Manuscript Details

Manuscript number	CHEMOLAB_2019_449
Title	Colourgrams GUI: a graphical user-friendly interface for the analysis of large datasets of RGB images
Article type	Original software publication
Keywords	Colourgrams; RGB images; Multivariate image analysis; Data exploration; Multivariate calibration; MATLAB Graphical User Interface
Taxonomy	Interactive Data Visualization, Multivariate Regression, 3D Data Compression, Multivariate Statistics
Corresponding Author	Rosalba Calvini
Order of Authors	Rosalba Calvini, Giorgia Orlandi, Giorgia Foca, Alessandro Ulrici
Suggested reviewers	Aoife Gowen, Jose Manuel Amigo, davide ballabio

Submission Files Included in this PDF

File Name [File Type]

ColGUI_coverLetter_CHEMOLAB.pdf [Cover Letter]

ColGUI_presentation_ok.docx [Manuscript File]

Figure01_keyStepsColourgrams.tif [Figure]

Figure02_OperantingProcedure.tif [Figure]

Figure03_mainWindow.tif [Figure]

Figure04_loadClasses.tif [Figure]

Figure05_clearModify.tif [Figure]

Figure06_selectClass.tif [Figure]

Figure07_plotSelectedCurves.tif [Figure]

Figure08_PCAwindow.tif [Figure]

Figure09_PCAselObj.tif [Figure]

Figure10_PCAloadings.tif [Figure]

Figure11_PLS_CV.tif [Figure]

Figure12_PLS_prediction.tif [Figure]

Figure13_ImageReconstruction.tif [Figure]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Dipartimento di Scienze della Vita

Reggio Emilia, August 6th, 2019

Dear Editor,

please find enclosed a copy of the manuscript:

R. Calvini, G. Orlandi, G. Foca, A. Ulrici

Colourgrams GUI: a graphical user-friendly interface for the analysis of large datasets of RGB images

that we would like to be considered for publication in Chemometrics and Intelligent Laboratory Systems.

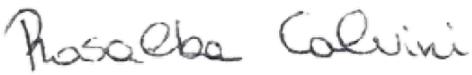
This article is the software description of Colourgrams GUI, a MATLAB based graphical user interface (GUI) for the analysis of RGB images through the colourgrams approach that we have developed.
The colourgrams approach allows the simultaneous analysis of a large number of RGB images altogether and it is essentially based on the conversion of each RGB image of the dataset into a one-dimensional signal, the colourgram, which acts like a fingerprint of the colour properties of the corresponding image. In this manner, a large dataset of images is converted into a matrix of signals, which in turn can be analysed with common chemometric methods for data exploration or for the quantification of colour-related properties of interest.
Among the different features, Colourgrams GUI allows to easily calculate the colourgrams matrix from a dataset of RGB images, to interactively visualize the signals coloured according to qualitative or quantitative properties of the samples and to analyse the colourgram matrix by means of Principal Component Analysis and Partial Least Squares algorithms.
In the article, the different functionalities of Colourgrams GUI are illustrated using a benchmark dataset, which is downloadable with the software.

The software has also been independently tested by Prof. Jose Manuel Amigo (University of Basque Country, Spain) and Prof. Davide Ballabio (University of Milano Bicocca, Italy). Their comments are reported in Section 7 “Independent testing”.

The article is original, unpublished and not being considered for publication elsewhere.

Best regards,

Rosalba Calvini



Dr. Rosalba Calvini, PhD
Department of Life Sciences
University of Modena and Reggio Emilia
Padiglione Besta - Via Amendola, 2 - 42122 Reggio Emilia - Italy
Phone: +39 0522 522048
email: rosalba.calvini@unimore.it

Colourgrams GUI: a graphical user-friendly interface for the analysis of large datasets of RGB images

Rosalba Calvini [a, b, *], Giorgia Orlandi [a], Giorgia Foca [a, b], Alessandro Ulrici [a, b]

[a] *Department of Life Sciences, University of Modena and Reggio Emilia, Padiglione Besta, Via Amendola, 2, Reggio Emilia, 42122, Italy*

[b] *Interdepartmental Centre BIOGEST-SITEIA, University of Modena and Reggio Emilia, Piazzale Europa, 1, Reggio Emilia, 42122, Italy*

* Corresponding author: Rosalba Calvini, rosalba.calvini@unimore.it

Abstract

Colourgrams GUI is a graphical user-friendly interface developed in order to facilitate the analysis of large datasets of RGB images through the colourgrams approach. Briefly, the colourgrams approach consists in converting a dataset of RGB images into a matrix of one-dimensional signals, the *colourgrams*, each one codifying the colour content of the corresponding original image. This matrix of signals can be in turn analysed by means of common multivariate statistical methods, such as Principal Component Analysis (PCA) for exploratory analysis of the image dataset, or Partial Least Squares (PLS) regression for the quantification of colour-related properties of interest.

Colourgrams GUI allows to easily convert the dataset of RGB images into the colourgrams matrix, to interactively visualize the signals coloured according to qualitative and/or quantitative properties of the corresponding samples and to visualize the colour features corresponding to selected colourgram regions into the image domain. In addition, the software also allows to analyse the colourgrams matrix by means of PCA and PLS.

Keywords

Colourgrams; RGB images; Multivariate image analysis; Data exploration; Multivariate calibration; MATLAB Graphical User Interface

1. Introduction

The colour properties of an object are related to its chemical and physical attributes, which in turn characterize the whole product. As an example, the visual aspect of food commodities plays a paramount role in the final decision of consumers, since the colour of food is strongly related with quality factors, such as freshness, safety and ripening among others [1].

Colour can be defined as the perceived response resulting from the interaction between the human eye and the visible light emitted or reflected by an object [2]. Colour is not an intrinsic property of an object, since it is strongly influenced from different factors such as type and angle of illumination, and characteristics of the perceiving subject [3, 4]. For these reasons, in the past decades a great effort has been devoted towards the implementation of automated measurement systems able to give in short times an objective evaluation of the colour properties of a sample, being at the same time cheap and easily transferrable between different laboratories or production plants [5 - 8].

While colour measurement of homogeneous samples (like e.g. a tomato sauce) can be performed quite easily by means of colorimeters [9 – 11], in the case of heterogeneous samples (like e.g. not completely ripe tomatoes) it is generally not sufficient to quantify the colour of local areas or to determine the average sample colour. In the case of unevenly distributed pigments or when dealing with defect detection issues, it is in fact necessary to quantify the value and the extent of all the colours which are present on the sample surface. In addition, in some applications the evaluation of chromatic variations occurring over time may help to monitor the evolution of a given process, like e.g. in the case of ripening of fruits and vegetables.

The colour measurement systems that allow to account for colour variability are based on the use of Red-Green-Blue (RGB) sensor arrays, whose photosensitive elements are able to mimic the response of human vision. Indeed, each pixel of an RGB image contains the information about the amount of light captured by the sensor in the red (λ = 700.0 nm), green (λ = 546.1 nm) and blue (λ = 435.8 nm) regions of the visible spectrum [12], which lead to a different response of the L, M and S types of the cone cells of the human eye . When using constant and uniform illumination conditions, RGB imaging systems allow to measure the colour properties of the analysed sample for each individual pixel, giving the possibility to effectively analyse surfaces with inhomogeneous colour [13].

The acquisition of an RGB image is very quick and can be done using cheap instrumentations such as cameras, flatbed scanners, smartphones or webcams. The most complex task consists in the subsequent extraction and quantification of the useful information from the large amount of data contained in each RGB image, corresponding to the R, G and B values measured for tens of

millions of pixels. This issue can be obviously solved by using proper Multivariate Image Analysis (MIA) methods, which should be conceived in a way to make the image analysis phase as much easy and effective as possible.

The most widespread MIA strategies are based on a pixel-level approach, which consist in considering each single pixel of the analysed image as a separate object [14 - 18].

However, for practical applications of computer vision systems it is often necessary to evaluate the variability of large sets of samples, which implies the need to consider large datasets of images. In these situations, hundreds or even thousands of images must be compared simultaneously, thus the analysis at the pixel-level would require strong computation efforts, or it could be even unfeasible [19].

In order to overcome this issue, some of us have proposed a data reduction method which consists in converting each three-dimensional RGB image into a one-dimensional signal, named colourgram [13]. Basically, each colourgram is obtained by merging in sequence the frequency distribution curves of the R, G and B channels of the original image and of additional colour parameters derived from the R, G and B values. In this manner, the whole dataset of images is converted into a matrix of signals, where each row corresponds to a signal codifying the colour properties of a specific image in the dataset. The basic idea of the colourgrams approach is to furnish a comprehensive investigation of all the colour aspects of an image, without requiring any a priori assumption about the colour parameters of interest for the problem under investigation.

The matrix of colourgrams can then be analysed by means of multivariate statistical methods, allowing to visualize the overall structure of the dataset, to highlight clusters of similar images, to identify the presence of outlier images, to calculate calibration or classification models, and to select the colour features relevant to a specific problem.

The colourgrams approach was successfully applied to different issues, mainly related to food matrices. The first application of colourgrams was aimed at discriminating different pesto brands [13] and the promising results suggested to use the same strategy for the quantification of pigments and sensory attributes in pesto sauce [20]. Further applications related to the quantification of properties of interest include the determination of *Lactobacillus* concentration in fermented milk [21], the estimate of maize nitrogen-related variables [22] and the prediction of parameters related to grape ripening [23]. Thanks to its ability to preserve the information related to colour variability within the images, the colourgrams approach resulted to be particularly effective when dealing with the identification of spatially-resolved features, such as the presence of defects. In this context, colourgrams were successfully applied for the detection of red skin defect in raw hams [24], the evaluation of the effects caused by bacterial infection in citrus plant leaves [25], and the automated

identification of defective kernels in maize [26] and hazelnuts [27]. In addition, the possibility of converting the colour content of RGB images into one-dimensional signals also allowed to fuse these "electronic eye" data with electronic tongue measurements to monitor grape ripening [28].

The effectiveness of this approach induced us to adapt it also to the analysis of large datasets of hyperspectral images, leading to the hyperspectrograms approach, which allowed to obtain satisfactory results in various applications [19, 29 - 33].

In order to facilitate the analysis of large datasets of RGB images using the colourgrams approach, we have developed an easy-to-use graphical user interface, named *Colourgrams GUI*. The software includes all the features necessary to convert the images into colourgrams, and to further analyse the matrix of signals by means of PCA to gain an overview of the whole dataset of images and by means of PLS for the quantification and prediction of properties of interest.

2. The *Colourgrams* approach

The colourgrams approach basically consists in converting each RGB image into a one-dimensional signal, i.e. the colourgram, which summarizes the colour content of the corresponding image.

The main steps involved in the conversion of the image into the respective colourgram are schematically reported in Figure 1.

At first, the RGB image, with size $\{r, c, 3\}$ (where r is the number of pixel rows, c is the number of pixels columns and 3 corresponds to the R, G and B channels), is unfolded into a two-dimensional matrix with as many rows as the number of pixels ($r \times c$) and 3 columns corresponding to the R, G and B values (Step 1 in Figure 1).

Afterwards, the unfolded RGB matrix is further expanded by calculating for each pixel a series of colour-related parameters derived from the R, G and B values (Step 2 in Figure 1). These parameters include: lightness (L), calculated as the sum of R, G and B values; the relative colours (rR, rG and rB), calculated as the ratio between each RGB value and lightness; the hue (H), saturation (S) and intensity (I) values, obtained by converting the RGB colour space into the HSI colour space; the score vectors derived from three PCA models calculated on the raw (SC1R, SC2R, SC3R), mean centered (SC1M, SC2M, SC3M) and autoscaled (SC1A, SC2A, SC3A) RGB data matrix, respectively. In this manner, a matrix containing 19 variables for each pixel is obtained.

Then, for each variable a 256 bins-long frequency distribution vector is calculated (Step 3 in Figure 1), and the 19 frequency distribution vectors are merged in sequence. Finally, the loading vectors and the eigenvalues of the PCA models are added at the end of the signal (Step 4 in Figure 1).

In this manner, the original RGB image is converted into a one-dimensional signal with size equal to 4900 points [= 256 bins × 19 variables + 3 PCA models × 3 PCs × (3 loading values + 1 eigenvalue)], which codifies the colour content of the image.

For further details about the algorithm the reader is referred to [13].

Figure 1 near here

3. Software specifications and requirements

Colourgrams GUI has been developed under MATLAB environment (ver. 7.11, The MathWorks, USA) and it is freely downloadable from http://www.chimslab.unimore.it/downloads/ . The download package includes all the routines necessary to run the GUI and also the Fast User Manual as .pdf file, which guides the user through the main features of the software.

Colourgrams GUI requires the MATLAB Image Processing Toolbox (The MathWorks, USA) and the DataSet Object Toolbox (Eigenvector Research Inc., USA); this latter one is freely downloadable from https://www.mathworks.com/matlabcentral/fileexchange/39336-dataset-object [34]. The software works under different versions of the Microsoft Windows operating system (7 / 8 / 8.1 / 10) and has been tested using different MATLAB versions (R2010/R2014/R2015/R2016/R2017).

Once all the Colourgrams GUI functions are added to the MATLAB path, the software can be easily started by typing *colourgrams_GUI* in the MATLAB command window.

4. Dataset

In order to describe the use of Colourgrams GUI, the *GrapeMusts* benchmark dataset containing 240 images of grape must is employed, downloadable with the MATLAB code.

The must samples were obtained from grape berries belonging to Lambrusco Marani grape variety and collected in the vineyard from 3 different plants at 5 subsequent harvest times, from veraison until complete ripening. After homogenization and maceration of the grapes, the corresponding must was extracted and, for each sample, 8 aliquots of 50 µL of must were deposited on a white sheet of absorbent paper. RGB images of the absorbent paper sheets were acquired with a common flatbed scanner (CanoScan Lide 220, Canon). The same procedure was repeated in two different measurement sessions. Then, the image area of each spot of must was cropped and saved as a separate image with constant pixel size, for a total of 240 RGB images with size equal to 900 × 900 pixels (= 3 plants × 5 harvest times × 8 aliquots × 2 measurement sessions). Together with image acquisition, the must samples were also analysed by means of reference analytical methods,

based on UV–Vis spectrophotometry and high-performance liquid chromatography, for the determination of a series of parameters usually considered to assess phenolic maturity of grapes. These parameters include colour index, total flavonoids content and total anthocyanins content, among others. The goal of this study, reported in [23], was to relate the images of the must samples with the level of phenolic maturity and to predict the corresponding analytical parameters of interest.

In the present paper, this dataset is used to illustrate the main features of Colourgrams GUI.

5. Operating procedure

The operating procedure generally applied when analysing datasets of RGB images by means of Colourgrams GUI is summarized in Figure 2.

Figure 2 near here

Briefly, once the dataset of RGB images has been converted into the corresponding matrix of colourgrams, additional information can be added to each image of the dataset, like e.g. presence/absence of defects, variety, brand etc. Then, as a first approach, the whole matrix of signals can be simply visualized in order to observe possible outlier signals, trends of interest or differences between groups of signals. For a more comprehensive overview of the dataset, the colourgrams matrix can be analysed by means of PCA, allowing to better identify groups of similar images and/or to highlight outliers. In addition, the RGB images and the corresponding colourgrams can be used to quantify some properties of interest of the samples using the Partial Least Squares (PLS) algorithm. Finally, in order to gain a better understanding of the results of PCA and/or of PLS, image reconstruction allows to visualize the relevant colour features back into the original image domain.

Colourgrams GUI is designed in order to easily manage the operating procedure reported in Figure 2 by means of three windows: the main window, the PCA window and the PLS window.

The main window of Colourgrams GUI (Figure 3), that can be opened by typing *colourgrams_GUI* in the command window of MATLAB, is subdivided into different sections according to the steps of the operating procedure that will be described in the following sections.

Figure 3 near here

5.1 Create or load colourgrams

Once Colourgrams GUI is launched from the MATLAB command window, it is possible to create a new dataset of signals (radio button “Create Colourgrams”) or to load an existing one (radio button

"Load Colourgrams"). Firstly, the user has to select the folder containing the images or a previously saved dataset by clicking on the folder icon in the toolbar.

To create a new dataset of colourgrams, different options are available. At first, the user has to select the image extension, choosing between the JPG, BMP and TIF image file formats. In addition, it is possible to decide whether calculating the signals from the whole original images ("Normal" mode) or from the images after background removal ("Keep only" mode). Background removal is performed by a thresholding procedure based on a defined colour parameter, like e.g. red, blue, hue, etc. In particular, in the "Keep only" mode the colourgrams are calculated including only the pixels with values greater than (*">"* mode) or less than (*"<"* mode) the threshold value.

Concerning the *GrapeMusts* dataset, colourgrams were calculated after background removal, that was performed by keeping only the pixels with relative red values greater than 0.34 (i.e., selecting: "Keep only" "Rel. Red" ">" "0.34"). The procedure that was followed to select this threshold value consisted in a first analysis of the colourgrams calculated on the whole images ("Normal" mode). Then, through the visualization of the frequency distribution curves of the single colour-related parameters (*see* section 5.3), and by reconstructing the images corresponding to selected portions of the frequency distribution curves (*see* section 5.6), it was possible to identify the relative red value equal to 0.34 as the optimal threshold for background removal.

Once all the RGB images are converted into the corresponding colourgrams, the matrix of signals is automatically saved as DataSet Object into a folder named "Generated_Colourgrams" and followed by date and time of creation (e.g. Generated_Colourgrams_20190611_T124823). In addition, in the same folder also an MS-Excel file named "Colourgrams_info.xls" is saved, which contains the names of the original RGB images and that can be modified to include additional information about the dataset, like e.g. class vectors.

Once the matrix of colourgrams is calculated or loaded, the signals are plotted on the main window and all the tools for data visualization and data analysis are enabled.

5.2 Add or modify supplementary information to the dataset

The "Class Manager" section of Colourgrams GUI main window can be used to load, modify or save class identifier sets (or class sets) in the current DataSet Object of colourgrams.

More in detail, the push button "Load class" opens a specific window which allows to load a new class vector (Figure 4a). The class vector can be loaded from a MATLAB workspace variable, from a .mat file or from an .xls / .xlsx / .csv file. The length of the class vectors must be equal to the number of signals and class assignment can be expressed using either vectors of numbers (e.g. 1, 2, 3, …) or character arrays (e.g. 'Class A', 'Class B', Class C', …), except for .csv files, from which

only numeric values can be loaded. When the class vector is loaded from an .xls or .xlsx file, the user is required to specify the worksheet number and the range of the class vector (Figure 4b), while if the class vector is loaded from a .csv file the user has to specify the delimiter type and the range of the class vector using an Excel-like notation (Figure 4c).

Figure 4 near here

An already existing class set can be deleted or modified using the push button "Clear/Modify", which opens the window shown in Figure 5.

