CHROMATOGRAPHIC FINGERPRINTING OF CHINESE GRAPE BERRIES

PEDRO F. M. DE SOUSA

Thesis submitted in fulfilment of the requirements for the degree of Master in Science

Supervised by:

Prof. Dr. Bjørn Grung (University of Bergen, Norway)

Prof. Dr. Yizeng Liang (Central South University, China)

ERASMUS MUNDUS MASTER IN QUALITY IN ANALYTICAL LABORATORIES

UNIVERSITY OF CADIZ - SPAIN

CENTRAL SOUTH UNIVERSITY - CHINA

UNIVERSITY OF BERGEN - NORWAY

2014

(This page was left blank intentionally)

ORIGINALITY STATEMENT

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma from EMQAL or any other master from other educational institutions, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked in EMQAL or from elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

(This page was left blank intentionally)

ABSTRACT

Chromatographic fingerprints from three varieties of grapes produced in China (Giant Rose, Red Globe and Summer Black), were obtained by gas chromatography coupled with mass spectrometry. These grapes were subjected to three different production treatments. Two pattern recognition techniques, PCA and PLS-DA were employed to verify the possibility of the creation of a model suitable for the classification of these samples. By means of PCA was verified that the samples could be decomposed according to the grape variety. Also, the varieties of the grapes could be discriminated by the means of PLS-DA (PLS2). Moreover, from PLS-DA (PLS1) models from the “Red Globe” variety samples it was verified that it is possible to classify these samples according to the one of the treatments (C), and some trends were observed on the classification of the others (A and B). The other varieties (“Giant Rose” and “Summer Black”) couldn’t be studied as regards the treatments due to the low number of samples.

CONTENTS

AKNOWLEDGEMENTS

This manuscript reflects part of the knowledge acquired during the Erasmus Mundus Master in Quality in Analytical Laboratories. The theoretical part of this master was attended at the University of Cadiz (Spain), the experimental part was performed in the Central South University (China), and the data analysis was performed in the University of Bergen, in Norway. I would like thank my supervisors, Prof. Dr. Yi-Zeng Liang, from the Central South University in China, and Prof. Dr. Bjørn Grung, from the University of Bergen in Norway. Also want to thank the director of the master in the University of Cadiz Dr. Miguel Palma, and the EMQAL program coordinator Dr. Isabel Cavaco, from the University of Algarve (Portugal), all the master lecturers, my friends and colleagues from the master and from CSU in China, specially my co-supervisors Fang Fang and Yonghuan Yun, who helped with the instrumental and chemometrics techniques in the CSU laboratory. A special thanks to my parents and my girlfriend. I dedicate this manuscript to my son André, who was born while I wrote this thesis.

André, Zhaneta and me (on the screen).

ABREVIATIONS

AC	– Alternate Current
DC	– Direct Current
CSU	– Central South University
CV	– Cross Validation
EI	– Electron Ionization
GC	– Gas Chromatography
LV	– Latent Variable
MS	– Mass spectrometry
m/z	– Mass to charge ratio
PCA	– Principal Component Analysis
PR	– Pattern Recognition
PC	– Principal Component
PCA	– Principal Component Analysis
PLS-DA	– Partial Least Squares Discriminant Analysis
UIB	– University of Bergen
PLS1(2)	– Partial Least Squares 1 or 2
RI	– Retention index
RMSEC	– Root Mean Square Error of Calibration
RMSECV	– Root Mean Square Error of Cross-Validation
TIC	– Total Ion Count

1. INTRODUCTION

1.1 Objectives

This research was a preliminary investigation on a potential classification method for certain varieties of grapes according to pre-harvest treatments. Three varieties of grape berries produced in China (Summer Black, Giant Rose, and Red Globe), subjected to different treatments during their growth, were sampled and their volatile contents were extracted. The sampling and the extraction procedures were both performed by another laboratory (in another CSU campus in Changsha-China). To accomplish this study, the following goals were established:

- Analysis of the grape samples by the means of gas chromatography coupled with mass spectrometry.
- Treatment of the chromatographic fingerprint data with baseline correction, peak resolution and identification of fingerprint markers.
- Analysis of the resolved fingerprint data by means of Principal Component Analysis for identification of similarities among samples.
- Creation of a classification model, by means of Partial Least Squares Discriminant Analysis, to discriminate the samples according to the grape variety and the treatments applied on the grapevines before sampling.

1.2 Theory

1.2.1 Food Quality Control

The need of authenticity of food products control required by consumers has led companies to adopt food safety and authenticity control strategies. The production of fake products, including food, is nowadays a worldwide problem. Some examples can be found on olive oils, honeys or alcoholic beverages, such as table wines and spirituous drinks. It is a concern of both authorities and food processors to avoid the unfair competition from counterfeiters who exploit the economy with the production and commerce of fake food products. Hence, the need of food companies to adopt methods which may improve their brands in the market. This may include the identification and reduction of forbidden compounds but also the monitoring of compounds which enhance the food value. [1]

In this work, three varieties of grapes subjected to different treatments during their maturing were analysed. Although the information about the exact nature of these treatments was not provided, the results of the analysis of these samples were studied in an attempt to find any possible trends in the chromatographic fingerprint data according to these treatments. The results obtained in this preliminary work may lead to the implementation of a method which can identify the quality of grapes as regards these treatments.

1.2.2 Grapevines

Grapevine, or *Vitis*, is the major genus of the family *Vitaceae*, and has two subgenera designated as *Muscadinia* and *Euvitis*. Typically, it grows within the latitudes of 50°N and 40°S and at altitudes under 3000 m. The most economically important varieties are the European grape (Vitis vinifera) and American grape (Vitis labrusca), which belong to the *Euvitis* subgenera. [2]

The grapevine *Vitis vinifera* is one of the most widely cultivated and economically important fruit crops worldwide. About 71% of the production is used in the production of wine, 27% as table grape, and 2% for raisins. [3] However, it can be difficult to grow due to high susceptibility to diseases (e.g., powdery mildew) and poor cold hardiness. Native American species and hybrids with *Vinifera* have better resistance, hence their popularity in areas with continental and humid climates. [4]

The hybridization of grapes has had a great development since countries with climates not suitable for grapevine production decided to produce wine or simply table grapes, such as Canada or North China. These grapes, designated as cold hardy varieties have a recent economic impact in the global market. [5]

Wine has archaeological records dating more than 7.5 thousand years. According to literature, it is suspected that wine residues were found in Iran, from the early mid-fifth millennium. Others suggest that Neolithic pottery (roughly the same time) revealed signs of beverage distribution. Older examples of fermented beverages were discovered, however produced from rice, honey, and fruit (hawthorn and/or grape). Such beverages were being produced in China as early as 7000 BC. [6]

In this work, three varieties of grapes produced in China were analysed. Giant Rose, “Jumeigui” in Chinese, is a hybrid derived from the *Vitis vinifera and Vitis labrusca* varieties. [7,8] Summer Black is also a hybrid of the varieties *Vitis vinifera and Vitis labrusca*. [9] Red Globe is a variety of *Vitis vinifera*. [10]

Grape analyses are essential for optimizing the harvest time, and eventually ensure the good quality of a wine or even the fruit itself as a table grape. The traditional analytical methods applied to the quality control of grapes are slow, tedious and destructive, and can’t keep up with the demands of the modern global market. Therefore, fast and low-cost analyses, combined with non-invasive or minimal sample preparation methods, are very important in food industry nowadays. [11] The application of different techniques for classification of grape samples can be found in literature, such as NIR spectroscopy [11], GC-MS [12,13], etc.

1.2.3 Chemical Pattern Recognition

Chemical pattern recognition is one of the first and most successful applications of chemometrics in analytical chemistry. For example, it is possible to determine the origin of a wine using a chromatogram of a sample, and determine which components distinguish different wines, and even determine the time of year the vine was grown. [14] Many examples of the application of pattern recognition techniques are described in scientific journals, and often related to analyses of food products. [15,16] There are several techniques used in chemical pattern recognition, and their success depends on the kind of data

provided. These techniques are classified into two main groups, designated as supervised and unsupervised techniques (**Figure 1.1**).

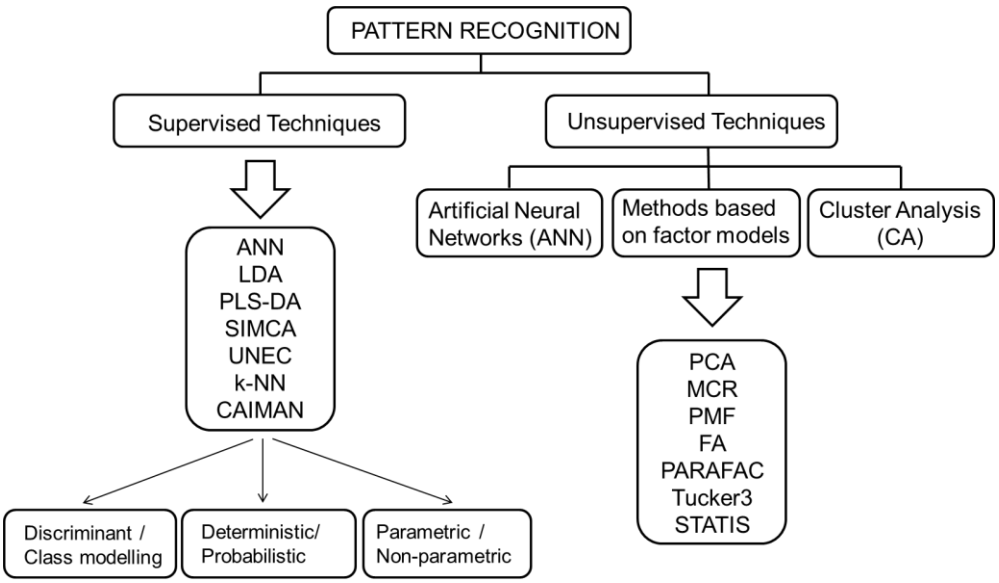

Figure 1.1 - General classification of pattern recognition techniques (adapted from **[17]**).

In unsupervised techniques, samples are decomposed taking into account the similarities between them, with no previous information provided about their classes. In supervised techniques, the samples are classified having previous knowledge about their classes. In this work, although there was a previous knowledge about the classes of the samples, Principal Component Analysis (PCA), an unsupervised technique, was employed to verify the possibility of classification. This technique is often used before applying any supervised technique to study the data's trends.

Because there was a previous knowledge about the classes of the samples (varieties and treatments), Partial Least Squares Discriminant Analysis (PLS-DA), a supervised technique was also employed to verify the possibility of the creation of a model for future classification of unknown the samples. [17]

1.2.4 Principal Component Analysis

The principal components concept has a great importance in chemometrics, since it is the basis of soft modelling and multivariate calibration methods. [18]

In this work GC-MS data was acquired, and a large amount of multivariate data was obtained when several compounds were taken into consideration simultaneously (several peaks areas from several samples).

This data, which can be arranged in a table (matrix), with rows as samples and columns as variables (compounds), may be virtually impossible to interpret due to its complexity. Principal Component Analysis (PCA) is simply a matrix algebra operation, easily performed by a computer, which allows the interpretation of multivariate complex data. [18] Basically, it reduces the amount of variables without losing important information. A straight line, designated as principal component (PC), is calculated so that it will have the direction of the maximum variance of the data. For each object (sample), the values obtained for n-variables are projected orthogonally onto this line (PC). These projections, designated as scores, are linear combinations of the original variables, and their values are the weighted sums of those variables. As represented by **Equation 1**, a matrix with the original data (**X**) is decomposed in the multiplication of a scores matrix (**T**) and a loadings (weights) matrix (**P**), plus a residual error (**E**). [14]

Equation 1 $$\boldsymbol{X} = \boldsymbol{T} \cdot \boldsymbol{P}^{\boldsymbol{T}} + \boldsymbol{E}$$

Figure 1.2 illustrates an example where seven samples can be visually divided into two groups taking into consideration three variables. The dots correspond to the samples and their coordinates have the variable values x_1, x_2 and x_3. The line bellow represents the orthogonal projection of the data on the PC (scores), where three dimensional data was transformed into one dimensional without losing information. Although this illustration only takes three dimensions (variables) into consideration, this can be performed for n-dimensional data.

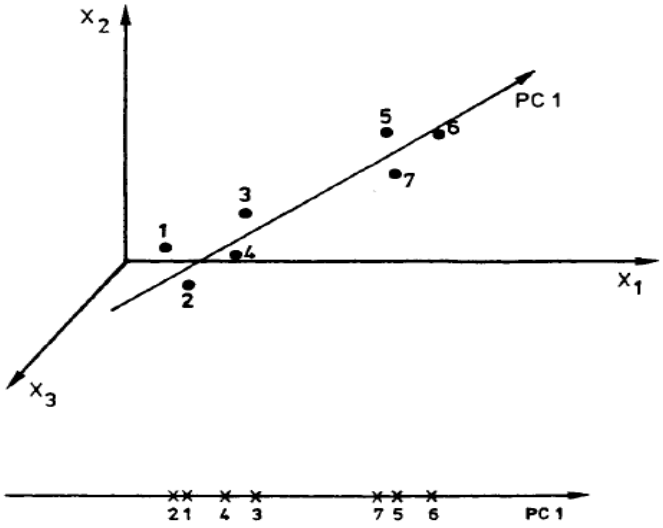

Figure 1.2 - Three dimensional data projected in one principal component. (Taken from **[18]**)

However, if the data has more complex trends it can't be just explained with one PC, and more PCs must be calculated, orthogonal to the previous ones and in the direction of the maximum variance of the data. This process goes on until a PC cannot explain more variance on the data. The first PC represents the direction in the data with the largest variation. The second PC, orthogonal to the first, represents the direction of the largest residual variation around the first PC (Figure 1.3). A third PC, orthogonal to the first two PC, will

represent the direction of highest residual variation around the plane formed by the first two PC. [18]

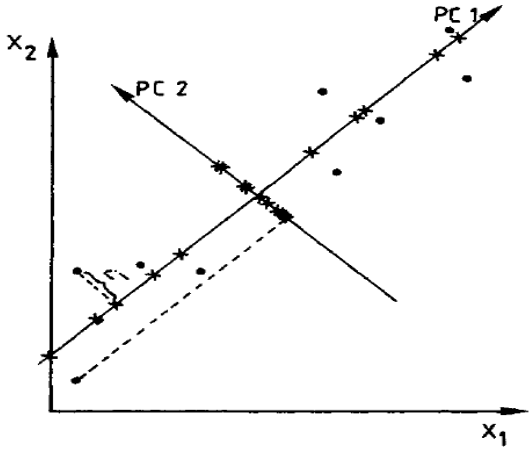

Figure 1.3 - Projection of two dimensional data in two principal components. (Taken from **[18]**)

After determining the needed number of PC and the scores, these can be plotted against each other. As illustrated in **Figure 1.4**, this allows visualizing clustering of samples, which means that they can be distinguished according to the studied variables. Additionally, the loadings also provide important information about the variables. From the analysis of loadings plots, the variables with more importance (weight) can be identified. [18]

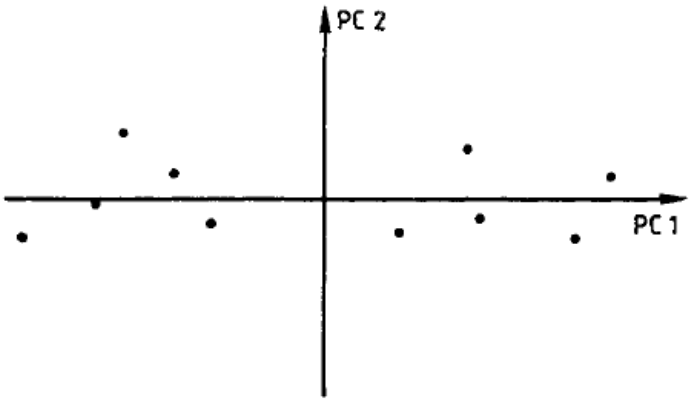

Figure 1.4 - Score plot of PC1 vs. PC2. (Taken from **[18]**)

1.2.5 Cluster Analysis

Cluster analysis is used to classify objects, characterized by the values of a set of variables, into groups. It is therefore an alternative to principal component analysis for describing the structure of a data table. There are many agglomerative methods described in literature. Software bundles, e.g. PLS Toolbox for MATLAB, have several clustering algorithms based on these algorithms. One of these methods, Ward's method, is based on a heterogeneity criterion. The heterogeneity is minimized when joining elements or clusters, and is defined as the sum of the squared distances of each member of a cluster to the cluster's centroid. This method can be used on original datasets or on the PCA reduced data. In this work, this method was employed for a better visualization of the PCA results observed in scores plots. [19,20]

1.2.6 Partial Least Squares Discriminant Analysis

Partial Least Squares (PLS), also designated as Projection to Latent Structure, is a multivariate regression algorithm based on latent variables designed to find important and related components between multivariate data, and is classified as a discriminant, probabilistic and parametric supervised pattern recognition technique. PLS regression combines features from PCA and Multiple Linear Regression. [17,20,21]

The strong point of this algorithm is that it can analyse high-correlated data, noisy data, and datasets with numerous variables. Also, it can model simultaneously several response variables. [22]

Two approaches arise from this method. In one, designated as PLS1, the relation between a data matrix and a response vector is studied. The other approach, designated as PLS2, handles several response variables simultaneously. There are many practical applications of this technique on analytical or statistical problems. As a general example related to analytical chemistry, PLS can be employed to study the relations between multivariate data, such as spectral data obtained from n-samples and the concentrations of a compound (PLS1). In the case of PLS2 the concentrations on n-compounds can be computed simultaneously. [20]

Although the PLS mathematical explanation is rather complex, basically it processes the data algebraically taking into account **Equation 2** and **3**. Assuming the same example from above, where the relation between spectral data and concentrations of compounds in samples is studied, in **Equation 2,** a matrix with spectral data (**X**) is decomposed in the product off a scores matrix (**T**) and a loadings (weights) matrix (**P**), plus a residual error (**E**). In **Equation 3**, a matrix or vector containing the concentrations of one or several compounds in the samples is decomposed in the multiplication of a scores matrix (**T**) and a loadings matrix (**q**), plus a residual error (**f**). This algorithm is essentially two PCA operations, where the scores matrix (**T**) is the same on both equations. This means that it is possible to obtain a scores matrix (**T**) that is common to both matrices "**X"** and "**c**". Therefore, the scores (**T**) model the spectral data

matrix (**X**) and also are good predictors of the concentration matrix (or vector) (**c**). [14]

Equation 2 $$\boldsymbol{X = T \cdot P^T + E}$$

Equation 3 $$\boldsymbol{c = T \cdot q^T + f}$$

Partial Least Squares Discriminant Analysis (PLSDA) is a classification modelling technique derived from PLS. The difference resides on the second matrix or vector (***c*** in **Equation 3**). In this case, instead of a matrix or vector containing measurement numerical values (i.e. concentrations, following the example described above), a "*dummy"* vector or matrix containing only categorical values -1 and 1 (or 0 and 1, depending on the software used) are used. These values represent the classes of the samples. The value 1 means that the sample belongs to one class and the other class takes the value 0 (or -1). If one vector is used (PLS1), only two classes can be compared. However, if a matrix is used (PLS2), several classes may be considered simultaneously, where each class is represented by a column, and the logical value 1 means that the sample belongs to a class and the value 0 is taken otherwise. After determining the scores and loadings, the score plots of the latent variables (instead of principal components) allow the visualization of clustering, and the loadings reveal the importance of the variables studied for the model. PLSDA has been applied successfully in many pattern recognition applications, such as food analysis. [14,20,23]

1.2.7 Peak resolution and baseline correction

Several problems arise in chromatography, especially when analysing complex mixtures, such as natural products like volatile compounds from grape samples. Similar compound migration rates and zone broadening are issues that may affect the interpretation of the chromatographic results. Similar or close migration rates result in peak overlapping, and also zone broadening often contributes to the same effect. If there are many compounds in a mixture, no matter how narrow the peaks may be, often it is virtually impossible obtain a total separation of peaks by simply changing chromatographic parameters. Multivariate curve resolution (MCR) has gained popularity recently because of the development of techniques that solve the problem of overlapping peaks (peak clusters). Several algorithms have been designed for this purpose and are reported in literature. Most of these algorithms can be used in MATLAB, and are also are available in user friendly software. [24]

In hyphenated techniques, such as GC-MS, multivariate data is collected in the form of a table or a data matrix, where one direction is related to the elution times, the other direction is related to the responses from the mass detector. In other words, one direction is related to the compositional variation of the system and the other to the variation in the response collected. These two variability directions can be used by chemometrics to differentiate overlapping peaks. [25]

Also, the extraction of qualitative or quantitative information from analytical signals, such as GC-MS, is difficult with the presence of drifting

baselines, particularly in multivariate analysis. Several background correction algorithms were developed and reported in literature. [26]

Often, background influences, such as baseline offset, baseline drift, or constant spectral background, are issues which may compromise the interpretation of chromatographic results. [27]

In this work, two UIB in-house programs were used for data treatment (section 2.3). Both programs, MS-Resolver 2.0 and Chrombox Q, perform baseline correction and peak resolution. However, these programs use different methods for the same purpose.

MS Resolver 2.0 was developed to automatically resolve peaks from complex GC-MS data. It is based on another program called Xtricator, which was also developed in UIB. While Xtricator resolves peak clusters individually, which is not very practical as regards time consumption when extracting fingerprints from tens or hundreds of chromatograms, MS Resolver resolves automatically all peak clusters from hyphenated chromatograms (GC-MS and LC-MS). [28] The peak resolution is performed by the Heuristic Evolving Latent Projections algorithm (HELP). The baseline detection and correction are performed by means of Latent Projection Graphs and Eigenstructure Tracking Analysis. [27]

Chrombox Q performs the resolution of overlapped peaks by means of Multivariate Curve Resolution-Alternating Least Squares algorithm (MCR-ALS). [29,25] The baseline correction is performed with CODA (Component Detection Algorithm) [30].

