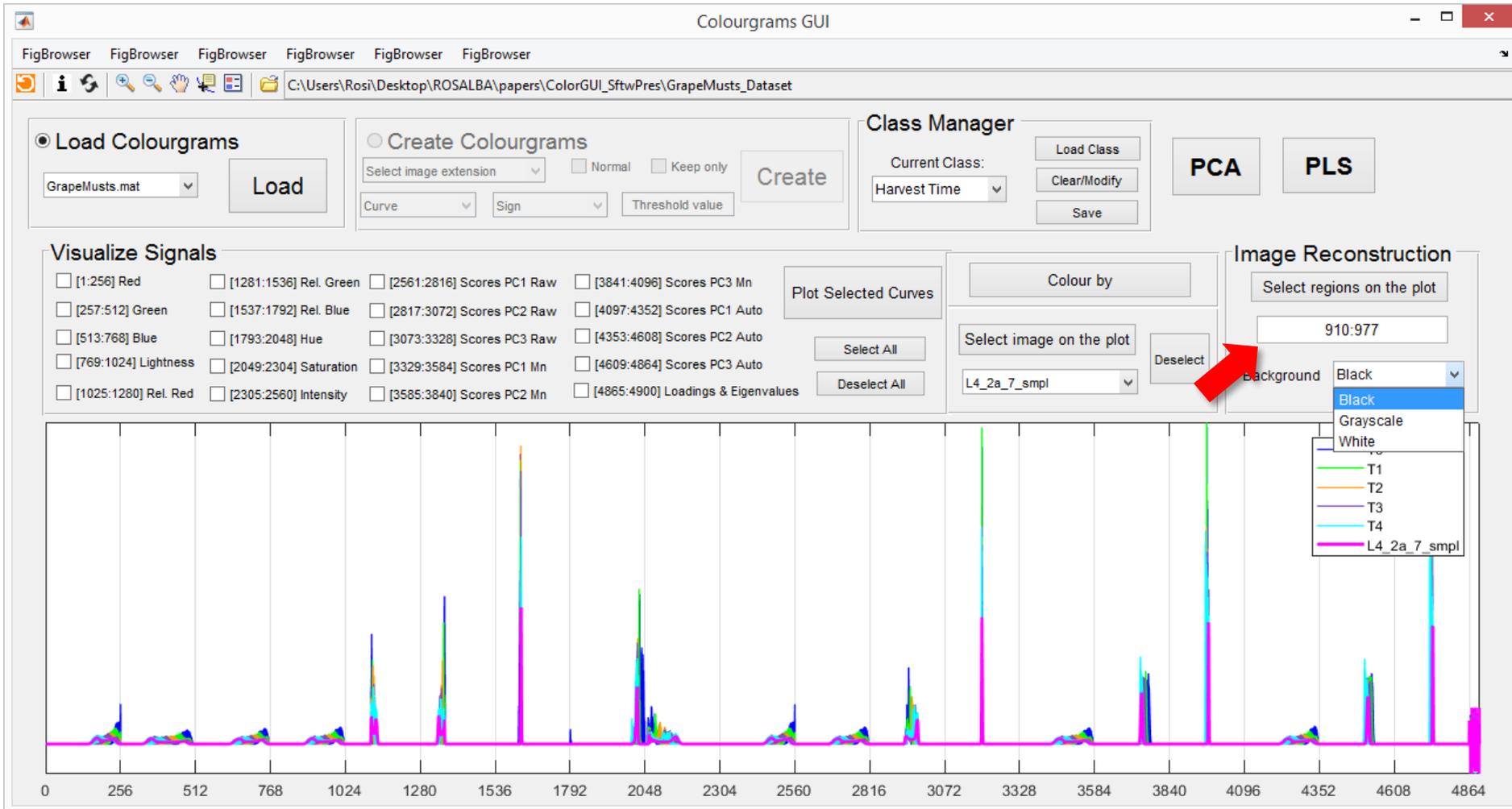
# Save Data

By clicking on the “Save Data”, the user can choose to save new (i.e., without the eliminated samples) Dataset or a structure array, which contains all the information about the calculated PCA model.



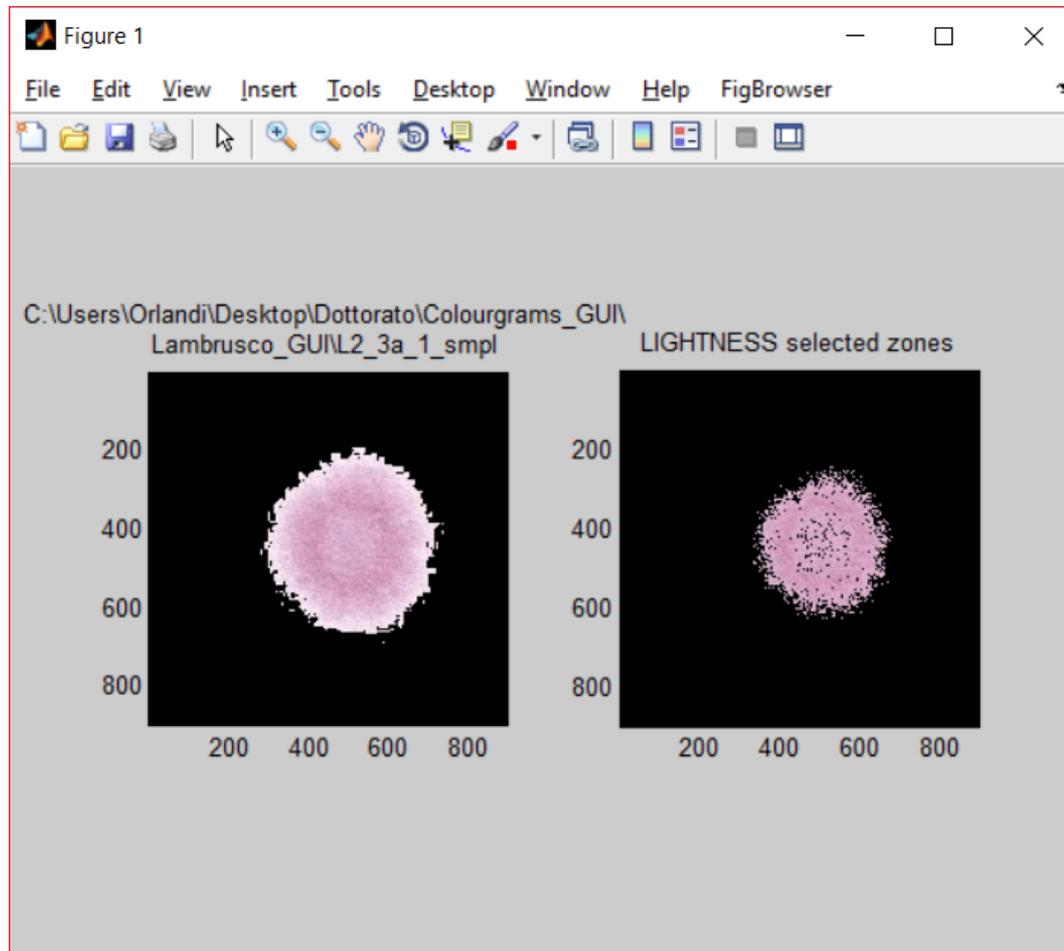
# Image reconstruction

The user can either type the selected intervals in the edit box of the Colourgrams GUI main window or click the “Select regions on the plot” push button to activate manual selection.