Figure 5 near here

On the left of the "Clear or Modify Classes" window, a list box reporting all the class sets saved in the current DataSet Object is displayed. When the user selects one of the class sets, a summary table reporting the corresponding information (i.e. class names and number of samples belonging to each class) is shown on the right part of the window. Furthermore, the "Clear or Modify Classes" windows also permits to delete the selected class set and to edit the class set name and the single class names.

If the class sets of the current DataSet Objects have been modified, it is possible to save the updated DataSet Object to a .mat file using the push button "Save" of the Class Manager section of the main window.

Furthermore, the Class Manager allows to select the class identifiers set to be used to define the colours of the signals in the plot of the main window, by using the "Current class" drop-down menu. For example, in Figure 6 the colourgrams shown in the plot are coloured according to the "Harvest Time" class set.

Figure 6 near here

5.3. Visualize signals

The "Visualize Signals" section has three main features:

a) Separate visualization of the frequency distribution curves of the colour-related parameters included in the colourgrams

b) Representation of the signals according to defined properties of the corresponding samples/images

c) Identification of a specific signal/sample

This section has 20 check boxes, each one corresponding to one colour parameter included in the colourgrams. Each check box has a label reporting the name of the corresponding colour parameter

and the respective colourgram region (e.g., 1793:2048 [Hue]). When the user selects one or more check boxes, the corresponding areas of the main plot are highlighted in light blue colour (Figure 7a) and the push button "Plot Selected Curves" allows to plot the frequency distribution curves of the selected colour parameters in a separate figure (Figure 7b).

Figure 7 near here

In addition to plotting the frequency distribution curves coloured according to the selected class set like in Figure 7b, it is also possible to colour the signals based on a continuous variable like in Figure 7c, where the colour coding reflects the total anthocyanins content measured on the corresponding grape samples, and the colour scale reports the total anthocyanins content value corresponding to each colour. To do this, the "Colour By" push button allows to load the vector containing the values of the property of interest by means of a window, named "Load Vector", similar to the one reported in Figure 4a. A new figure is then displayed, where the signals are coloured according to the corresponding values of the property of interest. If one or more check boxes have been previously selected in the "Visualize signals" section, only the frequency distribution curves of these parameters will be displayed, as shown in Figure 7c.

The analysis of plots like those in Figures 7b and 7c allows to make a preliminary assessment of the colour parameters that are mainly related to the problem at hand, i.e., of those parameters for which the change of the shape of the corresponding frequency distribution curves corresponds to different classes and/or to different values of a quantitative property.

Another feature of the "Visualize Signals" section is the possibility to easily retrieve information about which signal corresponds to a given image and, the other way around, about which image corresponds to a given signal. By selecting the name of the image of interest from the drop-down menu placed on the right side of the section, the corresponding signal is represented in magenta colour in the plot. Alternatively, in order to investigate which image corresponds to a given signal, it is possible to select the signal directly on the plot using the "Data Cursor" button in the toolbar and then, by clicking on the "Select image on the plot" button, the name of the corresponding image is automatically shown in the drop-down list.

5.4. Data exploration by PCA

By clicking on the "PCA" button of the main window, the PCA window reported in Figure 8 is displayed.

Figure 8 near here

By default, the PCA model is calculated on the colourgrams matrix considering mean center as signal preprocessing method and 2 principal components. Then, the user can select the proper data preprocessing method (none, mean center or autoscale) and the optimal number of PCs using the corresponding drop-down menus located in the upper left part of the PCA window. The proper number of PCs can be selected based on the scree plot reported in the upper right part; the drop-down menu associated with this plot allows to choose between visualizing the eigenvalues, the percentage of variance explained by each PC, the percentage of cumulative variance or the logarithm of the eigenvalues.

The PCA scores are reported in the two plots placed at the bottom of the PCA window. By default, the plot at the bottom left shows the Q residuals vs Hotelling T^2 plot, while the plot at the bottom right shows the PC1 vs PC2 score plot. However, it is possible to change the variables displayed in both plots using the "X" and "Y" drop-down menus associated with each plot. In addition, if the Q residuals and/or the Hotelling T^2 values are displayed in the plots, the corresponding confidence limits can be visualized by selecting the "Confidence limits" check box and typing the probability value (as shown in Figure 8). If a specific colourgram has been previously selected in the main window, the corresponding object is reported in the two plots as a magenta hexagram. Furthermore, the objects/images reported in the two plots can be coloured according to their class membership; the different class sets can be selected from the "Current class" drop-down menu.

Similarly to what described in Section 5.3, the objects reported in the score plots can also be coloured according to a quantitative property associated to the samples, like the total anthocyanins content in our example. This can be done by clicking on the "Colour by" button in the toolbar, which opens a window that allows to load the vector of numeric values, as described in Section 5.3.

The "Select objects" push buttons associated with each plot located at the bottom of PCA window allow to select directly on the plot some objects of interest, like for example outlier objects. The selected objects are then represented with magenta hexagrams in both plots and, furthermore, by right-clicking on the hexagrams a context menu is displayed which allows to visualize the labels of the selected objects, deselect the objects or eliminate them from the dataset (Figure 9).

Figure 9 near here

When the selected objects are deleted, the PCA model is automatically recalculated considering only the retained samples. If necessary, it is possible to re-include the eliminated samples by clicking on the "Reinclude Samples" button in the toolbar. Moreover, it is also possible to save the current PCA model or the modified (i.e. without the eliminated samples) DataSet Object to a .mat file, by clicking on the "Save Data" button in the toolbar.

By clicking on the "Only selected variables" radio button, it is possible to calculate the PCA model considering only a user-defined subset of the colourgram variables. The user can choose whether using the variables previously selected in the "Visualize Signals" section of the Colourgrams GUI main window, using the variables employed for image reconstruction (*see* section 5.6), or typing the indexes of the colourgram regions of interest.

The loading vectors of the current PCA model can be visualized using the "P" button in the toolbar. A new window will be displayed allowing to visualize one or more loading vectors in the plot (Figure 10).

Figure 10 near here

5.5 Calibration by PLS

The PLS window can be opened by clicking on the "PLS" push button in the Colourgrams GUI main window. Firstly, for the calculation of the PLS model, the user is required to define the training set: it is possible to use the whole colourgram matrix (radio button "Use current dataset") or a part of it. In this latter case, the training set objects can be selected either by typing the corresponding row indexes (radio button "Edit dataset") or by loading a column vector (radio button "Split dataset using vector") from a MATLAB workspace variable, from a .mat file or from an .xls / .xlsx / .csv file. The vector must have values equal to 1 for samples belonging to the training set and equal to 2 for samples belonging to the test set.

Then, the user has to define the preprocessing method of the X-block (mean centering or autoscaling). The training set response variable can be loaded from a MATLAB workspace variable, from a .mat file or from an .xls / .xlsx / .csv file. In the current version of Colourgrams GUI, only PLS1 models can be calculated on the autoscaled response variable.

Regarding cross-validation, it is possible to employ predefined cross-validation schemes (i.e., leave-one-out, contiguous blocks and venetian blinds) or to use a custom splitting of the samples by loading the corresponding cross-validation vector; then, the user must specify the maximum number of PLS latent variables and, in the case of contiguous blocks or venetian blinds CV, the number of data splits.

After cross-validation, a plot appears in the upper right of the PLS window, reporting the RMSEC and RMSECV values versus the number of LVs of the PLS model. This plot can be used to define the optimal dimensionality of the regression model (Figure 11), which can be specified using the "Selected number of LVs" drop-down menu.

Figure 11 near here

Then, after clicking on the "Calculate Model" button, the results of the selected PLS model appear: the software automatically shows a table with the model statistics (RMSEC, RMSECV, R^2_{Cal} and R^2_{CV}) and two graphs, reporting the Q residuals *vs.* Hotelling T^2 plot and the Y predicted *vs.* Y measured plot, respectively. Like for the PCA window, the variables displayed in the plots can be changed using the drop-down menus, and it is possible to select and remove samples.

Once the PLS model has been calculated, the model can be validated. If the option "Split dataset using vector" was previously selected to define the training set, the test set is automatically loaded and the "X ts" group is not active. Otherwise, the test set can be loaded using the "Load dataset" button or it can be defined as a subset of the current DataSet Object by specifying the row indexes of the corresponding samples.

Subsequently, the user can decide whether loading the Y measured values of the test set samples or performing the predictions with unknown Y values.

Once the current PLS model is applied to the test set by clicking on the "Apply Model" button, the outputs displayed in the PLS window are automatically updated: the table of results includes also the RMSEP and R^2_{pred} values, and the two plots in the bottom of the window report the values predicted for the test set samples (Figure 12). By using the "Show Tr with Ts Data" button, it is also possible to compare calibration and prediction results.

Figure 12 near here

Similarly to the PCA window, by clicking on the "P" button in the toolbar of the PLS window a new window appears allowing to visualize the loading vectors, the regression coefficients and the Variable Importance in Projection (VIP) scores.

Furthermore, PLS window allows also to save the current PLS model and/or the datasets (training set and test set) used for model calculation and validation.

5.6 Image reconstruction using the colour features of interest

To better understand the colour properties codified by the different regions of the colourgrams, and to interpret the results of PCA or PLS models calculated on the colourgrams matrix, it is possible to define intervals of colourgram variables and to visualize them back into the original image domain. These intervals of interest can be identified starting from a visual inspection of the signals, from the interpretation of a PCA and/or PLS model (loading vectors, regression coefficients, VIPs) or can be automatically selected using variable selection algorithms implemented in other software packages.

The user can either type the selected intervals in the corresponding edit box of the Colourgrams GUI main window, or click on the "Select regions on the plot" button for manual selection.

In both cases, for each colourgram region where at least one interval has been chosen, a figure is obtained where only the pixels falling in the selected interval(s) are visualized, while the remainder pixels are represented in black or white colour, or in grayscale tones of the original image. For comparison, also the original RGB image is reported together with the reconstructed image (Figure 13). Image reconstruction is performed considering the image currently displayed in the drop-down menu of "Visualize Signals" section; a different image can be selected from this drop-down menu.

As an example, Figure 13 shows the results of image reconstruction performed considering two intervals, corresponding to the relative green (Figure 13b) and relative blue regions (Figure 13c) of the colourgram, respectively. In this specific example, the colourgram intervals have been chosen based on the inspection of the regression vector related to the PLS model reported in Section 5.5 for the prediction of the total anthocyanins content.

Figure 13 near here

6. Conclusions

Colourgrams GUI is a user-friendly free graphical interface which allows the analysis of large datasets of RGB images through the colourgrams approach.

The GUI has been designed in order to easily manage in few windows the whole workflow, from the conversion of the images into colourgrams, to the visualization of the colourgrams matrix, its further statistical analysis by means of PCA and PLS, and the interpretation of the relevant colourgram features through image reconstruction. Furthermore, thanks to the easy-to-use interface, Colourgrams GUI does not need any prior knowledge of MATLAB programming language, and it can be used also by beginners with a basic background in PCA and PLS.

Further implementations are planned, in order to add new functionalities such as classification and variable selection tools. Moreover, an analogous GUI for the analysis of hyperspectral images through a similar approach (hyperspectrograms) is also under construction.

Colourgrams GUI is freely available from the website http://www.chimslab.unimore.it/downloads/ (last accessed, July 2019) and the software package is provided along with a user manual which illustrates the main features and operating procedures to run the GUI.

7. Independent testing

Prof. Davide Ballabio

Milano Chemometrics and QSAR Research Group, Department of Earth and Environmental Sciences, University of Milano Bicocca, P.zza della Scienza 1, 20126, Milano, Italy.

I have tested the "Colourgrams GUI" (graphical user-friendly interface for MATLAB for the analysis of RGB images) and I found that it appears to function as the authors described.

Prof. Jose Manuel Amigo

Ikerbasque, Basque Foundation for Science, 48011, Bilbao, Spain.

Department of Analytical Chemistry, University of Basque Country, Barrio Sarriena s/n. 48940, Leioa, Spain.

When analysing a batch of RGB images, it is important to standardize a tailored procedure and retrieve results that can answer the analytical problem. This is precisely what Colourgrams GUI does. Colourgrams GUI mixes different image features with some basic chemometric tools (e.g. PCA and PLS) in a dynamic and clever manner. Colourgrams GUI is thought for those researchers who need a fast and reliable way of handling many RGB images. Moreover, Colourgrams GUI is presented in a GUI format that highlights the visual results rather than the plane numbers. This is something important for what Image analysis concerns. Of course there is place for improvements. Nevertheless, I truly believe that this version of Colourgrams GUI is a round version that will grow with the time and with the suggestions of the users.

I have tried Colourgrams GUI myself in two different MATLAB versions (v 2018b and v 2019a) and it does not show any relevant issue.

References

[1] T.N. McCaig, Extending the use of visible/near-infrared reflectance spectrophotometers to measure colour of food and agricultural products. *Food Research International*, 35(8) (2002), 731-736.

[2] F. J. Francis, Quality as influenced by colour. *Food quality and preference*, 6(3) (1995), 149-155.

[3] A. J. Meléndez-Martínez, I. M. Vicario, F. J. Heredia, Instrumental measurement of orange juice colour: A review. *Journal of the Science of Food and Agriculture*, 85(6) (2005), 894-901.

[4] D. Wu, D. W. Sun, Colour measurements by computer vision for food quality control–A review. *Trends in Food Science & Technology*, 29(1) (2013), 5-20.

[5] S. Gunasekaran, Computer vision technology for food quality assurance. *Trends in Food Science & Technology*, 7(8) (1996), 245-256.

[6] T. Brosnan, D. W. Sun, Improving quality inspection of food products by computer vision—a review. Journal of food engineering, 61(1) (2004), 3-16.

[7] A. Vidal, P. Talens, J. M. Prats-Montalbán, S. Cubero, F. Albert, J. Blasco, In-line estimation of the standard colour index of citrus fruits using a computer vision system developed for a mobile platform. *Food and Bioprocess Technology*, 6(12) (2013), 3412-3419.

[8] P. B. Pathare, U. L. Opara, F. A. J. Al-Said, Colour measurement and analysis in fresh and processed foods: a review. *Food and Bioprocess Technology*, 6(1) (2013), 36-60.

[9] A. I. Carrapiso, C. Garcia, Instrumental colour of Iberian ham subcutaneous fat and lean (biceps femoris): Influence of crossbreeding and rearing system. *Meat Science*, 71 (2005), 284-290.

[10] G. Foca, D. Salvo, A. Cino, C. Ferrari, D. P. Lo Fiego, G. Minelli, A. Ulrici, Classification of pig fat samples from different subcutaneous layers by means of fast and non-destructive analytical techniques. *Food Research International*, 52 (1) (2013), 185-197.

[11] A. Herrera, D. Ballabio, N. Navas, R. Todeschini, C. Cardell, Principal Component Analysis to interpret changes in chromatic parameters on paint dosimeters exposed long-term to urban air. *Chemometrics and Intelligent Laboratory Systems*, 167 (2017), 113-122.

[12] N. Ohta, A. R. Robertson, Colorimetry: Fundamentals and Applications (2006).

[13] A. Antonelli, M. Cocchi, P. Fava, G. Foca, G. C. Franchini, D. Manzini, A. Ulrici, Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-based classification algorithm. *Analytica Chimica Acta*, 515(1) (2004), 3-13.

[14] K. Esbensen, P. Geladi, Strategy of multivariate image analysis (MIA). *Chemometrics and Intelligent Laboratory Systems*, 7(1-2) (1989), 67-86.

[15] P. Geladi, H. F. Grahn, Multivariate image analysis. *Encyclopedia of Analytical Chemistry* (1996).

[16] J. M. Prats-Montalbán, A. De Juan, A. Ferrer, Multivariate image analysis: a review with applications. Chemometrics and Intelligent Laboratory Systems, 107(1) (2011), 1-23.

[17] C. Duchesne, J. J. Liu, J. F. MacGregor, Multivariate image analysis in the process industries: A review, *Chemometrics and Intelligent Laboratory Systems*, 117 (2012), 116-128.

[18] J. M. Amigo, H. Babamoradi, S. Elcoroaristizabal, Hyperspectral image analysis. A tutorial. *Analytica Chimica Acta* (2015), 896, 34-51.

[19] R. Calvini, G. Foca, A. Ulrici, Data dimensionality reduction and data fusion for fast characterization of green coffee samples using hyperspectral sensors. *Analytical and bioanalytical chemistry*, 408(26) (2016), 7351-7366.

[20] G. Foca, F. Masino, A. Antonelli, A. Ulrici, Prediction of compositional and sensory characteristics using RGB digital images and multivariate calibration techniques. *Analytica chimica acta*, 706(2) (2011), 238-245.

[21] A. Borin, M. F. Ferrão, C. Mello, L. Cordi, L. C. Pataca, N. Durán, R. J. Poppi, Quantification of Lactobacillus in fermented milk by multivariate image analysis with least-squares support-vector machines. *Analytical and bioanalytical chemistry*, 387(3) (2007), 1105-1112.

[22] M. Corti, D. Cavalli, G. Cabassi, A. Vigoni, L. Degano, P. Marino Gallina, Application of a low-cost camera on a UAV to estimate maize nitrogen-related variables. *Precision Agriculture,* 20 (2019), 675-696.

[23] G. Orlandi, R. Calvini, L. Pigani, G. Foca, G. Vasile Simone, A. Antonelli, A. Ulrici, Electronic eye for the prediction of parameters related to grape ripening. Talanta, 186 (2018), 381-388.

[24] A. Ulrici, G. Foca, M. C. Ielo, L. A. Volpelli, D. P. Lo Fiego, Automated identification and visualization of food defects using RGB imaging: Application to the detection of red skin defect of raw hams. *Innovative Food Science & Emerging Technologies*, 16 (2012), 417-426.

[25] F. M. V. Pereira, D. M. B. P. Milori, E. R. Pereira-Filho, A. L. Venâncio, M. D. S. T. Russo, P. K. Martins, J. Freitas-Astúa, Fluorescence images combined to statistic test for fingerprinting of citrus plants after bacterial infection. *Analytical Methods*, 3(3) (2011), 552-556.

[26] G. Orlandi, R. Calvini, G. Foca, A. Ulrici, Automated quantification of defective maize kernels by means of Multivariate Image Analysis. *Food Control*, 85 (2018), 259-268.

[27] A. Giraudo, R. Calvini, G. Orlandi, A. Ulrici, F. Geobaldo, F. Savorani, Development of an automated method for the identification of defective hazelnuts based on RGB image analysis and colourgrams, *Food Control*, 94 (2018), 233-240.

[28] G. Orlandi, R. Calvini, G. Foca, L. Pigani, G. Vasile Simone, A. Ulrici, Data fusion of electronic eye and electronic tongue signals to monitor grape ripening. *Talanta*, 195 (2019),181-189.