1.2.8 Data pre-treatment

Before employing PCA or other modelling techniques, however, in order to obtain a suitable model which may describe any of the trends in the data analysed, often the data has to be submitted to some treatment. One reason for data pre-treatment resides in the fact that the magnitude of the values of the variables obtained often may differ drastically from each other. Large variables produce significantly larger variances when compared with smaller ones. As explained before, PCA is based on maximum variance projections of the data. Consequently, the variables with larger variance have more impact on the model than the ones with less variance, and this may compromise the results or the conclusions of the studies when applying this principle.

Chromatographic data, which is the case in this work, may contain variables (peak areas) with such drastic differences in magnitude. Therefore, the data obtained in this work had to be pre-treated. This process can be performed automatically depending on the software used to analyse the data. Eigenvector Research PLS Toolbox for MATLAB, for example, has two default pre-treatment methods, which are the most frequently used, designated as mean centring or auto-scaling (mean centring + standardization). These pre-treatment methods can also be customized by the user, where advanced scaling methods sometimes have to be employed. This choice depends on the kind of data analysed.

The mean centring, as the name suggests, centres the data by subtracting the average of each variable’s data from the all the data related to that variable.

Scaling, like explained previously, can be used when large numerical differences between variable values exist. In this case, each variable value is divided by the standard deviation of the variable values. The auto-scaling is the application of both mean centring and scaling on the data. [31]

1.2.9 Cross Validation

An important decision has to be made when performing PCA or other related soft modelling analysis, such as PLS. The number of components (or latent variables) used to create a model has to be defined by the user. This parameter will influence the model in terms of its degree of fit and also its predictive ability. The degree of fit is a number between 0 and 1 (or percentage) and represents the degree of explained variation of the data. With more complex data, more components have to be taken into consideration to explain the variation of the data. The predictive ability, which has a greater importance in modelling than the degree of fit, is given by the variation of prediction. In this case, when new data is tested by the model (or even the data used to create the model), samples may be classified correctly or incorrectly depending on the number of components chosen. A model tends to become less predictive as the number of components increases, because it will explain very well the data analysed (fit), but will eventually explain poorly data that was not used to build the model. [31]

Several methods have been developed for cross validation (CV), and they can be applied depending on the modelling method and also the software

used. Eigenvector Research PLS Toolbox 7.5.2 for MATLAB comes with four premade methods, and also allows the user to define parameters in a custom method.

These methods are all based on the same principle. Basically, for a given data set, a series of experiments, designated sub-validation experiments are undertaken. Each involves the removal of a subset of objects from a dataset, construction of a new model using the remaining objects, and subsequently the application of the resulting model to the removed objects.

These experiments are performed for each PC/LV and the results are expressed as Root Mean Square Error of Cross-Validation (RMSECV) and the Root Mean Square Error of Calibration (RMSEC) values. Other statistics often employed in CV studies, is Root Mean Squared Error of Prediction (RMSEP). However, RMSEP requires a set of data different from the one used in the calibration set, which was not provided in this work. RMSECV can be a good estimate of RMSEP.

All these methods are based on the calculation of Root Mean Squared Error (RMSE) (**Equation 4**). In RMSECV, $\hat{y}_i$ correspond to the CV values, y_i comes from model's calibration values, and n is the number of objects in the model.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}}$$

Equation 4

The RMSEC represent the fitting of the data and the values should always decrease as more PC/LV are added to the model. However, the RMSECV values are determined from the cross-validation experiments, and can actually increase as more PC/LV are added. The optimal number of PC/LV is usually determined when the RMSECV ceases to decrease, or starts to increase. This means that more PC/LV may not improve the performance of the model. [19,32]

The classification results can be interpreted visually but also as figures of merit, such as selectivity, specificity and misclassifications. Sensitivity is the number of true positives classified as positive in the model. Specificity is the number of true negatives classified as negatives Misclassifications are objects that were not classified correctly by the model. This can be calculated from the calibration and cross-validation data. [18,19,33]

1.2.10 Instrumentation

The instrumental technique employed on the analysis of the grape samples was gas chromatography coupled with mass spectrometry.

1.2.10.1 Gas Chromatography

Gas chromatography (GC) is a separation technique in which a vaporized sample is moved with a flowing gas (the mobile phase, e.g., nitrogen or helium) through a glass or metal column containing an immobilized stationary phase. This phase is generally a low-vapour-pressure liquid polymer, coated or chemically bonded to a stationary support, i.e. a capillary column. As the mobile or gas phase is pressured through the column, the components in a sample also flow through the column at speeds, which depend on their chemical structure, composition and amount of stationary phase, the temperature, and gas flow rate. The elution time of a compound depends on its partition coefficient, which is a ratio of its concentrations in the stationary and mobile phases. The separation of compounds is based on the differences in the partition or solubility of various analytes in the stationary phase. [34]

The injection system of a gas chromatographic system is designed to introduce a representative amount of sample into the chromatographic column (**Figure 1.5**). Amongst the several existent injection techniques, the most common are split and splitless. These techniques were employed in this work.

In split injection only a small portion of the vapour enters the column, the rest is purged. A split injector is required for this purpose. In splitless injection, nearly all of the sample vapour injected is transferred into the column. The split injection is usually applied for analyte concentrations greater than 50 ppm. Concentrations ranging from 0.1 to 200 ppm, can be detected with the splitless injection. [35,36]

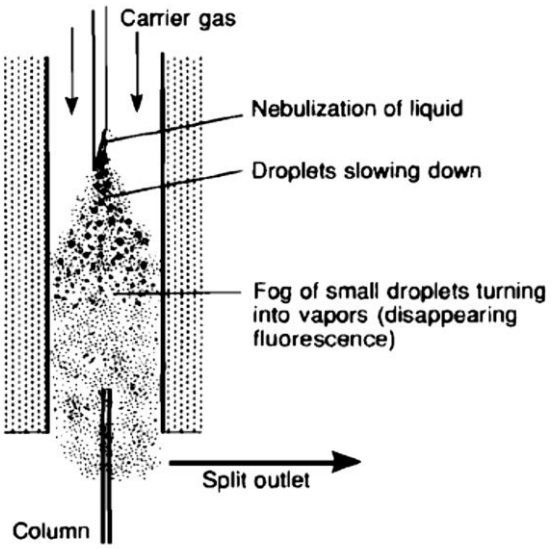

Figure 1.5 – Split injector scheme. Taken from **[36]**.

1.2.10.2 Mass spectrometry

Mass spectrometry (MS) is essentially the determination of the abundance of ions in gas phase according to their mass-to-charge ratio (m/z). The results are registered in the form of mass spectra, in which the relative intensities (ion abundances) are plotted against the m/z values of the ions. A mass spectrometer basically consists of a sample-inlet system, an ion source, a mass analyser for separating the ions according to their m/z values, and a detector (**Figure 1.6**). The ions are separated and detected in a high vacuum. Different ion sources can be used depending on the employed technique. However, electron ionization (EI) is the mostly used in GC-MS. [37]

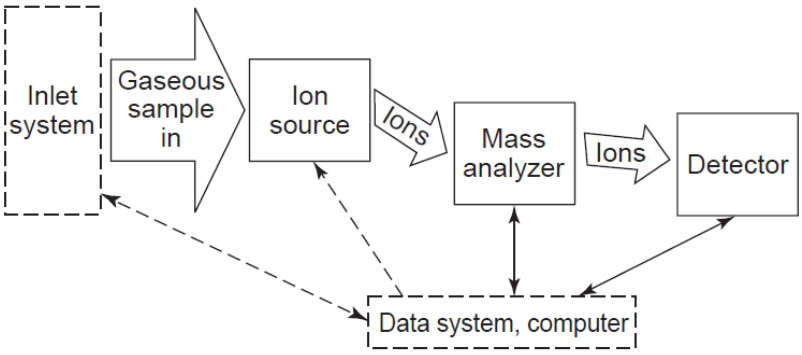

Figure 1.6 - Diagram of a mass spectrometer. (Taken from **[37]**)

An EI ion source (**Figure 1.7**) consists of a heated filament giving off electrons, which are accelerated towards an anode colliding with the gaseous molecules of the analysed sample. The collisions provoke the ionization of the molecules, and because the electron energy is higher (typically 70 eV) than the molecule ionization energy (about 10 eV), the remaining energy cause additional ionizations fragmenting the molecules even more. This fragmentation provides structural information for the elucidation of the analytes.

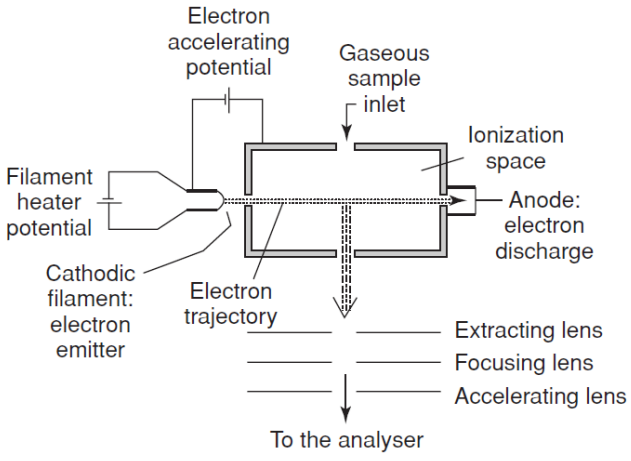

Figure 1.7 - Diagram of electron ionization source. (taken from **[37]**)

The mass analysers can be based on different principles, depending on their type. The mass spectrometer used in this work was equipped with a single

quadrupole analyser, which is composed of four circular section rods parallel to each other, with negative and positive charges (**Figure 1.8**). Positive ions entering the space between the rods will be drawn towards a negative rod. If the potential changes sign before they discharge themselves on this rod, the ions will change direction. Applying a radio frequency voltage (AC) superposed with a constant voltage (DC) will allow the ions either to reach the detector or not, depending on the ratio between both voltages. Changing this ratio allows the selective detection of ions according to their m/z. The mostly used detector in GC/MS systems is an electron multiplier, which basically converts the kinetic energy of the colliding ions into an electrical signal, which in its turn is processed by software into mass spectra. [38]

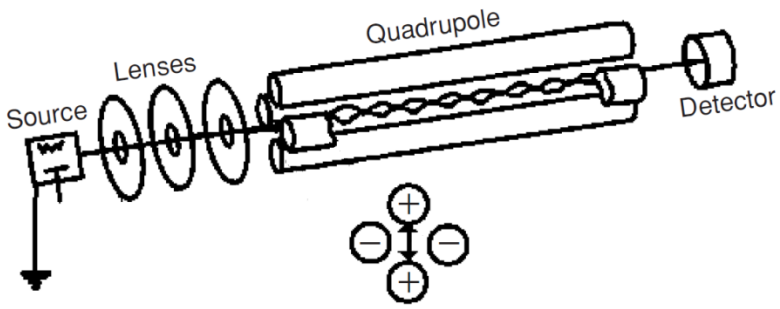

Figure 1.8 - Single quadrupole scheme. (Adapted from **[38]**)

1.2.10.3 Identification

One great feature of GC/MS data analysis procedures is the possibility of comparing the experimental mass spectra against library spectra. The National Institute of Standards and Technology (NIST) provides search routines on mass

spectra, continuously adding new spectra to their library and performing quality control over new and existing data. The mass spectra library contains more than 129,000 EI mass spectra of over 107,000 different compounds, and represents the most widely used mass spectral library in the world. [39]

2. EXPERIMENTAL

2.1 Sampling and extraction

The sampling and extraction of the samples were performed by another laboratory. Therefore, this section describes the sampling and extraction procedures according to the information provided by the laboratory where these procedures were performed.

The sampling was performed on three varieties of grapes produced in China (Summer Black, Giant Rose, and Red Grape) subjected to three different treatments in their production, which are designated as A, B and C. However, the description of these treatments was not provided until the present day.

In the extraction process, the grapes were crushed and distilled by micro-distillation and n-hexane was used as extracting solvent of the volatile compounds. The samples were kept as distilled/n-hexane mixtures in 15-ml polypropylene centrifuge tubes at –60 ºC until the time of analysis. The detailed information about the extraction procedures was not provided by the laboratory where the extraction was processed.

2.2 GC-MS Analysis

A Shimadzu gas chromatograph model GC-2010 coupled to a mass detector model QP-2010 and an auto sampler model AOC-20i was used. The column used was an OV-1 (100% dimethylpolysiloxane) capillary column (30 m × 0.25 mm i.d.; 0.25 μm film thickness). Helium was used as carrier gas under a flow rate of 1.0 mL/min. The volume of sample injection was 1 μL of in split mode (split ratio 2.0). The injector and interface temperatures were both at 250 °C. The oven temperature was programmed to hold at 50 °C for 5 min, rise until 100 ºC at 10 °C/min, and until 250 ºC at 5 °C/min. The mass detector worked in electronic impact (EI) mode, the ion source temperature was set at 200 °C, the detector voltage was set at 1.2 kV, and the solvent cut time was 4.5 min. The chromatograms were recorded in full scan mode (5 scan/s) with a mass acquisition range of 30-500 (m/z).

The GC-MS analysis conditions were adapted from a previously created method, applied in the fingerprinting of Traditional Herbal Medicines by the laboratory. These settings were loaded in the GC-MS software. However, several split ratios were studied before performing the definitive analysis in an attempt to improve the magnitude of smaller peaks. In splitless mode, although an improvement on the signals of smaller peaks was observed, it also caused excessive peak tailing. Therefore, different split ratios were studied and the resulting chromatograms were compared using the GC-MS software (GCMS Solution from Shimadzu).

Before analysing all the samples, which was performed during a week, one of the samples, the one with larger volume, was chosen as a control

sample. The analysis of this sample was performed every day of analysis, in order to detect any significant fluctuations in the signals and retention times. This helped to guarantee that conditions of analysis were maintained during the whole time of analysis.

Due to lack of time, because many students were queued to use the GC-MS and consequently the analysis had to be done in during a scheduled week, there was little time study of the chromatograms in a more profound fashion. The GC-MS analysis conditions could have better optimized, such as using different temperature programs to try achieving better peak separation. Different capillary columns could have been tested in order to try to obtain chromatograms with less overlapping peaks and peak tailing. Also, an internal standard should have been used, in order to correct the peaks areas, which may suffer variations due to loss of sample volume during the injection. However, this problem was minimized because an auto-sampler was used, and also the data was normalized before proceeding to the data analysis. The peak area data normalization was calculated in relation to the sum of all the peaks in each sample, i.e. the sum of the variable values equals 1 for each sample. This minimizes the effect of the differences in concentrations between samples when creating models.

2.3 Data analysis

The raw chromatographic data was pre-treated in order to make it suitable for pattern recognition analysis. Firstly, MS-Resolver 2.0 Pattern Recognition Systems AS, Bergen, Norway) was used for baseline correction, peak resolution and integration of the resolved profiles. Then, Chrombox Q, a MATLAB environment program, a UIB in-house program, which also performs baseline correction and peak resolution, was used to export the mass spectra to NIST 11 Mass Search 2.0.

The identification of the peaks by comparison with NIST 11 mass spectra library was just a mean to identify the variables (peaks) in the chromatograms. These compounds were not confirmed by comparison with standards or other means.

Although these programs work with different algorithms, they practically produced the same results. However, MS-Resolver is more user friendly as regards exporting the resolved profiles to a spreadsheet (MS Excel). On the other hand, Chrombox Q easily exports mass spectra data to NIST 11 Mass Search 2.0 for compound identification.

2.3.1 MS-Resolver 2.0

In MS-Resolver, some parameters can be adjusted in order to obtain reasonable peak resolutions. These parameters define how the resolution is processed. Very thorough resolutions result in an excess of resolved profiles,

including resolved unwished noisy zones. One parameter had to be adjusted. In the Xpert parameters tab, the “Minimum resolved intensity” was set to 0.01. This setting defines the lower limit of intensity of the peaks which should be resolved. All the other parameters were the program’s defaults (more detailed information in the software’s manual). [28]

In the background/resolution process, the “Zero Component Regions”, due to malfunctions in the automatic background detection method, were set manually from 1st to 100th and from 10300th to 10400th retention time. These regions were situated in the beginning and end of all chromatograms respectively, as shown in **Figure 2.1**. In each chromatogram 10425 retention times and 480 masses were analysed.

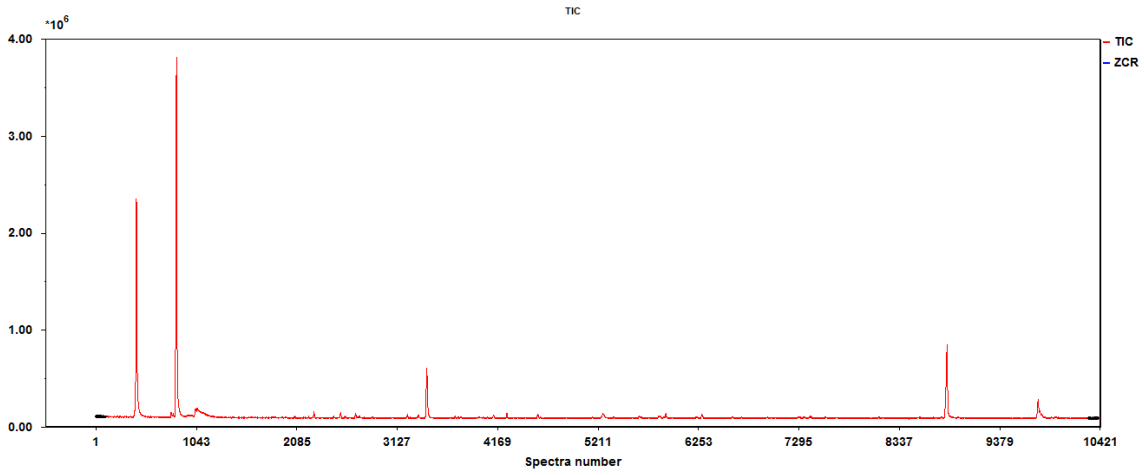

Figure 2.1 – Raw chromatogram from one of the samples (Summer Black CK39.6.7.8-1) with the TIC and the selected Zero Component Regions (beginning and end of chromatogram). This was performed on all chromatograms (60).

2.3.2 Chrombox Q

An automatic baseline correction was performed on all chromatograms. Since the resolved profiles' retention times were already known by means of "MS-Resolver", the peak detection "threshold" was set just to include these profiles (peaks). The mass spectra of the resolved peaks were exported to "NIST 11 Mass Search 2.0", where the mass spectra compared against a mass spectra library.

2.3.3 Eigenvector Research PLS Toolbox

The pattern recognition techniques, PCA and PLS-DA, were performed in MATLAB R2012a, with the aid of PLS Toolbox 7.5.2 (Eigenvector Research, Inc. 3905 West Eaglerock Drive, Wenatchee, WA 98801, USA). The results are described and discussed in the next section.

3. RESULTS AND DISCUSSION

3.1 Chromatographic data pre-treatment

The chromatographic data obtained from the analysis of 30 grape berry samples in duplicate analyses, which accounts for a total of 60 chromatograms, is represented in **Figure 3.1**. It is noticeable that there are some similarities and also some differences between the chromatograms. The samples are grouped in the three varieties: the upper 12 chromatograms belong to the “Summer Black” variety, the 36 in the middle to the “Red Globe”, and the 12 in the bottom to the “Giant Rose”. However, to clarify this assumption and to classify the samples according to the pre-harvest treatment, two pattern recognition techniques, PCA and PLSDA, were employed on the chromatographic pre-treated data to classify the grapes samples according to their variety and treatments.

Pattern recognition techniques were employed to classify the grapes samples according to their variety and treatments. However, the data had to be treated before proceeding to data analysis, and also a suitable and representative selection of peaks had to be performed.

Figure 3.2 illustrates a baseline drift, which was present in all the chromatograms. This drift is usually due to column bleeding and, in this case, as is can be observed by the mass spectrum of a zero component region of the chromatogram, a high peak with mass 32 indicates the presence of oxygen in the system due to some leak during the analysis. However, the baseline

correction algorithms from the software used (MS-Resolver and Chrombox Q) subtracted these masses from the all the spectra.

