Chromatographic Fingerprinting and Quality Control of Herbal Medicines:

Comparison of two officinal Chinese pharmacopoeia species of *Dendrobii* based on High-Performance Liquid Chromatography and Chemometric analysis

by

Débora Sara da Costa Mendes

Thesis for the degree of European Master in Quality in Analytical Laboratories

Supervisors:

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

Prof. Dr. Yizeng Liang, Central South University, PR China

Department of Chemistry
Faculty of Mathematics and Natural Sciences
University of Bergen, Norway

College of Chemistry and Chemical Engineering
Central South University
Changsha, P.R.China

ACKNOWLEDGMENTS

First of all, I would like to gratefully acknowledge the assistance, advice and guidance in University of Bergen of Prof. Bjørn Grung, Prof. Svein Mjøs, Terje Ligre, Bjarte Homelid and in Central South University of Prof. Yizeng Liang and Prof. Hongmei Lu.

I would also like to thank Erasmus Mundus Programme through Professor Isabel Cavaco and Professor José Paulo Pinheiro as coordinators of European Master in Quality in Analytical Laboratories and also to all the professors that work to make this Master very interesting and useful.

To all the friends for the guidance and support in China and Norway. In particular I would like to thank Yangchao Wei, Long Xuxia, Wei Fan, Yun Yonghuang, Leslie Euceda, Zhu Han, Marko Birkic, Alexandre Dias and all the friends that near or even far away were a great support

Finally, but not the least, my dear family, especially my father, my mother, my brother, Rui and also Sandrine and Lia for all the love, patience and support in every moment.

Débora Mendes

Bergen, August 2013

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS, ACRONYMS AND TERMINOLOGY

airPLS	adaptive iteratively reweighted Penalized Least Squares
COW	Correlation optimized warping
CSU	Central South University
D	*Dendrobii*
DAD	Diode-Array Detector
DO	*Dendrobii Officinalis*
GC-MS	Gas Chromatography coupled with Mass spectrometry
HM	Herbal Medicine(s)
HPLC-DAD	High-Performance Liquid Chromatography coupled with Diode-Array Detector
HPLC-MS	High-Performance Liquid Chromatography coupled with Mass Spectrometry
IUPAC	International Union of Pure and Applied Chemistry
LC	Liquid Chromatography
LV	Latent Variable(s)
MS	Mass Spectrometry
PC	Principal Component(s)
PCA	Principal Component Analysis
PLS	Partial Least Squares
PLS-DA	Partial Least Squares-discriminant analysis
SEP	Standard Error of Prediction
SSE	Sum Square Errors
SVD	Singular Value Decomposition
UiB	University of Bergen

ABSTRACT

The fingerprinting quantitative analysis combining similarity evaluation, Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA) is a valid method for classification of herbal medicine species. The main objective of this study was to investigate the chemical differences between two officinal Chinese pharmacopoeia species *Dendrobii Caulis* (Shihu) and *Dendrobii Officinalis Caulis* (Tiepi Shihu). As far as is known no systematic chemical differences study between the two species, especially based on *Dendrobii* whole profile, was done before. A total of twelve samples, six from each species collected from five different provinces in China were analyzed in China at Central South University (CSU) and in Norway at University of Bergen (UiB). The extraction method of flavonoids or other phenolic compounds present in the two different species of *Dendrobii* and the sample preparation were developed and were relatively simple processes. The main advantages of these processes were low solvent consumption, relatively short extraction time, good extraction efficiency, stability and repetitiveness. The HPLC-DAD method was developed to separate the components present in the two species of the Chinese herbal medicine (HM) *Dendrobii* with good resolution. Based on the optimization of the chromatography conditions, an efficient chromatography fingerprint of these species was established. It was verified that some compounds with retention times in the range from 40 to 50 min appeared in *Dendrobii* species but not in *Dendrobii Officinalis* species. All the samples were analyzed at four different wavelengths, the results obtained at 254 nm being the most useful. PCA results showed that the distribution of the samples in two groupings before and after peak alignment is almost the same revealing the similarity between the two species. Regarding PLS results, it was observed a regular relationship between the *Dendrobii* samples and between the *Dendrobii Officinalis* samples with a clear separation between the two different clusters. In the results obtained for one wavelength or even four wavelengths, the final predictive properties of the models were good due to the low values obtained for the Standard Error of Prediction (SEP). The selectivity ratio showed specific regions in the raw data that could help distinguish between the two *Dendrobii* species. The method established by this study could be applied to other similar *Dendrobii* species for the quality assessment.