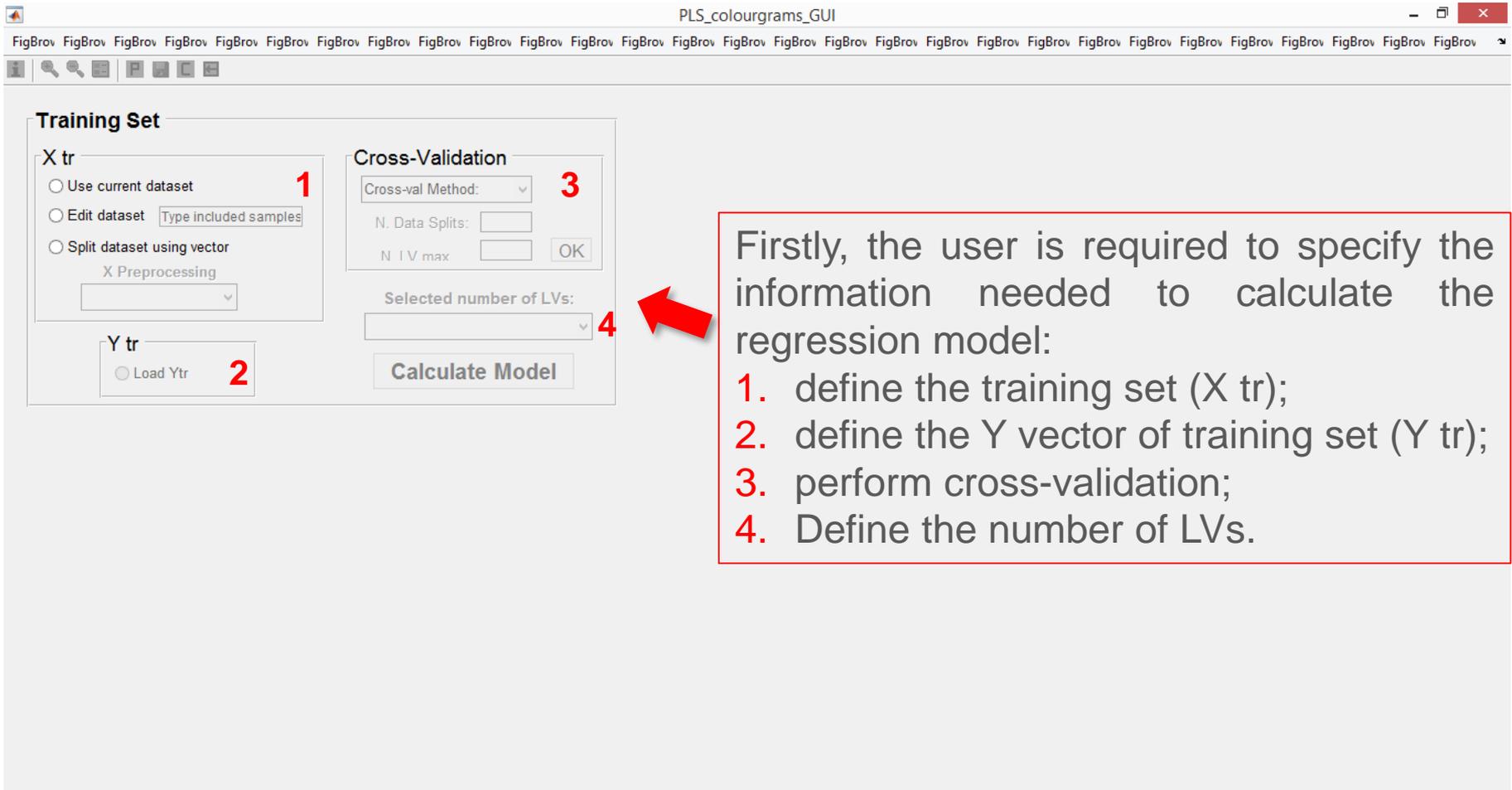
# Image reconstruction

In both cases, for each frequency distribution curve where at least one interval has been chosen, a figure is obtained where only the pixels falling in the selected interval(s) are displayed, while the remainder pixels can be represented in black or white colour, or in grayscale tones of the original image.



# PLS of the colourgrams matrix

By clicking the “PLS” button of the Colourgrams GUI main window, the PLS window will be automatically displayed.



The screenshot shows the PLS\_colourgrams\_GUI window with the following sections:

- Training Set**
  - X tr**
    - Use current dataset (1)
    - Edit dataset (Type included samples)
    - Split dataset using vector (X Preprocessing)
  - Y tr**
    - Load Ytr (2)
- Cross-Validation**
  - Cross-val Method: (3)
  - N. Data Splits: [ ]
  - N. I V max: [ ] [OK]
  - Selected number of LVs: [ ] (4)
  - Calculate Model

Firstly, the user is required to specify the information needed to calculate the regression model:

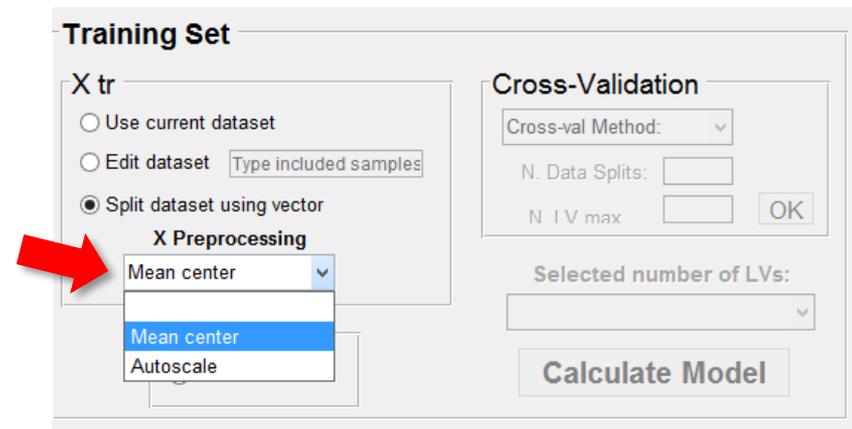
1. define the training set ( $X_{tr}$ );
2. define the Y vector of training set ( $Y_{tr}$ );
3. perform cross-validation;
4. Define the number of LVs.

# Define the training set - Xtr

The user can choose between three options:

1. use the whole dataset analyzed in the main window of colourgrams GUI (“Use current dataset”);
2. include only some samples in the training set (“Edit dataset”). By clicking on this radio button, it is required to specify the interval(s) of included samples in the edit box “Type included samples”;
3. split the whole dataset in training and test set using a vector (“Split dataset using vector”). By clicking on this radio button, the user can choose to load a vector of numeric values (1 for samples of the training set and 2 for samples of the test set) from a MATLAB workspace variable, from a .mat file or from an .xls / .xlsx / .csv file. The vector must have as many rows as the number of the samples analyzed in the main window of colourgrams\_GUI.

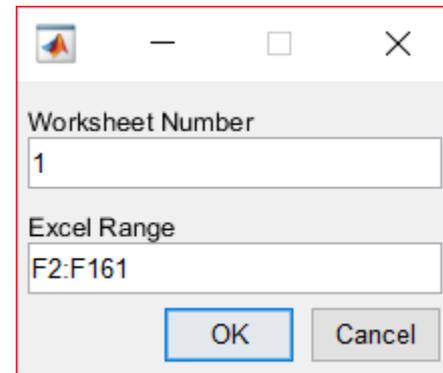
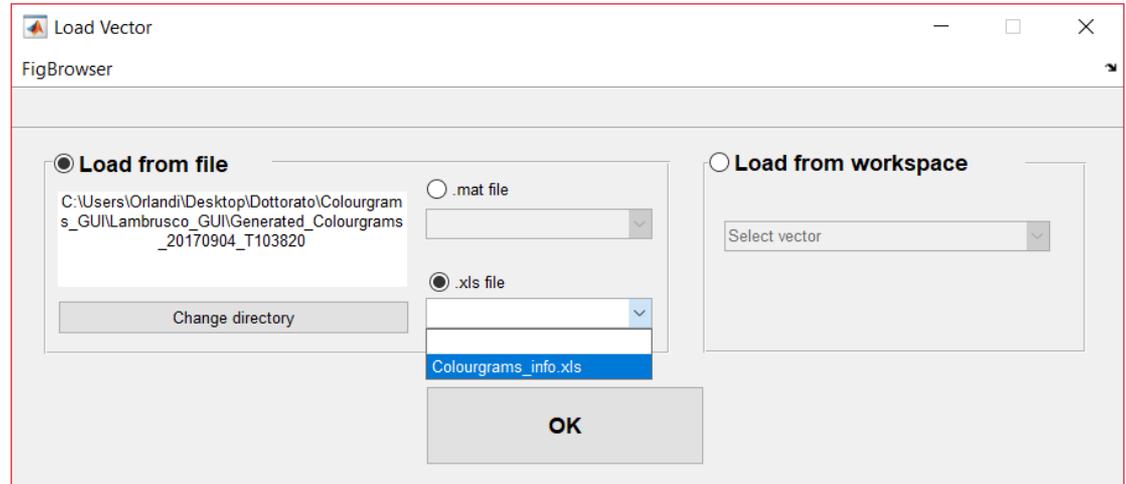
Then, the pretreatment of the colourgrams matrix has to be defined using the corresponding drop-down menu.



# Define $Y$ vector of the training set - $Y_{tr}$

The radio button “Load  $Y_{tr}$ ” opens a window to load a vector of numeric values from a MATLAB workspace variable, from a .mat file or from a .xls file.

If the vector is loaded from an .xls /.xlsx file, the user is required to specify the number of the Excel worksheet and the Excel range of the vector



# Perform Cross-Validation

Firstly, the cross-validation method has to be defined using the corresponding pop-up menu.