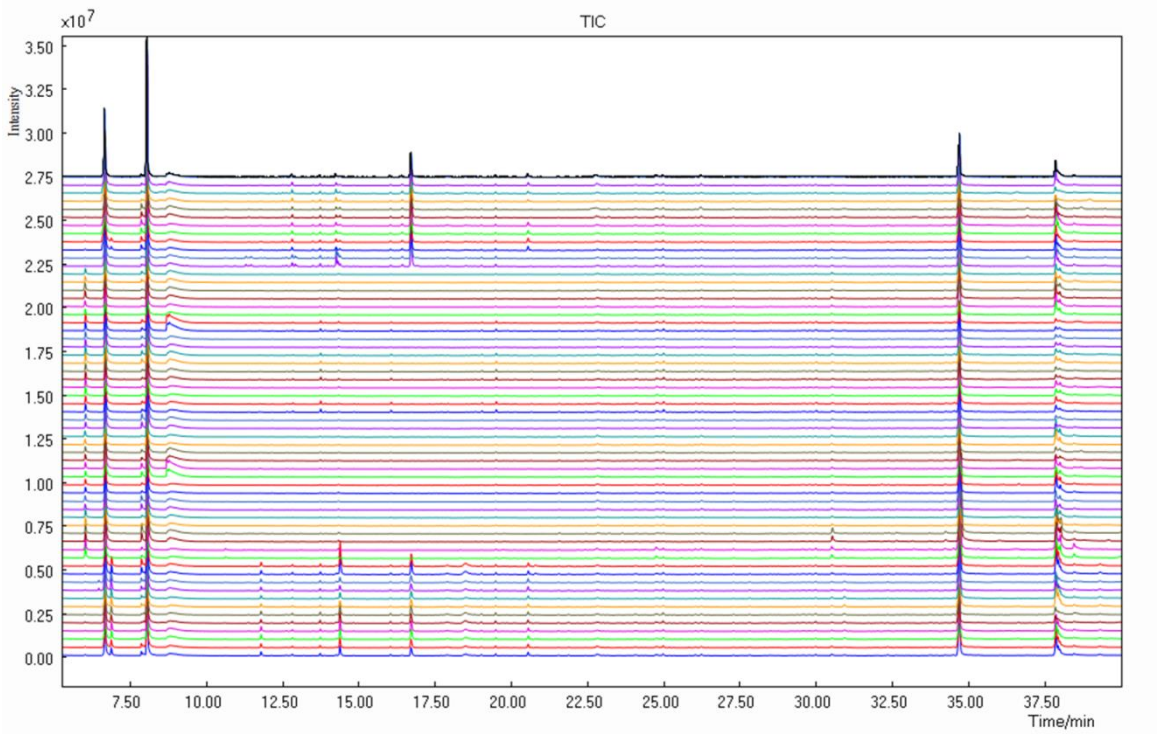

Figure 3.1 - Chromatographic fingerprints of 30 grape berry samples in duplicate. Obtained using Changde, an in-house software from CSU.

The criteria used for peak selection was established after data pre-treatment. For this purpose, the baseline was corrected and the peaks were resolved for each chromatogram using MS Resolver 2.0, as described in section 2.3.1.

Many resolved profiles were obtained on each chromatogram (from 30 to more than 100 peaks). However, only a few could be utilized for pattern recognition. Some problematic peaks, with unacceptable shapes due to improper resolution were discarded, and also many other peaks with very low intensities. The data was arranged in a MS Excel spreadsheet, in the form of a matrix with the samples as rows and the resolved profiles (peaks) as columns. The resolved profiles (peaks) of replicates and samples of the same variety of

grape were compared. Peaks occurring only once or few times in a grape variety were discarded. Additionally, some peaks were also discarded after verifying that ratios between replicates had exaggerated values. This was probably due to the small peak size and peak tailing seemed to affect the resolution in some cases. Within each grape variety, several variables were removed until the matrix contained no empty entries. This was made to improve the modelling, which can be greatly affected by bad consistency of the data.

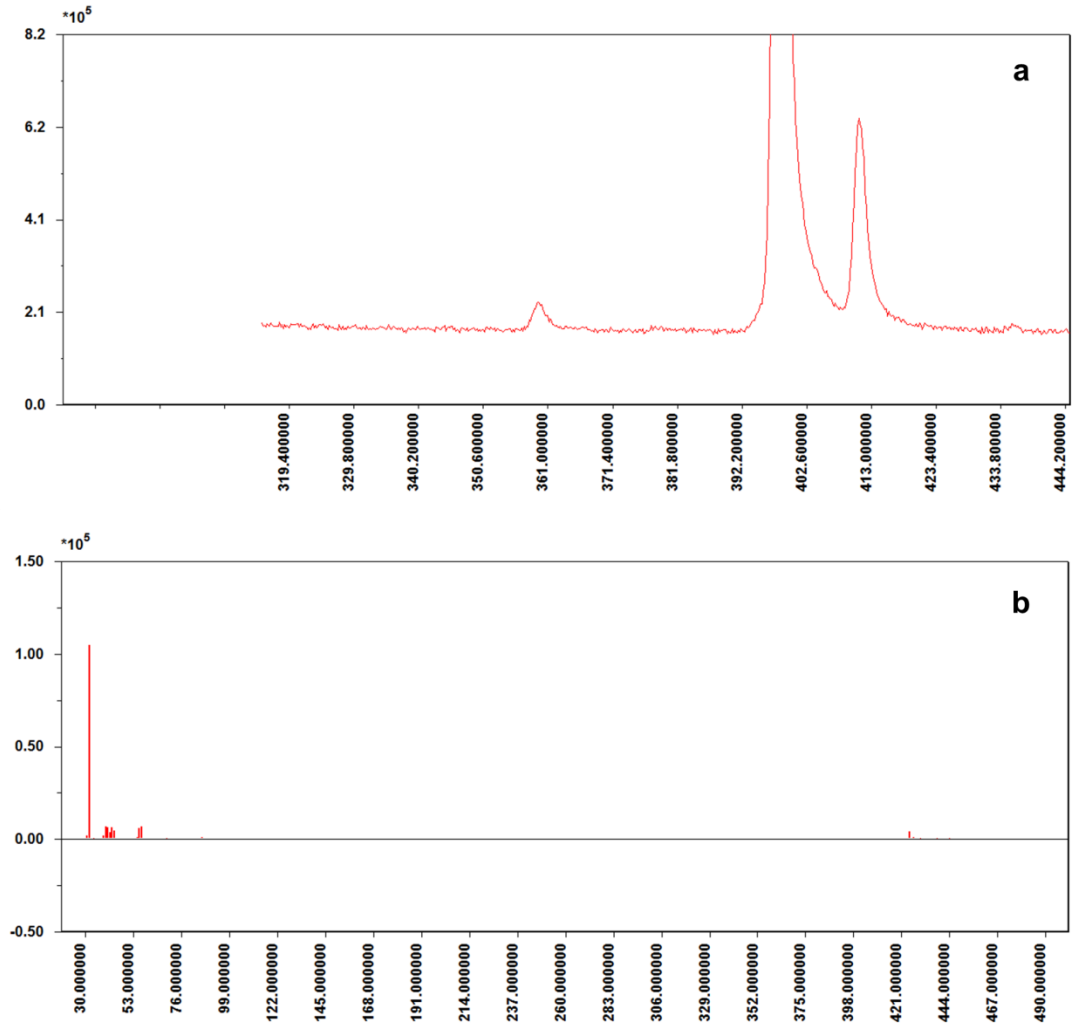

Figure 3.2 - Baseline drift observed in one of the chromatograms (TIC) (**a**). The retention times are expressed in seconds. Mass spectrum of a zero component region (**b**). The peak intensities are plotted against the m/z values. Obtained by means of MS-Resolver 2.0.

After removing variables that could have influenced negatively the models, a total of 20 resolved profiles (compounds) remained. **Figure 3.3** represents a chromatogram after pre-treatment, with the chosen resolved profiles.

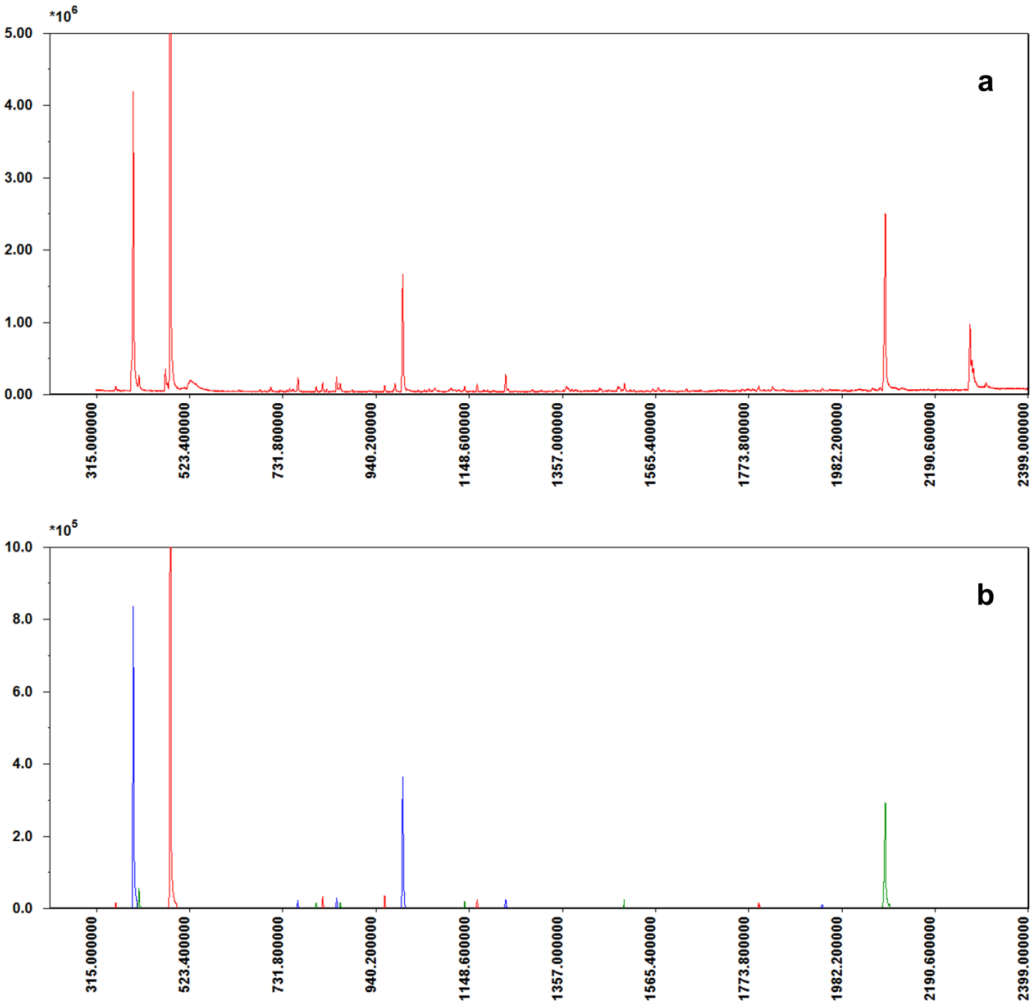

Figure 3.3 - Chromatograms of a sample Summer A38-4.5.6-1 obtained from raw data (**a**) and resolved profiles (**b**). The retention times are expressed in seconds.

The number of compounds was not the same for each variety of grapes. From the 20 compounds, the Giant Red variety had 18, the Red Globe had 9, and the Summer Black had 19. **Figure 3.4** illustrates the differences in the resolved profiles obtained from three samples of different varieties.

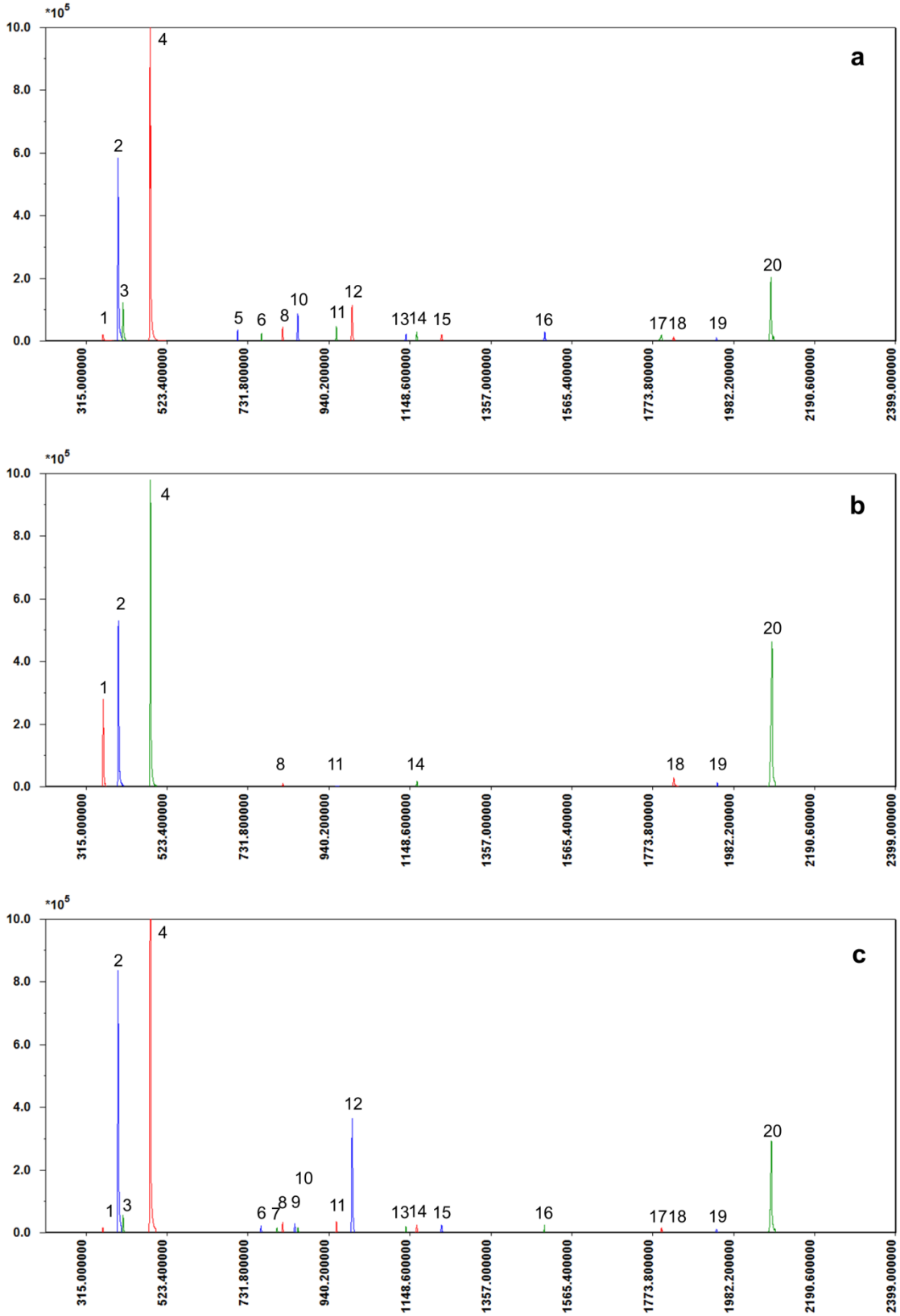

Figure 3.4 - Resolved profiles from three samples of different grape varieties: Giant Rose (**a**) Red Globe (**b**) and Summer Black (**c**). The retention times are expressed in seconds. The peak numbers correspond to the compounds found in **Table 1**. The identification of the compounds was not confirmed.

The compounds were identified by comparison of their mass spectra with the NIST 11 mass spectra library, by means of MS Search 2.0. All peaks were studied on each chromatogram. However, the identification of these compounds was not confirmed by comparing with a standard. Therefore, the identification of these compounds may be incorrect. These results are represented in **Table 1**. The similarities were verified for all chromatograms individually, and ranges presented are an approximation of the obtained values.

Table 1. Identified compounds by comparison with NIST 11 mass spectra library. None of these compounds was confirmed by comparison with standards. The similarities presented are approximations considering all chromatograms analysed.

Peak	Time (s)	Compound	Similarites (%)
1	359.4	Toluene	95 - 98
2	398.4	Hexanal	95 - 98
3	411.2	Ethyl butanoate	95 - 98
4	481.2	2-hexenal (E)	95 - 98
5	706.0	Ethyl hexanoate	95 - 98
6	767.0	Limonene	95 - 98
7	808.0	γ-Terpinene	95 - 98
8	822.6	Undecane	95 - 98
9	853.0	α- Terpinolen	95 - 98
10	861.4	β-Linalool	95 - 98
11	961.4	Methylcyclohexyldimethoxysilane	96-98
12	1001.2	α-Terpineol	95 - 98
13	1140.0	Tridecane	95 - 98
14	1166.8	2,7,10-Trimethyldodecane	95 - 98
15	1231.8	3-Isopentyl-2,4,4-trimethyl-2-cyclohexen-1-one	75-80
16	1497.4	Hexadecane	95 - 98
17	1798.2	Eicosane	95 - 98
18	1829.0	Tetradecanoic acid	95 - 98
19	1939.2	Isobutyl phthalate	95 - 98
20	2081.6	n-Hexadecanoic acid	95 - 98

Nevertheless, the spectra library comparison was just a mean to assure that the peaks from different chromatograms corresponded to the same compound. Most compounds had similarities with the spectra library above 95% in all

chromatograms. Also, most of them were reported in articles related to the analysis of grapes. There are reports of toluene (peak 1) being present in the grains and skin of grapes [40]. The presence of alkanes (peaks 8, 13, 14, 16 and 17) are also reported in literature [41]. Aldehydes, fatty acids, terpenes and terpene alchools (peaks 2, 3, 4, 5, 6, 7, 9, 10, 12, 18, 20) are typical grape components also reported in literature [42,43,44]. Isobutyl phthalate was also reported as an aroma volatile compound in fruits [45] . However, no reports were found about the presence of methyl-cyclohexyl-dimethoxysilane (peak 11) and the ketone 3-Isopentyl-2,4,4-trimethyl-2-cyclohexen-1one (peak 15) in grapes. The former had very high percentage of similarity (above 95%), which is a sign that it is very probable that this compound was actually present in the samples. However, the ketone (peak 15) had a low similarity in all samples (around 75%). Possibly this peak was not resolved properly due to its low intensity and it may correspond to a mixture of compounds.

The data matrix with the resolved profiles is represented in **Table A (ANNEXES)**, with the rows and columns corresponding respectively to the samples and the variables (compounds' peaks). A normalization of the data was performed to have all the chromatograms in the same scale. This is useful because of possible variations in the concentrations of the samples influence negatively the modelling. These variations may be due to the extraction process, or the maturity of the grape, and even loss of sample in the GC injection. The values (peak areas) were normalized relatively to the sum of the areas of the compounds in each sample (row), which means that the sum of all values in each row equals 1 (**Table B in ANNEXES**).

3.2 Principal Component Analysis

An unsupervised approach was carried out to investigate if the samples could be decomposed according to the varieties of the grapes and their treatments. However, the samples were only decomposed according to the variety of the grapes, and nothing notable as regards their treatments was verified by means of PCA. Two data pre-processing methods (described in section **1.2.8**) were applied and compared (**Figure 3.5** and **Figure 3.6**).

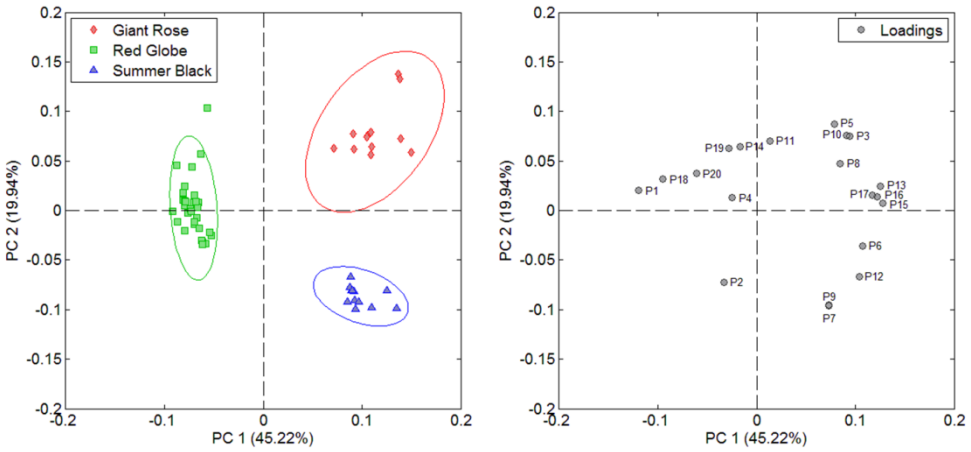

Figure 3.5 - PCA scores and loadings plots of the 30 grape samples and 30 duplicates using “Autoscale” pre-processing method.

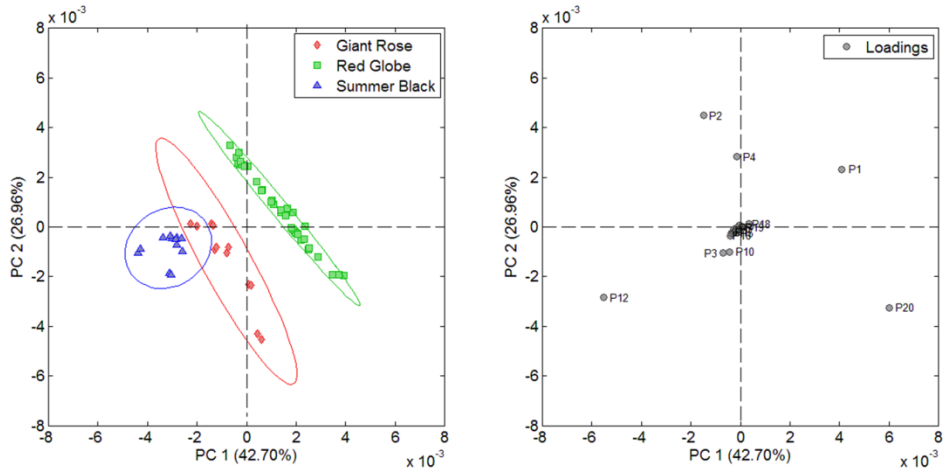

Figure 3.6 - PCA scores and loadings plots of the 30 grape samples and duplicates (30) using “Mean Centre” pre-processing method.

With just two PC it was possible to visualize clustering. However, the CV results (**Figure 3.7**) suggest that four PC should be taken into consideration in both models with “Autoscale” and “Mean-Centre” data pre-processing.