[29] C. Ferrari, G. Foca, A. Ulrici, Handling large datasets of hyperspectral images: Reducing data size without loss of useful information. *Analytica Chimica Acta*, 802 (2013), 29-39.

[30] C. Ferrari, G. Foca, R. Calvini, A. Ulrici, Fast exploration and classification of large hyperspectral image datasets for early bruise detection on apples. *Chemometrics and Intelligent Laboratory Systems*, 146 (2015), 108-119.

[31] G. Foca, C. Ferrari, A. Ulrici, G. Sciutto, S. Prati, S. Morandi, M. Brasca, P. Lavermicocca, S. Lanteri, P. Oliveri, The potential of spectral and hyperspectral-imaging techniques for bacterial detection in food: A case study on lactic acid bacteria. *Talanta*, 153 (2016), 111-119.

[32] J. L. Xu, C. Riccioli, D. W. Sun, Efficient integration of particle analysis in hyperspectral imaging for rapid assessment of oxidative degradation in salmon fillet. *Journal of Food Engineering*, 169 (2016), 259-271.

[33] M. Corti, P. Marino Gallina, D. Cavalli, G. Cabassi, Hyperspectral imaging of spinach canopy under combined water and nitrogen stress to estimate biomass, water, and nitrogen content. *Biosystems Engineering*, 158 (2017), 38-50.

[34] Koch, S. (2019). DataSet Object (https://www.mathworks.com/matlabcentral/fileexchange/39336-dataset-object), MATLAB Central File Exchange. Retrieved July 16, 2019.

Figure captions

Figure 1. Key steps involved in the conversion of an RGB image into the corresponding colourgram.

Figure 2. Operating procedure of Colourgrams GUI

Figure 3. Colourgrams GUI main window. The different sections are numbered according to the steps of the operating procedure reported in Figure 2: 1) create or load colourgrams, 2) add supplementary information to the dataset, 3) visualize signals, 4) perform PCA, 5) perform PLS and 6) image reconstruction using the features of interest.

Figure 4. **"**Load Vector" window (a) and windows for loading of the vector from MS-Excel files (b) or .csv files (c).

Figure 5. "Clear or Modify Classes" window

Figure 6. Selection of the class set to be used for the visualization of the colourgrams.

Figure 7. Colourgrams GUI main window, where the regions related to red, blue and hue are highlighted in light blue colour (a); frequency distribution curves of the selected parameters coloured according harvest time (b) and to total anthocyanins content (c).

Figure 8. PCA window

Figure 9. PCA window with selected objects

Figure 10. Window for the visualization of PCA loadings.

Figure 11. Definition of the model parameters using PLS window

Figure 12. PLS window in prediction mode.

Figure 13. Colourgrams GUI main window in image reconstruction mode (a) and reconstruction of the same image considering two colourgram intervals corresponding to portions of the relative green (b) and relative blue (c) frequency distribution curves.

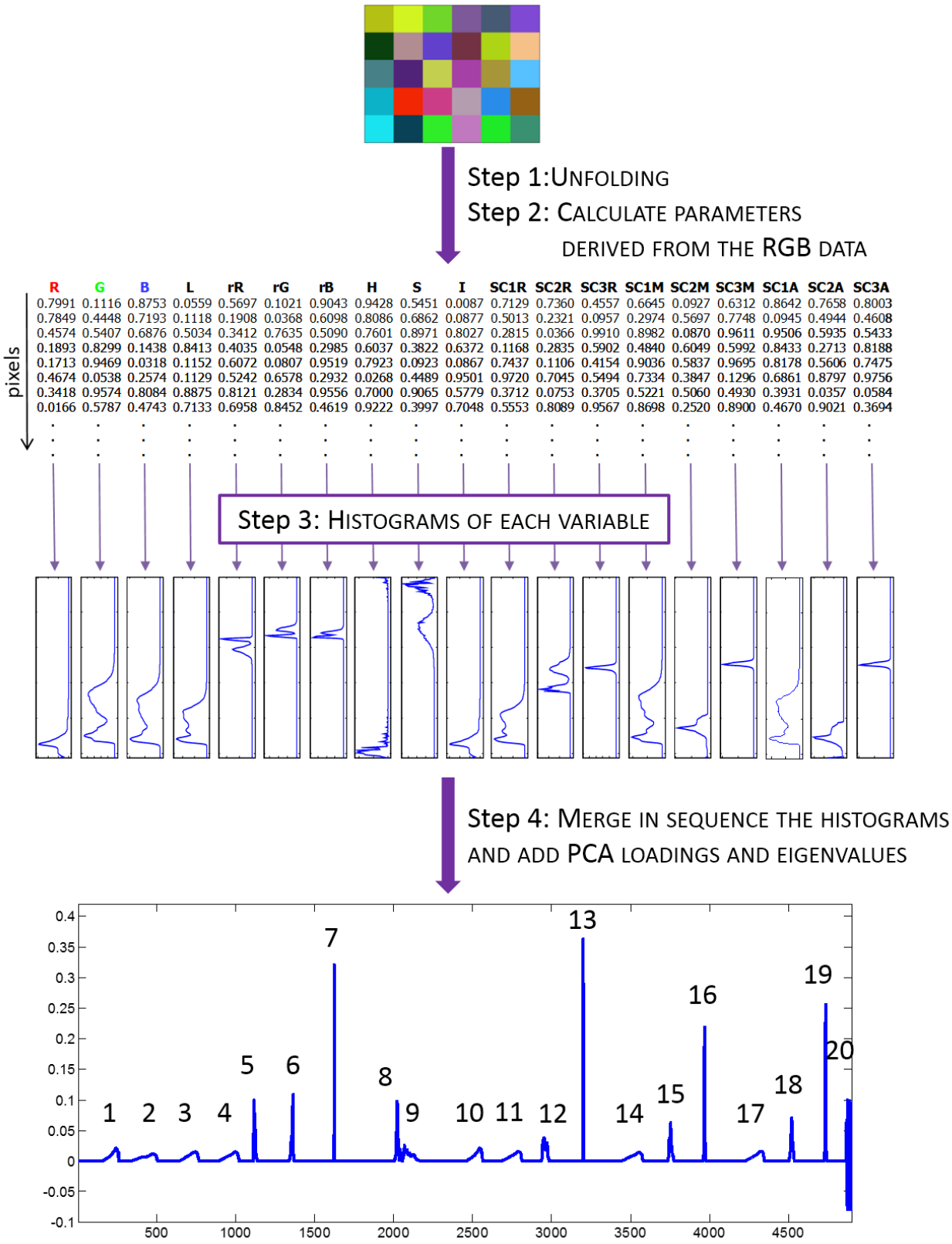
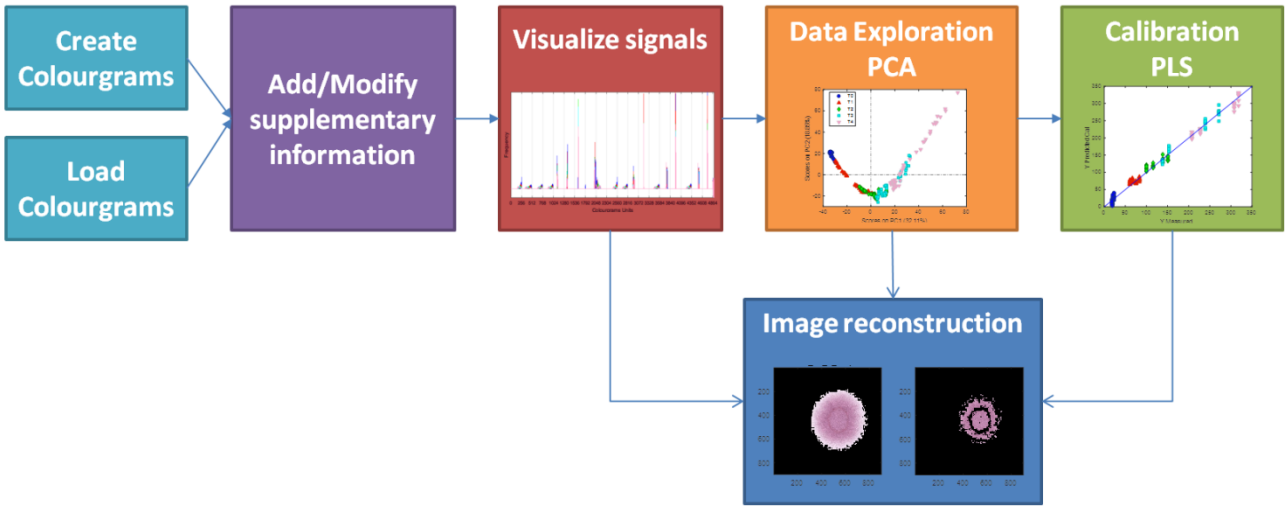

Step 1:UNFOLDING
Step 2: CALCULATE PARAMETERS
DERIVED FROM THE RGB DATA
R G B L rR rG rB H S I SC1R SC2R SC3R SC1M SC2M SC3M SC1A SC2A SC3A
0.7991 0.1116 0.8753 0.0559 0.5697 0.1021 0.9043 0.9428 0.5451 0.0087 0.7129 0.7360 0.4557 0.6645 0.0927 0.6312 0.8642 0.7658 0.8003
0.7849 0.4448 0.7193 0.1118 0.1908 0.0368 0.6098 0.8086 0.6862 0.0877 0.5013 0.2321 0.0957 0.2974 0.5697 0.7748 0.0945 0.4944 0.4608
0.4574 0.5407 0.6876 0.5034 0.3412 0.7635 0.5090 0.7601 0.8971 0.8027 0.2815 0.0366 0.9910 0.8982 0.0870 0.9611 0.9506 0.5935 0.5433
0.1893 0.8299 0.1438 0.8413 0.4035 0.0548 0.2985 0.6037 0.3822 0.6372 0.1168 0.2835 0.5902 0.4840 0.6049 0.5992 0.8433 0.2713 0.8188
0.1713 0.9469 0.0318 0.1152 0.6072 0.0807 0.9519 0.7923 0.0923 0.0867 0.7437 0.1106 0.4154 0.9036 0.5837 0.9695 0.8178 0.5606 0.7475
0.4674 0.0538 0.2574 0.1129 0.5242 0.6578 0.2932 0.0268 0.4489 0.9501 0.9720 0.7045 0.5494 0.7334 0.3847 0.1296 0.6861 0.8797 0.9756
0.3418 0.9574 0.8084 0.8875 0.8121 0.2834 0.9556 0.7000 0.9065 0.5779 0.3712 0.0753 0.3705 0.5221 0.5060 0.4930 0.3931 0.0357 0.0584
0.0166 0.5787 0.4743 0.7133 0.6958 0.8452 0.4619 0.9222 0.3997 0.7048 0.5553 0.8089 0.9567 0.8698 0.2520 0.8900 0.4670 0.9021 0.3694
pixels
Step 3: HISTOGRAMS OF EACH VARIABLE
Step 4: MERGE IN SEQUENCE THE HISTOGRAMS
AND ADD PCA LOADINGS AND EIGENVALUES
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
500
1000
1500
2000
2500
3000
3500
4000
4500
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Figure 1



561

562 **Figure 2**

563

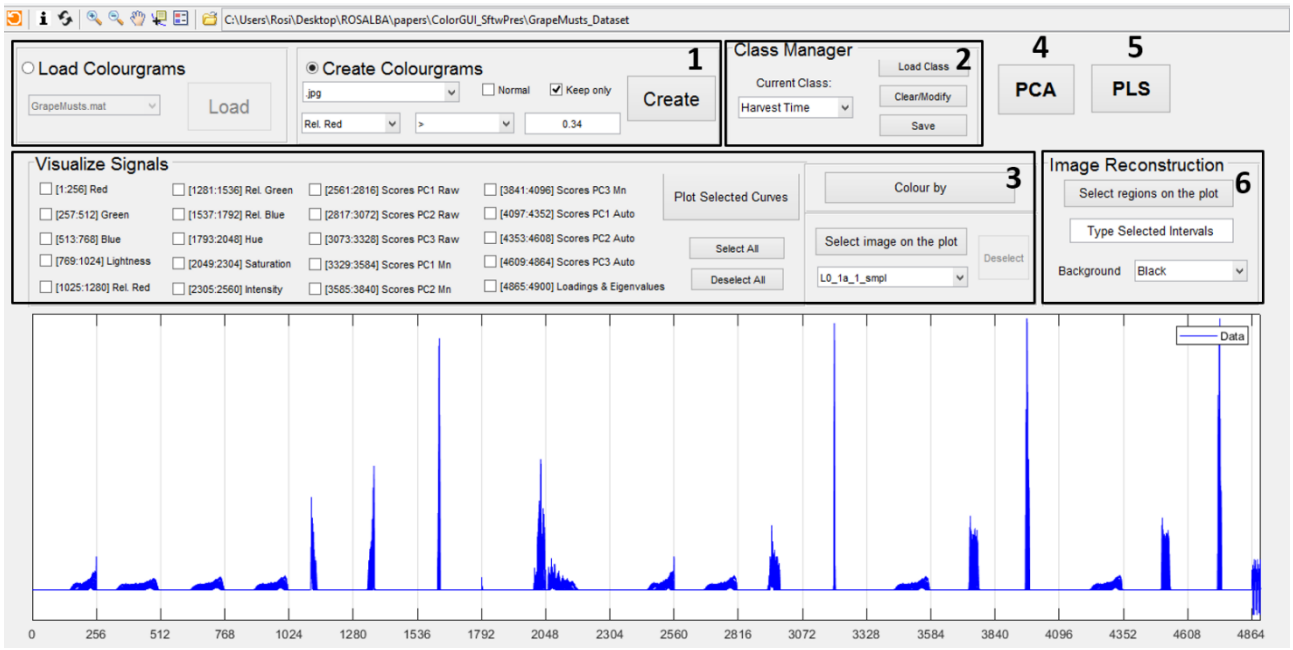

C:\Users\Rosi\Desktop\ROSALBA\papers\ColorGUI_SftwPres\GrapeMusts_Dataset
Load Colourgrams
GrapeMusts.mat
Load
Create Colourgrams
jpg
Normal
Keep only
Rel. Red
>
0.34
Create
1
Class Manager
Current Class:
Harvest Time
Load Class
Clear/Modify
Save
2
4
PCA
5
PLS
Visualize Signals
[1:256] Red
[257:512] Green
[513:768] Blue
[769:1024] Lightness
[1025:1280] Rel. Red
[1281:1536] Rel. Green
[1537:1792] Rel. Blue
[1793:2048] Hue
[2049:2304] Saturation
[2305:2560] Intensity
[2561:2816] Scores PC1 Raw
[2817:3072] Scores PC2 Raw
[3073:3328] Scores PC3 Raw
[3329:3584] Scores PC1 Mn
[3585:3840] Scores PC2 Mn
[3841:4096] Scores PC3 Mn
[4097:4352] Scores PC1 Auto
[4353:4608] Scores PC2 Auto
[4609:4864] Scores PC3 Auto
[4865:4900] Loadings & Eigenvalues
Plot Selected Curves
Select All
Deselect All
Colour by
3
Select image on the plot
L0_1a_1_smpl
Deselect
Image Reconstruction
6
Select regions on the plot
Type Selected Intervals
Background
Black
Data
0
256
512
768
1024
1280
1536
1792
2048
2304
2560
2816
3072
3328
3584
3840
4096
4352
4608
4864

Figure 3

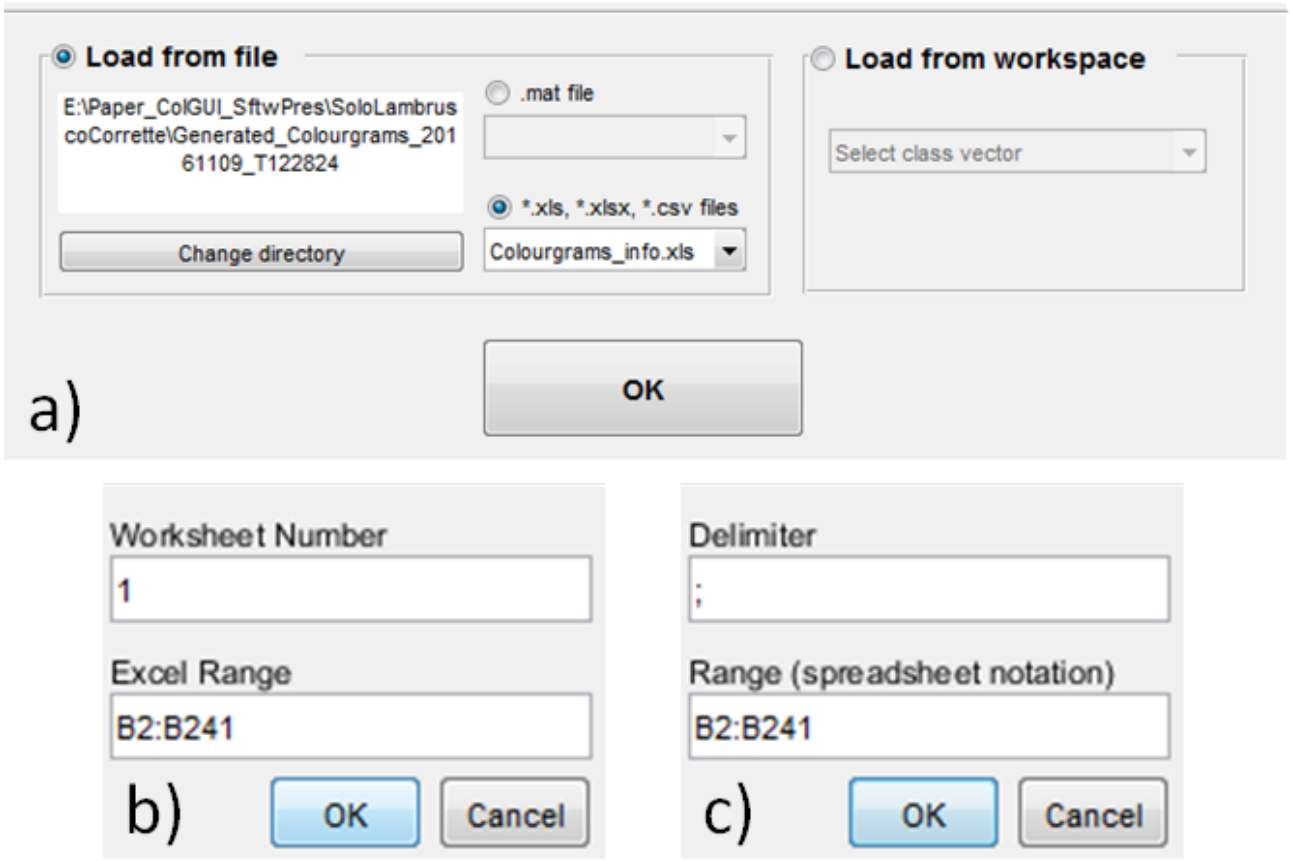
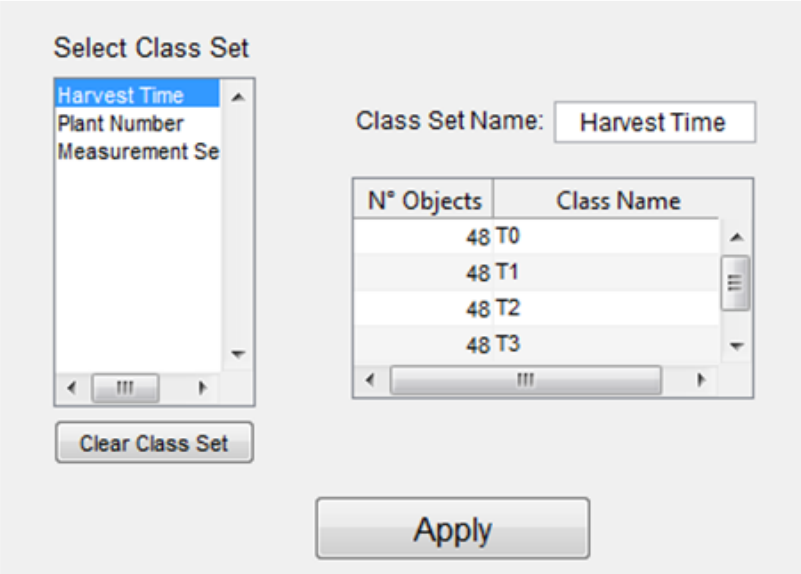