If the custom method is selected, it is required to load a vector of numeric values from a MATLAB workspace variable, from a .mat file or from a .xls / .xlsx / .csv file.

## Training Set

### X tr

- Use current dataset  
 Edit dataset

- Split dataset using vector

### X Preprocessing

Mean center

### Y tr

- Load Ytr

### Cross-Validation

Cross-val Method: 

- leave one out
- venetian blinds
- contiguous blocks
- custom

OK

Calculate Model

## Training Set

### X tr

- Use current dataset  
 Edit dataset

- Split dataset using vector

### X Preprocessing

Mean center

### Y tr

- Load Ytr

### Cross-Validation

custom 

N. Data Splits:

N. LV max

OK

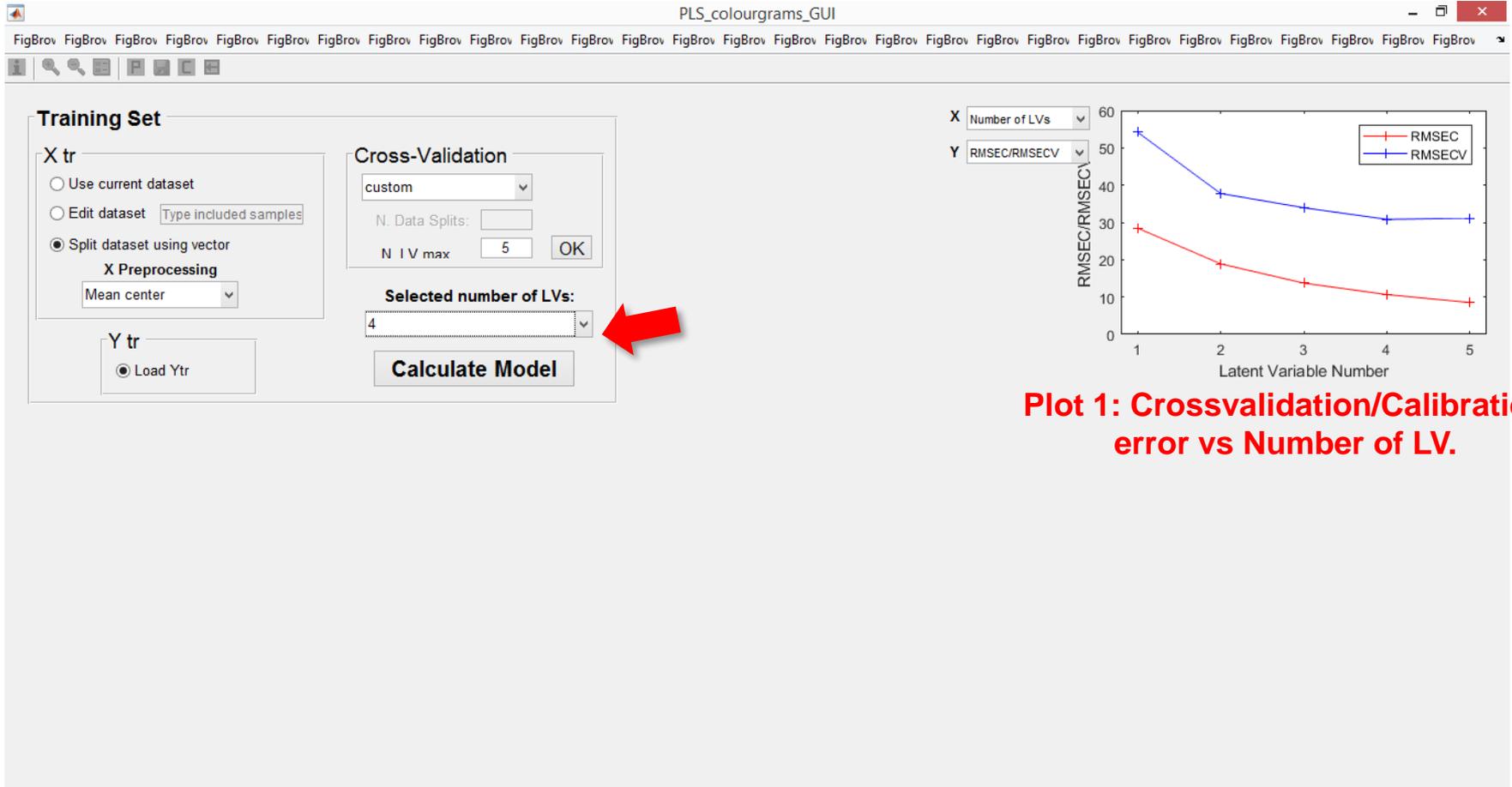
Selected number of LVs:

Calculate Model

The user is required to type the maximum number of Latent Variables (LVs) in the corresponding edit box. Then, by clicking on the push button "OK", the calculation of cross-validation will be started.

# Define the number of LVs

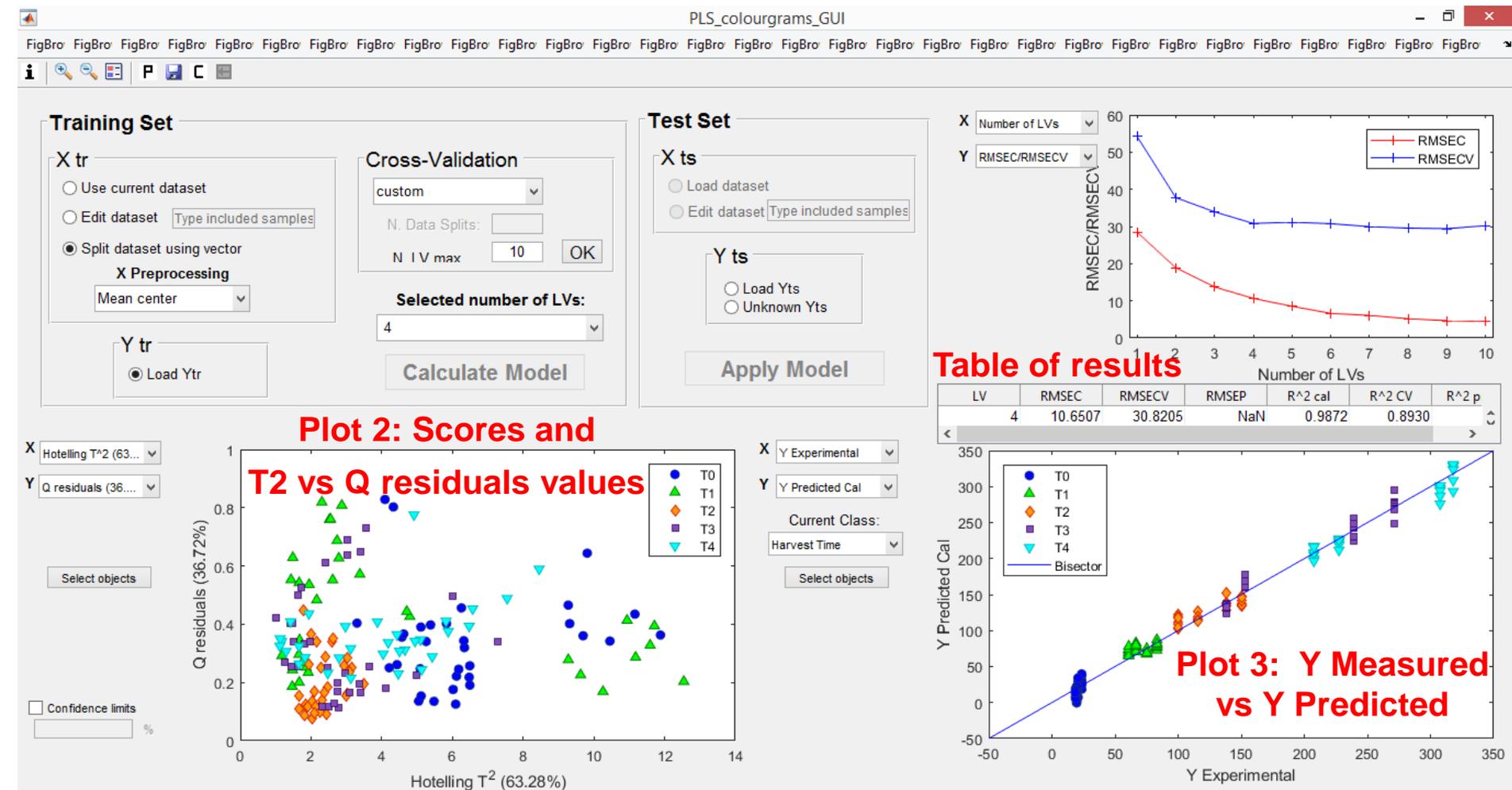
At the end of the cross-validation, the plot of Crossvalidation/Calibration error vs Number of LVs is displayed. Then, the user can select the optimal number of LVs using the corresponding drop-down menu.