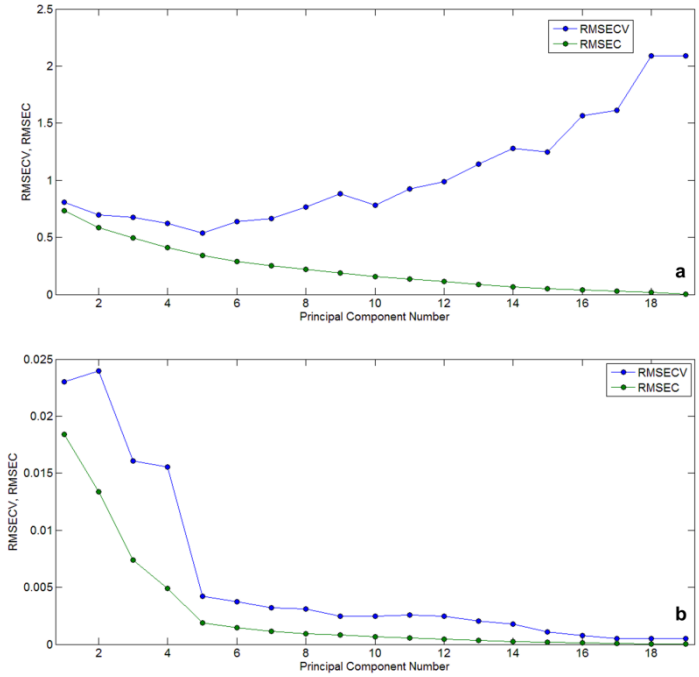

Figure 3.7 – Cross Validation plots from the PCA models from (**a**) “Autoscale” and (**b**) “Mean-Centre” data pre-processing methods.

The cluster separation and agglomeration in the PCA scores plot with the “Autoscale” data pre-processing method seems to be higher than with “Mean-centre”. Also, when comparing the loadings plots, it is noticeable that more variables define the model with “Autoscale”. With “Mean-centre” only five variables (the largest peaks) seem to have significant weight on the model (Peaks 1, 2, 4, 12, 20), whereas with “Autoscale” all variables seem to have significant weight. According to these results, the “Autoscale” pre-process method seems to produce a better PCA model. The dendrograms of the first four principal components with both pre-processing methods “Autoscale” and “Mean-Centre” are represented in **Figure 3.8** and **Figure 3.9**.

From the observation of the dendrograms, here it is also noticeable that the “Mean Centre” method resulted in a worse clustering. Also, nothing was observed as regards the grape treatments. Therefore, PCA cannot decompose the samples according to these treatments, and another technique is required for this purpose. This is described in the next section.

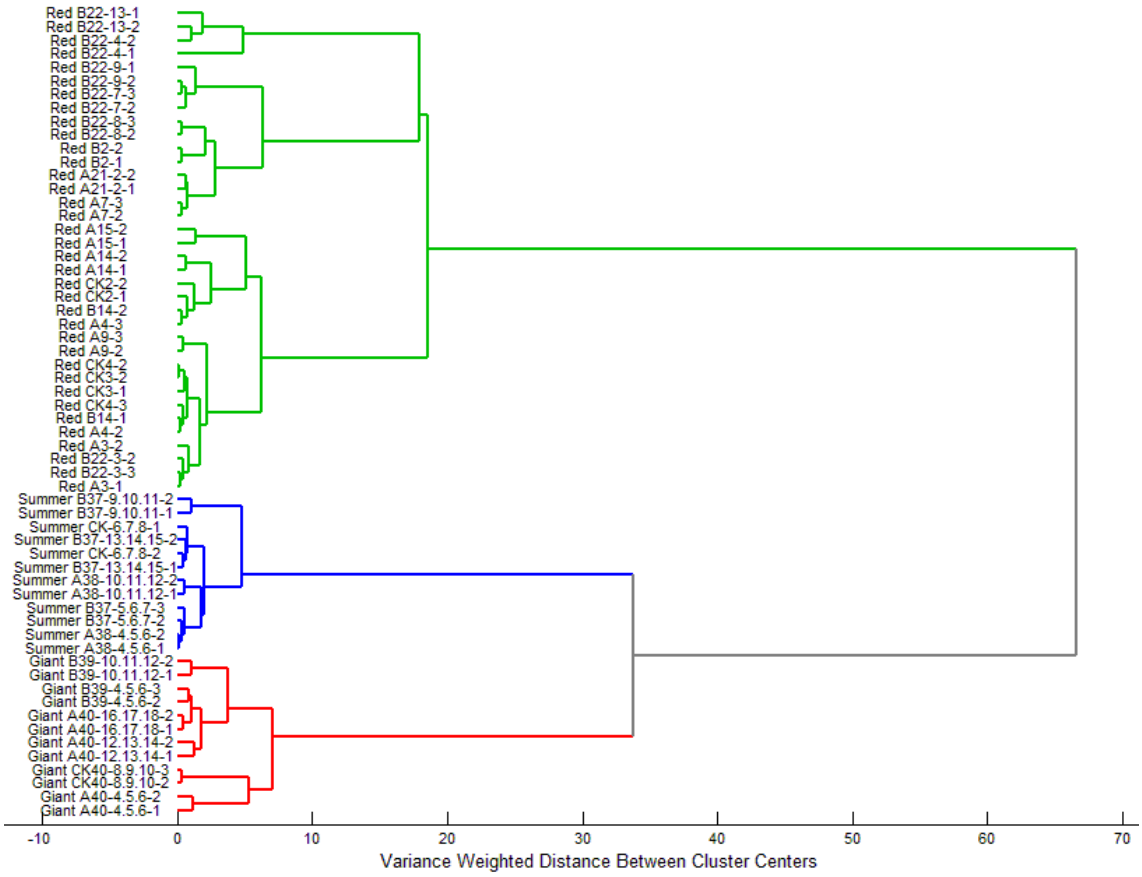

Figure 3.8 - Dendrogram of the samples using Ward’s Method using PCA (with four PC). The data was pre-processed with the “Autoscale” method.

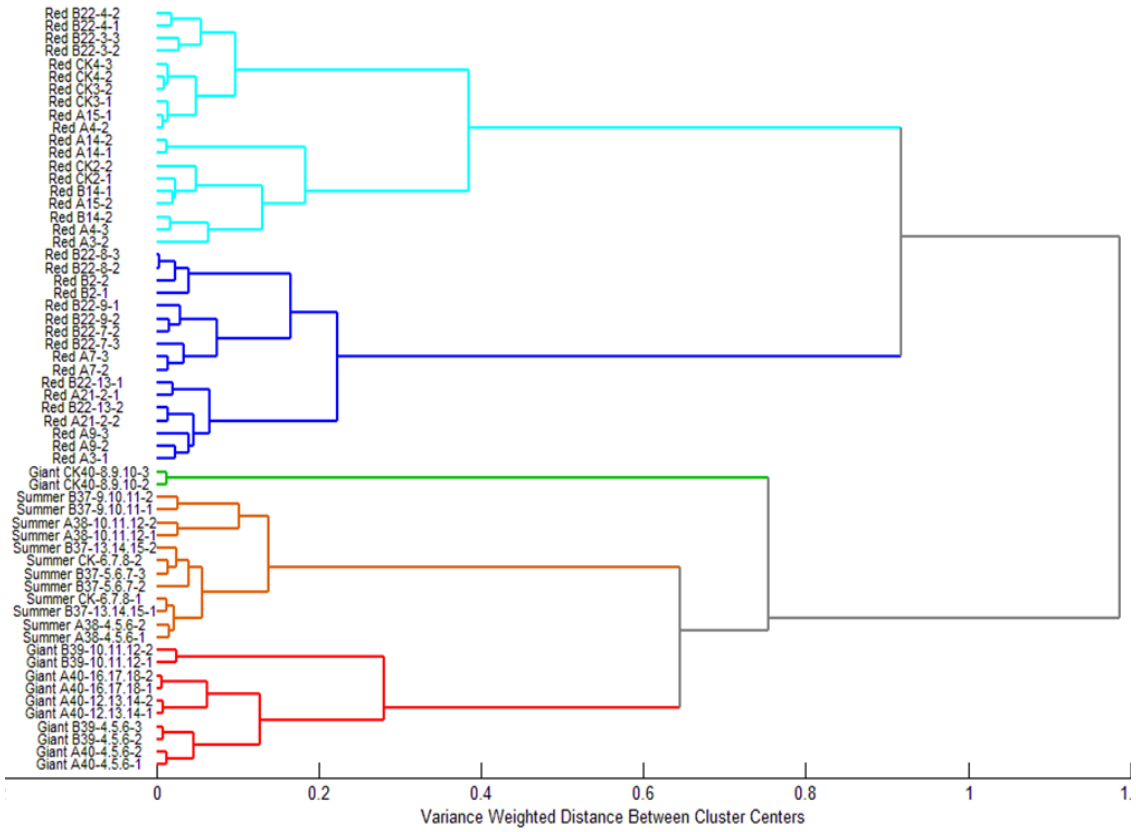

Figure 3.9 - Dendrogram of the samples using Ward's Method using PCA with the first four PC. The data was pre-processed with the "Mean-Centre" method.

3.3 Partial Least Squares Discriminant Analysis

3.3.1 Classification of grape varieties – PLS2

After the unsupervised decomposition of the samples by PCA, a supervised technique was employed in order to create a model to classify the grape samples according their variety. Two PLS-DA models (**Figure 3.10** and **Figure 3.11**) were created to classify the samples into three classes (varieties) simultaneously (PLS2 algorithm, described in section **1.2.5**) with two data pre-processing methods ("Autoscale" and "Mean-Centre").

Both models present good clustering, especially with the "Autoscale" pre-processing. In the model with "Mean-Centre" pre-processed data, the loadings from plots show that only six variables have significant weight on the model (Peaks 1, 2, 3, 4, 12 and 20), whereas all variables seem to have significant weight on the model with "Autoscale" pre-processing.

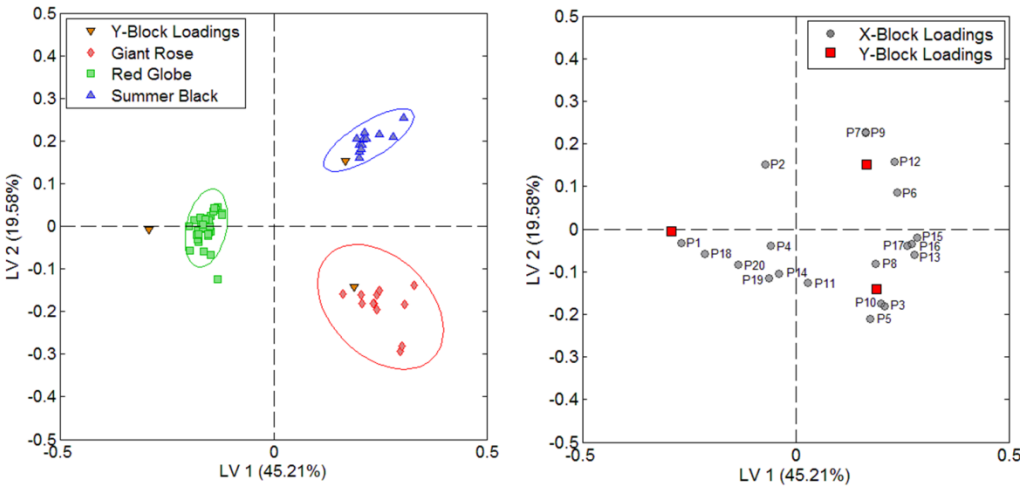

Figure 3.10 – PLS-DA (PLS2) scores and loading plots of the 30 grape samples and duplicates, classified according to the variety of the grapes. Data was pre-processed with "Autoscale" method.

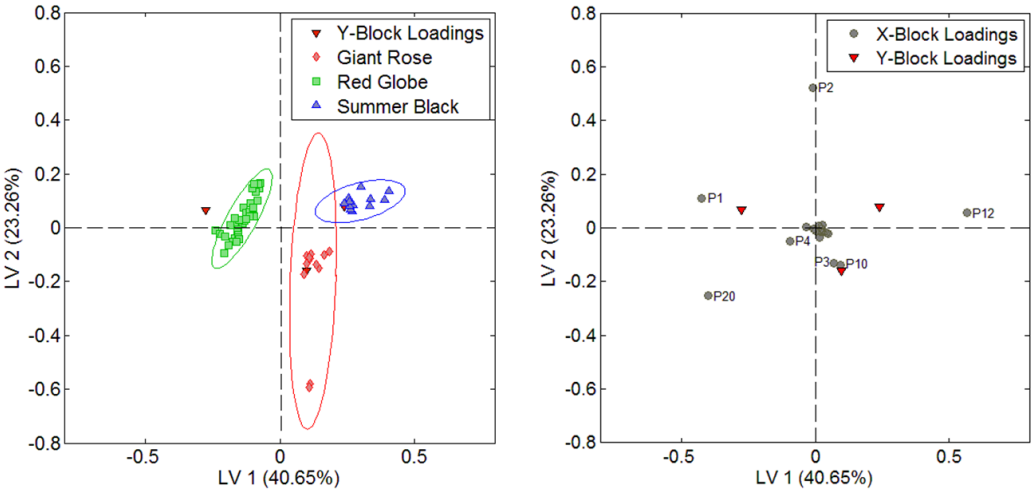

Figure 3.11 - PLS-DA (PLS2) scores and loading plots of the 30 grape samples and duplicates, classified according to the variety of the grapes. Data was pre-processed with “Mean-Centre” method.

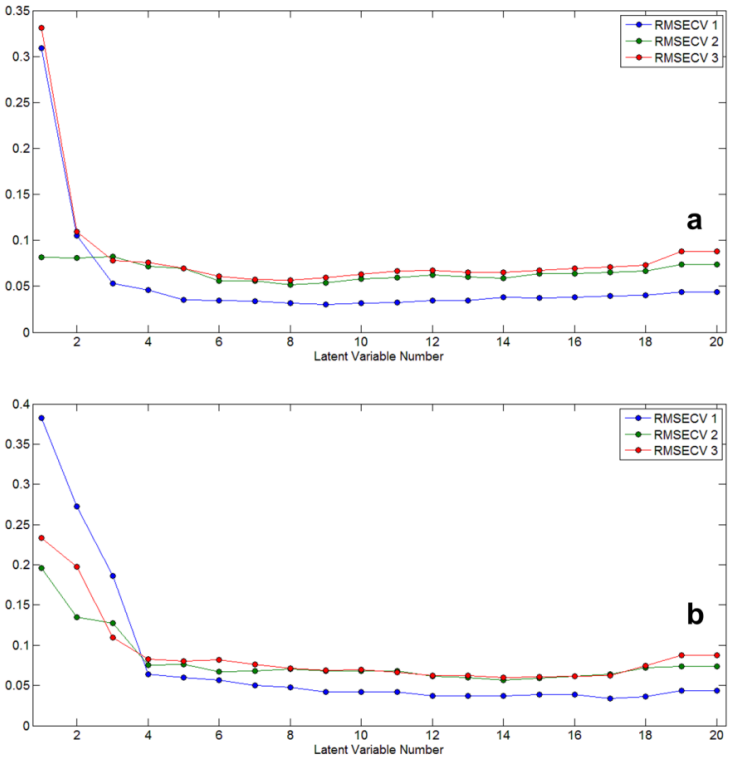

Figure 3.12 – Cross validation plots of the PLS-DA models with data (**a**) “Autoscale” and (**b**) “Mean-Centre” pre-processing considering thee classes simultaneously (PLS2). The RMSECV lines 1, 2, 3 correspond to each class of grape variety “Giant Rose”, “Red Globe” and “Summer Black” respectively.

From the cross validation plots (**Figure 3.12**), where the RMSECV values for each class studied (grape variety) are plotted against the number of LV, in the model with “Autoscale” pre-treatment no more than two LV should be considered in the model, whereas with “Mean-Centre”, four LV should be taken into consideration.

The classification results from the calibration and cross-validation of these models are represented in **Table 2**. According to these results and the samples analysed it was verified that both data pre-processing methods resulted in good classification models, with no misclassifications and with 100% selectivity and specificity for all the classes studied.

Table 2 – Classification results of the PLS2 model for the discrimination of the samples according to the grape variety (Giant Rose, Red Globe, and Summer Black).

				Calibration			Cross-Validation		
	Variety	**Samples**	**LV**	**TP[a]**	**TN[b]**	**Miss.[c]**	**TP[a]**	**TN[b]**	**Miss.[c]**
Autoscale data pre-processing	Giant Rose	12	2	100.00	100.00	0	100.00	100.00	0
	Red Globe	36		100.00	100.00	0	100.00	100.00	0
	Summer Black	12		100.00	100.00	0	100.00	100.00	0
Mean-Centre data pre-processing	Giant Rose	12	4	100.00	100.00	0	100.00	100.00	0
	Red Globe	36		100.00	100.00	0	100.00	100.00	0
	Summer Black	12		100.00	100.00	0	100.00	100.00	0

[a] True Positives (%) (Selectivity)
[b] True Negatives (%) (Specificity)
[c] Misclassifications

The variables that have more weight on the classification of the “Giant Rose” class are P3 and P10 (intense peaks) when the data was pre-processed with “Mean Centre”. With “Autoscale” the variables P3, P5, P8 and P10 contribute for the classification. These variables P5 and P8 are low intense peaks.

For the “Red Globe” variety, P1 is the most productive variable with “Mean-Centre” data pre-processing. With “Autoscale”, the low intensity peak P18 also contributes for this classification.

For the “Summer Black” variety, P6, P7, P9 and P12 are important low intense variables for the classification using “Autoscale” data pre-processing. With “Mean-Centre” only the intensity variable P12 contributes for the classification of this variety.

3.3.2 Classification according to grape treatments – PLS2

Besides the information about the variety of the grapes, some information about the pre-harvest treatments was also provided. Unfortunately, the information about the exact nature of these treatments was not provided. Nevertheless, a study was performed to verify if these treatments, designated as A, B, and C, somehow altered the composition of the grapes, and a classification model could be created from the obtained chromatographic fingerprint data.

Due to the low number of samples from the “Giant Pink” and “Summer Black” grape varieties, especially because only one sample with the treatment C was analysed from these two varieties, only the chromatographic fingerprint data from the “Red Globe” variety could be used in this study.

Two PLS-DA models of the “Red Globe” samples taking into consideration three classes (treatments) simultaneously (PLS2 algorithm) were studied, one with “Autoscale” and other with “Mean-Centre” data pre-processing.

The optimal number of LV was determined from the cross-validation of both models (**Figure 3.13**). The model with “Mean-Centre” data pre-processing seems be optimal with no more than two LV. However, four LV should be taken in consideration with “Autoscale” data pre-processed model. The PLS-DA (PLS2) score plots modelled with both pre-processing methods present a certain level of clustering, as shown in **Figure 3.14** and **Figure 3.15**. However, in any of the cases is possible to observe good clusters.

The classification results from the calibration and cross-validation of these models are represented in **Table 3**. According to these results it seems that only treatment C had a good classification. The other classes have too many misclassifications, considering the number of samples analysed. Also, the “Autoscale” data pre-processing seem to have resulted in one less misclassification on treatment “A”. The selectivity and specificity of the models studied are similar with both pre-processing methods.

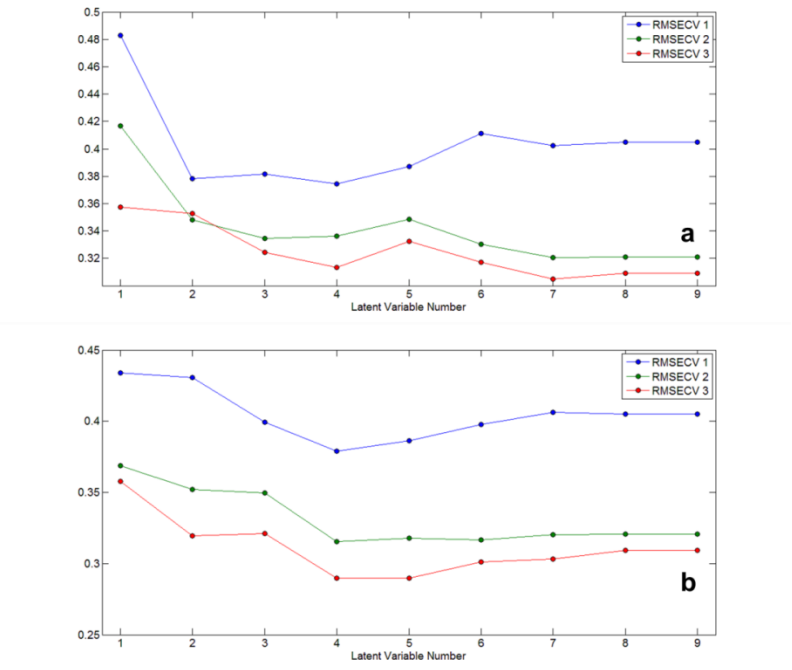

Figure 3.13 – Cross validation plots from the “Red Globe” discrimination PLS2 models, using (a) “Mean-Centre” and (b) “Autoscale” data pre-processing methods.

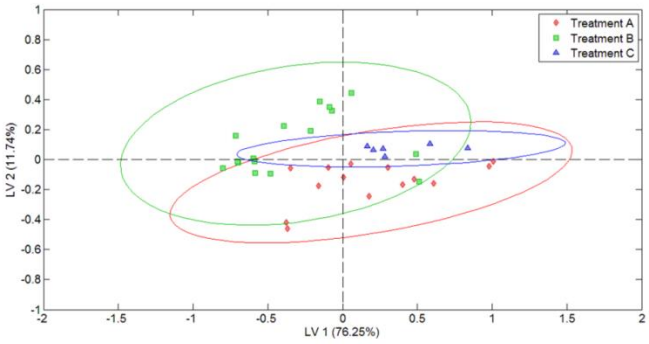

Figure 3.14 - PLS-DA score plots of the first two LV using PLS2 method. Three treatments classes are considered simultaneously. Data pre-processed with “Mean-Centre” method.