Load from file
E:\Paper_ColGUI_SftwPres\SoloLambruscoCorrette\Generated_Colourgrams_20161109_T122824
Change directory
.mat file
*.xls, *.xlsx, *.csv files
Colourgrams_info.xls
Load from workspace
Select class vector
a)
OK
Worksheet Number
1
Excel Range
B2:B241
b)
OK
Cancel
Delimiter
;
Range (spreadsheet notation)
B2:B241
c)
OK
Cancel

Figure 4



570

571 **Figure 5**

572

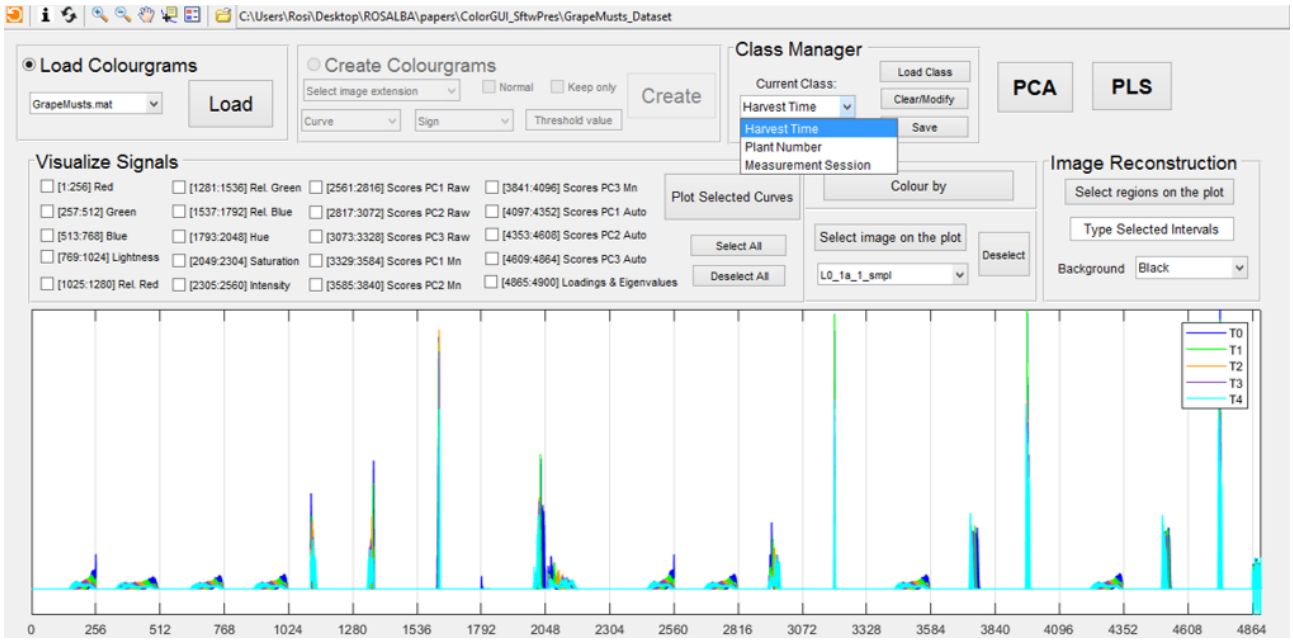

C:\Users\Ros\Desktop\ROSALBA\papers\ColorGUI_SftwPres\GrapeMusts_Dataset
Load Colourgrams
GrapeMusts.mat
Load
Create Colourgrams
Select image extension
Normal
Keep only
Create
Curve
Sign
Threshold value
Class Manager
Current Class:
Harvest Time
Plant Number
Measurement Session
Load Class
Clear/Modify
Save
PCA
PLS
Visualize Signals
[1:256] Red
[257:512] Green
[513:768] Blue
[769:1024] Lightness
[1025:1280] Rel. Red
[1281:1536] Rel. Green
[1537:1792] Rel. Blue
[1793:2048] Hue
[2049:2304] Saturation
[2305:2560] Intensity
[2561:2816] Scores PC1 Raw
[2817:3072] Scores PC2 Raw
[3073:3328] Scores PC3 Raw
[3329:3584] Scores PC1 Mn
[3585:3840] Scores PC2 Mn
[3841:4096] Scores PC3 Mn
[4097:4352] Scores PC1 Auto
[4353:4608] Scores PC2 Auto
[4609:4864] Scores PC3 Auto
[4865:4900] Loadings & Eigenvalues
Plot Selected Curves
Select All
Deselect All
Colour by
Select image on the plot
L0_1a_1_smpl
Deselect
Image Reconstruction
Select regions on the plot
Type Selected Intervals
Background
Black
T0
T1
T2
T3
T4
0
256
512
768
1024
1280
1536
1792
2048
2304
2560
2816
3072
3328
3584
3840
4096
4352
4608
4864

Figure 6

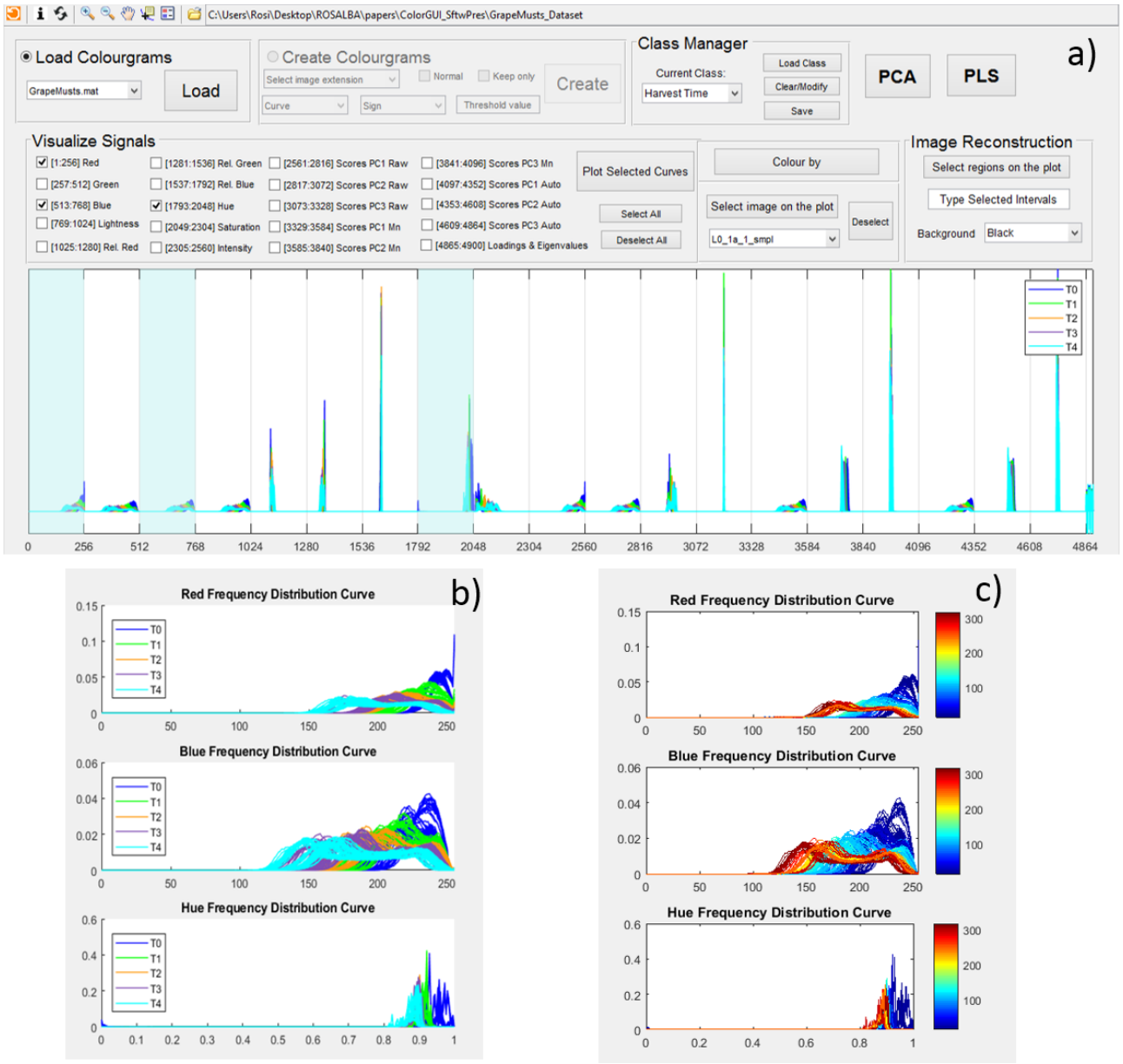

a)
Load Colourgrams
GrapeMusts.mat
Load
Create Colourgrams
Select image extension
Normal
Keep only
Curve
Sign
Threshold value
Create
Class Manager
Current Class:
Harvest Time
Load Class
Clear/Modify
Save
PCA
PLS
Visualize Signals
[1:256] Red
[257:512] Green
[513:768] Blue
[769:1024] Lightness
[1025:1280] Rel. Red
[1281:1536] Rel. Green
[1537:1792] Rel. Blue
[1793:2048] Hue
[2049:2304] Saturation
[2305:2560] Intensity
[2561:2816] Scores PC1 Raw
[2817:3072] Scores PC2 Raw
[3073:3328] Scores PC3 Raw
[3329:3584] Scores PC1 Mn
[3585:3840] Scores PC2 Mn
[3841:4096] Scores PC3 Mn
[4097:4352] Scores PC1 Auto
[4353:4608] Scores PC2 Auto
[4609:4864] Scores PC3 Auto
[4865:4900] Loadings & Eigenvalues
Plot Selected Curves
Select All
Deselect All
Colour by
Select image on the plot
L0_1a_1_smpl
Deselect
Image Reconstruction
Select regions on the plot
Type Selected Intervals
Background
Black
T0
T1
T2
T3
T4
b)
Red Frequency Distribution Curve
Blue Frequency Distribution Curve
Hue Frequency Distribution Curve
c)
Red Frequency Distribution Curve
Blue Frequency Distribution Curve
Hue Frequency Distribution Curve

Figure 7

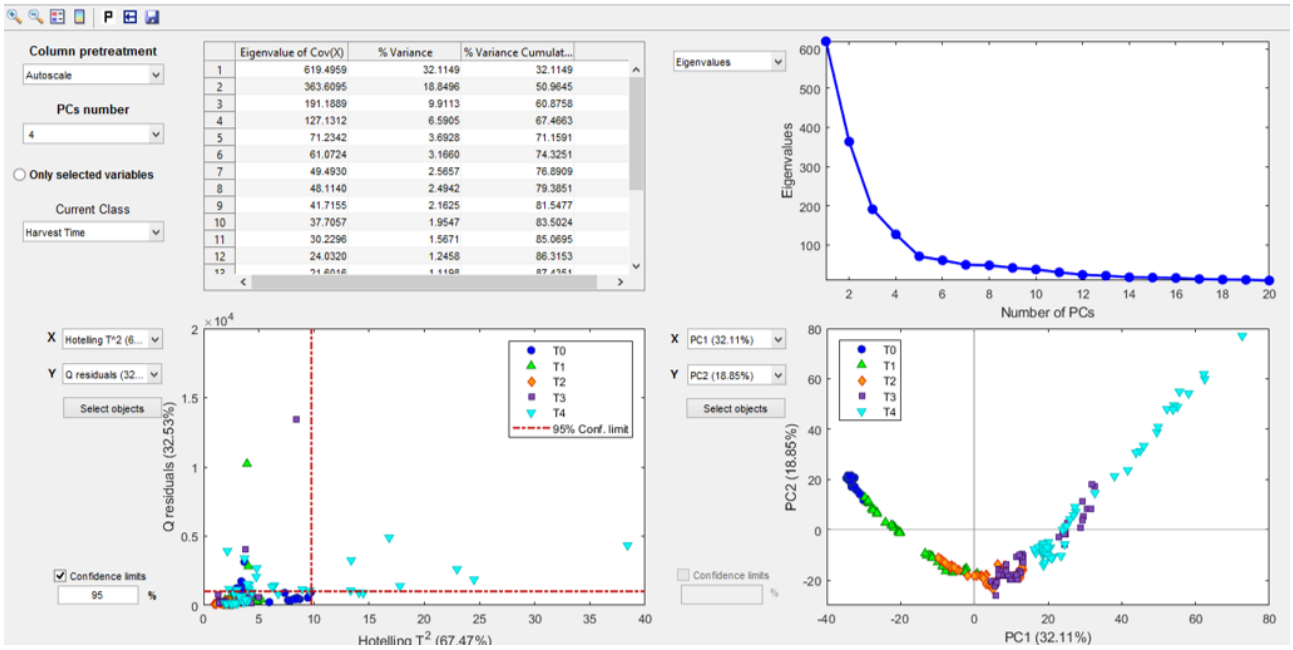

Column pretreatment
Autoscale
PCs number
4
Only selected variables
Current Class
Harvest Time
Eigenvalue of Cov(X) % Variance % Variance Cumulat...
1 619.4959 32.1149 32.1149
2 363.6095 18.8496 50.9645
3 191.1889 9.9113 60.8758
4 127.1312 6.5905 67.4663
5 71.2342 3.6928 71.1591
6 61.0724 3.1660 74.3251
7 49.4930 2.5657 76.8909
8 48.1140 2.4942 79.3851
9 41.7155 2.1625 81.5477
10 37.7057 1.9547 83.5024
11 30.2296 1.5671 85.0695
12 24.0320 1.2458 86.3153
Eigenvalues
Eigenvalues
Number of PCs
X Hotelling T^2 (6...
Y Q residuals (32...
Select objects
Confidence limits
95 %
Q residuals (32.53%)
Hotelling T² (67.47%)
T0
T1
T2
T3
T4
95% Conf. limit
X PC1 (32.11%)
Y PC2 (18.85%)
Select objects
Confidence limits
PC2 (18.85%)
PC1 (32.11%)

Figure 8

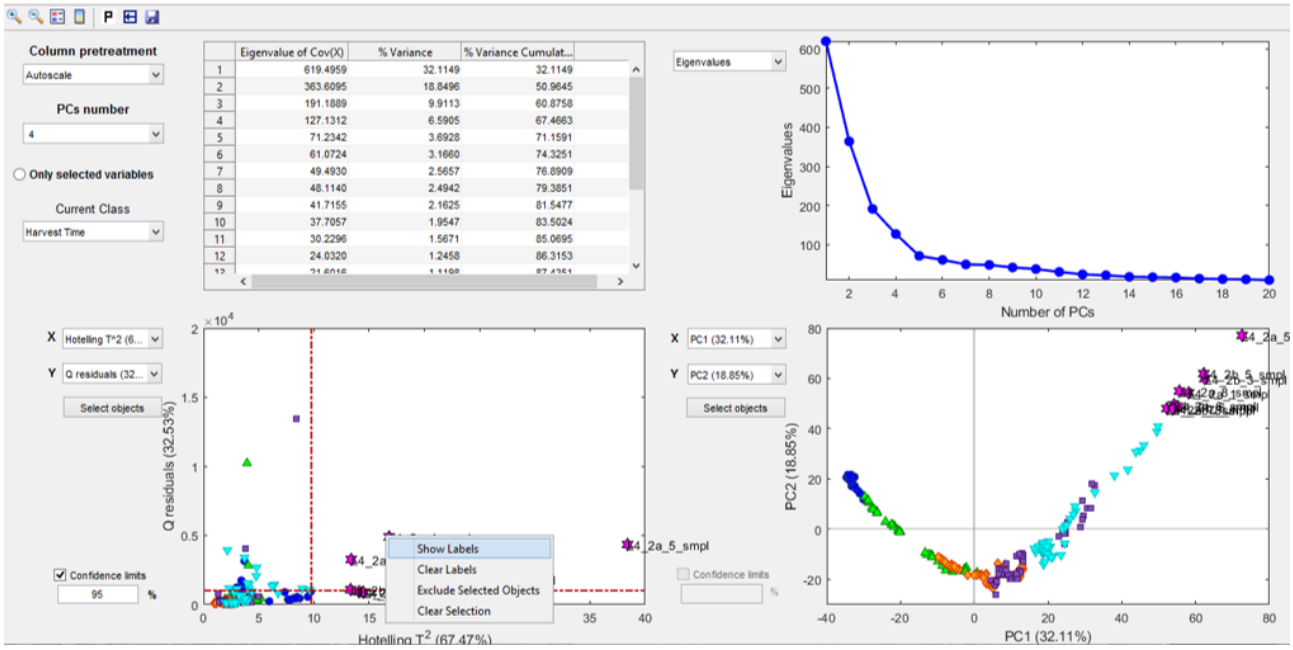
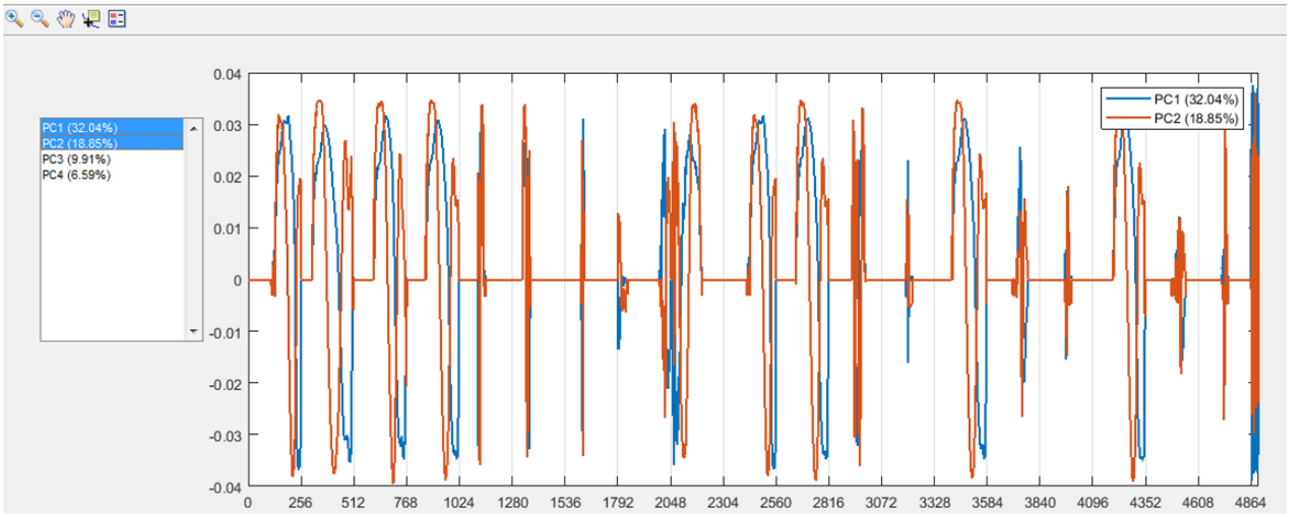