**Plot 1: Crossvalidation/Calibration error vs Number of LV.**

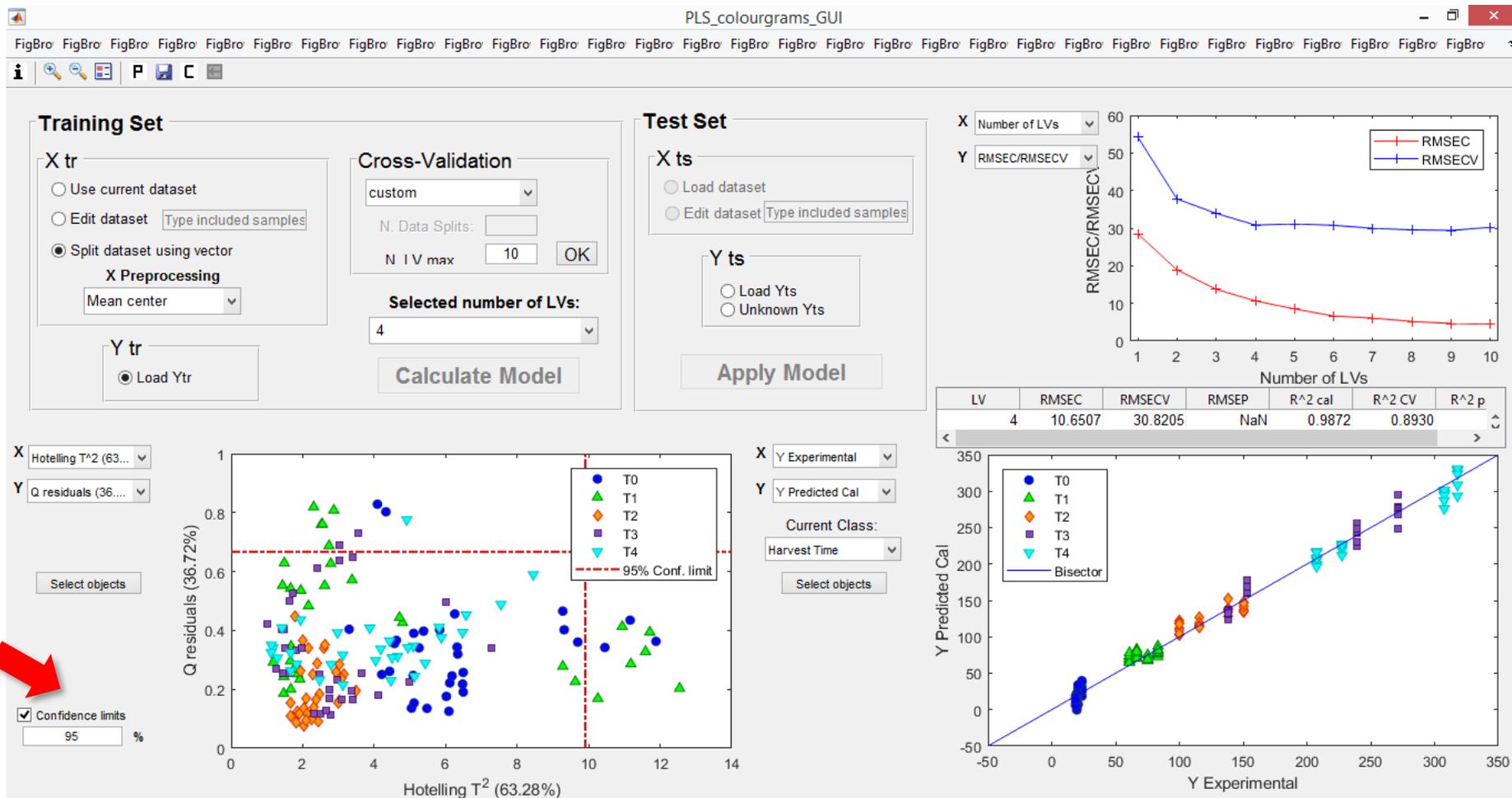
# Calibrate model

By clicking on the push button “Calibrate Model”, the calibration model will be automatically calculated.

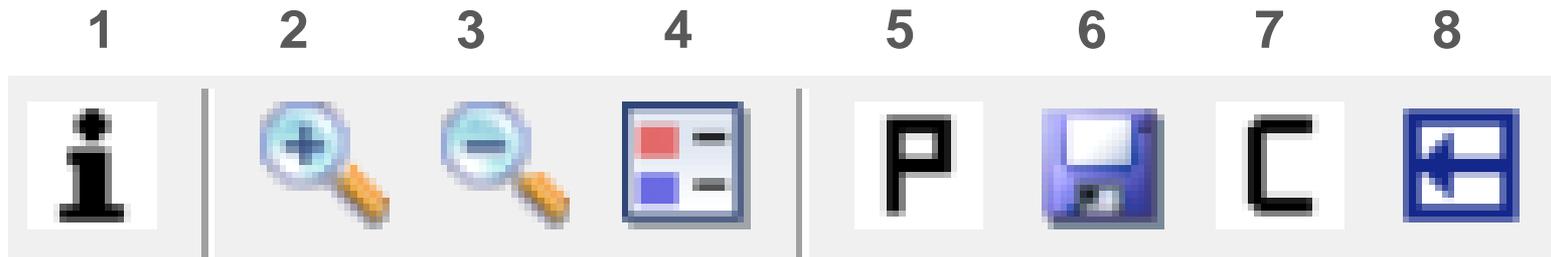


# Plot manager

The user can choose the quantities to be displayed in the plots using the corresponding “X” and “Y” drop-down menus. Furthermore, if Q residuals and/or Hotelling  $T^2$  values are represented in Plot 2, it is also possible to visualize the corresponding confidence limits by clicking the “Confidence limits” check box and typing the probability value in the edit box (the default value is 95 %).