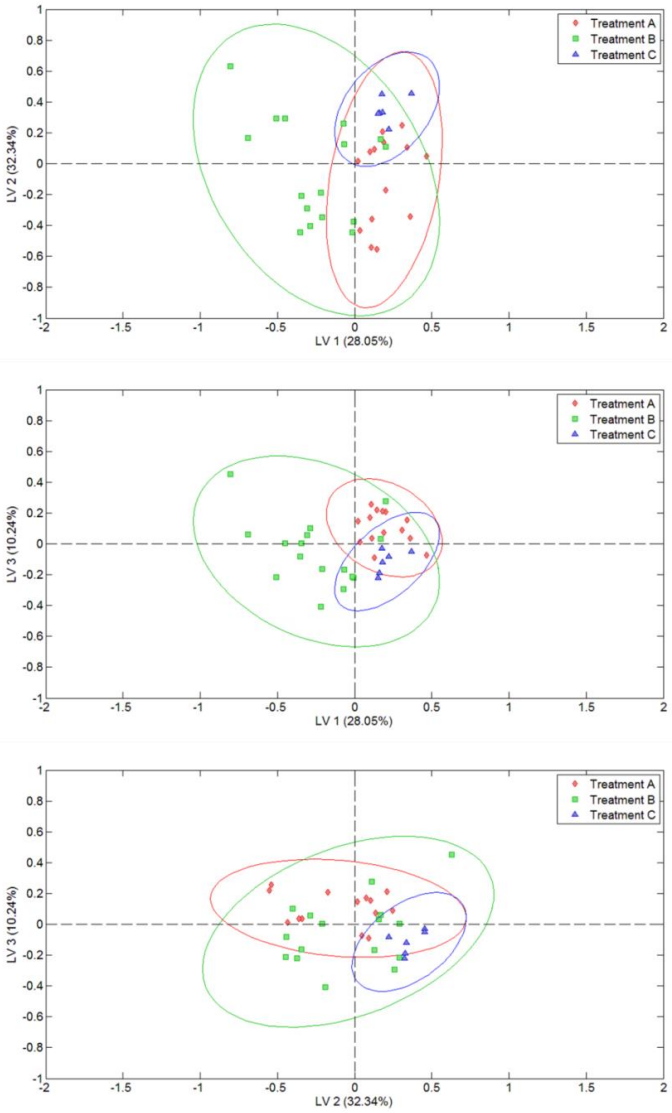

Figure 3.15 – PLS-DA score plots combining the first three LV, using PLS2 method. Three classes are considered simultaneously. Data was pre-processed with “Autoscale” method.

Table 3 – Classification results of the PLS2 model for the discrimination of the "Red Globe" samples according to the grape treatments (A, B, C).

				Calibration			Cross-Validation		
	Treatment	Samples	LV	TP[a]	TN[b]	Miss.[c]	TP[a]	TN[b]	Miss.[c]
Autoscale data pre-processing	A	14	4	100.00	90.91	0	85.71	86.36	2
	B	16		87.50	100.00	2	87.50	90.00	2
	C	6		100.00	100.00	0	83.33	100.00	1
Mean-Centre data pre-processing	A	14	2	100.00	90.91	1	85.71	86.36	3
	B	16		87.50	100.00	2	87.50	90.00	2
	C	6		100.00	100.00	0	83.33	100.00	0

[a] True Positives (%) (Selectivity)
[b] True Negatives (%) (Specificity)
[c] Misclassifications

3.3.3 Classification according to grape treatments – PLS1

Another approach made was to discriminate the treatments individually (PLS1 algorithm). Since there were three varieties of grapes and three treatments on each variety, the data had to be analysed in subsets. This should have been performed for all varieties of grapes. However, due to the low number of samples from the "Giant Rose" and "Summer Black" varieties, this study was performed only on the "Red Globe" samples. In these PLS1 models, each treatment (A, B, C) is discriminated from the rest.

Once again, two pre-processing methods were employed in this study. **Figure 3.16** and **Figure 3.17** represent the cross validation plots of the PLS-DA models with "Autoscale" and "Mean-Centre" data pre-processing respectively. The optimal number of latent variables to use in the models studied was determined from the observation of these plots.

In the models with "Autoscale" data pre-processing (**Figure 3.16**) it was verified that three LV should be taken into consideration for the treatments A and C, and only two for the treatment B. In the latter, the RMSECV with three

LV is slightly lower than with two. However this difference is small, and the third LV does not improve the prediction of the model considerably. Therefore, two variables suffice. In the models with “Mean-Centre” data pre-processing (**Figure 3.17**) it was verified that two LV should be taken into consideration when discriminating the three treatments.

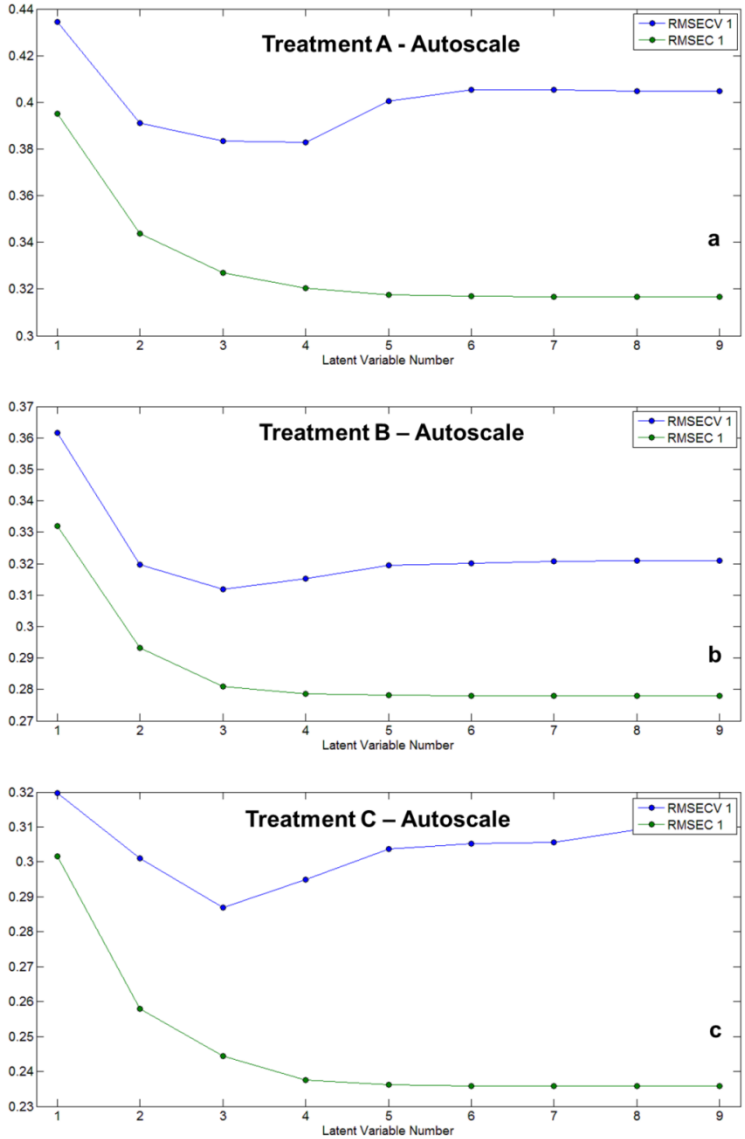

Figure 3.16 – Cross validation plots of the PLS1 models to classify each Treatment (A, B, C), with “Autoscale” data pre-processing. The RMSECV and RMSEC are plotted against each latent variable added.

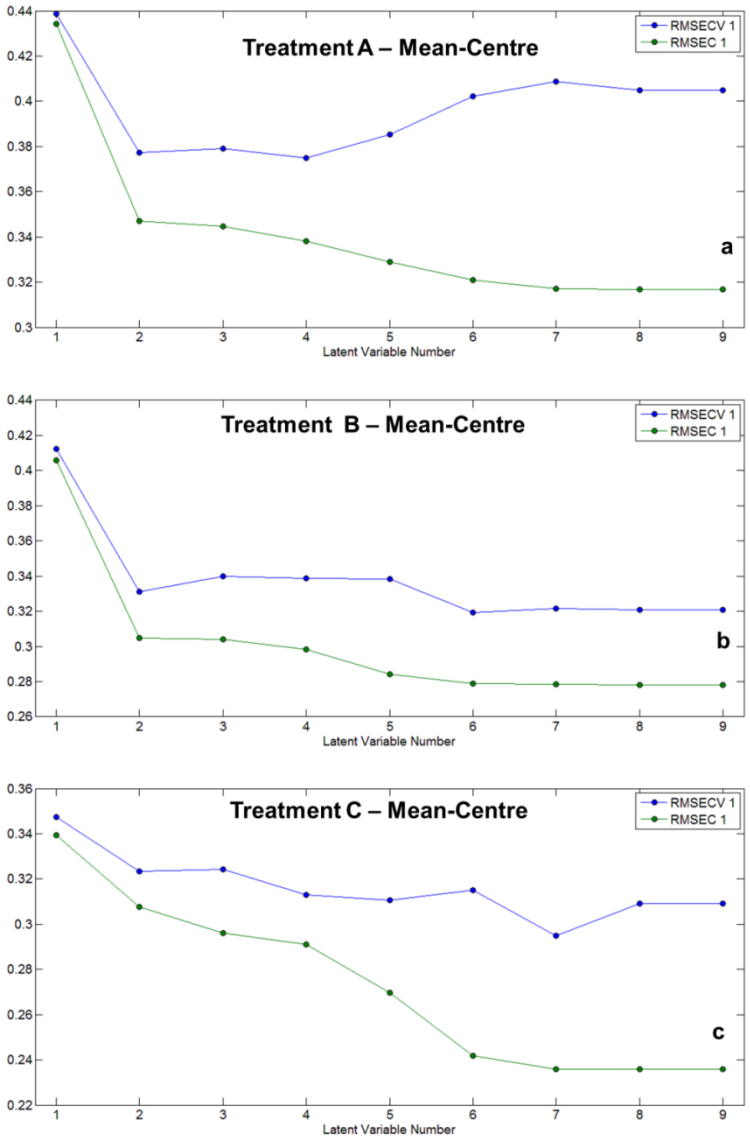

Figure 3.17 – Cross validation plots of the PLS1 models to classify each Treatment (A, B, C), with "Mean-Centre" data pre-processing. The RMSECV and RMSEC are plotted against each latent variable added to the models.

The PLS1 scores and loadings from the discrimination of the treatment “A” samples from the other treatments are represented in **Figure 3.18**, **Figure 3.19** and **Figure 3.20**, using the “Autoscale” and “Mean-Centre” data pre-processing methods.

Although the clustering in the scores plots is not very clear in any case (red dots on the figures), it is noticeable that there may be some cluster separation when considering the three score plots simultaneously (combination of three LV) from the “Autoscale” pre-processed data (**Figure 3.18**). However, no evident cluster separation was observed in the score plot with just two LV from the “Mean-Centre” data pre-processing method (**Figure 3.20**).

The loadings plots using “Autoscale” data pre-processing (**Figure 3.19**) show that all variables seem to have impact on the model. The variables P2 and P20 seem to appear close to the Treatment “A” cluster in all the LV combination score plots. Therefore, these variables may be responsible for the classification of this treatment.

In the loadings plots using “Mean-Centre” data pre-processing (**Figure 3.20b**) it is noticeable that only a few variables seem to affect the model. However, in this case it was also verified that the variables P2 and P20 are closer to the Treatment “A” cluster.

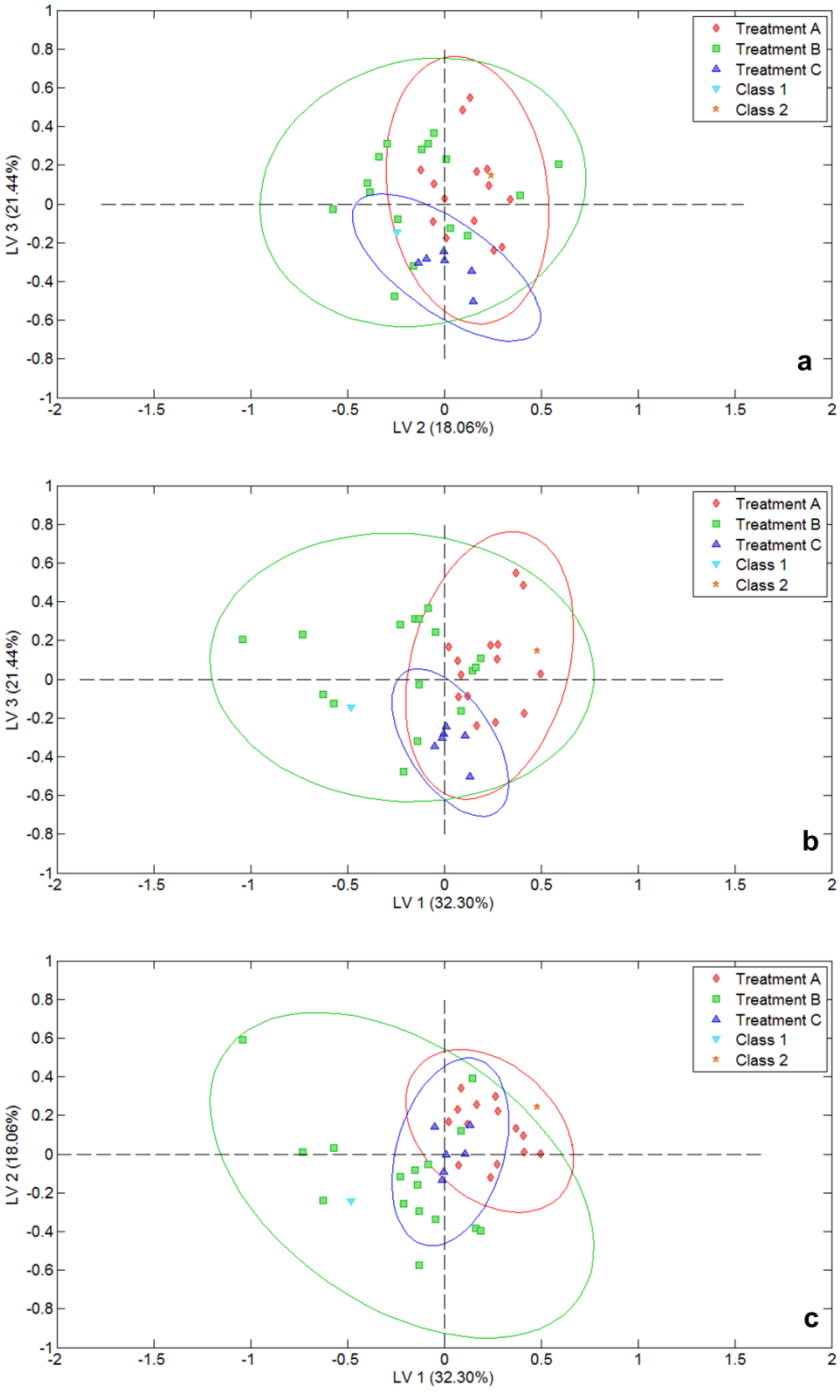

Figure 3.18 – PLS1 scores of the Red Globe samples discriminating the treatment A, with “Autoscale” data pre-treatment, considering the first three latent variables.

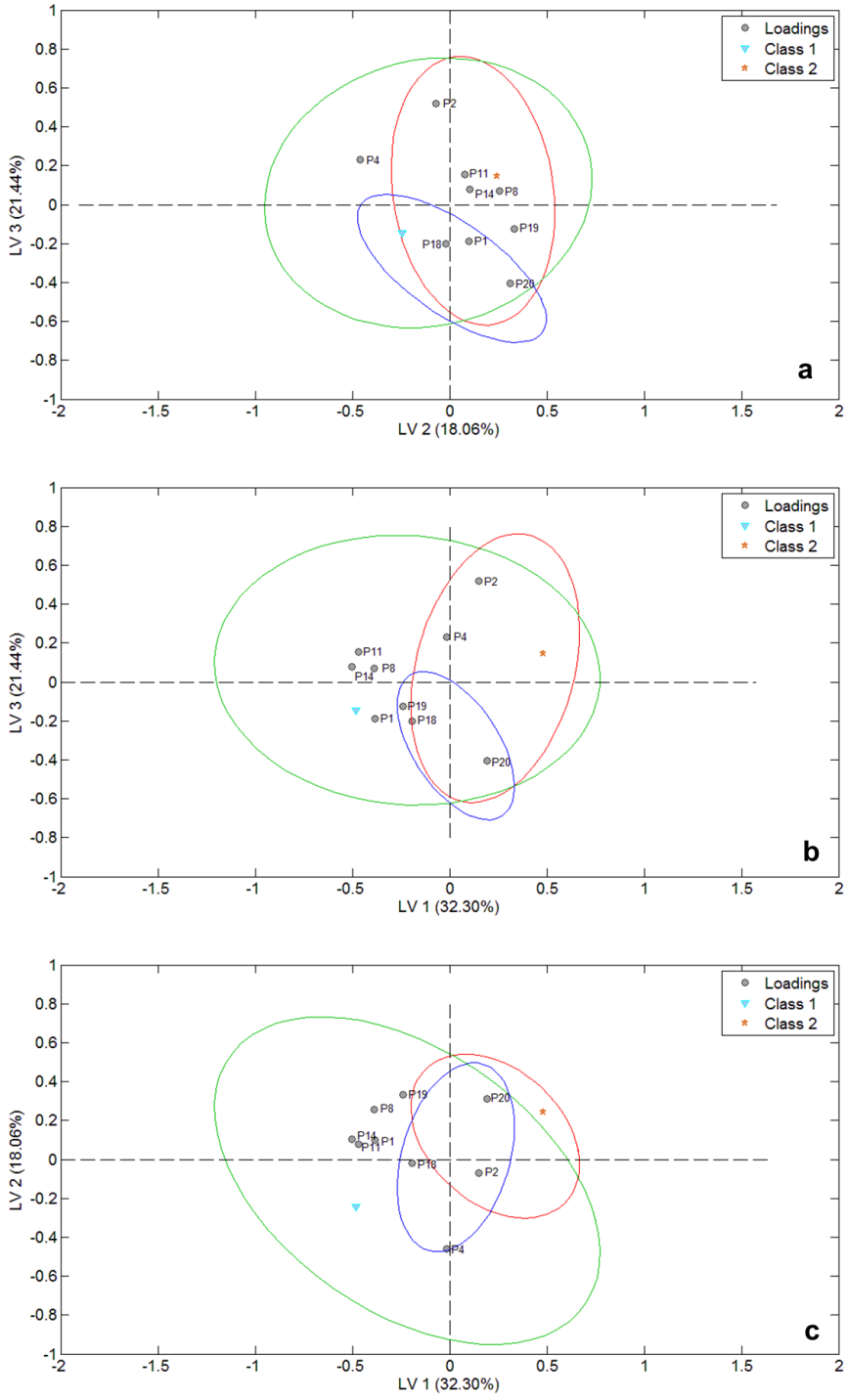

Figure 3.19 – PLS1 loadings of the Red Globe samples discriminating the treatment A, with “Autoscale” data pre-treatment, considering the first three latent variables.

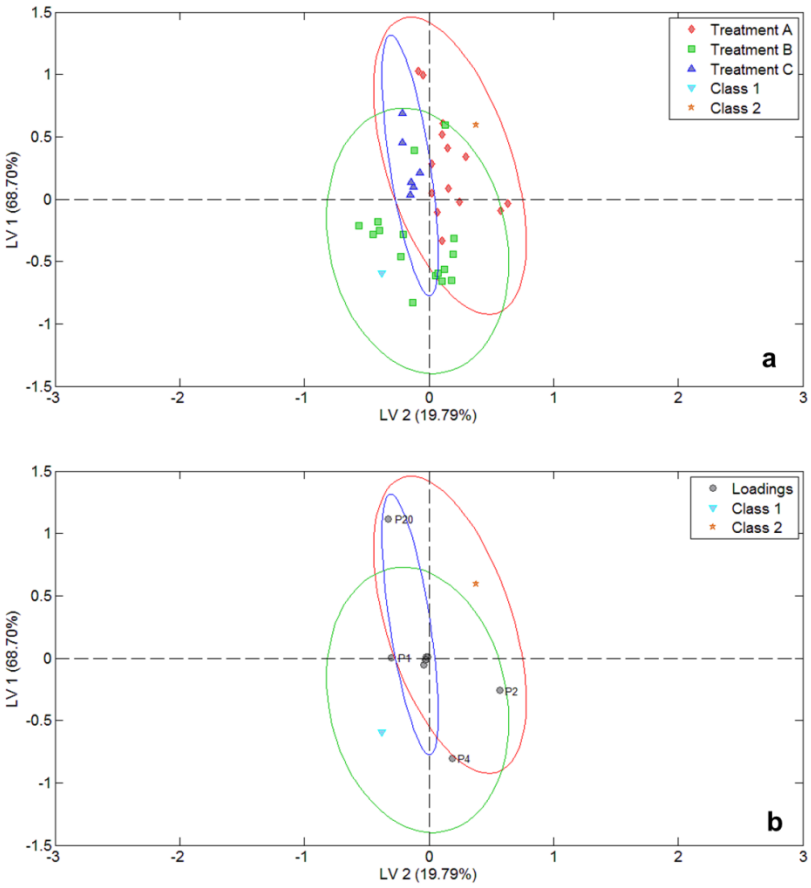

Figure 3.20 – PLS1 scores (**a**) and loadings (**b**) plots of the Red Globe samples discriminating the treatment A, with "Mean-Centre" data pre-treatment, considering the first two latent variables.