Column pretreatment
Autoscale
PCs number
4
Only selected variables
Current Class
Harvest Time
Eigenvalue of Cov(X) % Variance % Variance Cumulat...
1 619.4959 32.1149 32.1149
2 363.6095 18.8496 50.9645
3 191.1889 9.9113 60.8758
4 127.1312 6.5905 67.4663
5 71.2342 3.6928 71.1591
6 61.0724 3.1660 74.3251
7 49.4930 2.5657 76.8909
8 48.1140 2.4942 79.3851
9 41.7155 2.1625 81.5477
10 37.7057 1.9547 83.5024
11 30.2296 1.5671 85.0695
12 24.0320 1.2458 86.3153
Eigenvalues
Eigenvalues
Number of PCs
X Hotelling T^2 (6...
Y Q residuals (32...
Select objects
Q residuals (32.53%)
Hotelling T^2 (67.47%)
Show Labels
Clear Labels
Exclude Selected Objects
Clear Selection
4_2a_5_smpl
Confidence limits
95 %
X PC1 (32.11%)
Y PC2 (18.85%)
Select objects
PC2 (18.85%)
PC1 (32.11%)
Confidence limits

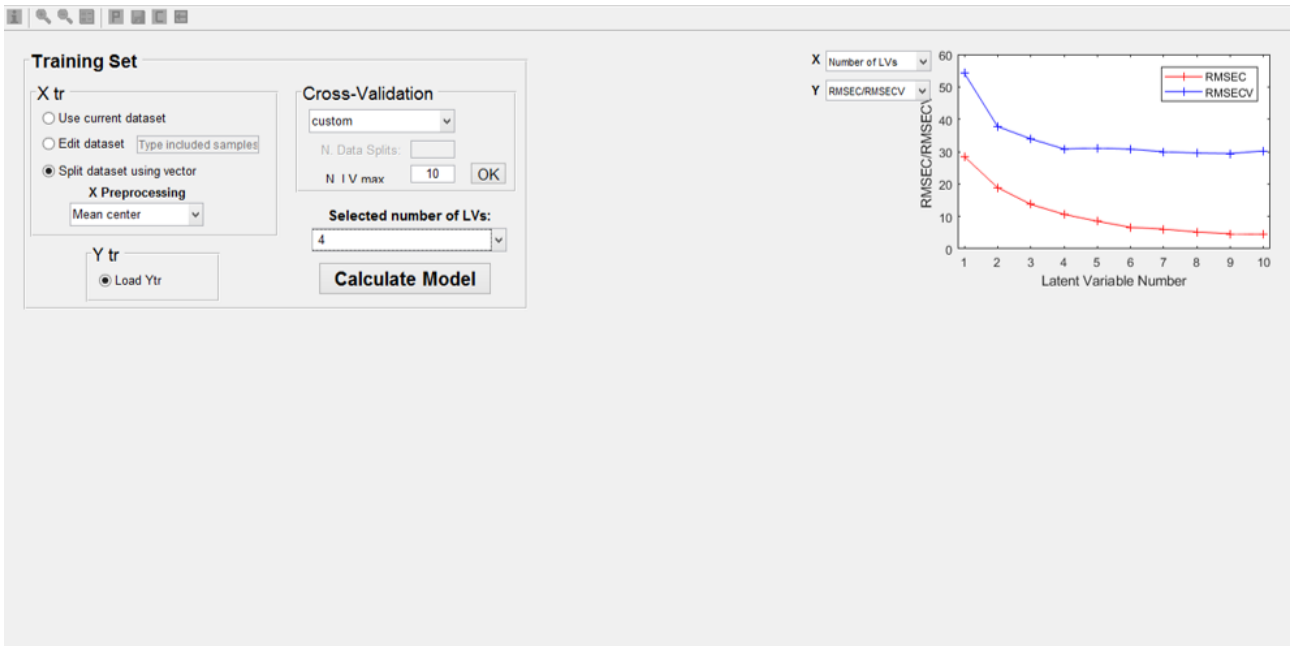
Figure 9



585

586 **Figure 10**

587



588

589 **Figure 11**

590

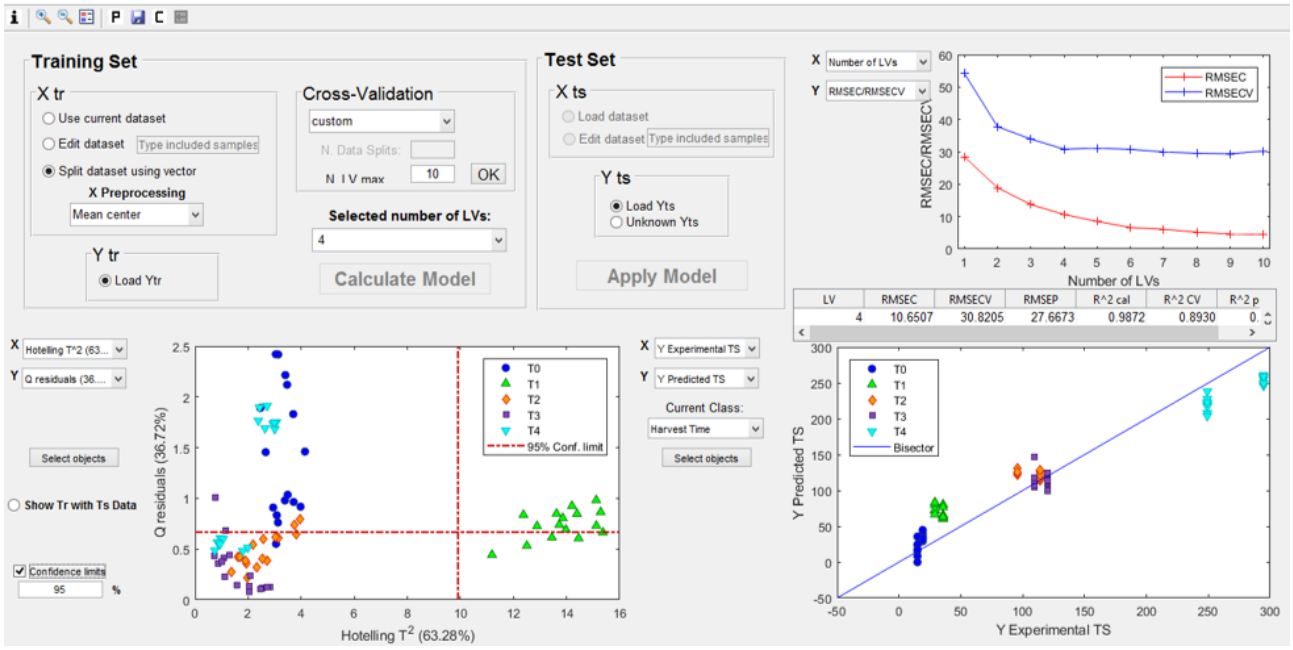

Training Set
X tr
Use current dataset
Edit dataset
Type included samples
Split dataset using vector
X Preprocessing
Mean center
Y tr
Load Ytr
Cross-Validation
custom
N. Data Splits:
N LV max
10
OK
Selected number of LVs:
4
Calculate Model
Test Set
X ts
Load dataset
Edit dataset
Type included samples
Y ts
Load Yts
Unknown Yts
Apply Model
X Number of LVs
Y RMSEC/RMSECV
RMSEC
RMSECV
Number of LVs
LV RMSEC RMSECV RMSEP R^2 cal R^2 CV R^2 p
4 10.6507 30.8205 27.6673 0.9872 0.8930 0.
X Hotelling T^2 (63...
Y Q residuals (36....
Select objects
Show Tr with Ts Data
Confidence limits
95
%
Q residuals (36.72%)
Hotelling T2 (63.28%)
T0
T1
T2
T3
T4
95% Conf. limit
X Y Experimental TS
Y Y Predicted TS
Current Class:
Harvest Time
Select objects
Y Predicted TS
Y Experimental TS
Bisector

Figure 12

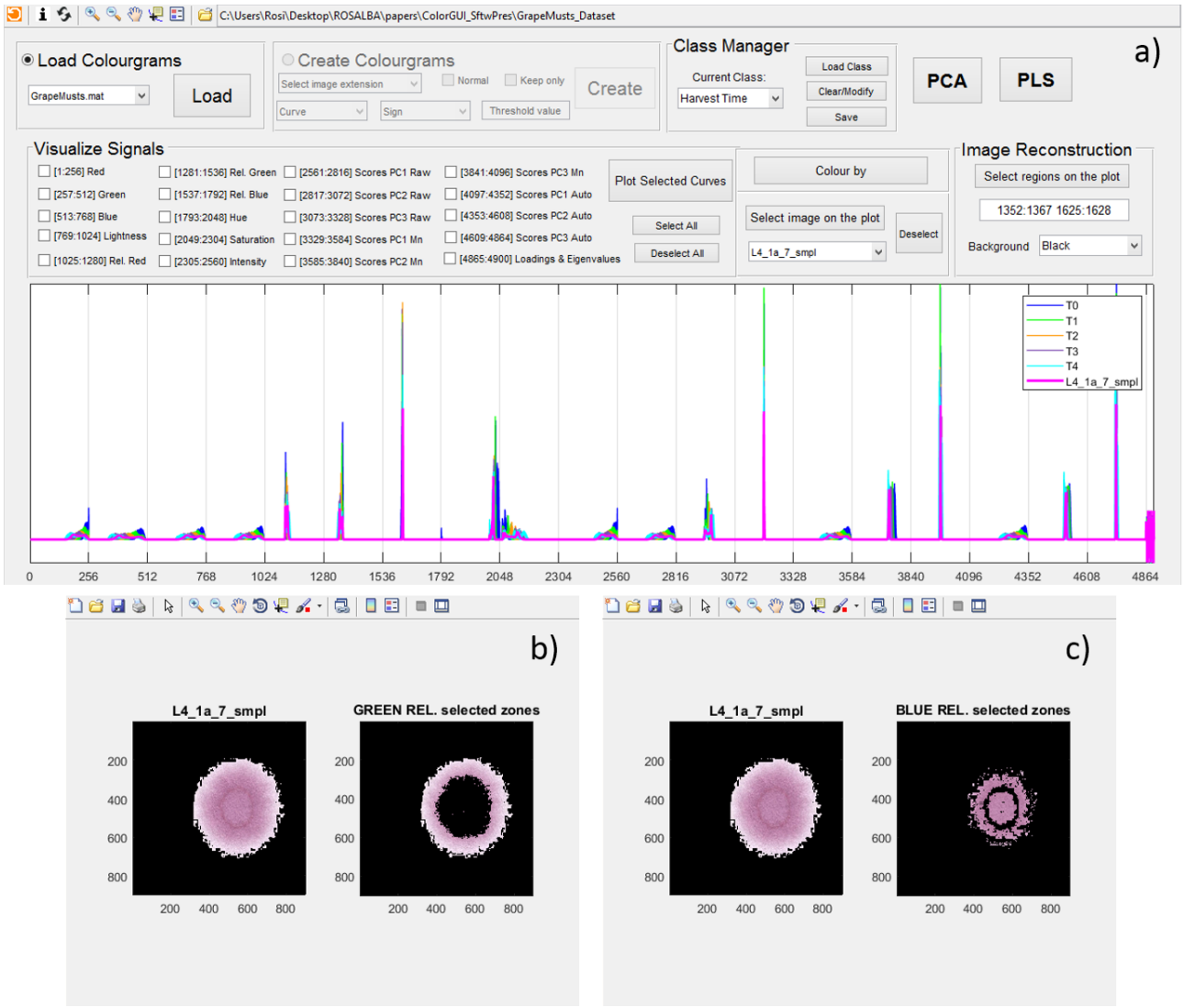

a)
Load Colourgrams
GrapeMusts.mat
Load
Create Colourgrams
Select image extension
Normal
Keep only
Create
Curve
Sign
Threshold value
Class Manager
Current Class:
Harvest Time
Load Class
Clear/Modify
Save
PCA
PLS
Visualize Signals
[1:256] Red
[257:512] Green
[513:768] Blue
[769:1024] Lightness
[1025:1280] Rel. Red
[1281:1536] Rel. Green
[1537:1792] Rel. Blue
[1793:2048] Hue
[2049:2304] Saturation
[2305:2560] Intensity
[2561:2816] Scores PC1 Raw
[2817:3072] Scores PC2 Raw
[3073:3328] Scores PC3 Raw
[3329:3584] Scores PC1 Mn
[3585:3840] Scores PC2 Mn
[3841:4096] Scores PC3 Mn
[4097:4352] Scores PC1 Auto
[4353:4608] Scores PC2 Auto
[4609:4864] Scores PC3 Auto
[4865:4900] Loadings & Eigenvalues
Plot Selected Curves
Select All
Deselect All
Colour by
Select image on the plot
L4_1a_7_smpl
Deselect
Image Reconstruction
Select regions on the plot
1352:1367 1625:1628
Background
Black
T0
T1
T2
T3
T4
L4_1a_7_smpl
0 256 512 768 1024 1280 1536 1792 2048 2304 2560 2816 3072 3328 3584 3840 4096 4352 4608 4864
b)
L4_1a_7_smpl
GREEN REL. selected zones
200 400 600 800
c)
L4_1a_7_smpl
BLUE REL. selected zones
200 400 600 800