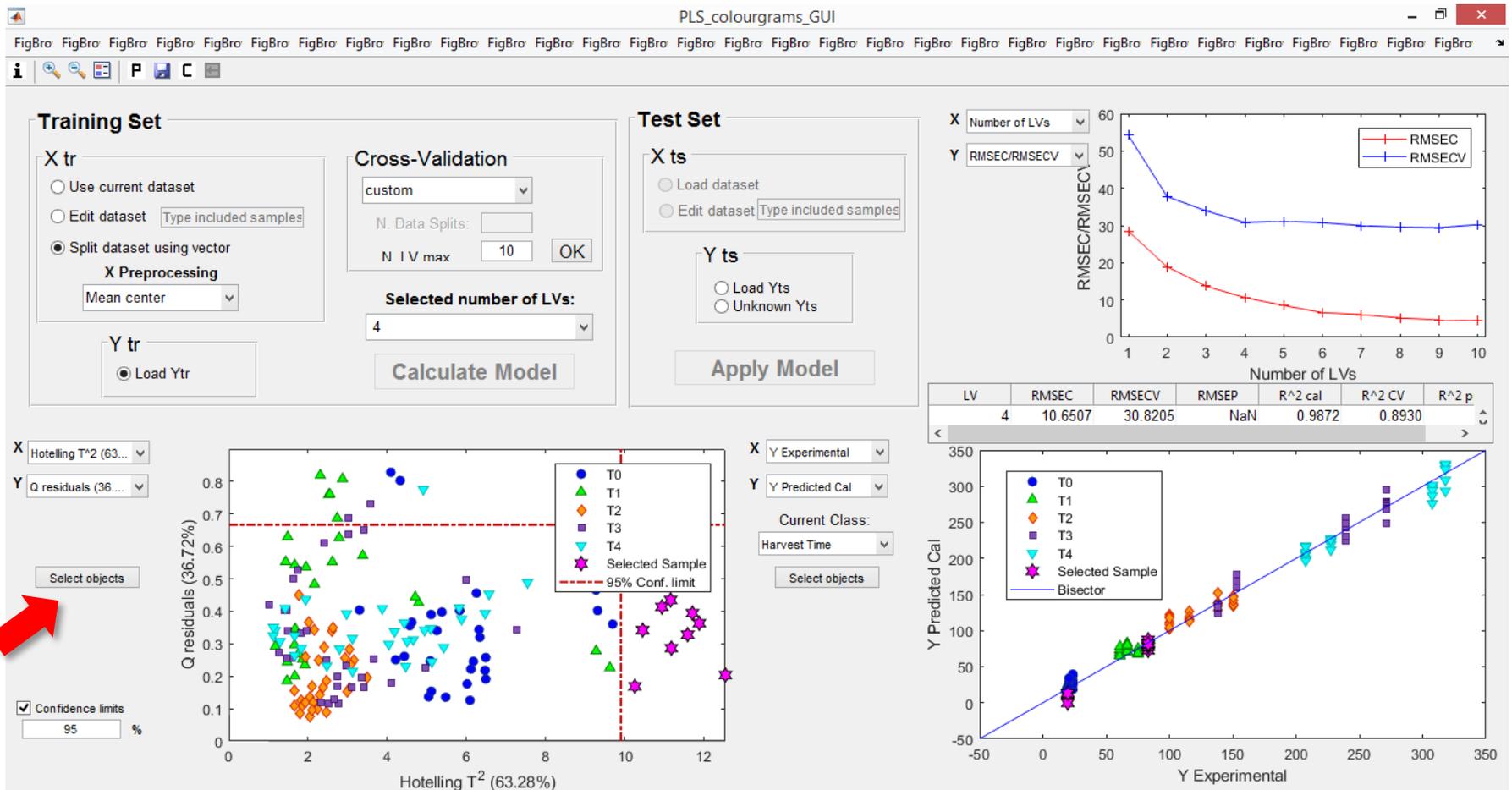
# Toolbar – PLS colourgrams GUI



1. Model Info
2. Zoom in
3. Zoom out
4. Insert Legend
5. View Loadings
6. Save Data or Model
7. Clear model
8. Reininclude samples

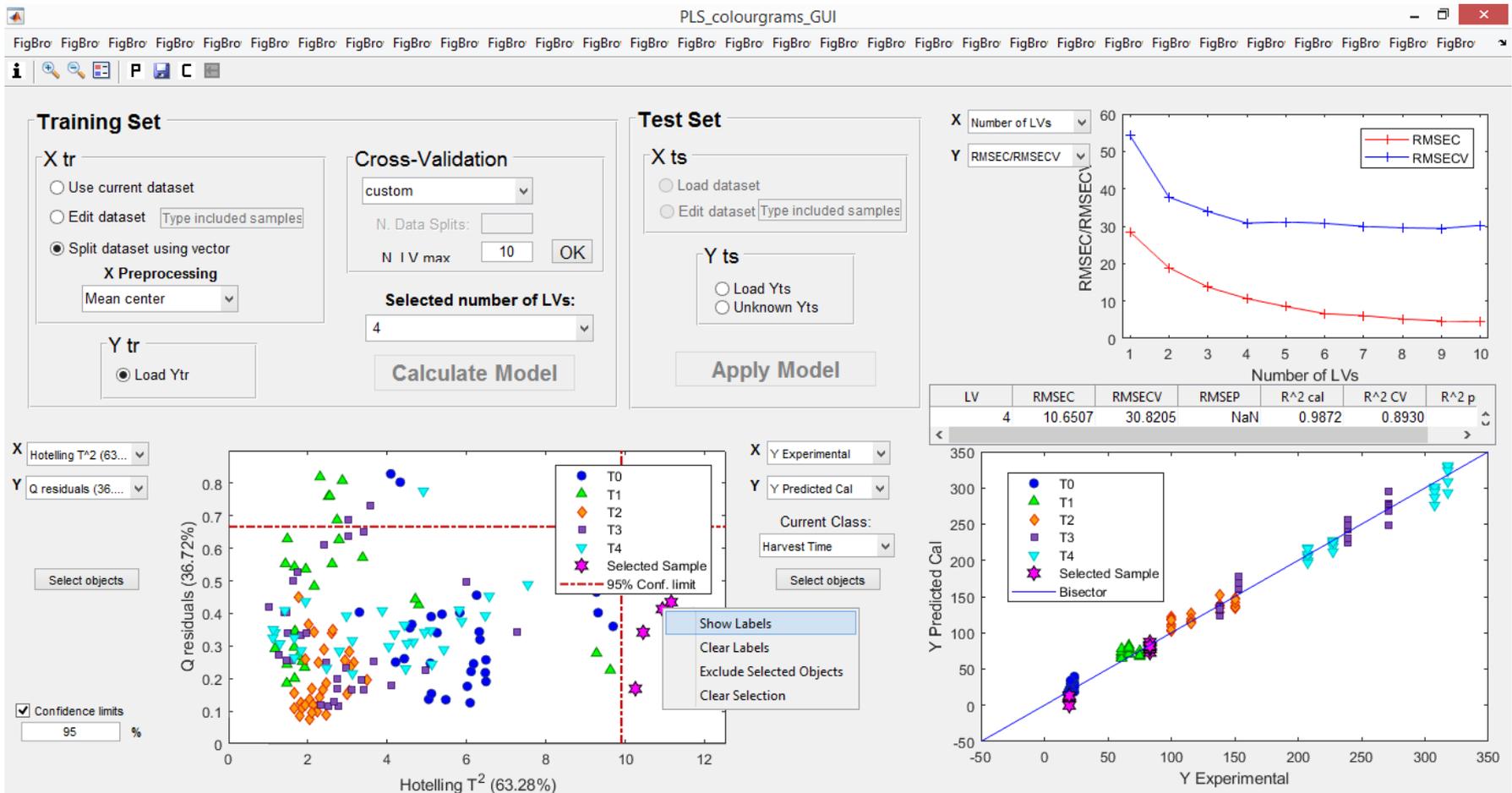
# Select objects

The “Select objects” push buttons associated with Plot 2 and Plot 3 allow to select some objects of interest directly on the plots, for example objects identified as outliers.



# Select objects

By right-clicking on the magenta hexagrams and using the resulting context menu, the user can decide whether visualizing the labels of the selected objects, deselecting the objects or eliminating them from the dataset. If the selected objects are eliminated, the PLS model is automatically updated but it is important to verify the new optimal number of LVs.



# Apply Model

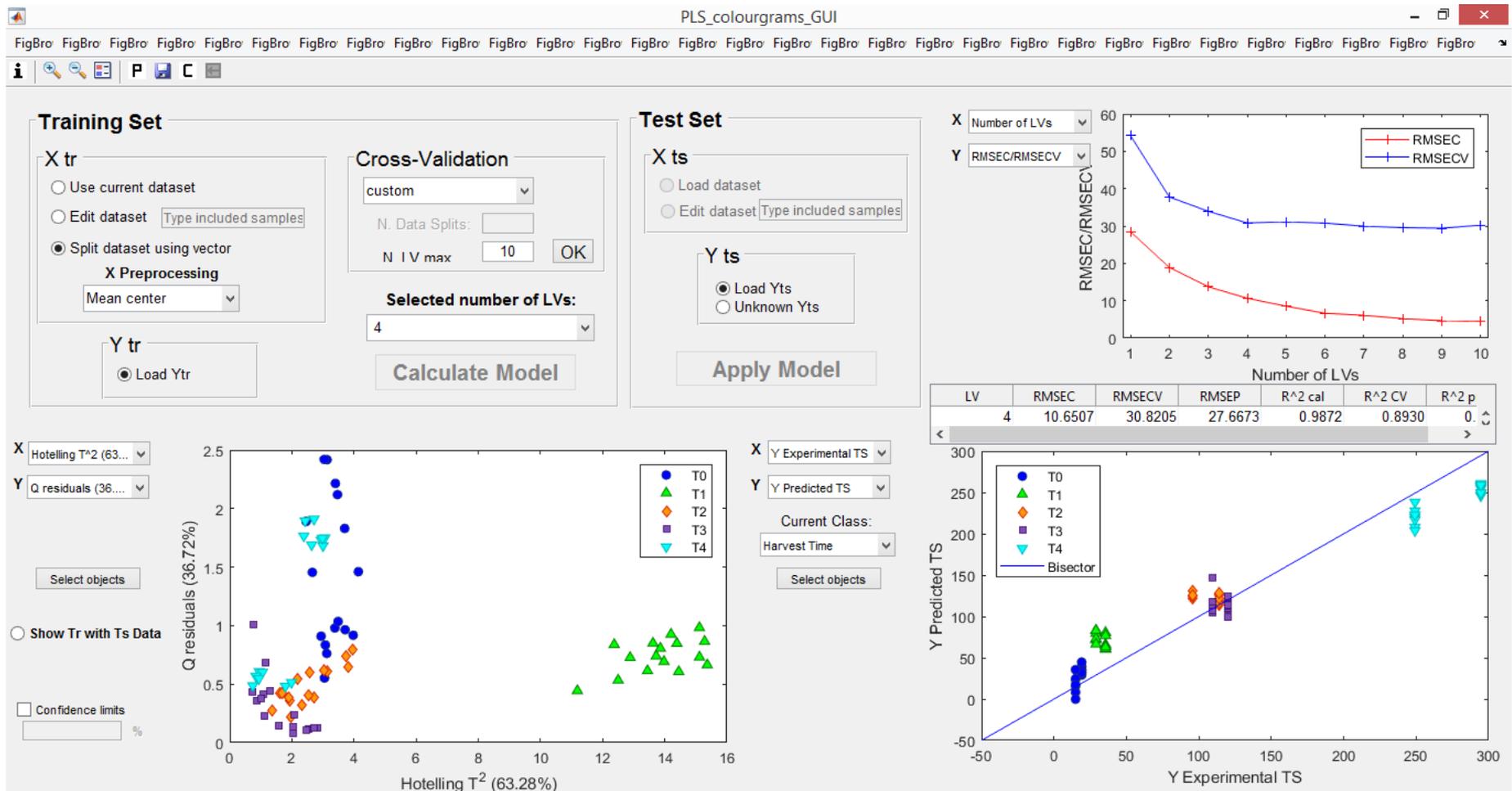
Before enabling the push button “Apply Model”, the information needed for the prediction must be defined. If the whole dataset has been splitted using a vector, it will be necessary to define only the Y of test set. The user can choose between loading a vector of numeric values (“Load Yts”) or applying the model without the value of Yts (“Unknown Yts”).