The PLS1 scores and loadings from the discrimination of the treatment "B" samples, using the "Autoscale" and "Mean-Centre" data pre-processing methods are represented in **Figure 3.21** and **Figure 3.22**.

A good cluster separation was observed with both pre-processing methods. The treatment "B" samples (green squares) seem to be clustered and well separated from the rest. However, two of these samples appeared inside the other clusters in both cases. There samples are replicates (Red Globe B-14-1 and B-14-2) and may have been wrongly labelled. Nevertheless, the clustering observed is quite satisfactory as regards visual interpretation.

The loadings plots (**Figure 3.21b** and **Figure 3.22b**) from both data pre-processing methods show that the variables P1 and P4 are closer to the treatment “B” cluster centre than the other variables. This is a sign that these variables have more weight on the classification of this treatment.

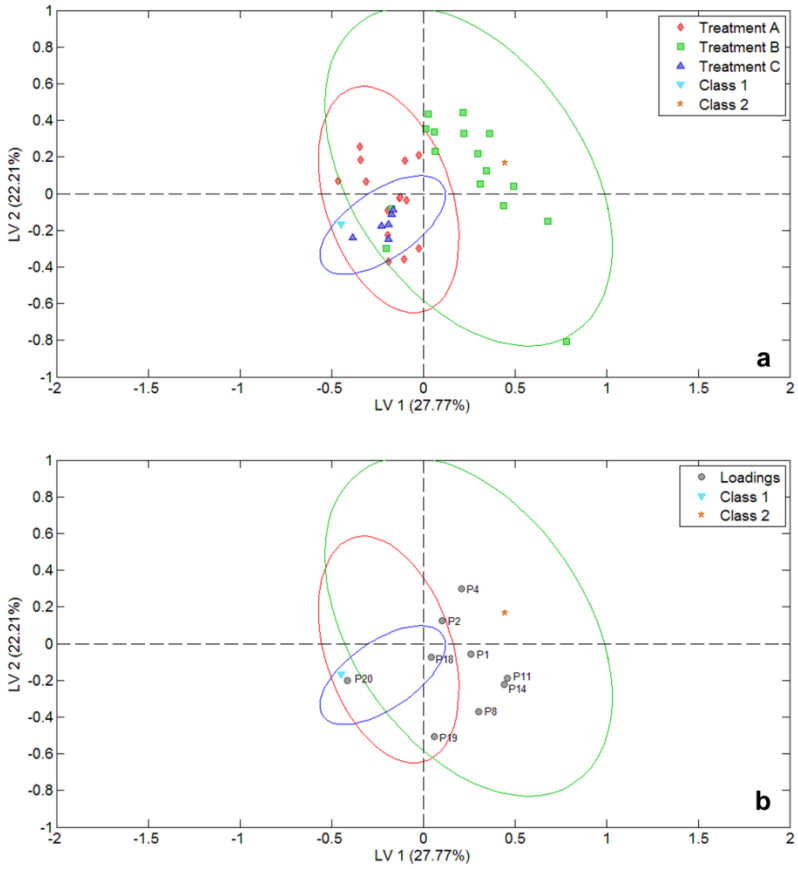

Figure 3.21 – PLS1 scores (**a**) and loadings (**b**) plots of the Red Globe samples discriminating the treatment B, with “Autoscale” data pre-treatment, considering the first two latent variables.

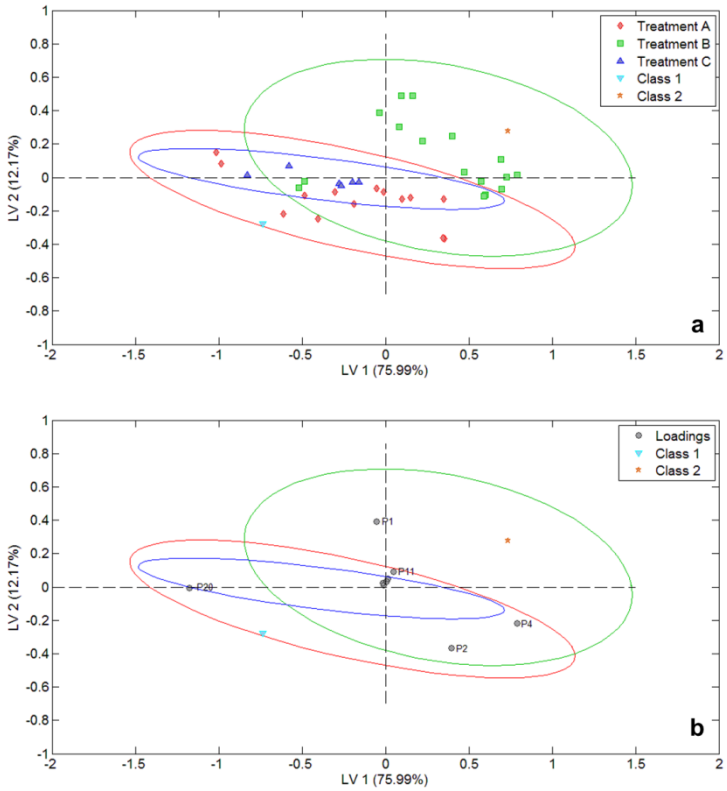

Figure 3.22 – PLS1 scores (**a**) and loadings (**b**) plots of the Red Globe samples discriminating the treatment B, with "Mean-Centre" data pre-treatment, considering the first two latent variables.

The PLS1 scores and loadings from the discrimination of the treatment "C" samples from the other treatments are represented in **Figure 3.23**, **Figure 3.24** and **Figure 3.25**, using the "Autoscale" and "Mean-Centre" data pre-processing methods.

The cluster separation (blue triangles in the figures) observed in the scores plots from the model with "Autoscale" data pre-processing is more evident when combining LV1/LV2 and LV1/LV3 (**Figure 3.23b** and **Figure 3.23c**). However, with "Mean Centre" data pre-processing, the clustering observed was not satisfactory (**Figure 3.25a**).

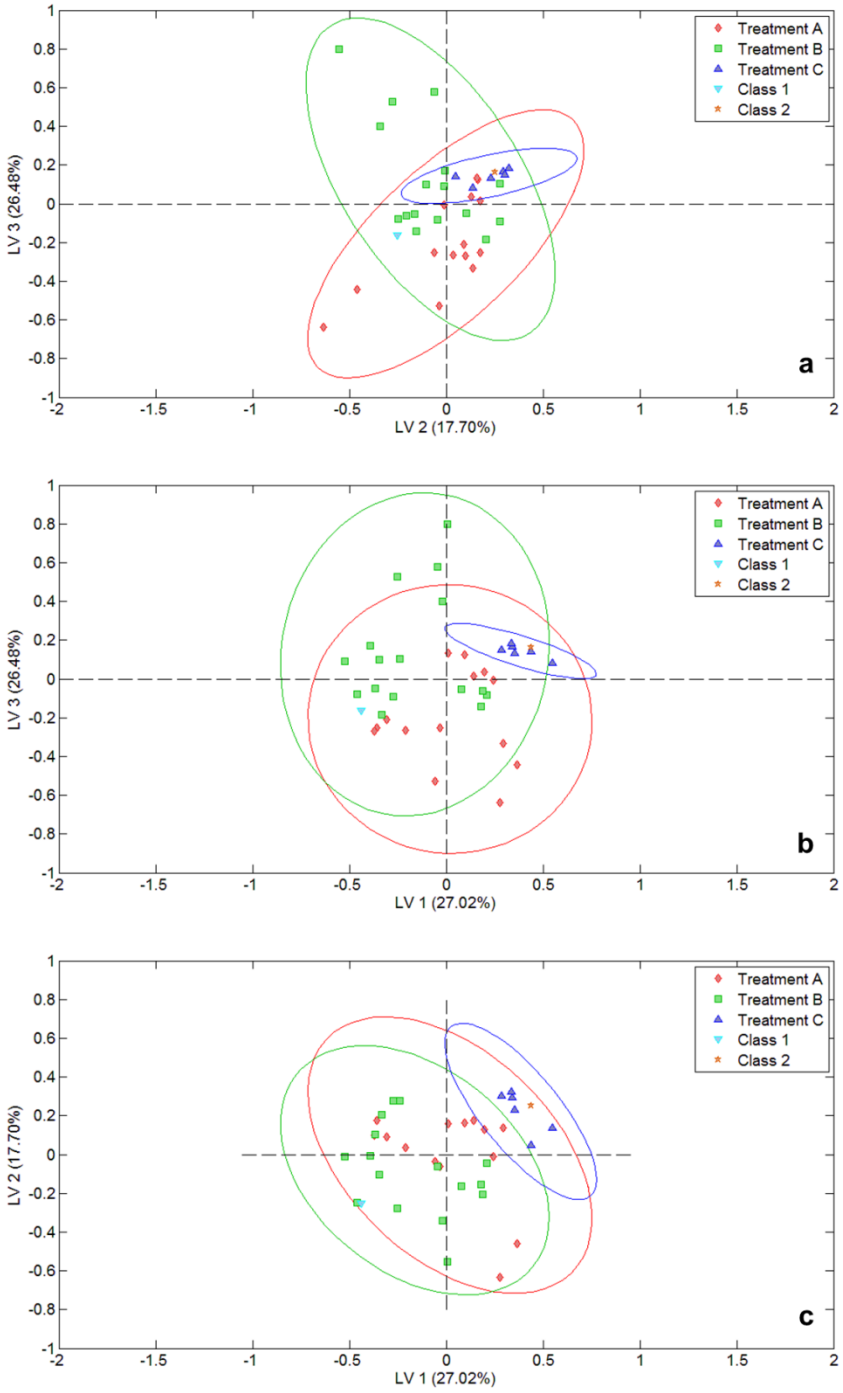

Figure 3.23 – PLS1 scores of the Red Globe samples discriminating the treatment C, with "Autoscale" data pre-treatment, considering the first three latent variables.

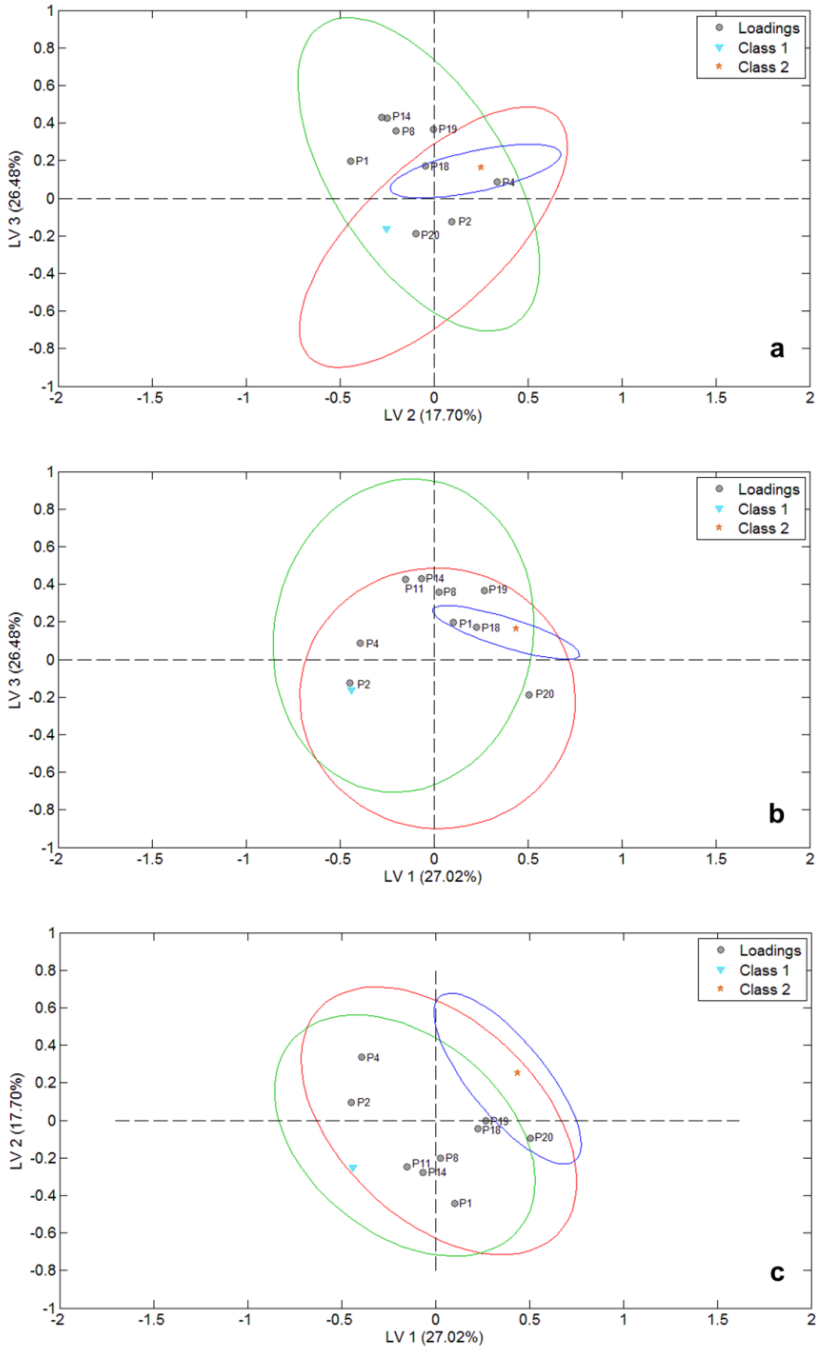

Figure 3.24 – PLS1 loadings of the Red Globe samples discriminating the treatment C, with “Autoscale” data pre-treatment, considering the first three latent variables.

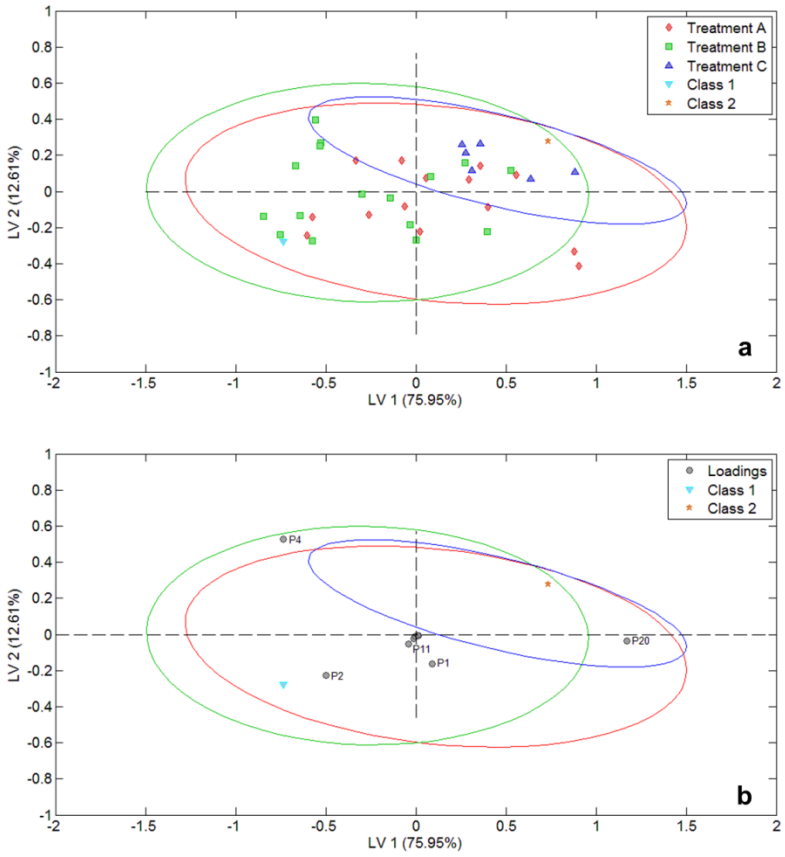

Figure 3.25 – PLS1 loadings of the Red Globe samples discriminating the treatment C, with "Mean-Centre" data pre-treatment, considering the first two latent variables.

As regards the variables of more importance in the classification of the treatment "C" samples, P20 appears close to the cluster centre when using both pre-treatment methods (**Figure 3.24** and **Figure 3.25b**). This suggests that this variable is the most responsible for the classification of the samples with treatment "C". However, because "Autoscale" makes all variables more influent in the model, the less intense variables P18 and P19 seem to have also contributed for the classification with this is data pre-processing method (**Figure 3.24**).

The classification results (**Table 4)** show that similar results were obtained from data pre-processing methods. However, the “Autoscale” method resulted in a very good discrimination between the treatment “C” and the rest of the treatments. Although some trends of discrimination were also observed for the treatments “A” and “B”, the results were not as satisfactory as for the treatment C, especially because of the misclassifications observed relatively to the number of samples analysed.

Table 4 – Classification results of the PLS1 models for the discrimination of the “Red Globe” samples according to the grape treatments (A, B, C).

				Calibration			Cross-Validation		
	Treatment	Samples	LV	TP[a]	TN[b]	Miss.[c]	TP[a]	TN[b]	Miss.[c]
Autoscale data pre-processing	A	14	3	92.86	90.91	1	92.86	81.82	1
	B+C	22		90.91	92.86	2	81.82	92.86	4
	B	16	2	87.50	95.00	2	87.50	95.00	2
	A+C	20		95.00	87.50	1	95.00	87.50	1
	C	6	3	100.00	100.00	0	100.00	100.00	0
	A+B	30		100.00	100.00	0	100.00	100.00	0
Mean-Centre data pre-processing	A	14	2	78.57	86.36	3	78.57	81.82	3
	B+C	22		86.36	78.57	3	81.82	78.57	4
	B	16	2	87.50	95.00	2	87.50	95.00	2
	A+C	20		95.00	87.50	1	95.00	87.50	1
	C	6	2	100.00	86.67	0	100.00	86.67	0
	A+B	30		86.67	100.00	0	86.67	100.00	0

[a] True Positives (%) (Selectivity)
[b] True Negatives (%) (Specificity)
[c] Misclassifications

Better results would have been achieved if more samples had been analysed, especially from the Giant Rose and Summer Black varieties, where such a few number of samples were not enough to create an efficient model for an eventual supervised classification of unknown samples as regards the treatments. The study of on the “Red Globe” variety revealed that the obtained fingerprints can be used to discriminate the treatment “C”, and some trends were also observed as regards the other treatments, however, with the data

provided nothing can be concluded about treatments. Part of the misclassifications may be due to the replicates (Red Globe B-14-1 and B-14-2), which were observed far from the Treatment “B” cluster centre. The removal of these samples from the dataset could improve the classification results. Nevertheless, the obtained results already demonstrate that the method employed in this research may eventually classify grape samples studied according to the pre-harvest treatments.

4. CONCLUSIONS

The results of this experimental work revealed that the analytical method employed in the analysis of the three varieties grapes (Giant Rose, Red Globe and Summer Black) has potential to be applied in quality control of these samples. However, more work is necessary to develop a suitable and consistent method, especially, when a significantly larger amount of samples should have been studied in order to obtain suitable classification models.

Despite the low number of samples analysed, some trends were observed, which may be a foresight for a successful application of this method in quality control processes. From the information obtained by PCA it is clear that the chromatographic fingerprinting data obtained can be decomposed into the three varieties of grapes studied. The PLS-DA (PLS2) model also classified all the samples according to the grape variety, and good clustering was also verified. These models may be useful if, in practice, one desires to analyse unknown grape samples to determine their varieties. However, by means of PCA and PLS-DA (PLS2) nothing could be concluded as regards the treatments applied to the grapevines.

Due to the low number of samples, the study of the treatments by the means of PLS-DA (PLS1) could only be performed on the “Red Globe” variety. These results indicate that it is possible to create a model for discrimination of the samples from this variety according to the treatment “C”. As regards the treatments “A” and “B”, some trends were observed that suggest the possibility of discrimination. However, the results were not as good as for the treatment “C”.

As regards the data pre-processing methods studied, apparently “Autoscale” produced better PCA and PLS-DA (PLS2) models when all the varieties were discriminated simultaneously. When classifying the “Red Grape” variety according to the treatments (A, B and C), it was verified that the “Autoscale” data pre-processing method produced better classification results. This means that the some low intense peaks contributed positively for the classification.

5. COMMENTS AND FUTURE WORK

Some issues and suggestions about this work and its followings are presented in this section.

As regards sampling, it was performed elsewhere by another laboratory. The missing information about the nature of the treatments applied on the grapevines makes this work rather meaningless as regards the practical understanding of the analytical problem. Therefore, this information has should be provided.

The number of samples provided was not sufficient to produce suitable classification models, which could discriminate the grape samples according to the three grapevine treatments applied on the three varieties studied. Only one grape variety could be studied as regards treatments. Therefore, more samples should have been collected for this study.

Considering the results and a possible improvement in the quality of the chromatographic fingerprint models, it would be advisable to acquire, at least,

batches of 20 samples, from grapevines subjected to each treatment grape variety and subjected to each of the three treatments. This would result in a total of 180 samples.

To improve the chromatographic results and the quality of the models, samples should have been pre-concentrated before the GC-MS analysis. This might have improved the quality of chromatographic hyphenated data, minimizing the variability effects caused by instrumental noise or column bleeding on very low intense peaks.

The identification of the compounds by means of spectra library comparison should have been confirmed with analyses of respective standards.