Figure 13

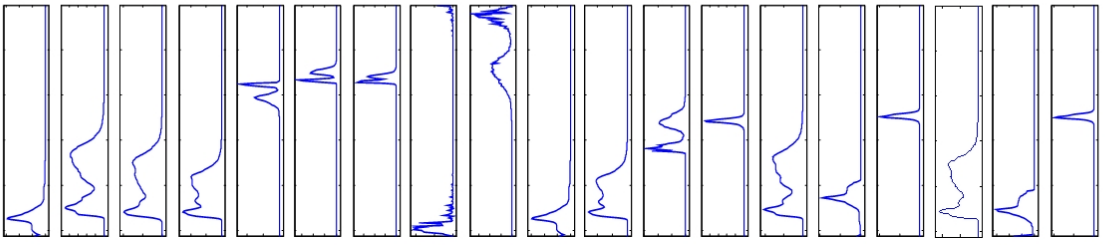
Step 1:UNFOLDING

Step 2: CALCULATE PARAMETERS DERIVED FROM THE RGB DATA

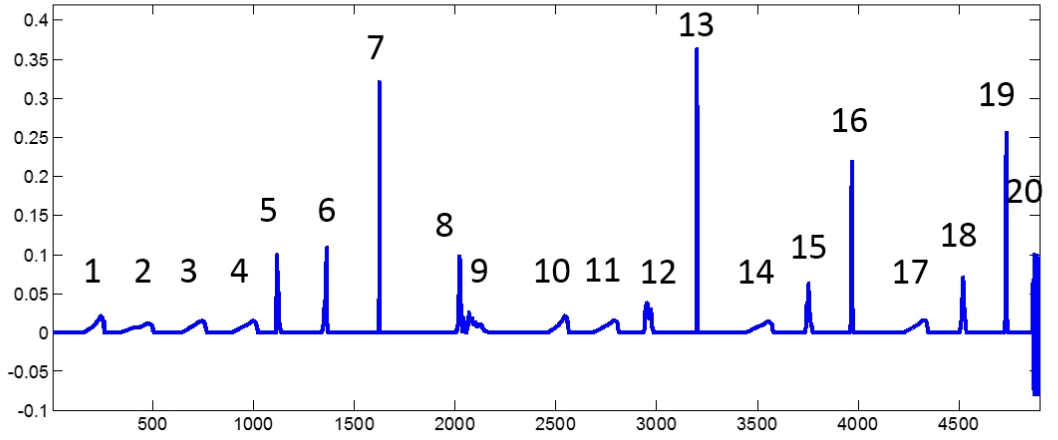
pixels

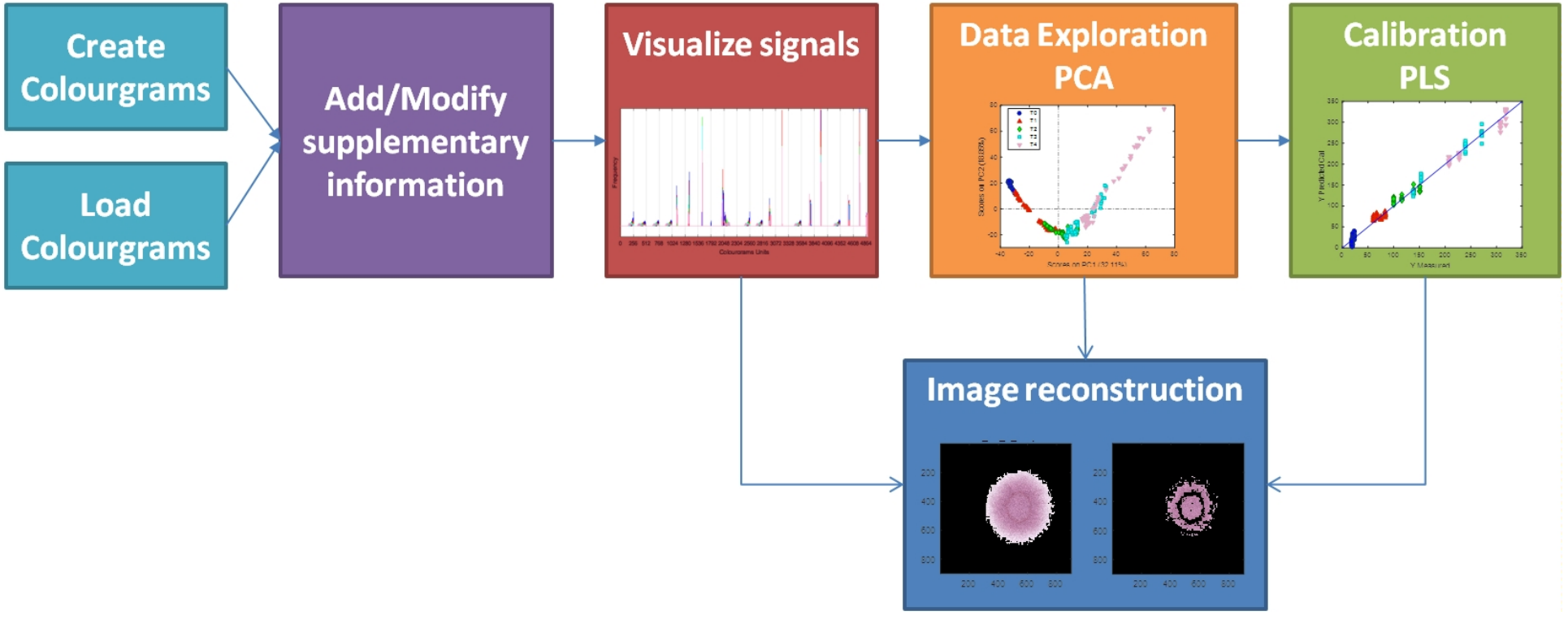
R	G	B	L	rR	rG	rB	H	S	I	SC1R	SC2R	SC3R	SC1M	SC2M	SC3M	SC1A	SC2A	SC3A
0.7991	0.1116	0.8753	0.0559	0.5697	0.1021	0.9043	0.9428	0.5451	0.0087	0.7129	0.7360	0.4557	0.6645	0.0927	0.6312	0.8642	0.7658	0.8003
0.7849	0.4448	0.7193	0.1118	0.1908	0.0368	0.6098	0.8086	0.6862	0.0877	0.5013	0.2321	0.0957	0.2974	0.5697	0.7748	0.0945	0.4944	0.4608
0.4574	0.5407	0.6876	0.5034	0.3412	0.7635	0.5090	0.7601	0.8971	0.8027	0.2815	0.0366	0.9910	0.8982	0.0870	0.9611	0.9506	0.5935	0.5433
0.1893	0.8299	0.1438	0.8413	0.4035	0.0548	0.2985	0.6037	0.3822	0.6372	0.1168	0.2835	0.5902	0.4840	0.6049	0.5992	0.8433	0.2713	0.8188
0.1713	0.9469	0.0318	0.1152	0.6072	0.0807	0.9519	0.7923	0.0923	0.0867	0.7437	0.1106	0.4154	0.9036	0.5837	0.9695	0.8178	0.5606	0.7475
0.4674	0.0538	0.2574	0.1129	0.5242	0.6578	0.2932	0.0268	0.4489	0.9501	0.9720	0.7045	0.5494	0.7334	0.3847	0.1296	0.6861	0.8797	0.9756
0.3418	0.9574	0.8084	0.8875	0.8121	0.2834	0.9556	0.7000	0.9065	0.5779	0.3712	0.0753	0.3705	0.5221	0.5060	0.4930	0.3931	0.0357	0.0584
0.0166	0.5787	0.4743	0.7133	0.6958	0.8452	0.4619	0.9222	0.3997	0.7048	0.5553	0.8089	0.9567	0.8698	0.2520	0.8900	0.4670	0.9021	0.3694