1. INTRODUCTION

1.1 Theory and Background

A great number of oriental countries have extensively used traditional HM and their preparations for many centuries [1, 2].

The quality control of traditional HM is one of the main concerns for its application and development, so to obtain additional evidence of its safety and efficacy more scientific research and improvement of the quality of the research is needed. This fact is recognized by World Health Organization: “Despite its existence and continued use over many centuries, and its popularity and extensive use during the last decade, traditional medicine has not been officially recognized in most countries. Consequently, education, training and research in this area have not been accorded due attention and support. The quantity and quality of the safety and efficacy data on traditional medicine are far from sufficient to meet the criteria needed to support its use worldwide. The reasons for the lack of research data are due not only to health care policies, but also to a lack of adequate or accepted research methodology for evaluating traditional medicine” [3].

1.1.1 Chromatographic Fingerprints and Quality Control of Herbal Medicines

A chromatographic fingerprint of a HM is, by definition, “a chromatographic pattern of the extract of some common chemical components of pharmacologically active and or chemically characteristics [1, 4-6]. This chromatographic profile should be featured by the fundamental attributions of ‘integrity’ and ‘fuzziness’ or ‘sameness’ and ‘differences’ so as to chemically represent the HM investigated” [1, 6, 7]. So, using the chromatographic fingerprints it is possible to do accurately the authentication and identification of the HM (‘integrity’) even if the concentration/amount of the characteristic constituents are slightly different for the same HM (‘fuzziness’) and the

chromatographic fingerprints can also effectively show the ‘sameness’ and the ‘differences’ between several samples [1, 6, 8].

In every HM and its extract there is a great number of components that are unknown and most of them are in low amount and even in the same HM samples it is frequently observed some variability [1, 9, 10]. Therefore, to obtain a chromatographic fingerprint that represents the pharmacologically active and chemically characteristic constituents is not very simple [1].

To ensure the consistency of HM products, the phytoequivalence concept was developed. The full HM product can be seen as the active compound, because the several constituents act together being responsible for its therapeutic effect. According to the phytoequivalence concept, “a chemical profile, such as a chromatographic fingerprint, for an herbal product should be constructed and compared with the profile of a clinically proven reference product” [1, 11].

So, an extract of the HM should be prepared and its activity by pharmacological and clinical methods should be determined. A qualitative and quantitative profile of all the constituents should be obtained by using a hyphenated technique with high efficiency and sensitive detection, such as HPLC-DAD, HPLC-MS or GC-MS. These hyphenated techniques used to obtain the chromatographic fingerprints and further combined with chemometric approaches are the perfect tools for quality control and authenticity of HM [1, 2, 11].

To obtain a good chromatographic fingerprint that represents the phytoequivalence of a HM depends on many factors, such as extraction methods, measurement instruments, measurement conditions, etc. The chemical constituents in the HM may also vary depending on plant origins, harvest seasons, drying processes and even possible contaminations such as excessive or banned pesticides, microbial contaminants, heavy metals, chemical toxins, etc. [1, 2, 12].

Since a single HM may contain a great number of natural constituents, obtaining a good fingerprint is dependent on the method of extraction and the sample preparation.

A powerful tool for the quality control of herbal medicines is the combination of chromatographic fingerprints of HM with chemometric approaches.