The screenshot displays the PLS\_colourgrams\_GUI interface. The **Training Set** section includes options for X tr (Use current dataset, Edit dataset, or Split dataset using vector), X Preprocessing (Mean center), Y tr (Load Ytr), and Cross-Validation (custom, N. Data Splits, N I V max, Selected number of LVs: 4, Calculate Model). The **Test Set** section includes options for X ts (Load dataset, Edit dataset) and Y ts (Load Yts, Unknown Yts), with a red arrow pointing to the Y ts options. A table below the test set options shows RMSEC, RMSECV, RMSEP, R<sup>2</sup> cal, R<sup>2</sup> CV, and R<sup>2</sup> p for LV 4. The bottom left shows a scatter plot of Q residuals (36.72%) vs Hotelling T<sup>2</sup> (63.28%) with a 95% confidence limit. The bottom right shows a plot of Y Predicted Cal vs Y Experimental with a 1:1 line and data points for T0, T1, T2, T3, and T4.

LV	RMSEC	RMSECV	RMSEP	R <sup>2</sup> cal	R <sup>2</sup> CV	R <sup>2</sup> p
4	10.6507	30.8205	NaN	0.9872	0.8930	

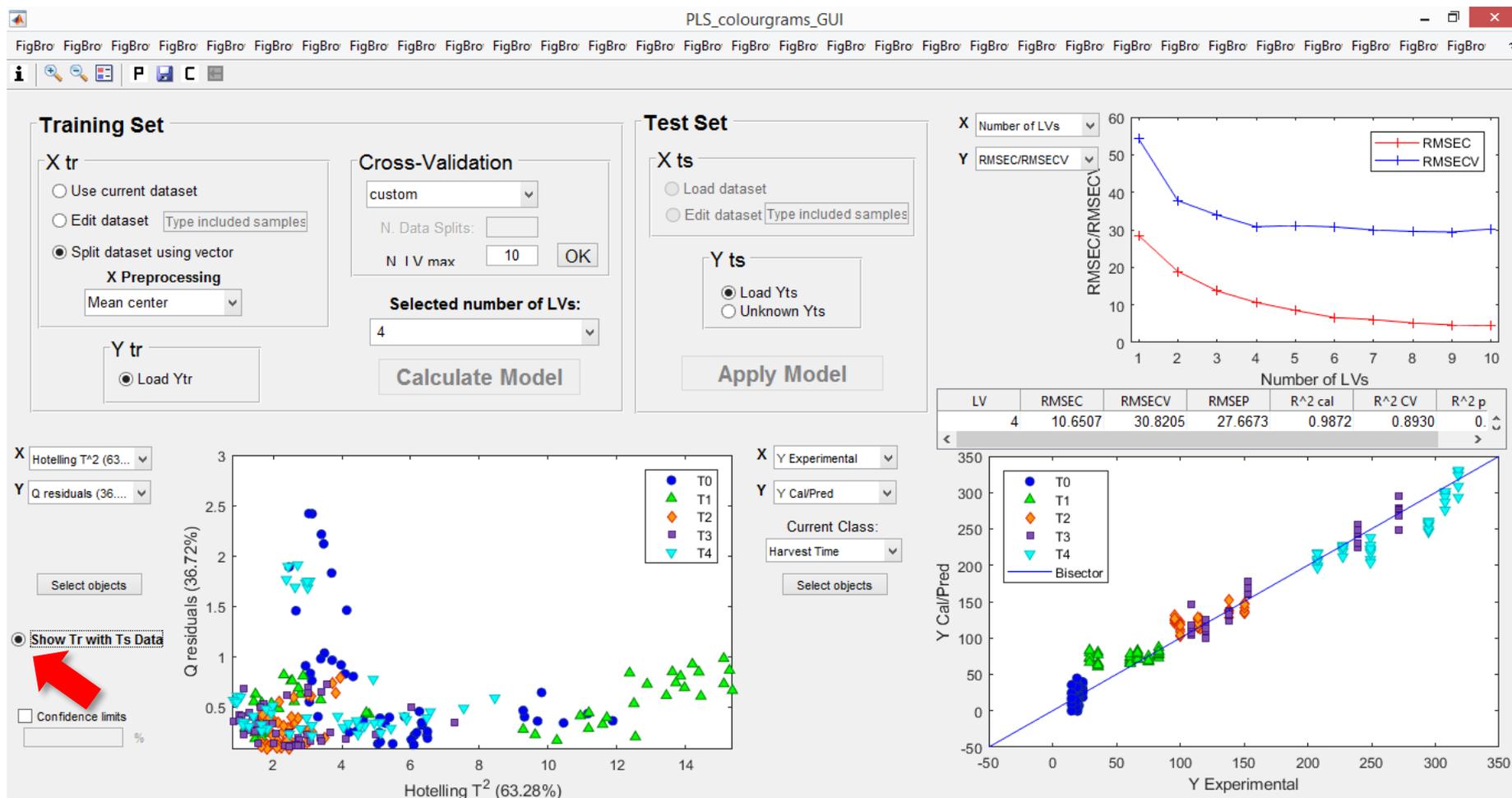
# Apply Model

By clicking on the push button “Apply Model”, the prediction model will be automatically calculated.