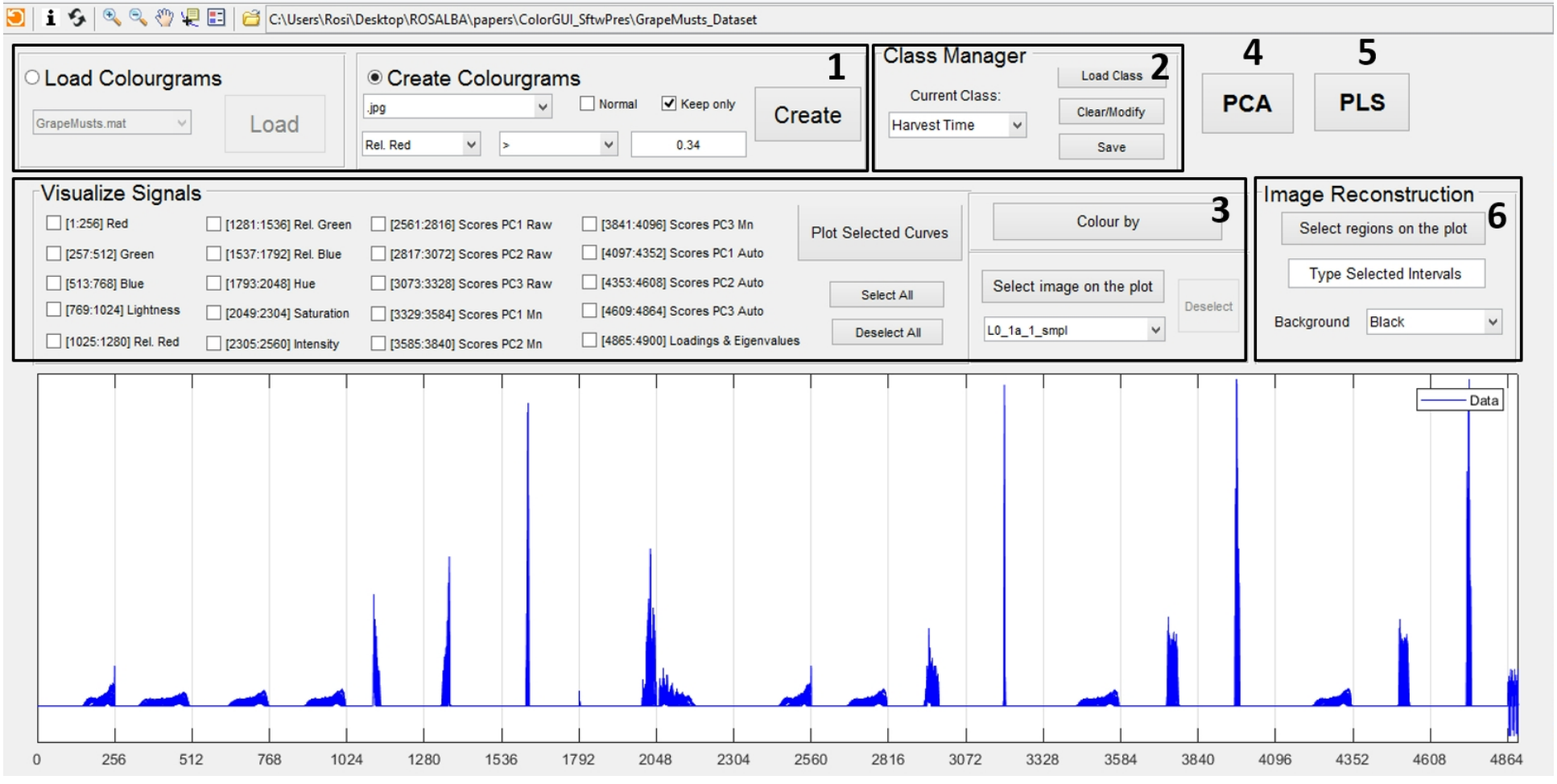
Step 3: HISTOGRAMS OF EACH VARIABLE

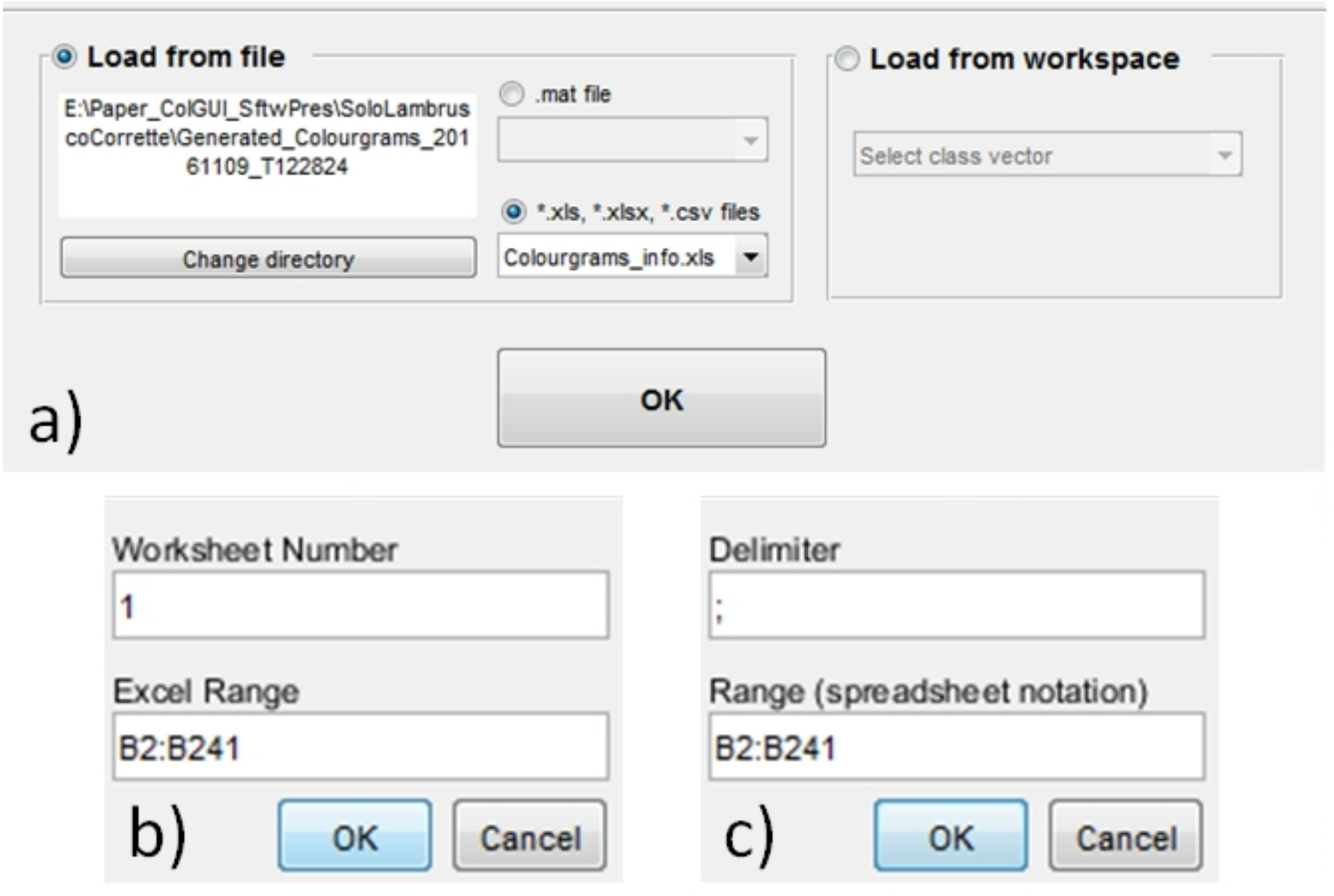


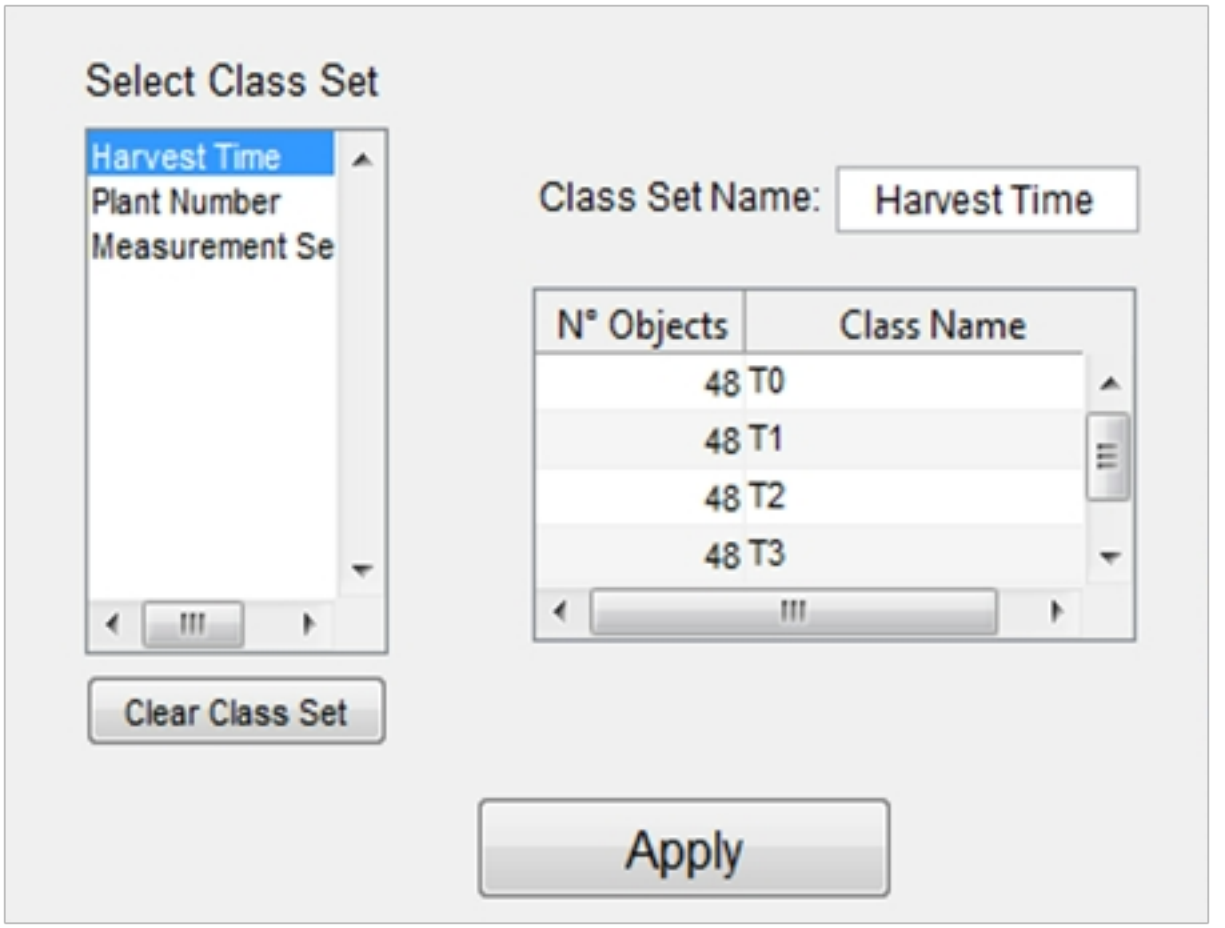
Step 4: MERGE IN SEQUENCE THE HISTOGRAMS AND ADD PCA LOADINGS AND EIGENVALUES

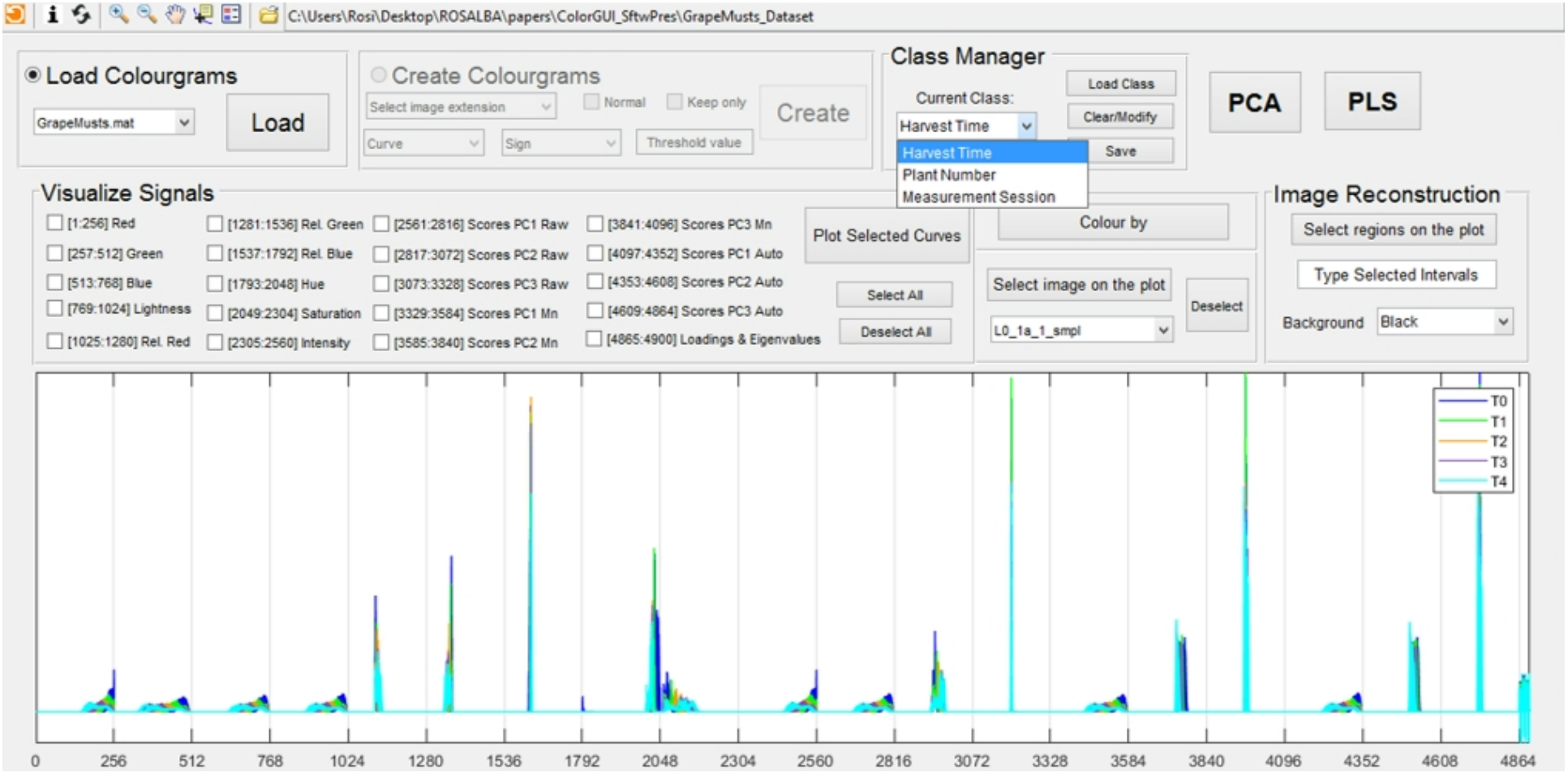


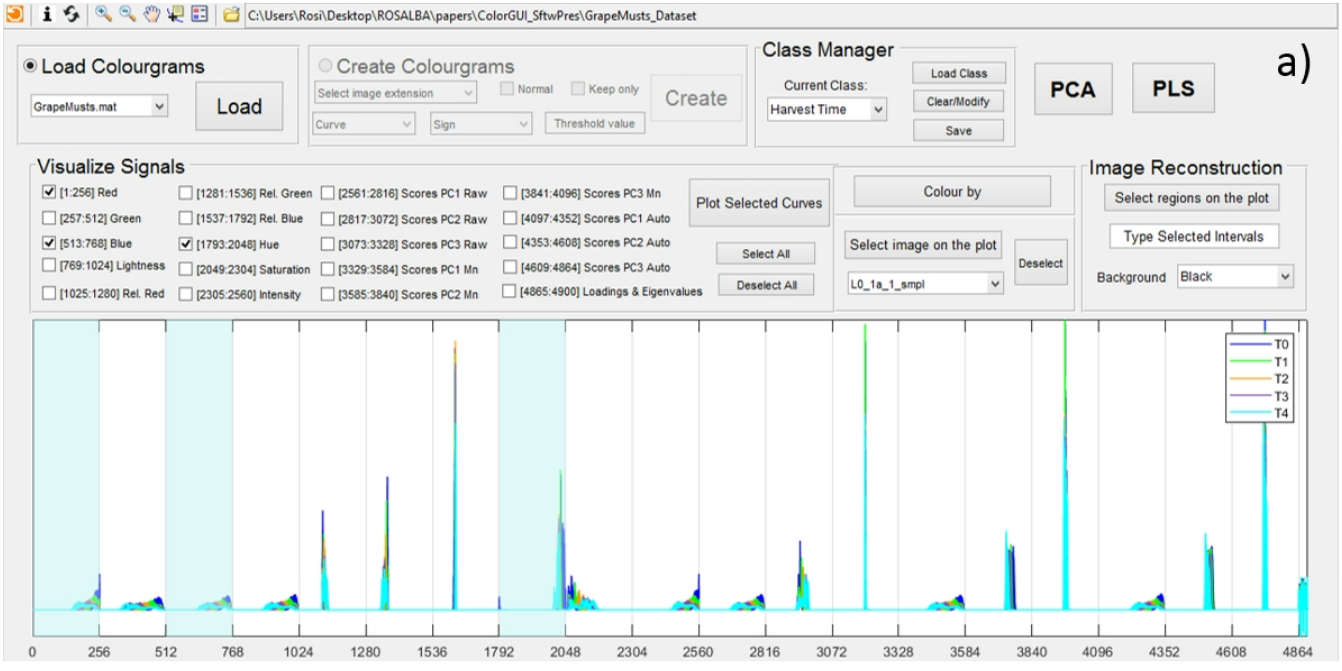

Create Colourgrams
Load Colourgrams
Add/Modify supplementary information
Visualize signals
Data Exploration PCA
Calibration PLS
Image reconstruction

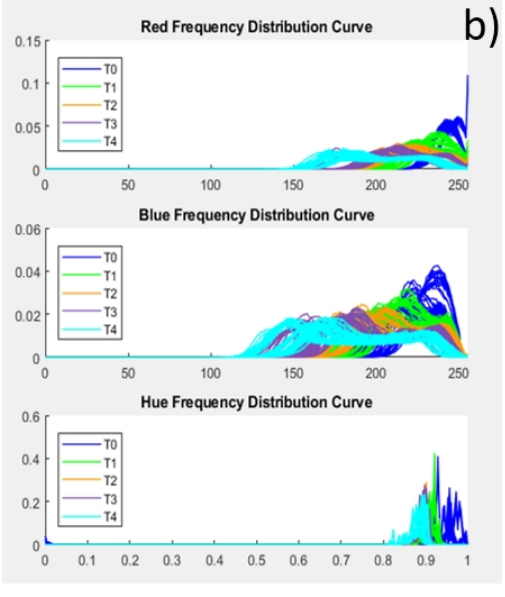

C:\Users\Rosi\Desktop\ROSALBA\papers\ColorGUI_SftwPres\GrapeMusts_Dataset
Load Colourgrams
GrapeMusts.mat
Load
Create Colourgrams
.jpg
Normal
Keep only
Rel. Red
>
0.34
Create
1
Class Manager
Current Class:
Harvest Time
Load Class
Clear/Modify
Save
2
4
PCA
5
PLS
Visualize Signals
[1:256] Red
[257:512] Green
[513:768] Blue
[769:1024] Lightness
[1025:1280] Rel. Red
[1281:1536] Rel. Green
[1537:1792] Rel. Blue
[1793:2048] Hue
[2049:2304] Saturation
[2305:2560] Intensity
[2561:2816] Scores PC1 Raw
[2817:3072] Scores PC2 Raw
[3073:3328] Scores PC3 Raw
[3329:3584] Scores PC1 Mn
[3585:3840] Scores PC2 Mn
[3841:4096] Scores PC3 Mn
[4097:4352] Scores PC1 Auto
[4353:4608] Scores PC2 Auto
[4609:4864] Scores PC3 Auto
[4865:4900] Loadings & Eigenvalues
Plot Selected Curves
Select All
Deselect All
Colour by
3
Select image on the plot
Deselect
L0_1a_1_smpl
Image Reconstruction
Select regions on the plot
6
Type Selected Intervals
Background
Black
Data
0
256
512
768
1024
1280
1536
1792
2048
2304
2560
2816
3072
3328
3584
3840
4096
4352
4608
4864

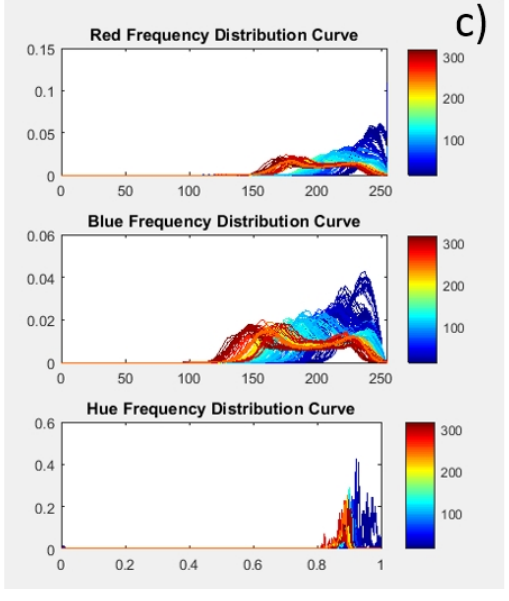

Load from file
E:\Paper_ColGUI_SftwPres\SoloLambruscoCorrette\Generated_Colourgrams_20161109_T122824
Change directory
.mat file
*.xls, *.xlsx, *.csv files
Colourgrams_info.xls
Load from workspace
Select class vector
a)
OK
Worksheet Number
1
Excel Range
B2:B241
b)
OK
Cancel
Delimiter
;
Range (spreadsheet notation)
B2:B241
c)
OK
Cancel

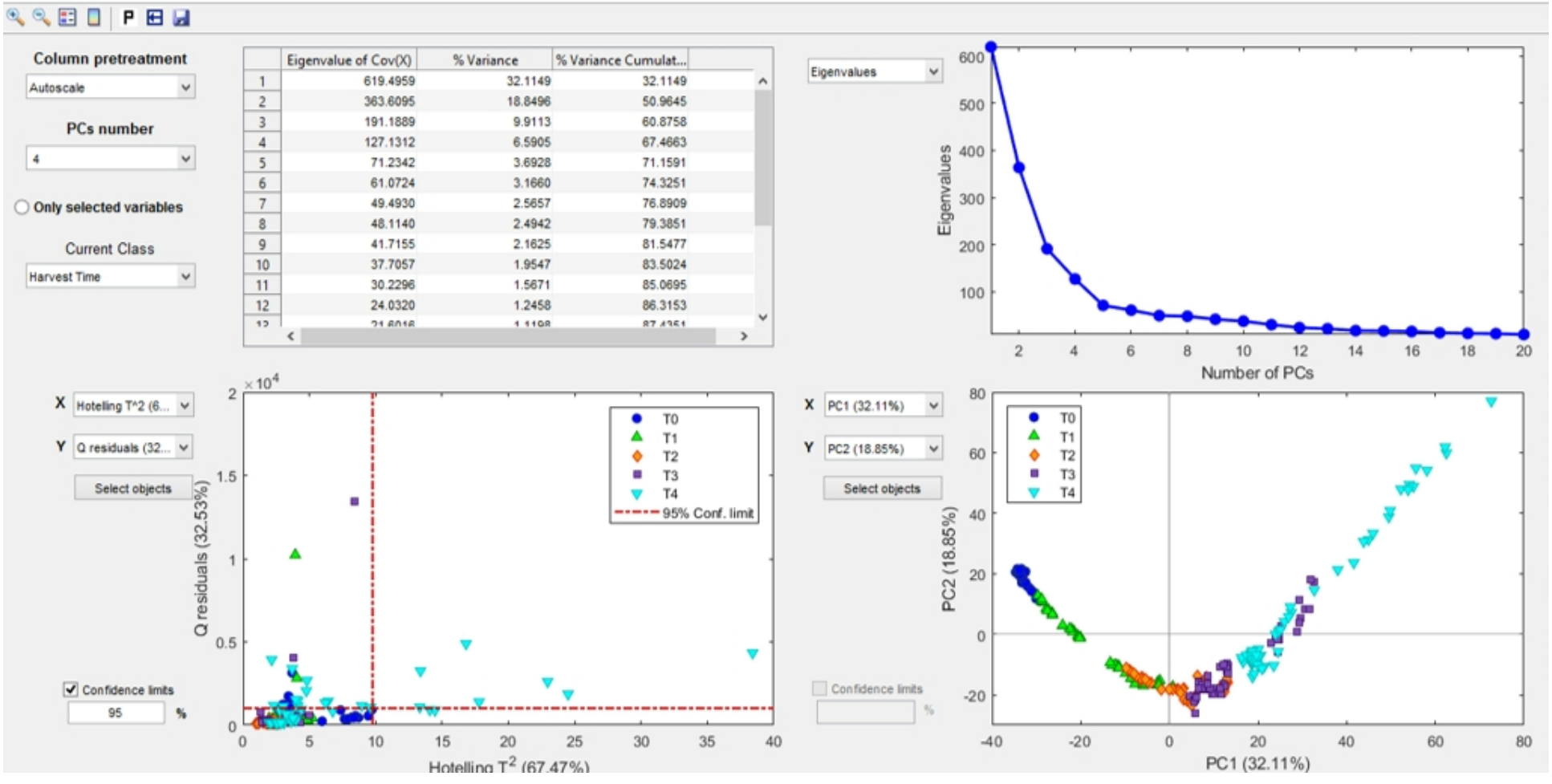

Select Class Set
Harvest Time
Plant Number
Measurement Se
Class Set Name:
Harvest Time
N° Objects
Class Name
48 T0
48 T1
48 T2
48 T3
Clear Class Set
Apply

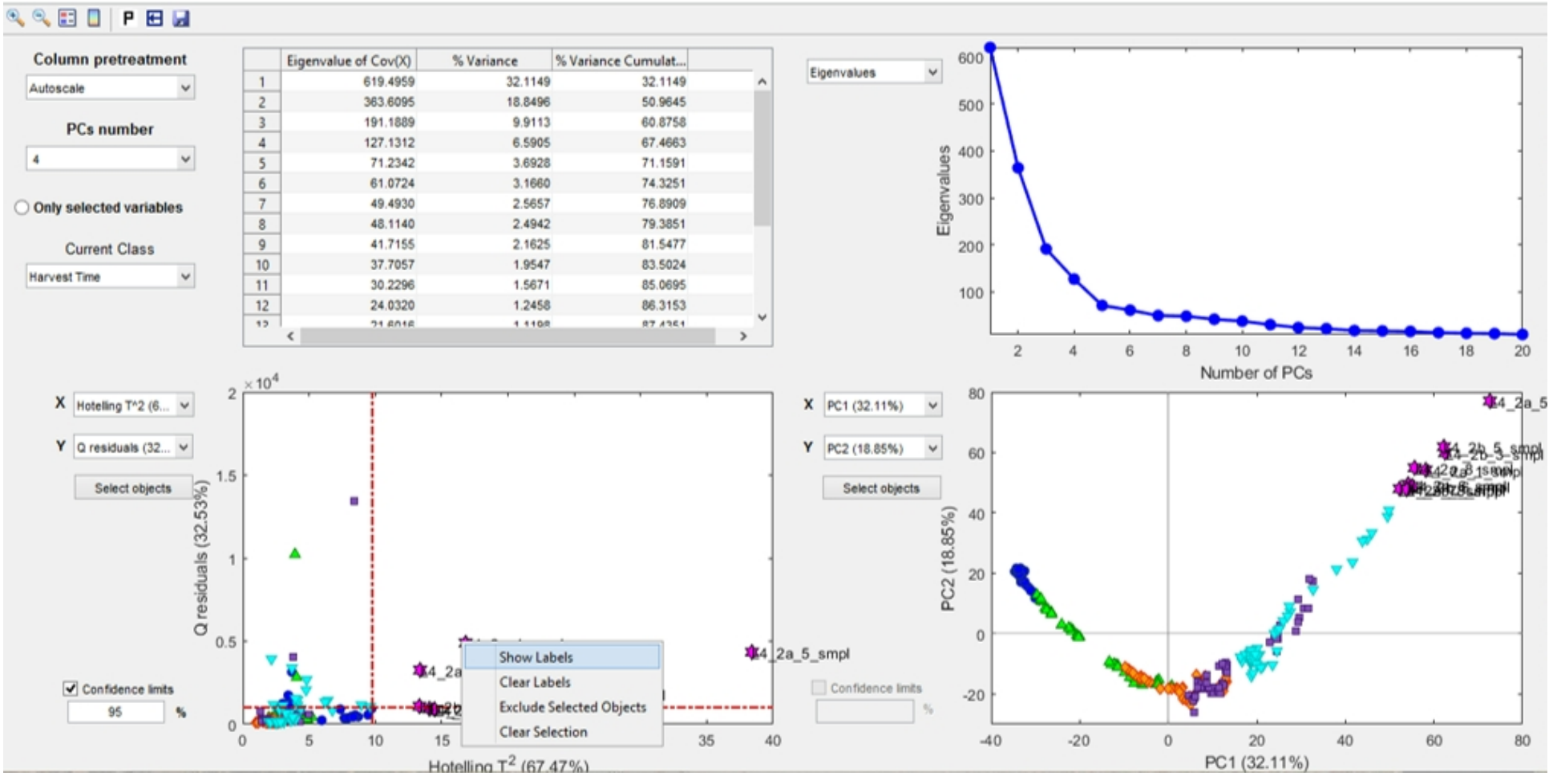

C:\Users\Rosi\Desktop\ROSALBA\papers\ColorGUI_SftwPres\GrapeMusts_Dataset
Load Colourgrams
GrapeMusts.mat
Load
Create Colourgrams
Select image extension
Normal
Keep only
Create
Curve
Sign
Threshold value
Class Manager
Current Class:
Harvest Time
Plant Number
Measurement Session
Load Class
Clear/Modify
Save
PCA
PLS
Visualize Signals
[1:256] Red
[257:512] Green
[513:768] Blue
[769:1024] Lightness
[1025:1280] Rel. Red
[1281:1536] Rel. Green
[1537:1792] Rel. Blue
[1793:2048] Hue
[2049:2304] Saturation
[2305:2560] Intensity
[2561:2816] Scores PC1 Raw
[2817:3072] Scores PC2 Raw
[3073:3328] Scores PC3 Raw
[3329:3584] Scores PC1 Mn
[3585:3840] Scores PC2 Mn
[3841:4096] Scores PC3 Mn
[4097:4352] Scores PC1 Auto
[4353:4608] Scores PC2 Auto
[4609:4864] Scores PC3 Auto
[4865:4900] Loadings & Eigenvalues
Plot Selected Curves
Select All
Deselect All
Colour by
Select image on the plot
L0_1a_1_smpl
Deselect
Image Reconstruction
Select regions on the plot
Type Selected Intervals
Background
Black
T0
T1
T2
T3
T4
0
256
512
768
1024
1280
1536
1792
2048
2304
2560
2816
3072
3328
3584
3840
4096
4352
4608
4864