In quality control of herbal medicines, the research field is very interdisciplinary because it uses knowledge from chemistry, biochemistry, pharmacology, medicine and also statistics [1, 2].

1.1.2 Herbal Medicine *Dendrobii*

The second largest group of the family Orchidaceae is the genus *Dendrobii or Dendrobium*, which comprises approximately 1400 species [13, 14].

In Pharmacopoeia of the People's Republic of China 2005 edition also known as the Chinese Pharmacopoeia 2005, *Dendrobii Caulis* – Shihu – is officially recorded as the fresh or dried stem of *Dendrobium nobile* Lindl, *Dendrobium officinale* Kimura et Migo, *Dendrobium fimbriatum* Hook. var. *oculatum* Hook and similar species [15].

In Chinese history and literature, *Dendrobii Officinale* Kimura et Migo was described as a miraculous drug. Its antitumor [16], cardio-protective [17, 18], immunomodulatory [19] and hepatoprotective [20] effects have recently been confirmed by modern research. Recently, in the Chinese herbal medicine market, the price of *Dendrobium officinale* Kimura et Migo has increased one hundred times more in relation to other *Dendrobii* species. *Dendrobium officinale* Kimura et Migo is even considered the precious wild "Tiepi Shihu" in traditional conception and due to the increasing demand and price is often adulterated by other related species [21].

Figure 1. *Dendrobium* fresh(left) **[22]** and *Dendrobium* dry stems (right) **[23]**

Therefore, in current Chinese Pharmacopoeia 2010, *Dendrobii Officinalis Caulis* (Tiepi Shihu) is separated from *Dendrobii Caulis* (Shihu) and recorded as the dried stem of *Dendrobium officinale* Kimura et Migo, while *Dendrobii Caulis* (Shihu) is officially recorded as the fresh or dried stem of *Dendrobium nobile* Lindl., *Dendrobium chrysotoxum* Lindl. and *Dendrobium fimbriatum* Hook. [24]. However, from the data base of Chinese State Food and Drug Administration for registered medicines entitled with "Shihu" and produced by more than 190 factories, no species designation was clarified [25].

And to the best of our knowledge, no systematic chemical differences study, especially based on *Dendrobii* whole profile, between the two species has been done.

1.1.2.1 Flavonoids and Herbal Medicine *Dendrobii*

Flavonoids are phenolic compounds widely present in an extensive range of natural plants, with over 8000 individual substances known. The flavonoids are classified as flavones, flavanones, catechins and anthocyanins. The basic structure of flavonoids is shown in **Figure 2**. These type of compounds show different functions in plants, such as antioxidants, antimicrobials, photoreceptors, visual attractors, feeding repellants, and light screening. Studies on pharmacological effects of flavonoids have shown that these compounds have extensive biological activities and significant pharmacological effects on cardiovascular, digestive and nervous systems. They also have anti-inflammatory, antiallergenic, antiviral, vasodilatory, immunoregulator, anti-tumor, analgesic, liver-protecting, aging-delaying, antidepressive and immunity-improving effects [26-28]. The role of flavonoids as antioxidants that reduce free radical formation and quench free radicals has been the subject of many studies. The antioxidant activity is observed in both the absorbed flavonoids and their metabolites. It is very common that the flavonoids occur in plants as glycosylated derivatives and they also give a contribution to the colors in leaves, flowers, and fruits [29]. Significant sources of flavonoids are the medicinal plants and their phytomedicines [27, 30].