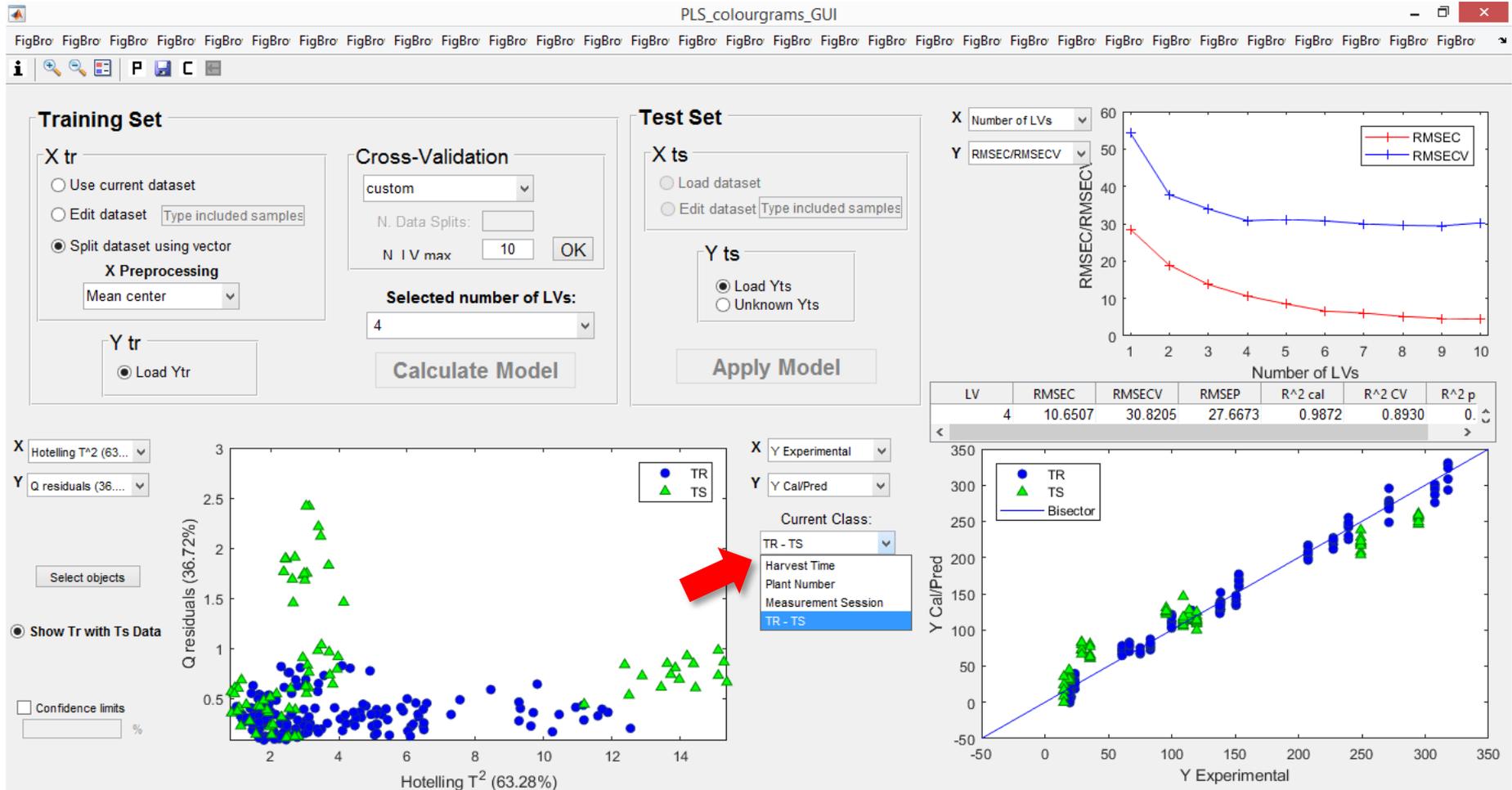
# Show Tr with Ts Data

By clicking on the radio button “Show Tr with Ts Data”, it is possible to visualize in Plot 2 and Plot 3 both training and test data.



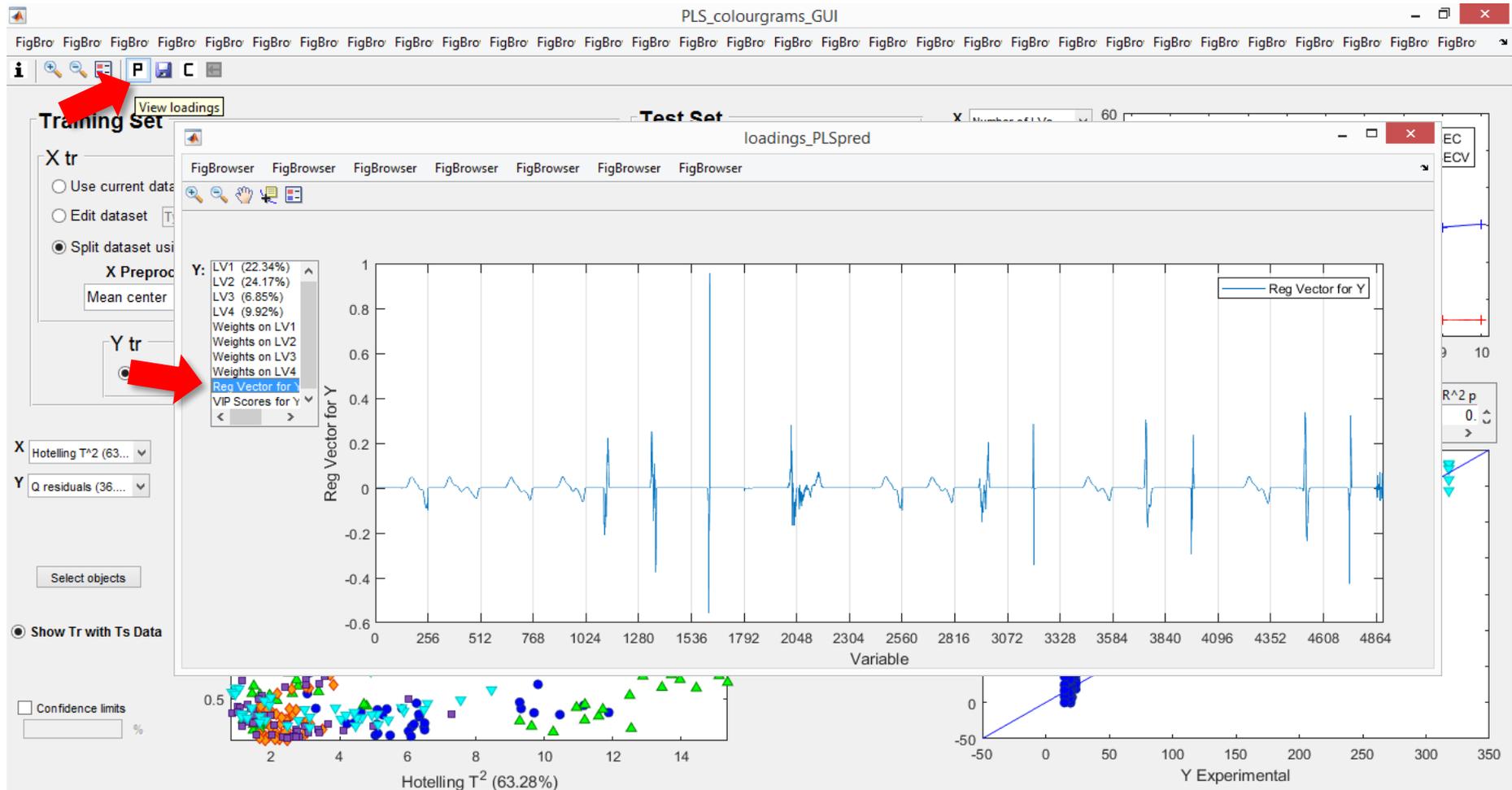
# Current class

By using the “Current Class” drop-down menu, it is possible to change the class identifiers set used to colour the samples. If the radio button “Show Tr with Ts Data” is selected, the class “TR – TS” is automatically generated.



# Loadings, weights, regression vector and VIPs

By using the corresponding “Y” List box, it is possible to chose the quantities to be displayed in the plot.



# Save Data or Model

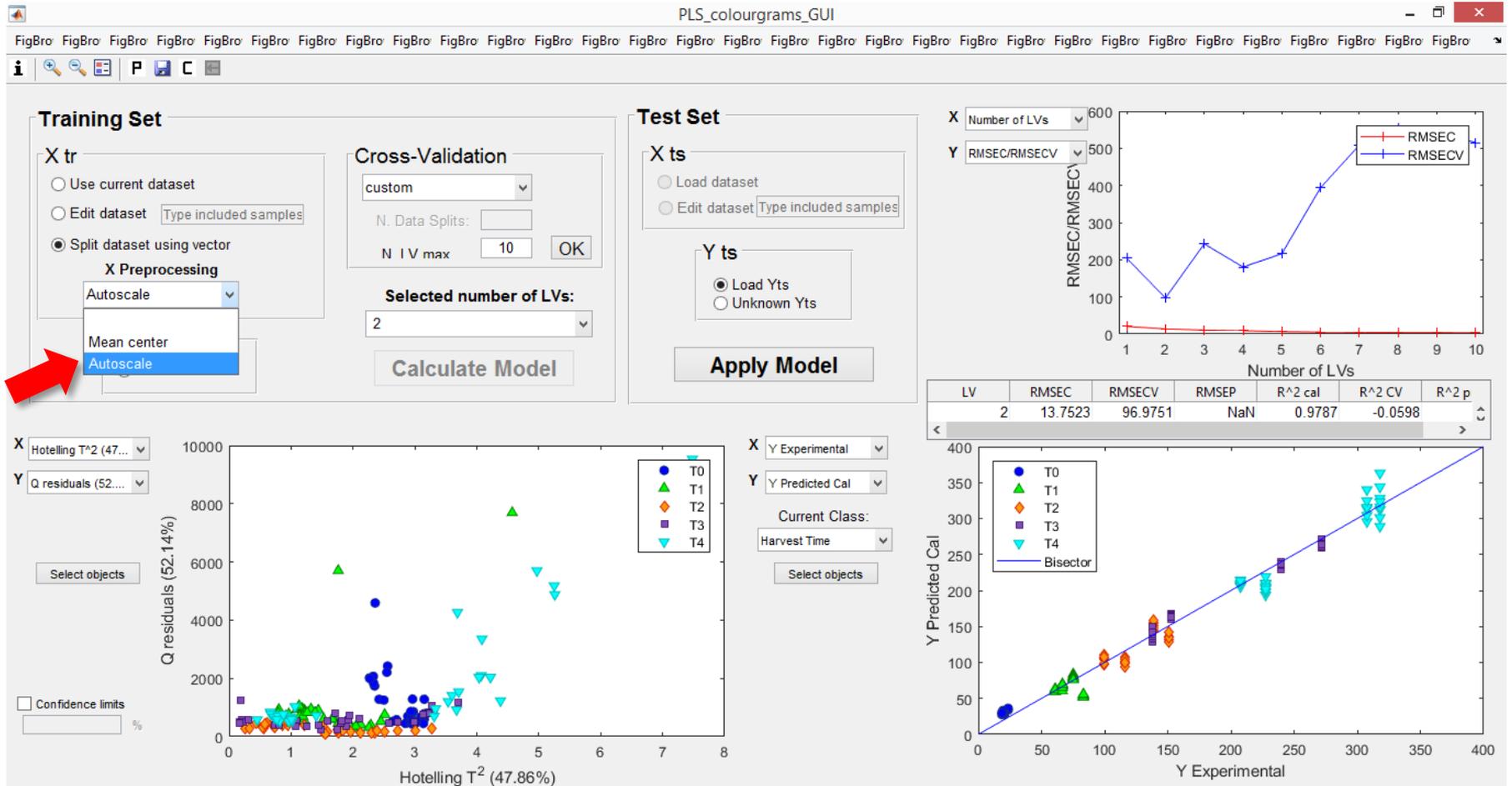
By clicking on the “Save” button, the user can choose whether to save X and Y of training set and test set or a structure array, which contains all the information about the calculated PLS model.