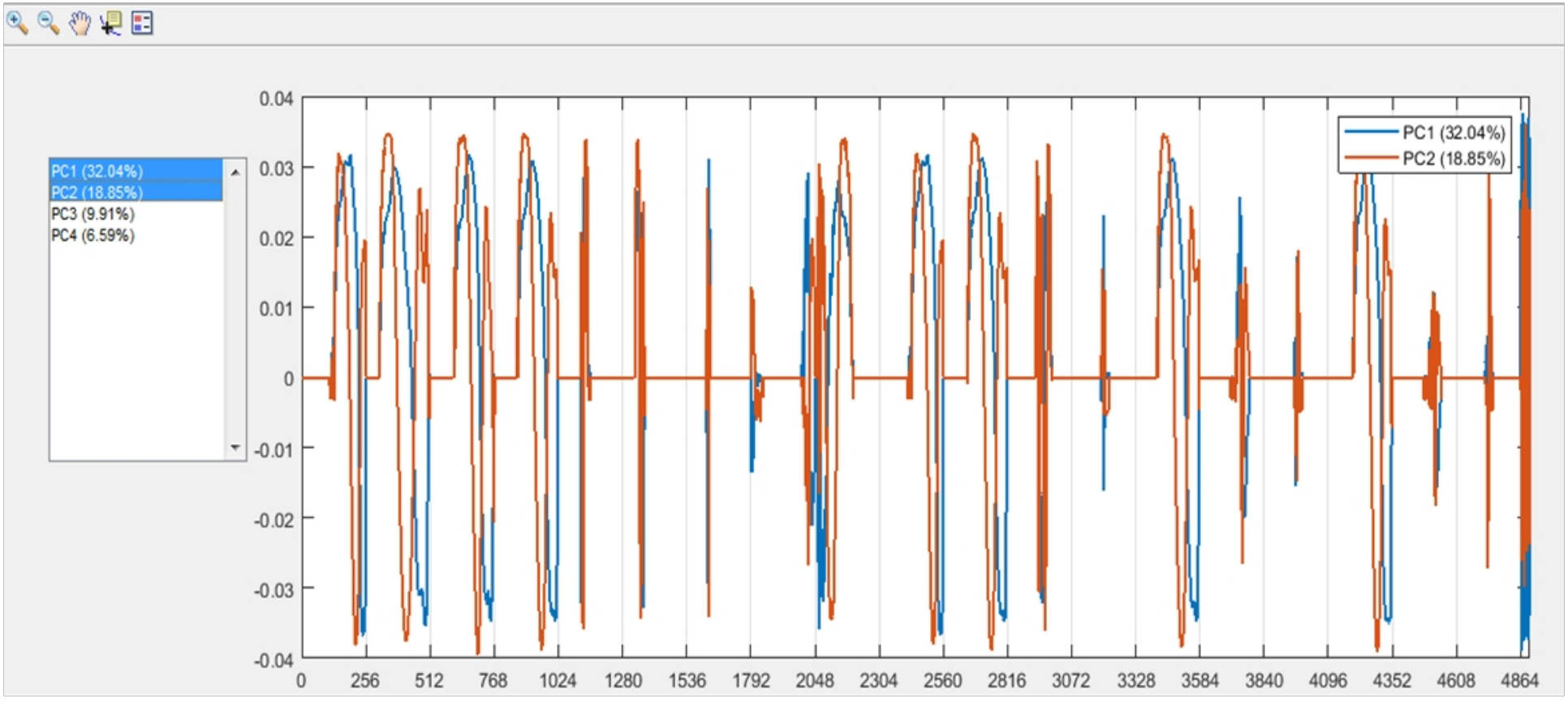

C:\Users\Rosi\Desktop\ROSALBA\papers\ColorGUI_SftwPres\GrapeMusts_Dataset
Load Colourgrams
GrapeMusts.mat
Load
Create Colourgrams
Select image extension
Normal
Keep only
Create
Curve
Sign
Threshold value
Class Manager
Current Class:
Harvest Time
Load Class
Clear/Modify
Save
PCA
PLS
a)
Visualize Signals
[1:256] Red
[257:512] Green
[513:768] Blue
[769:1024] Lightness
[1025:1280] Rel. Red
[1281:1536] Rel. Green
[1537:1792] Rel. Blue
[1793:2048] Hue
[2049:2304] Saturation
[2305:2560] Intensity
[2561:2816] Scores PC1 Raw
[2817:3072] Scores PC2 Raw
[3073:3328] Scores PC3 Raw
[3329:3584] Scores PC1 Mn
[3585:3840] Scores PC2 Mn
[3841:4096] Scores PC3 Mn
[4097:4352] Scores PC1 Auto
[4353:4608] Scores PC2 Auto
[4609:4864] Scores PC3 Auto
[4865:4900] Loadings & Eigenvalues
Plot Selected Curves
Select All
Deselect All
Colour by
Select image on the plot
L0_1a_1_smpl
Deselect
Image Reconstruction
Select regions on the plot
Type Selected Intervals
Background
Black
T0
T1
T2
T3
T4
0
256
512
768
1024
1280
1536
1792
2048
2304
2560
2816
3072
3328
3584
3840
4096
4352
4608
4864


b)
Red Frequency Distribution Curve
Blue Frequency Distribution Curve
Hue Frequency Distribution Curve
T0
T1
T2
T3
T4


c)
Red Frequency Distribution Curve
Blue Frequency Distribution Curve
Hue Frequency Distribution Curve


Column pretreatment
Autoscale
PCs number
4
Only selected variables
Current Class
Harvest Time
Eigenvalue of Cov(X) % Variance % Variance Cumulat...
1 619.4959 32.1149 32.1149
2 363.6095 18.8496 50.9645
3 191.1889 9.9113 60.8758
4 127.1312 6.5905 67.4663
5 71.2342 3.6928 71.1591
6 61.0724 3.1660 74.3251
7 49.4930 2.5657 76.8909
8 48.1140 2.4942 79.3851
9 41.7155 2.1625 81.5477
10 37.7057 1.9547 83.5024
11 30.2296 1.5671 85.0695
12 24.0320 1.2458 86.3153
Eigenvalues
Eigenvalues
Number of PCs
X Hotelling T^2 (6...
Y Q residuals (32...
Select objects
Confidence limits
95 %
×10^4
Q residuals (32.53%)
Hotelling T^2 (67.47%)
T0
T1
T2
T3
T4
95% Conf. limit
X PC1 (32.11%)
Y PC2 (18.85%)
Select objects
Confidence limits
%
PC2 (18.85%)
PC1 (32.11%)
T0
T1
T2
T3
T4


Column pretreatment
Autoscale
PCs number
4
Only selected variables
Current Class
Harvest Time
Eigenvalue of Cov(X) % Variance % Variance Cumulat...
1 619.4959 32.1149 32.1149
2 363.6095 18.8496 50.9645
3 191.1889 9.9113 60.8758
4 127.1312 6.5905 67.4663
5 71.2342 3.6928 71.1591
6 61.0724 3.1660 74.3251
7 49.4930 2.5657 76.8909
8 48.1140 2.4942 79.3851
9 41.7155 2.1625 81.5477
10 37.7057 1.9547 83.5024
11 30.2296 1.5671 85.0695
12 24.0320 1.2458 86.3153
Eigenvalues
Eigenvalues
Number of PCs
X Hotelling T^2 (6...
Y Q residuals (32...
Select objects
Q residuals (32.53%)
Hotelling T2 (67.47%)
Show Labels
Clear Labels
Exclude Selected Objects
Clear Selection
Confidence limits
95 %
X PC1 (32.11%)
Y PC2 (18.85%)
Select objects
Confidence limits
PC2 (18.85%)
PC1 (32.11%)


PC1 (32.04%)
PC2 (18.85%)
PC3 (9.91%)
PC4 (6.59%)
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
0
256
512
768
1024
1280
1536
1792
2048
2304
2560
2816
3072
3328
3584
3840
4096
4352
4608
4864
PC1 (32.04%)
PC2 (18.85%)

Training Set

X tr

- ○ Use current dataset
- ○ Edit dataset — Type included samples
- ◉ Split dataset using vector

X Preprocessing

Mean center

Y tr

◉ Load Ytr

Cross-Validation

custom

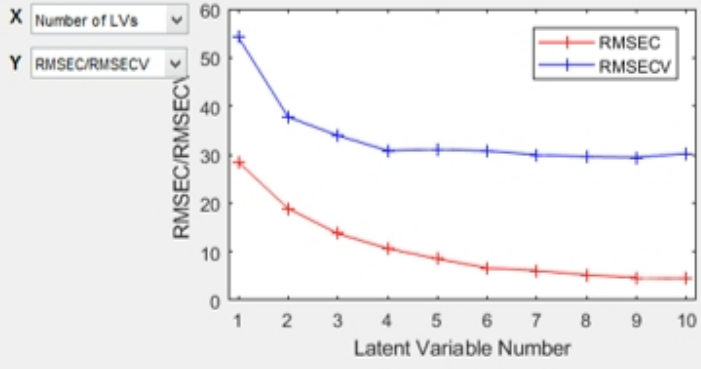
N. Data Splits:

N. LV max: 10 — OK

Selected number of LVs:

4

Calculate Model



Training Set

X tr

- Use current dataset
- Edit dataset — Type included samples
- Split dataset using vector

X Preprocessing

Mean center

Y tr

- Load Ytr

Cross-Validation

custom

N. Data Splits:

N. LV max: 10 — OK

Selected number of LVs:

4

Calculate Model

Test Set

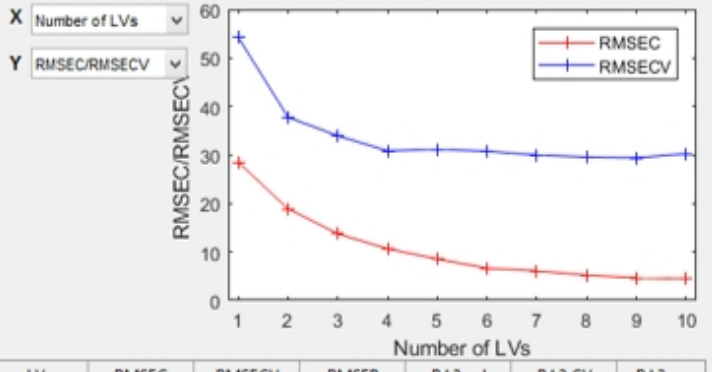
X ts

- Load dataset
- Edit dataset — Type included samples

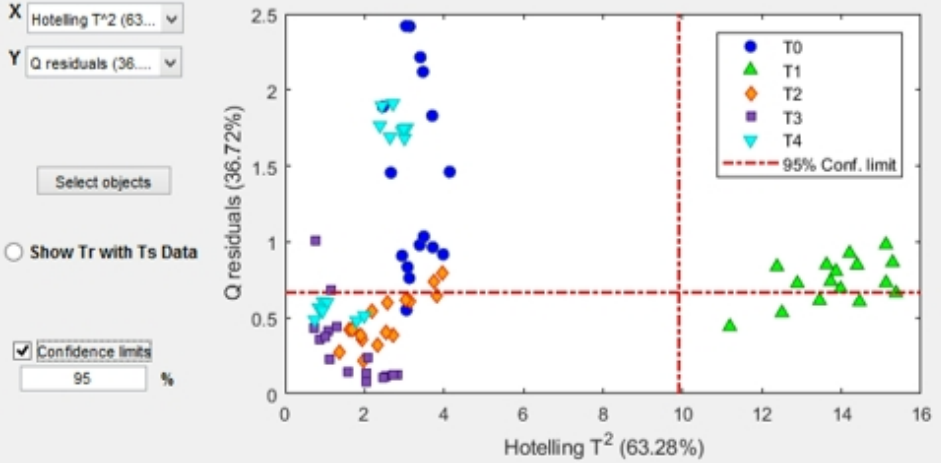
Y ts

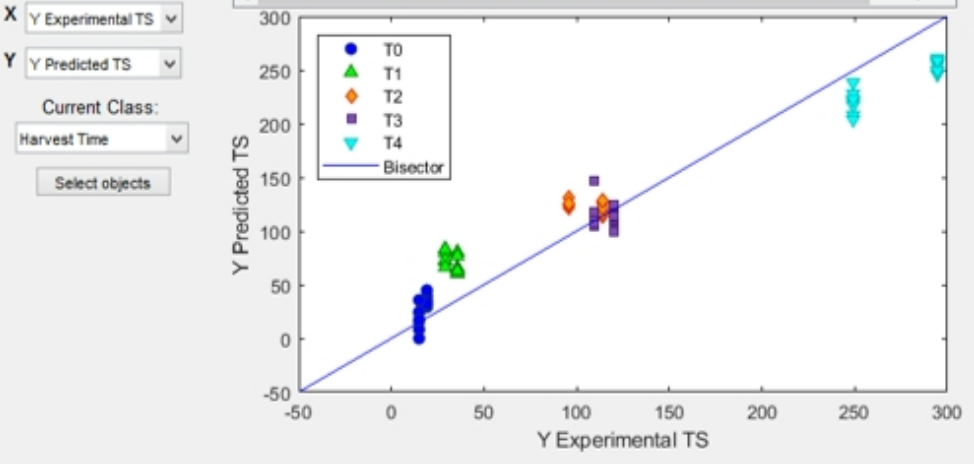
- Load Yts
- Unknown Yts

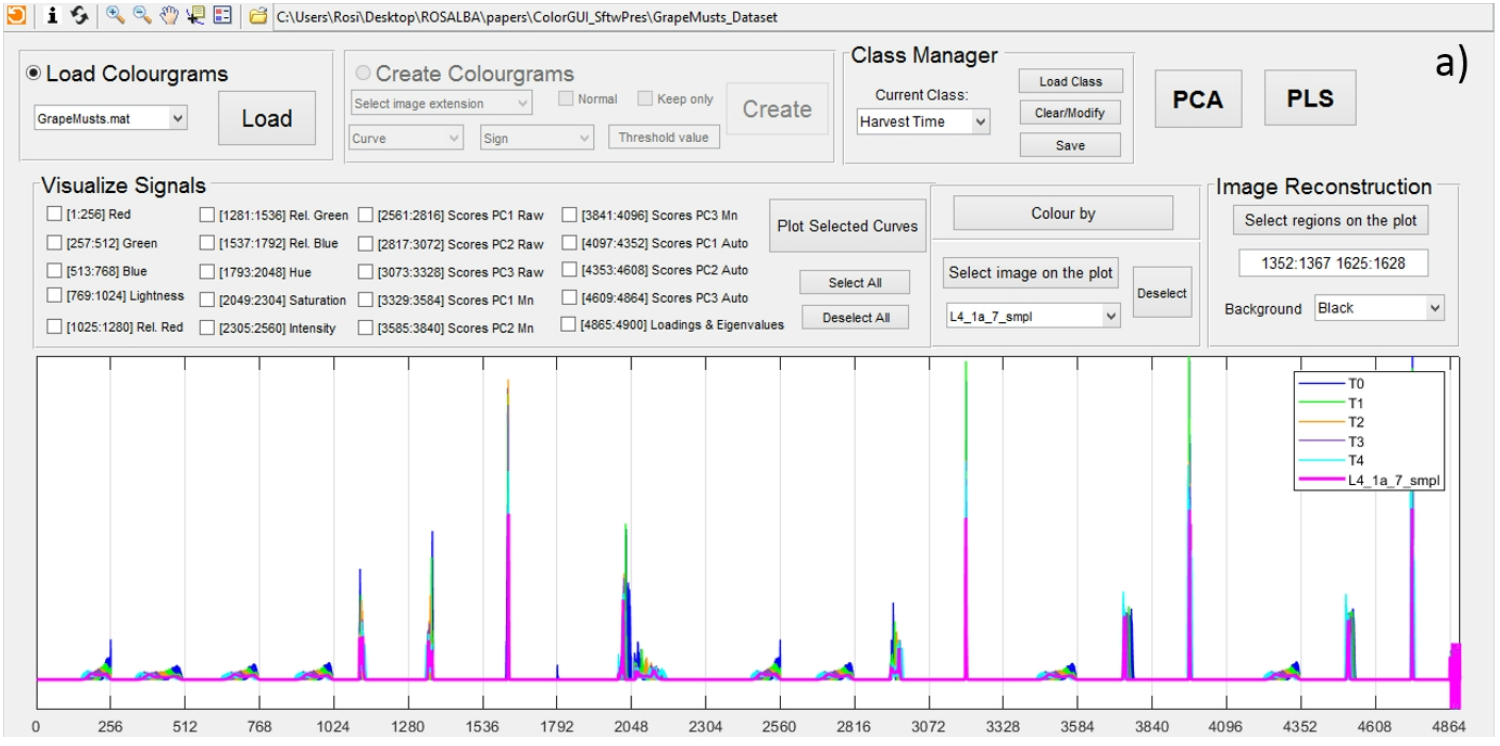
Apply Model

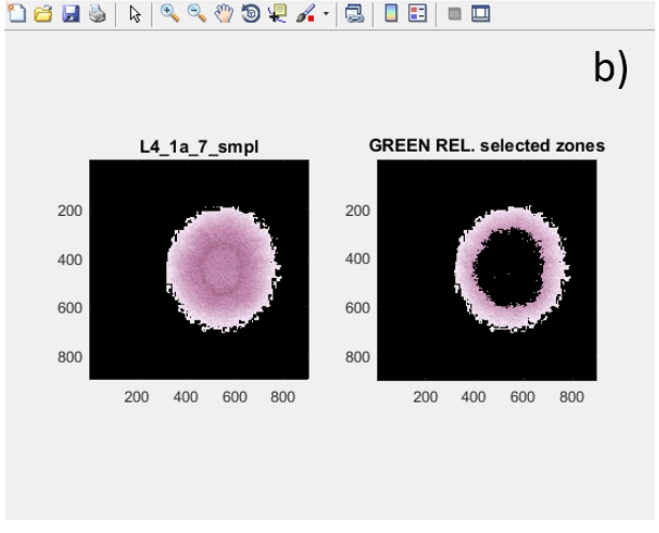


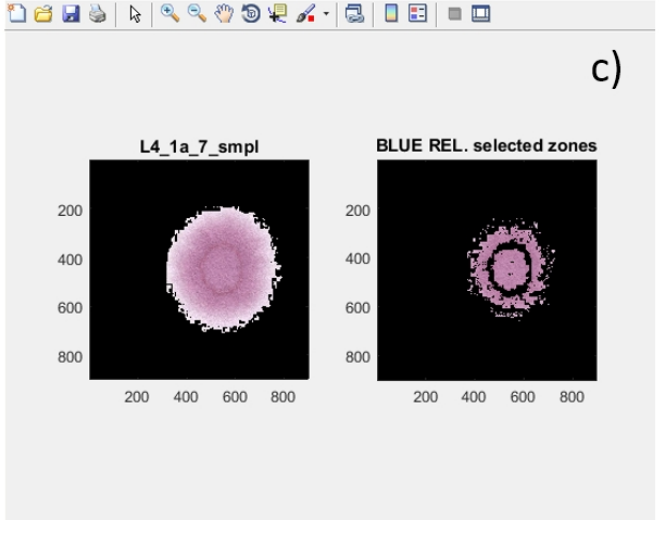
LV	RMSEC	RMSECV	RMSEP	R^2 cal	R^2 CV	R^2 p
4	10.6507	30.8205	27.6673	0.9872	0.8930	0.






C:\Users\Rosi\Desktop\ROSALBA\papers\ColorGUI_SftwPres\GrapeMusts_Dataset
a)
Load Colourgrams
GrapeMusts.mat
Load
Create Colourgrams
Select image extension
Normal
Keep only
Create
Curve
Sign
Threshold value
Class Manager
Current Class:
Harvest Time
Load Class
Clear/Modify
Save
PCA
PLS
Visualize Signals
[1:256] Red
[257:512] Green
[513:768] Blue
[769:1024] Lightness
[1025:1280] Rel. Red
[1281:1536] Rel. Green
[1537:1792] Rel. Blue
[1793:2048] Hue
[2049:2304] Saturation
[2305:2560] Intensity
[2561:2816] Scores PC1 Raw
[2817:3072] Scores PC2 Raw
[3073:3328] Scores PC3 Raw
[3329:3584] Scores PC1 Mn
[3585:3840] Scores PC2 Mn
[3841:4096] Scores PC3 Mn
[4097:4352] Scores PC1 Auto
[4353:4608] Scores PC2 Auto
[4609:4864] Scores PC3 Auto
[4865:4900] Loadings & Eigenvalues
Plot Selected Curves
Select All
Deselect All
Colour by
Select image on the plot
L4_1a_7_smpl
Deselect
Image Reconstruction
Select regions on the plot
1352:1367 1625:1628
Background
Black
T0
T1
T2
T3
T4
L4_1a_7_smpl
0
256
512
768
1024
1280
1536
1792
2048
2304
2560
2816
3072
3328
3584
3840
4096
4352
4608
4864


b)
L4_1a_7_smpl
GREEN REL. selected zones
200
400
600
800


c)
L4_1a_7_smpl
BLUE REL. selected zones
200
400
600
800