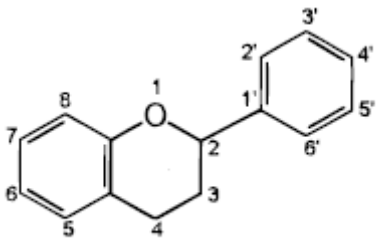


Figure 2. Basic flavonoid structure **[27]**

It is convenient to use dry, lyophilized or frozen samples because when the plant material to be analyzed is fresh or non-dried, the flavonoids (especially glycosides) can be decomposed by enzymatic action. Therefore, the dry samples are grinded into a powder and the extraction solvent is selected according to the polarity of the flavonoids present in the sample to be analyzed. For less polar flavonoids such as isoflavones, flavanones, flavonols and methylated flavones the extraction is done with chloroform, dichloromethane, diethyl ether or ethyl acetate whereas more polar aglycones and flavonoid glycosides are extracted with alcohols or alcohol–water mixtures. Direct solvent extraction is still the most used method [31]. This kind of medicinal material is usually extracted using ultrasonic methods and with alcohols [32-34] and also in order to remove saccharides since they are not soluble in this type of solvents [35, 36].

The application of standardized UV/UV–Vis spectroscopy has been applied for years in the analyses of this kind of polyphenolic compounds. This type of compounds has two characteristic UV absorption bands, with maxima within an interval that varies from 240 to 285 and from 300 to 550 nm.

It is possible to recognize the different flavonoid classes by their UV spectra characteristics that include the effects of the number of aglycone hydroxyl groups, glycosidic substitution pattern and the nature of aromatic acyl groups [31, 37].

The type of flavonoids existent in *Dendrobium* species are anthocyanins (anthocyanidins), flavonol glycosides (based on kaempferol, quercetin, myricetin and methylated derivatives) and flavonol aglycones [38-43].

1.1.3 Analytical techniques

Chromatography is defined by International Union of Pure and Applied Chemistry (IUPAC) as “a physical method of separation in which the components to be separated are distributed between two phases, one of which is stationary (stationary phase) while the other (the mobile phase) moves in a definite direction.”

Liquid Chromatography (LC) is also defined by IUPAC as “A separation technique in which the mobile phase is a liquid. LC can be carried out either in a column or on a plane. Present-day liquid chromatography generally utilizing very small particles and a relatively high inlet pressure is often characterized by the term high-performance or high-pressure liquid chromatography, and the acronym HPLC” [44].

The hyphenated analytical technique used in this work, both in China and Norway, was High-Performance Liquid Chromatography with Diode-Array Detection (HPLC-DAD).

1.1.3.1 HPLC-DAD

·High-performance (or High-pressure) Liquid Chromatography

As mentioned before, the mobile and stationary phases are the two parameters related to the separation that is carried out in a chromatographic system.

In HPLC, the stationary phase is packed into a column capable to support high pressures while the mobile phase is a liquid supplied under high pressure (up to 400 bar/4×10^7 Pa) to guarantee a constant flow rate and consequently reproducible chromatography.

Therefore, the sample is dissolved in the mobile phase and after it is forced to pass through the stationary phase by means of high pressure so that chromatographic separation occurs because the different components of the sample have different affinity with the stationary or the mobile phase and consequently take different times

to move from the position of sample introduction to the position where they are detected.

With previous knowledge about the analytes under investigation it is possible to change the properties of the stationary and/or mobile phases to achieve the desired separation.

Different kinds of detectors can be coupled to HPLC and its type is chosen according to the sort of analysis being performed, for instance qualitative (identification) or quantitative [2, 45, 46].

The HPLC system is schematized in **Figure 3**.