The screenshot displays the PLS\_colourgrams\_GUI software interface. A red arrow points to the 'Save' button in the top toolbar. A 'Save data' dialog box is open in the center, asking 'What do you want to save?' with three options: 'Xtr and Ytr', 'Xts and Yts', and 'PLS model'. The background interface includes sections for 'Training Set', 'Test Set', and 'Cross-Validation'. The 'Training Set' section has options for 'X tr' (Use current dataset, Edit dataset, Split dataset using vector) and 'Y tr' (Load Ytr). The 'Cross-Validation' section has a 'custom' dropdown and 'N. Data Splits' input. The 'Test Set' section has options for 'X ts' (Load dataset, Edit dataset). A plot on the right shows RMSEC and RMSECV vs. Number of LVs. A table below the plot shows RMSEP, R<sup>2</sup> cal, R<sup>2</sup> CV, and R<sup>2</sup> p. The bottom left plot shows Q residuals vs. Hotelling T<sup>2</sup>, and the bottom right plot shows Y Call/Pred vs. Y Experimental.

Number of LVs	RMSEP	R <sup>2</sup> cal	R <sup>2</sup> CV	R <sup>2</sup> p
3	27.6673	0.9872	0.8930	0.8930

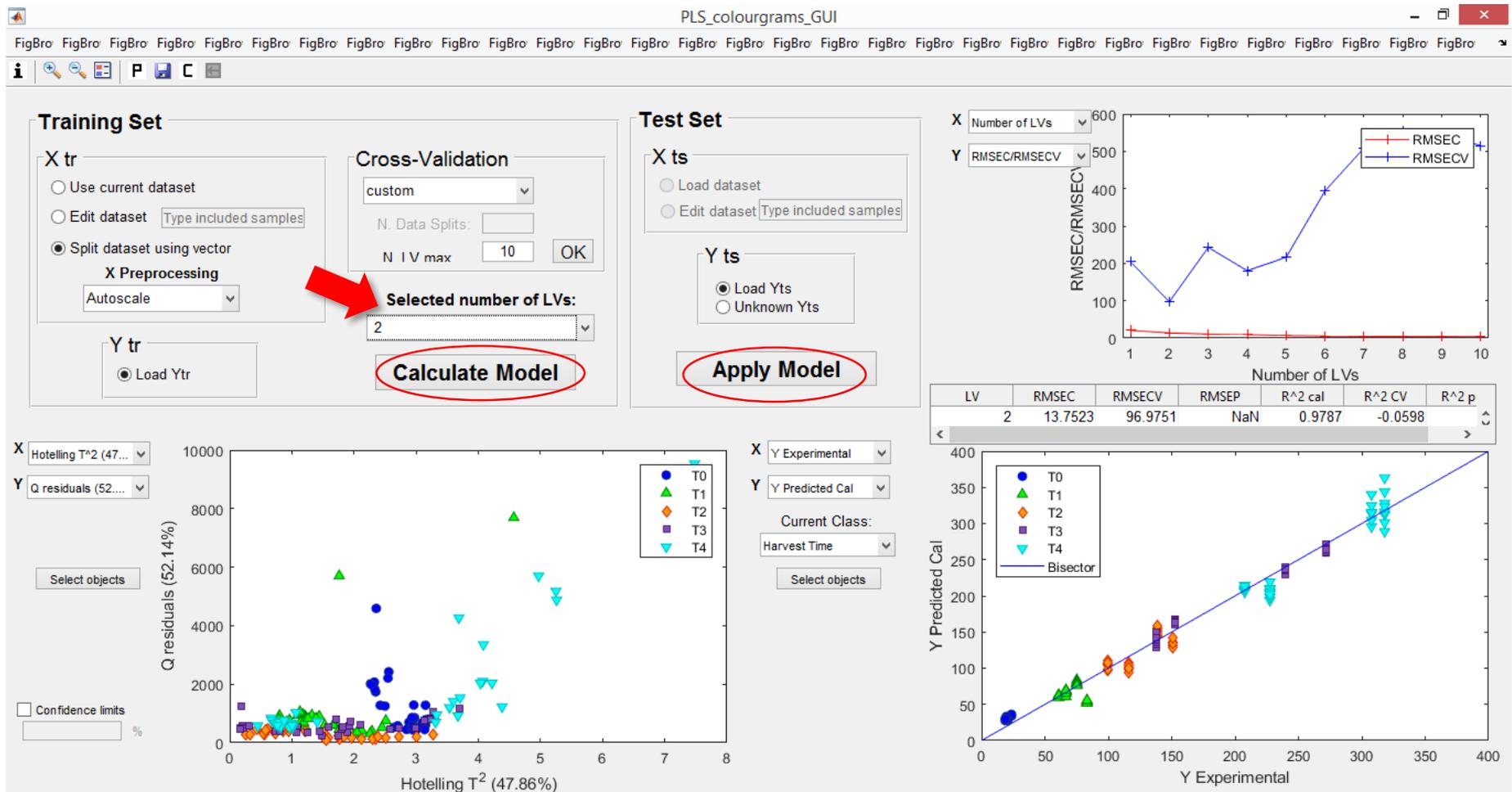
# Change pretreatment of colourgrams matrix

By changing the X preprocessing using the corresponding drop-down menu, the PLS model is automatically updated but it is important to verify the new optimal number of LVs.



# Change selected number of LVs

If the selected number of LVs is changed, the user is required to click on the push buttons “Calculate Model” and “Apply Model” in order to update the PLS model.



# Clear Model

By clicking on the “Clear Model”, it is possible to chose which model the user want to delete in order to start a new session.

The screenshot displays the PLS\_colourgrams\_GUI interface. A red arrow points to the 'Clear model' button in the top toolbar. A dialog box titled 'Clear' is open, asking 'Which model do you want to clear?' with three buttons: 'Clear All', 'Clear Prediction', and 'Cancel'. The background interface includes sections for 'Training Set', 'Cross-Validation', and 'Test Set'. The 'Training Set' section has 'X tr' set to 'Split dataset using vector' and 'X Preprocessing' set to 'Autoscale'. The 'Cross-Validation' section has 'N. Data Splits' set to 'custom' and 'N. I V max' set to '10'. The 'Test Set' section has 'X ts' set to 'Load dataset' and 'Y ts' set to 'Load Yts'. A plot on the right shows 'RMSEC/RMSECV' vs 'Number of LVs' with a legend for 'RMSEC' (red line) and 'RMSECV' (blue line). A table below the plot shows the following data:

MSECV	RMSEP	R <sup>2</sup> cal	R <sup>2</sup> CV	R <sup>2</sup> p
96.9751	32.2149	0.9787	-0.0598	0.0

Another plot at the bottom right shows 'Y Predictor' vs 'Y Experimental TS' with a legend for 'T0' (blue circle), 'T1' (green triangle), 'T2' (orange diamond), 'T3' (purple square), 'T4' (cyan inverted triangle), and 'Bisector' (blue line). A scatter plot at the bottom left shows 'Q residuals (52.14%)' vs 'Hotelling T<sup>2</sup> (47.86%)' with a legend for 'Show Tr with Ts Data' and 'Confidence limits'.