These suggestions represent a rather more time consuming method, especially in the sampling and extraction processes. The time available to perform this work in China was only three months. The fact that the sampling and extraction was already performed saved plenty of time. Otherwise, it would be virtually impossible to perform the whole analytical process in such a short time. Nevertheless, this work was rather gratifying in the sense than it contributed for a future research in this matter.

6. ANNEXES

Table A. Matrix with the areas of the peaks of the compounds (columns) for every sample and replicate (rows).

SAMPLES	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12	Peak 13	Peak 14	Peak 15	Peak 16	Peak 17	Peak 18	Peak 19	Peak 20
Giant A40-4.5.6-1	1.41E-02	2.23E-01	4.92E-02	4.42E-01	4.73E-03	5.64E-03	0	1.27E-02	0	2.92E-02	1.45E-02	4.62E-02	7.39E-03	8.86E-03	8.58E-03	1.19E-02	1.27E-02	4.80E-03	3.59E-03	1.01E-01
Giant A40-4.5.6-2	7.62E-03	2.19E-01	4.13E-02	4.53E-01	1.04E-02	2.01E-03	0	1.28E-02	0	3.04E-02	1.53E-02	4.16E-02	6.77E-03	5.19E-03	7.66E-03	1.54E-02	9.66E-03	3.67E-03	4.55E-03	1.13E-01
Giant A40-12.13.14-1	9.43E-03	2.18E-01	3.25E-02	4.16E-01	1.10E-02	1.93E-03	0	1.01E-02	0	3.10E-02	1.28E-02	3.74E-02	7.01E-03	5.81E-03	7.24E-03	9.77E-03	8.38E-03	5.72E-03	6.30E-03	1.70E-01
Giant A40-12.13.14-2	1.30E-02	2.23E-01	4.10E-02	4.21E-01	1.07E-02	1.90E-03	0	4.40E-03	0	2.22E-02	1.12E-02	3.54E-02	6.00E-03	8.51E-03	7.33E-03	9.46E-03	2.76E-03	6.38E-03	5.87E-03	1.70E-01
Giant A40-16.17.18-1	5.63E-03	2.04E-01	4.46E-02	4.38E-01	1.38E-02	2.23E-03	0	9.66E-03	0	3.17E-02	1.16E-02	4.36E-02	5.64E-03	7.15E-03	8.70E-03	9.52E-03	4.15E-03	5.31E-03	4.83E-03	1.49E-01
Giant A40-16.17.18-2	6.18E-03	2.06E-01	4.04E-02	4.42E-01	1.24E-02	2.35E-03	0	1.00E-02	0	2.93E-02	1.34E-02	4.22E-02	4.29E-03	6.48E-03	7.11E-03	8.44E-03	1.02E-02	5.94E-03	5.15E-03	1.52E-01
Giant B39-4.5.6-2	1.17E-02	2.29E-01	5.44E-02	4.51E-01	1.03E-02	1.55E-03	0	9.49E-03	0	1.25E-02	8.98E-03	2.99E-02	6.43E-03	4.27E-03	1.03E-02	7.99E-03	9.61E-03	4.99E-03	5.38E-03	1.33E-01
Giant B39-4.5.6-3	1.18E-02	2.23E-01	5.47E-02	4.56E-01	9.33E-03	1.49E-03	0	9.14E-03	0	1.28E-02	1.07E-02	2.91E-02	6.49E-03	7.60E-03	9.93E-03	8.07E-03	4.68E-03	5.79E-03	4.92E-03	1.35E-01
Giant B39-10.11.12-1	1.70E-02	2.19E-01	5.00E-02	3.50E-01	3.41E-03	1.77E-03	0	1.02E-02	0	2.08E-02	1.20E-02	4.10E-02	6.90E-03	7.94E-03	7.27E-03	9.58E-03	1.49E-02	1.19E-02	6.84E-03	2.14E-01
Giant B39-10.11.12-2	6.37E-03	2.08E-01	5.17E-02	3.75E-01	1.01E-02	2.10E-03	0	9.53E-03	0	1.98E-02	1.00E-02	3.63E-02	7.38E-03	8.50E-03	7.78E-03	9.58E-03	4.02E-03	8.85E-03	5.12E-03	2.19E-01
Giant CK40-8.9.10-2	6.63E-03	8.82E-04	4.21E-02	5.43E-01	1.12E-02	2.53E-03	0	1.20E-02	0	5.52E-02	1.33E-02	5.63E-02	6.61E-03	8.09E-03	9.40E-03	1.01E-02	6.63E-03	5.45E-03	4.88E-03	2.06E-01
Giant CK40-8.9.10-3	6.50E-03	8.64E-04	4.12E-02	5.32E-01	1.10E-02	5.93E-03	0	1.09E-02	0	6.52E-02	1.20E-02	5.45E-02	6.47E-03	7.93E-03	7.45E-03	8.89E-03	5.70E-03	6.68E-03	4.74E-03	2.12E-01
Red A3-1	7.36E-02	2.28E-01	0	4.63E-01	0	0	0	5.01E-03	0	0	5.02E-03	0	0	5.28E-03	0	0	0	1.38E-02	5.32E-03	2.01E-01
Red A3-2	7.39E-02	2.58E-01	0	4.31E-01	0	0	0	4.64E-03	0	0	5.11E-03	0	0	5.08E-03	0	0	0	9.57E-03	5.12E-03	2.08E-01
Red A4-2	7.15E-02	2.20E-01	0	4.45E-01	0	0	0	5.83E-03	0	0	6.95E-03	0	0	6.21E-03	0	0	0	9.00E-03	6.63E-03	2.29E-01
Red A4-3	7.14E-02	2.26E-01	0	4.18E-01	0	0	0	7.74E-03	0	0	7.87E-03	0	0	7.13E-03	0	0	0	1.18E-02	6.51E-03	2.44E-01
Red A7-2	5.05E-02	2.95E-01	0	4.83E-01	0	0	0	4.20E-03	0	0	4.13E-03	0	0	4.15E-03	0	0	0	6.83E-03	4.62E-03	1.47E-01
Red A7-3	5.28E-02	3.03E-01	0	4.72E-01	0	0	0	4.51E-03	0	0	4.75E-03	0	0	4.74E-03	0	0	0	9.16E-03	4.32E-03	1.44E-01
Red A9-2	7.37E-02	2.45E-01	0	4.52E-01	0	0	0	7.33E-03	0	0	9.31E-03	0	0	4.54E-03	0	0	0	9.80E-03	7.83E-03	1.91E-01
Red A9-3	7.21E-02	2.60E-01	0	4.60E-01	0	0	0	7.03E-03	0	0	4.57E-03	0	0	7.79E-03	0	0	0	1.15E-02	7.42E-03	1.70E-01
Red A14-1	1.15E-01	2.10E-01	0	3.48E-01	0	0	0	2.97E-03	0	0	3.46E-03	0	0	5.01E-03	0	0	0	1.26E-02	2.80E-03	3.01E-01
Red A14-2	1.03E-01	2.09E-01	0	3.57E-01	0	0	0	3.36E-03	0	0	2.93E-03	0	0	7.35E-03	0	0	0	1.61E-02	3.32E-03	2.98E-01
Red A15-1	5.87E-02	2.26E-01	0	4.53E-01	0	0	0	2.55E-03	0	0	2.45E-03	0	0	4.10E-03	0	0	0	5.60E-03	1.93E-03	2.45E-01
Red A15-2	6.10E-02	2.18E-01	0	4.31E-01	0	0	0	2.52E-03	0	0	2.21E-03	0	0	2.43E-03	0	0	0	1.41E-02	3.44E-03	2.66E-01
Red A21-2-1	6.76E-02	2.44E-01	0	5.05E-01	0	0	0	4.97E-03	0	0	5.35E-03	0	0	4.52E-03	0	0	0	3.38E-03	3.84E-03	1.60E-01
Red A21-2-2	6.77E-02	2.34E-01	0	4.88E-01	0	0	0	4.79E-03	0	0	5.16E-03	0	0	4.84E-03	0	0	0	3.26E-03	3.71E-03	1.89E-01
Red B2-1	7.09E-02	2.27E-01	0	5.46E-01	0	0	0	4.63E-03	0	0	5.43E-03	0	0	7.20E-03	0	0	0	1.33E-02	2.60E-03	1.23E-01
Red B2-2	7.00E-02	2.55E-01	0	5.23E-01	0	0	0	5.94E-03	0	0	6.12E-03	0	0	6.89E-03	0	0	0	1.27E-02	2.49E-03	1.18E-01
Red B14-1	7.68E-02	2.01E-01	0	4.35E-01	0	0	0	5.90E-03	0	0	7.09E-03	0	0	7.46E-03	0	0	0	9.24E-03	4.98E-03	2.53E-01
Red B14-2	8.18E-02	2.33E-01	0	4.03E-01	0	0	0	6.71E-03	0	0	7.08E-03	0	0	6.76E-03	0	0	0	9.82E-03	6.34E-03	2.45E-01
Red B22-3-2	1.22E-01	1.77E-01	0	4.62E-01	0	0	0	6.10E-03	0	0	1.17E-03	0	0	7.02E-03	0	0	0	1.17E-02	4.47E-03	2.08E-01
Red B22-3-3	1.16E-01	1.98E-01	0	4.66E-01	0	0	0	6.26E-03	0	0	1.04E-03	0	0	7.09E-03	0	0	0	8.97E-03	5.34E-03	1.92E-01
Red B22-4-1	1.23E-01	2.04E-01	0	4.17E-01	0	0	0	1.82E-02	0	0	2.68E-02	0	0	1.94E-02	0	0	0	1.00E-02	1.01E-02	1.72E-01
Red B22-4-2	1.29E-01	2.03E-01	0	4.36E-01	0	0	0	7.26E-03	0	0	2.83E-02	0	0	7.31E-03	0	0	0	1.06E-02	6.85E-03	1.72E-01
Red B22-7-2	7.94E-02	2.69E-01	0	4.87E-01	0	0	0	4.36E-03	0	0	1.37E-02	0	0	1.07E-02	0	0	0	8.49E-03	5.11E-03	1.23E-01
Red B22-7-3	9.01E-02	2.72E-01	0	4.66E-01	0	0	0	4.00E-03	0	0	1.57E-02	0	0	8.26E-03	0	0	0	9.34E-03	3.35E-03	1.32E-01
Red B22-8-2	6.77E-02	2.44E-01	0	5.31E-01	0	0	0	3.54E-03	0	0	3.50E-03	0	0	3.83E-03	0	0	0	7.82E-03	3.57E-03	1.35E-01
Red B22-8-3	6.71E-02	2.46E-01	0	5.29E-01	0	0	0	3.46E-03	0	0	3.87E-03	0	0	3.64E-03	0	0	0	8.21E-03	2.55E-03	1.35E-01
Red B22-9-1	8.42E-02	2.75E-01	0	5.02E-01	0	0	0	1.89E-03	0	0	1.61E-02	0	0	8.49E-03	0	0	0	9.63E-03	3.54E-03	9.93E-02
Red B22-9-2	9.66E-02	2.71E-01	0	4.82E-01	0	0	0	6.13E-03	0	0	1.08E-02	0	0	8.16E-03	0	0	0	1.33E-02	4.88E-03	1.07E-01
Red B22-13-1	9.39E-02	2.20E-01	0	4.71E-01	0	0	0	1.06E-02	0	0	2.56E-02	0	0	1.63E-02	0	0	0	1.28E-02	5.01E-03	1.45E-01
Red B22-13-2	9.40E-02	2.18E-01	0	4.58E-01	0	0	0	3.33E-03	0	0	2.38E-02	0	0	1.58E-02	0	0	0	1.75E-02	4.87E-03	1.66E-01
Red CK2-1	8.63E-02	1.91E-01	0	4.20E-01	0	0	0	5.78E-03	0	0	7.55E-03	0	0	7.83E-03	0	0	0	1.25E-02	6.99E-03	2.62E-01
Red CK2-2	8.29E-02	1.83E-01	0	4.09E-01	0	0	0	2.68E-03	0	0	6.02E-03	0	0	6.92E-03	0	0	0	9.25E-03	7.55E-03	2.92E-01
Red CK3-1	7.46E-02	2.11E-01	0	4.49E-01	0	0	0	5.53E-03	0	0	5.23E-03	0	0	5.87E-03	0	0	0	1.57E-02	6.36E-03	2.28E-01
Red CK3-2	7.26E-02	2.05E-01	0	4.63E-01	0	0	0	5.38E-03	0	0	5.09E-03	0	0	5.71E-03	0	0	0	1.53E-02	6.19E-03	2.22E-01
Red CK4-2	7.16E-02	2.02E-01	0	4.70E-01	0	0	0	5.31E-03	0	0	5.02E-03	0	0	5.63E-03	0	0	0	1.51E-02	6.10E-03	2.19E-01
Red CK4-3	7.12E-02	2.00E-01	0	4.65E-01	0	0	0	4.72E-03	0	0	6.20E-03	0	0	5.58E-03	0	0	0	7.52E-03	7.24E-03	2.32E-01
Summer A38-4.5.6-1	2.55E-03	2.32E-01	1.30E-02	4.51E-01	0	4.35E-03	3.46E-03	7.45E-03	7.15E-03	0.00480632	7.93E-03	1.10E-01	4.46E-03	6.50E-03	7.45E-03	3.40E-03	4.00E-03	4.45E-03	3.39E-03	1.24E-01
Summer A38-4.5.6-2	5.42E-03	2.42E-01	1.34E-02	4.40E-01	0	4.30E-03	3.62E-03	9.12E-03	6.46E-03	5.20E-03	8.18E-03	1.11E-01	4.35E-03	5.75E-03	6.37E-03	3.52E-03	4.14E-03	3.53E-03	3.16E-03	1.20E-01
Summer A38-10.11.12-1	2.85E-03	2.20E-01	2.24E-03	4.08E-01	0	4.31E-04	3.52E-03	1.14E-02	4.07E-03	5.73E-03	8.40E-03	1.48E-01	4.04E-03	4.49E-03	5.91E-03	9.58E-03	4.68E-03	4.76E-03	2.93E-03	1.48E-01
Summer A38-10.11.12-2	2.71E-03	2.35E-01	2.13E-03	3.88E-01	0	4.10E-04	3.50E-03	8.68E-03	4.38E-03	6.85E-03	8.43E-03	1.49E-01	3.84E-03	4.27E-03	5.62E-03	1.26E-02	3.44E-03	3.49E-03	2.78E-03	1.55E-01
Summer B37-5.6.7-2	3.44E-03	2.63E-01	2.76E-03	4.16E-01	0	4.06E-03	3.23E-03	8.53E-03	5.39E-03	5.54E-03	1.25E-01	4.11E-03	6.34E-03	6.88E-03	4.61E-03	5.03E-03	2.41E-03	1.92E-03	1.24E-01
Summer B37-5.6.7-3	2.64E-03	2.52E-01	2.03E-03	4.28E-01	0	3.83E-03	3.25E-03	8.10E-03	5.54E-03	5.70E-03	8.33E-03	1.07E-01	4.23E-03	4.70E-03	7.08E-03	6.23E-03	7.62E-03	6.10E-03	2.80E-03	1.34E-01
Summer B37-9.10.11-1	9.96E-03	2.30E-01	4.96E-03	4.10E-01	0	5.61E-03	5.02E-03	8.94E-03	9.58E-03	8.20E-03	1.60E-01	4.05E-03	6.67E-03	6.78E-03	1.11E-02	9.47E-03	6.04E-03	5.70E-03	9.14E-02
Summer B37-9.10.11-2	4.33E-03	2.20E-01	5.25E-03	4.34E-01	0	8.12E-03	2.88E-03	8.72E-03	8.65E-03	7.01E-03	9.03E-03	1.52E-01	4.29E-03	7.62E-03	7.18E-03	8.00E-03	8.82E-03	8.54E-03	4.41E-03	9.04E-02
Summer B37-13.14.15-1	8.42E-03	2.29E-01	1.03E-02	4.43E-01	0	4.61E-03	3.72E-03	6.67E-03	6.44E-03	4.85E-03	1.26E-03	1.15E-01	4.38E-03	4.54E-03	7.32E-03	5.59E-03	2.63E-03	4.47E-03	4.28E-03	1.33E-01
Summer B37-13.14.15-2	9.00E-03	2.37E-01	9.75E-03	4.21E-01	0	5.03E-03	3.76E-03	5.85E-03	6.94E-03	4.61E-03	1.20E-03	1.15E-01	4.16E-03	4.40E-03	6.96E-03	9.58E-03	2.48E-03	3.05E-03	4.13E-03	1.46E-01
Summer CK-6.7.8-1	1.05E-02	2.38E-01	3.37E-03	4.38E-01	0	6.88E-03	3.19E-03	7.70E-03	6.59E-03	3.44E-03	2.73E-03	1.13E-01	4.33E-03	3.38E-03	7.24E-03	1.10E-02	5.02E-03	2.44E-03	4.01E-03	1.29E-01
Summer CK-6.7.8-2	4.38E-03	2.44E-01	4.14E-03	4.39E-01	0	5.48E-03	3.38E-03	8.21E-03	6.41E-03	3.30E-03	3.15E-03	1.04E-01	4.35E-03	4.45E-03	7.27E-03	6.90E-03	4.88E-03	2.80E-03	4.10E-03	1.39E-01

Table B. Matrix with the normalized data.