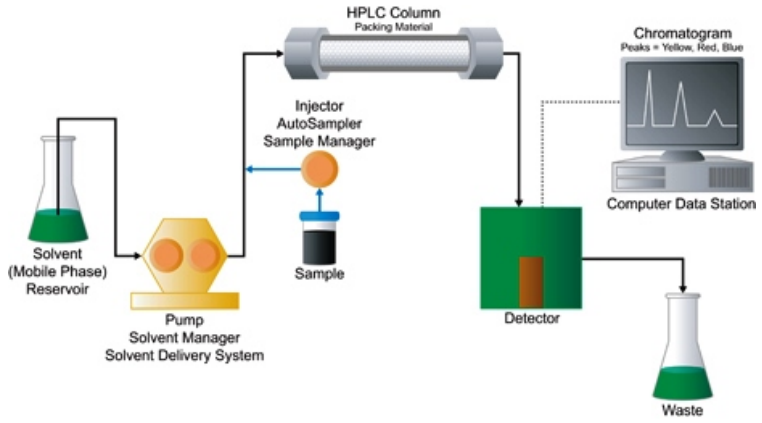


Figure 3. Schematic representation of HPLC system **[47]**

·Diode Array Detector (DAD)

A detector is a device that is used to sense each solute as it is eluted from a chromatography column.

The diode array detector can use a deuterium or xenon lamp that emits light over the UV spectrum range or a tungsten lamp for the visible region. The light from the lamp is focused by a lens through the sample cell and onto a holographic grating. Therefore, the sample is subjected to light of all wavelengths produced by the lamp. The dispersed light from the grating is able to reach a diode array. The array can have many hundreds of diodes and the output from each diode is regularly sampled and

stored in a computer. At the end of the run, it is possible to select the output from any diode and to produce a chromatogram using the UV wavelength that was falling on that particular diode [46].

The diode array detector is schematized in **Figure 4**.

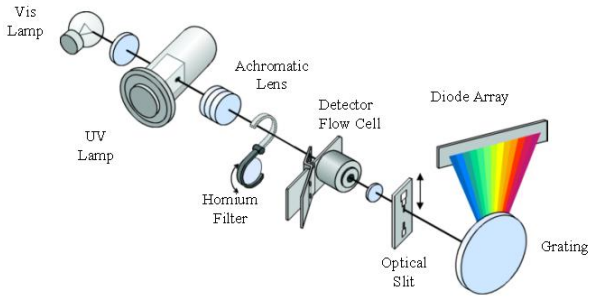


Figure 4. Schematic representation of a Diode Array Detector (DAD) **[48]**

This analytical technique HPLC-DAD was previously used in the analysis of *Dendrobium* species, as was HPLC coupled with Mass Spectrometry (MS), an extremely versatile technique when it comes to analyze this kind of samples [49-53].

1.1.4 Chemometric techniques

With the use of chromatographic instrumentation two goals can be achieved, such as quantitative analysis and qualitative (identification) analysis. In quantitative analysis it is possible to determine how much of a substance is present in a mixture and the data is obtained from peak height or peak area measurements. In qualitative analysis the solutes present in a mixture can be identified and the data is often obtained from retention measurements [46].

Pattern recognition tools as Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA) are very useful and widely used chemometric techniques to visualize and summarize the very large amount of data obtained from multivariate measurements in chemistry. By using the proper

mathematical approaches, pattern recognition is used to identify patterns in large data sets [54].

1.1.4.1 Principal Component Analysis

In chromatography, very often the chromatographic peaks are partially overlapping, so chemometric methods help to resolve the chromatogram into individual components. In order to obtain predictions, first the chromatogram is treated as a multivariate data matrix and then PCA is performed. In the mixture, each compound is a (chemical) factor with its spectra and elution profile which by a mathematical transformation can be related to principal components. After performing PCA, there is a reduction of the original variables to a number of significant principal components (e.g. two). In this way, PCA is used as a form of variable reduction, reducing the large original dataset to a much smaller manageable dataset more easily interpreted [54].

In the case of coupled chromatography like HPLC-DAD, the essential dataset for a single chromatogram can be described as a sum of responses for each significant compound in the data, characterized by an elution profile and a spectrum plus noise or instrumental error. Using matrix notation it can be written as:

$$\boldsymbol{X} = \boldsymbol{CS} + \boldsymbol{E}$$

Equation 1

where $\boldsymbol{X}$ is the original data matrix or coupled chromatogram, $\boldsymbol{C}$ is a matrix consisting of the elution profiles of each compound, $\boldsymbol{S}$ is a matrix consisting of the spectra of each compound and $\boldsymbol{E}$ is an error matrix.