SAMPLES	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12	Peak 13	Peak 14	Peak 15	Peak 16	Peak 17	Peak 18	Peak 19	Peak 20
Giant A40-4.5.6-1	1.41E-02	2.23E-01	4.92E-02	4.42E-01	4.73E-03	5.64E-03	0	1.27E-02	0	2.92E-02	1.45E-02	4.62E-02	7.39E-03	8.86E-03	8.58E-03	1.19E-02	1.27E-02	4.80E-03	3.59E-03	1.01E-01
Giant A40-4.5.6-2	7.62E-03	2.19E-01	4.13E-02	4.53E-01	1.04E-02	2.01E-03	0	1.28E-02	0	3.04E-02	1.53E-02	4.16E-02	6.77E-03	5.19E-03	7.66E-03	1.54E-02	9.66E-03	3.67E-03	4.55E-03	1.13E-01
Giant A40-12.13.14-1	9.43E-03	2.18E-01	3.25E-02	4.16E-01	1.10E-02	1.93E-03	0	1.01E-02	0	3.10E-02	1.28E-02	3.74E-02	7.01E-03	5.81E-03	7.24E-03	9.77E-03	8.38E-03	5.72E-03	6.30E-03	1.70E-01
Giant A40-12.13.14-2	1.30E-02	2.23E-01	4.10E-02	4.21E-01	1.07E-02	1.90E-03	0	4.40E-03	0	2.22E-02	1.12E-02	3.54E-02	6.00E-03	8.51E-03	7.33E-03	9.46E-03	2.76E-03	6.38E-03	5.87E-03	1.70E-01
Giant A40-16.17.18-1	5.63E-03	2.04E-01	4.46E-02	4.38E-01	1.38E-02	2.23E-03	0	9.66E-03	0	3.17E-02	1.16E-02	4.36E-02	5.64E-03	7.15E-03	8.70E-03	9.52E-03	4.15E-03	5.31E-03	4.83E-03	1.49E-01
Giant A40-16.17.18-2	6.18E-03	2.06E-01	4.04E-02	4.42E-01	1.24E-02	2.35E-03	0	1.00E-02	0	2.53E-02	1.34E-02	4.22E-02	4.29E-03	6.48E-03	7.11E-03	8.44E-03	1.02E-02	5.94E-03	5.15E-03	1.52E-01
Giant B39-4.5.6-2	1.17E-02	2.29E-01	5.44E-02	4.51E-01	1.03E-02	1.55E-03	0	9.49E-03	0	1.25E-02	8.98E-03	2.99E-02	6.43E-03	4.27E-03	1.03E-02	7.99E-03	9.61E-03	4.99E-03	5.38E-03	1.33E-01
Giant B39-4.5.6-3	1.18E-02	2.23E-01	5.47E-02	4.56E-01	9.33E-03	1.49E-03	0	9.14E-03	0	1.28E-02	1.07E-02	2.91E-02	6.49E-03	7.60E-03	9.93E-03	8.07E-03	4.68E-03	5.79E-03	4.92E-03	1.35E-01
Giant B39-10.11.12-1	1.70E-02	2.19E-01	5.00E-02	3.50E-01	3.41E-03	1.77E-03	0	1.02E-02	0	2.08E-02	1.20E-02	4.10E-02	6.90E-03	7.94E-03	7.27E-03	1.49E-02	1.19E-02	6.84E-03	5.30E-03	2.14E-01
Giant B39-10.11.12-2	6.37E-03	2.08E-01	5.17E-02	3.75E-01	1.01E-02	2.10E-03	0	9.53E-03	0	1.98E-02	1.00E-02	3.63E-02	7.38E-03	8.50E-03	7.78E-03	9.58E-03	4.02E-03	8.85E-03	5.12E-03	2.19E-01
Giant CK40-8.9.10-2	6.63E-03	8.82E-04	4.21E-02	5.43E-01	1.12E-02	2.53E-03	0	1.20E-02	0	5.52E-02	1.33E-02	5.63E-02	6.61E-03	8.09E-03	9.40E-03	1.01E-02	6.63E-03	5.45E-03	4.88E-03	2.06E-01
Giant CK40-8.9.10-3	6.50E-03	8.64E-04	4.12E-02	5.32E-01	1.10E-02	5.93E-03	0	1.09E-02	0	6.52E-02	1.20E-02	5.45E-02	6.47E-03	7.93E-03	7.45E-03	8.89E-03	5.70E-03	6.68E-03	4.74E-03	2.12E-01
Red A3-1	7.36E-02	2.28E-01	0	4.63E-01	0	0	0	5.01E-03	0	0	5.02E-03	0	0	5.28E-03	0	0	0	1.38E-02	5.32E-03	2.01E-01
Red A3-2	7.39E-02	2.58E-01	0	4.31E-01	0	0	0	4.64E-03	0	0	5.11E-03	0	0	5.08E-03	0	0	0	9.57E-03	5.12E-03	2.08E-01
Red A4-2	7.15E-02	2.20E-01	0	4.45E-01	0	0	0	5.83E-03	0	0	6.95E-03	0	0	6.21E-03	0	0	0	9.00E-03	6.63E-03	2.29E-01
Red A4-3	7.14E-02	2.26E-01	0	4.18E-01	0	0	0	7.74E-03	0	0	7.87E-03	0	0	7.13E-03	0	0	0	1.18E-02	6.51E-03	2.44E-01
Red A7-2	5.05E-02	2.95E-01	0	4.83E-01	0	0	0	4.20E-03	0	0	4.13E-03	0	0	4.15E-03	0	0	0	6.83E-03	4.62E-03	1.47E-01
Red A7-3	5.28E-02	3.03E-01	0	4.72E-01	0	0	0	4.51E-03	0	0	4.75E-03	0	0	4.74E-03	0	0	0	9.16E-03	4.32E-03	1.44E-01
Red A9-2	7.37E-02	2.45E-01	0	4.52E-01	0	0	0	7.33E-03	0	0	9.31E-03	0	0	4.54E-03	0	0	0	9.80E-03	7.83E-03	1.91E-01
Red A9-3	7.21E-02	2.60E-01	0	4.60E-01	0	0	0	7.03E-03	0	0	4.57E-03	0	0	7.79E-03	0	0	0	1.15E-02	7.42E-03	1.70E-01
Red A14-1	1.15E-01	2.10E-01	0	3.48E-01	0	0	0	2.97E-03	0	0	3.46E-03	0	0	5.01E-03	0	0	0	1.26E-02	2.80E-03	3.01E-01
Red A14-2	1.03E-01	2.09E-01	0	3.57E-01	0	0	0	3.36E-03	0	0	2.93E-03	0	0	7.35E-03	0	0	0	1.61E-02	3.32E-03	2.98E-01
Red A15-1	5.87E-02	2.26E-01	0	4.53E-01	0	0	0	2.55E-03	0	0	2.45E-03	0	0	4.10E-03	0	0	0	5.60E-03	1.93E-03	2.45E-01
Red A15-2	6.10E-02	2.18E-01	0	4.31E-01	0	0	0	2.52E-03	0	0	2.21E-03	0	0	2.43E-03	0	0	0	1.41E-02	3.44E-03	2.66E-01
Red A21-2-1	6.76E-02	2.44E-01	0	5.05E-01	0	0	0	4.97E-03	0	0	5.35E-03	0	0	4.52E-03	0	0	0	3.38E-03	3.84E-03	1.60E-01
Red A21-2-2	6.77E-02	2.34E-01	0	4.88E-01	0	0	0	4.79E-03	0	0	5.16E-03	0	0	4.84E-03	0	0	0	3.26E-03	3.71E-03	1.89E-01
Red B2-1	7.09E-02	2.27E-01	0	5.46E-01	0	0	0	4.63E-03	0	0	5.43E-03	0	0	7.20E-03	0	0	0	1.33E-02	2.60E-03	1.23E-01
Red B2-2	7.00E-02	2.55E-01	0	5.23E-01	0	0	0	5.94E-03	0	0	6.12E-03	0	0	6.89E-03	0	0	0	1.27E-02	2.49E-03	1.18E-01
Red B14-1	7.68E-02	2.01E-01	0	4.35E-01	0	0	0	5.90E-03	0	0	7.09E-03	0	0	7.46E-03	0	0	0	9.24E-03	4.98E-03	2.53E-01
Red B14-2	8.18E-02	2.33E-01	0	4.03E-01	0	0	0	6.71E-03	0	0	7.08E-03	0	0	6.76E-03	0	0	0	9.82E-03	6.34E-03	2.45E-01
Red B22-3-2	1.22E-01	1.77E-01	0	4.62E-01	0	0	0	6.10E-03	0	0	1.17E-03	0	0	7.02E-03	0	0	0	1.17E-02	4.47E-03	2.08E-01
Red B22-3-3	1.16E-01	1.98E-01	0	4.66E-01	0	0	0	6.26E-03	0	0	1.04E-03	0	0	7.09E-03	0	0	0	8.97E-03	5.34E-03	1.92E-01
Red B22-4-1	1.23E-01	2.04E-01	0	4.17E-01	0	0	0	1.82E-02	0	0	2.68E-02	0	0	1.94E-02	0	0	0	1.00E-02	1.01E-02	1.72E-01
Red B22-4-2	1.29E-01	2.03E-01	0	4.36E-01	0	0	0	7.26E-03	0	0	2.83E-02	0	0	7.31E-03	0	0	0	1.06E-02	6.85E-03	1.72E-01
Red B22-7-2	7.94E-02	2.69E-01	0	4.87E-01	0	0	0	4.36E-03	0	0	1.37E-02	0	0	1.07E-02	0	0	0	8.49E-03	5.11E-03	1.23E-01
Red B22-7-3	9.01E-02	2.72E-01	0	4.66E-01	0	0	0	4.00E-03	0	0	1.57E-02	0	0	8.26E-03	0	0	0	9.34E-03	3.35E-03	1.32E-01
Red B22-8-2	6.77E-02	2.44E-01	0	5.31E-01	0	0	0	3.54E-03	0	0	3.50E-03	0	0	3.83E-03	0	0	0	7.82E-03	3.57E-03	1.35E-01
Red B22-8-3	6.71E-02	2.46E-01	0	5.29E-01	0	0	0	3.46E-03	0	0	3.87E-03	0	0	3.64E-03	0	0	0	8.21E-03	2.55E-03	1.35E-01
Red B22-9-1	8.42E-02	2.75E-01	0	5.02E-01	0	0	0	1.89E-03	0	0	1.61E-02	0	0	8.49E-03	0	0	0	9.63E-03	3.54E-03	9.93E-02
Red B22-9-2	9.66E-02	2.71E-01	0	4.82E-01	0	0	0	6.13E-03	0	0	1.08E-02	0	0	8.16E-03	0	0	0	1.33E-02	4.88E-03	1.07E-01
Red B22-13-1	9.39E-02	2.20E-01	0	4.71E-01	0	0	0	1.06E-02	0	0	2.56E-02	0	0	1.63E-02	0	0	0	1.28E-02	5.01E-03	1.45E-01
Red B22-13-2	9.40E-02	2.18E-01	0	4.58E-01	0	0	0	3.33E-03	0	0	2.38E-02	0	0	1.58E-02	0	0	0	1.75E-02	4.87E-03	1.65E-01
Red CK2-1	8.63E-02	1.91E-01	0	4.20E-01	0	0	0	5.78E-03	0	0	7.55E-03	0	0	7.83E-03	0	0	0	1.25E-02	6.99E-03	2.62E-01
Red CK2-2	8.29E-02	1.83E-01	0	4.09E-01	0	0	0	2.68E-03	0	0	6.02E-03	0	0	6.92E-03	0	0	0	9.25E-03	7.55E-03	2.92E-01
Red CK3-1	7.46E-02	2.11E-01	0	4.49E-01	0	0	0	5.53E-03	0	0	5.23E-03	0	0	5.87E-03	0	0	0	1.57E-02	6.36E-03	2.28E-01
Red CK3-2	7.26E-02	2.05E-01	0	4.63E-01	0	0	0	5.38E-03	0	0	5.09E-03	0	0	5.71E-03	0	0	0	1.53E-02	6.19E-03	2.22E-01
Red CK4-2	7.16E-02	2.02E-01	0	4.70E-01	0	0	0	5.31E-03	0	0	5.02E-03	0	0	5.63E-03	0	0	0	1.51E-02	6.10E-03	2.19E-01
Red CK4-3	7.12E-02	2.00E-01	0	4.65E-01	0	0	0	4.72E-03	0	0	6.20E-03	0	0	5.58E-03	0	0	0	7.52E-03	7.24E-03	2.32E-01
Summer A38-4.5.6-1	2.55E-03	2.32E-01	1.30E-02	4.51E-01	0	4.35E-03	3.46E-03	7.45E-03	7.15E-03	0.00480632	7.93E-03	1.10E-01	4.46E-03	6.50E-03	7.45E-03	3.40E-03	4.00E-03	4.45E-03	3.39E-03	1.24E-01
Summer A38-4.5.6-2	5.42E-03	2.42E-01	1.34E-02	4.40E-01	0	4.30E-03	3.62E-03	9.12E-03	6.46E-03	5.20E-03	8.18E-03	1.11E-01	4.35E-03	5.75E-03	6.37E-03	3.52E-03	4.14E-03	3.53E-03	3.16E-03	1.20E-01
Summer A38-10.11.12-1	2.85E-03	2.20E-01	2.24E-03	4.08E-01	0	4.31E-04	3.52E-03	1.14E-02	4.07E-03	5.73E-03	8.40E-03	1.48E-01	4.04E-03	4.49E-03	5.91E-03	9.58E-03	4.68E-03	4.76E-03	2.93E-03	1.48E-01
Summer A38-10.11.12-2	2.71E-03	2.35E-01	2.13E-03	3.88E-01	0	4.10E-04	3.50E-03	8.68E-03	4.38E-03	6.85E-03	8.43E-03	1.49E-01	3.84E-03	4.27E-03	5.62E-03	1.26E-02	3.44E-03	3.49E-03	2.78E-03	1.55E-01
Summer B37-5.6.7-2	3.44E-03	2.63E-01	2.76E-03	4.16E-01	0	4.06E-03	3.23E-03	8.53E-03	5.39E-03	5.54E-03	8.55E-03	1.25E-01	4.11E-03	6.34E-03	6.88E-03	4.61E-03	5.03E-03	2.41E-03	1.92E-03	1.24E-01
Summer B37-5.6.7-3	2.64E-03	2.52E-01	2.03E-03	4.28E-01	0	3.83E-03	3.25E-03	8.10E-03	5.54E-03	5.70E-03	8.33E-03	1.07E-01	4.23E-03	4.70E-03	7.08E-03	6.23E-03	7.62E-03	6.10E-03	2.80E-03	1.34E-01
Summer B37-9.10.11-1	9.96E-03	2.30E-01	4.96E-03	4.10E-01	0	5.61E-03	5.02E-03	8.94E-03	9.58E-03	6.30E-03	8.20E-03	1.60E-01	4.05E-03	6.67E-03	6.78E-03	1.11E-02	9.47E-03	6.04E-03	5.70E-03	9.14E-02
Summer B37-9.10.11-2	4.33E-03	2.20E-01	5.25E-03	4.34E-01	0	8.12E-03	2.88E-03	8.72E-03	8.65E-03	7.01E-03	9.03E-03	1.52E-01	4.29E-03	7.62E-03	7.18E-03	8.00E-03	8.82E-03	8.54E-03	4.41E-03	9.04E-02
Summer B37-13.14.15-1	8.42E-03	2.29E-01	1.03E-02	4.43E-01	0	4.61E-03	3.72E-03	6.67E-03	6.44E-03	4.85E-03	1.26E-03	1.15E-01	4.38E-03	4.54E-03	7.32E-03	5.59E-03	2.63E-03	4.47E-03	4.28E-03	1.33E-01
Summer B37-13.14.15-2	9.00E-03	2.37E-01	9.75E-03	4.21E-01	0	5.03E-03	3.76E-03	5.85E-03	6.94E-03	4.61E-03	1.20E-03	1.15E-01	4.16E-03	4.40E-03	6.96E-03	9.58E-03	2.48E-03	3.05E-03	4.13E-03	1.46E-01
Summer CK-6.7.8-1	1.05E-02	2.38E-01	3.37E-03	4.38E-01	0	6.88E-03	3.19E-03	7.70E-03	6.59E-03	3.44E-03	2.73E-03	1.13E-01	4.33E-03	3.38E-03	7.24E-03	1.10E-02	5.02E-03	2.44E-03	4.01E-03	1.29E-01
Summer CK-6.7.8-2	4.38E-03	2.44E-01	4.14E-03	4.39E-01	0	5.48E-03	3.38E-03	8.21E-03	6.41E-03	3.30E-03	3.15E-03	1.04E-01	4.35E-03	4.45E-03	7.27E-03	6.90E-03	4.88E-03	2.80E-03	4.10E-03	1.39E-01

7. REFERENCES

[1] N. O. Soto et al., "Procedures of Food Quality Control: Analysis Methods, Sampling and Sample Pretreatment," in *Quality Control of Herbal Medicines and Related Areas*, Prof. Yukihiro Shoyama, Ed.: ISBN: 978-953-307-682-9, InTech, 2011. [Online accessed in March 2014]. http://www.intechopen.com/books/quality-control-of-herbal-medicines-and-related-areas/procedures-of-food-quality-control-analysis-methods-sampling-and-sample-pretreatment

[2] K.R. Lee et al., "Molecular cloning and functional analysis of two FAD2 genes from American grape (Vitis labrusca L.)," *Gene*, vol. 509, pp. 189-194, 2012.

[3] X. L. Li et al., "Modifications of Kyoho grape berry quality under long-term NaCl treatment," *Food Chemistry*, vol. 139, pp. 931-937, 2013.

[4] Q. Sun et al., "Comparison of Odor-Active Compounds in Grapes and Wines from Vitis vinifera and Non-Foxy American Grape Species," *Journal of Agricultural and Food Chemistry*, vol. 59, pp. 10657-10664, 2011.

[5] K. Pedneault et al., "Flavor of Cold-Hardy Grapes: Impact of Berry Maturity and Environmental Conditions," *Journal of Agricultural and Food Chemistry*, vol. 61, p. 10418−10438, 2013.

[6] R. S. Jackson, *Wine Science, Principles and Applications*, 3rd ed.: Elsevier Academic Press, 2008.

[7] L.C. Yu et al., "Effects of Root Restriction on Soluble Sugar Contents and Related Enzyme Activities in‘Jumeigui’ Grape Berries," *Acta Horticulturae*

Sinica, vol. 38, pp. 825–832, 2011.

[8] Pinghu Municipal Government. (2014, March) Government, Pinghu Municipal. [Online accessed in March 2014]. http://english.pinghu.gov.cn/docs/Updating/2012-08-14/1345079728117.html

[9] B. Wang et al., "Root restriction affected anthocyanin composition andup-regulated the transcription of their biosynthetic genesduring berry development in ‘Summer Black’ grape," *Acta Physiologiae Plantarum*, vol. 35, pp. 2205-2217, 2013.

[10] S. R. Segade et al., "Impact of different advanced ripening stages on berry texture properties of ‘Red Globe’ and ‘Crimson Seedless’ table grape cultivars (Vitis vinifera L.)," *Scientia Horticulturae*, vol. 160, pp. 313–319, 2013.

[11] V. G. Caballero et al., "First steps towards the development of a non-destructive technique for the quality control of wine grapes during on-vine ripening and on arrival at the winery," *Journal of Food Engineering*, vol. 101, pp. 158-165, 2010.

[12] C. Yang et al., "Volatiles of grape berries evaluated at the germplasm level by headspace-SPME with GC–MS," *Food Chemistry*, vol. 114, pp. 1106-1114, 2009.

[13] C. Yang et al., "Volatile compounds evolution of three table grapes with different flavour during and after maturation," *Food Chemistry*, vol. 128, pp. 823-830, 2011.

[14] R. G. Brereton, *Chemometrics: Data Analysis for the Laboratory and*

Chemical Plant.: John Wiley & Sons, Ltd, 2003.

[15] L. G. Zhang et al., "Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy," *Food Chemistry*, vol. 145, pp. 342–348, 2014.

[16] Manuel Urbano et al., "Ultraviolet–visible spectroscopy and pattern recognition methods for differentiation and classification of wines," *Food Chemistry*, vol. 97, pp. 166–175, 2006.

[17] A. Gredilla et al., "Unsupervised pattern-recognition techniques to investigate metal pollution in estuaries," *Trends in Analytical Chemistry*, vol. 46, pp. 59-69, 2013.

[18] D. L. Massart et al., "Handbook of Chemometrics and Qualimetrics: Part A," in *Data Handling in Science and Technology*.: Elsevier, 1997.

[19] Eigenvector Research. (2014, March) Eigenvector Wiki. [Online accessed in March 2014]. http://wiki.eigenvector.com/

[20] B.G.M. Vandeginste et al., "Handbook of Chemometrics and Qualimetrics: Part B," in *Data Handling in Science and Technology*.: Elsevier, 1998.

[21] C. S. Cinca et al., "Partial Least Square Discriminant Analysis for bankruptcy prediction," *Decision Support Systems*, vol. 54, pp. 1245-1255, 2013.

[22] S. Wold et al., "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, pp. 109-130, 2001.

[23] L. A. Berrueta et al., "Supervised pattern recognition in food analysis," *Journal of Chromatography A*, vol. 1158, pp. 196-214, 2007.

[24] L. W. Hantao et al., "Multivariate curve resolution combined with gas chromatography to enhance analytical separation in complex samples: A review," *Analytica Chimica Acta*, vol. 731, pp. 11-23, 2012.

[25] A. de Juan et al., "Chemometrics applied to unravel multicomponent processes and mixtures Revisiting latest trends in multivariate resolution," *Analytica Chimica Acta*, vol. 500, pp. 195-210, 2003.

[26] L. Ximbo et al., "Selective iteratively reweighted quantile regression for baseline correction," *Analytical and Bioanalytical Chemistry*, vol. 406, pp. 1985-1998, 2014.

[27] R. Ameberg et al., *SirExtricate*.: Pattern Recognition Systems, 1997.

[28] B. Grung et al., *MS-Resolver*.: Pattern Recognition Systems, 2002.

[29] Z. Wasta et al., "A database of chromatographic properties and mass spectra of fatty acid methyl esters from omega-3 products," *Journal of Chromatography A*, vol. 1299, pp. 94-102, 2013.

[30] W. Windig et al., "A Noise and Background Reduction Method for Component Detection in Liquid Chromatography/Mass Spectrometry," *Analytical Chemistry*, vol. 68, pp. 3602-3606, 1996.

[31] L. Eriksson et al. (2006) Multi- And Megavariate Data Analysis : Basic Principles And Applications. Ebookdb. [Online accessed in March 2014]. http://www.ebookdb.org/reading/276F7E3CG96729GE201D7F69/Multi--And-Megavariate-Data-Analysis--Basic-Principles-And-Applications

[32] M. K. Boysworth et al., "Aspects of Multivariate Calibration Applied to Near-Infrared Spectroscopy," in *Handbook of Near Infrared Analysis*, 3rd ed., D. A. Burns et al., Ed.: CRC Press, 2008, ch. 10.

[33] L. Yi et al., "Plasma fatty acid metabolic profiling and biomarkers of type 2 diabetes mellitus based on GC/MS and PLS-LDA," *FEBS Letters*, vol. 580, pp. 6837–6845, 2006.

[34] G. A. Eiceman, "Gas Chromatography: Introduction," in *Encyclopedia of Analytical Chemistry*.: Wiley, 2000.

[35] W.M. Coleman et al., "Hyphenated Gas Chromatography," in *Encyclopedia of Analytical Chemistry*, R.A. Meyers, Ed.: Wiley, 2000.

[36] K. Grob, *Split and Splitless Injection for Quantitative Gas Chromatography*, 4th ed.: Wiley VCH, 2001.

[37] O. D. Sparkman, "Mass Spectrometry: Overview and History," in *Encyclopedia of Analytical Chemistry*.: Wiley, 2000.

[38] V. Stroobant et al., *Mass Spectrometry: Principles and Applications*, 3rd ed.: Wiley, 2007.

[39] P. T. Palmer, "Gas Chromatography/Mass Spectrometry," in *Encyclopedia of Analytical Chemistry*, R. A. Meyers, Ed., 2000.

[40] A. G.-Binkul et al., "Determination of monocyclic aromatic hydrocarbons in fruit and vegetables by gas chromatography-mass spectrometry," *Journal of Chromatography A*, vol. 734, pp. 297-302, 1996.

[41] N. Radulović et al., "Volatiles of the Grape Hybrid Cultivar Othello (Vitis vinifera x (Vitis labrusca x Vitis riparia)) Cultivated in Serbia," *Journal of Essential Oil Research*, vol. 22, no. 6, pp. 616-619, 2011.

[42] A. Genovese et al., "Aroma of Aglianico and Uva di Troia grapes by aromatic series," *Food Research International*, vol. 53, pp. 15-23, 2013.

[43] J. E. Welke et al., "Differentiation of wines according to grape variety using

multivariate analysis of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection data," *Food Chemistry*, vol. 141, pp. 3897-3905, 2013.

[44] L. P. Santos et al., "Phenolic compounds and fatty acids in different parts of Vitis labrusca and V. vinifera grapes," *Food Research International*, vol. 44, pp. 1414-1418, 2011.

[45] X.l. Li et al., "Aroma Volatile Compound Analysis of SPME Headspace and Extract Samples from Crabapple (Malus sp.) Fruit Using GC-MS," *Agricultural Sciences in China*, vol. 7, pp. 1451-1457, 2008.

[46] J. Fenoll et al., "Changes in the aromatic composition of the Vitis vinifera grape Muscat Hamburg during ripening," *Food Chemistry*, vol. 114, pp. 420-428, 2009.