In summary, PCA is a way of identifying patterns in data and expressing it in a way to emphasize their differences and similarities. Since patterns in data of high dimension can be hard to find and where graphical representation is not available, PCA is a powerful tool for analyzing data. Another advantage of PCA is that there is not much loss of information once the patterns are found in the data and this data is compressed by reducing dimensions [54].

·Scores and Loadings

The abstract mathematical transformation of the original data matrix in PCA is:

$$\boldsymbol{X = TP + E}$$

Equation 2

where ***T*** are called scores, and have as many rows as the original data matrix, ***P*** are the loadings, and have as many columns as the original data matrix and the number of columns in the matrix ***T*** equals the number of rows in the matrix ***P***.

The principal components are vectors of loadings or scores where variables or objects with largest variance will make the greatest impact. The scores, in the case of chromatography, relate to elution profiles and the loadings relate to the spectra.

The schematic representation of PCA is showed in **Figure 5**.

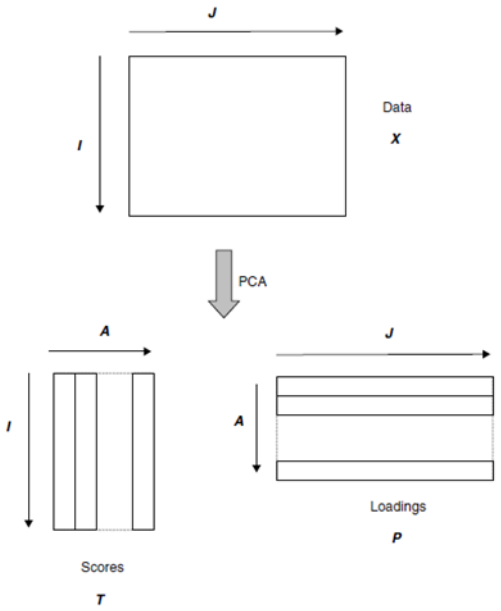


Figure 5. Schematic representation of PCA **[54]**

The aim of PCA is to obtain a description of a data table in terms of uncorrelated new variables called principal components (PCs). The PCs are linear combinations of all the original variables subjected to two restrictions: First they are located in the direction explaining most of the variation in the data table and second they are

orthogonal with respect to each other (angle between them is 90°) and to the residual matrix.

As said before, the PCs are given as vectors of loadings or scores. The loading vectors represent a basis for the variable space, while the score vectors represent a basis for the object space. Plotting the objects on the loading vectors shows the relationships between objects, while plotting the variables on the score vectors shows the relationships between variables.

The significance of a PC is measured through the ratio of its variance to the total variance contained in the original variables. A PC with small variance usually means that it carries little information. The variance explained is most often used as the criterion for deciding on the number of PCs needed to obtain a data table [54].

1.1.4.2 Partial Least Squares

Partial Least Squares Regression (PLS) is a calibration method based on finding the model relating the components of $\boldsymbol{X}$ to the components in $\boldsymbol{Y}$. PLS components are calculated finding the directions of maximum covariance between $\boldsymbol{X}$ and $\boldsymbol{Y}$, i.e., the maximum variation of $\boldsymbol{X}$ correlated to $\boldsymbol{Y}$. PLS Component is different from the PCA component that means the scores obtained in PLS are different from the ones obtained in PCA. In PLS1 a single $\boldsymbol{y}$ variable is predicted and in PLS2 a block of $\boldsymbol{Y}$ variables is predicted.

In **Figure 6** it is represented the decomposition of $\boldsymbol{X}$ ($\boldsymbol{T}, \boldsymbol{P^T}$) and $\boldsymbol{Y}$ ($\boldsymbol{U}, \boldsymbol{Q^T}$) matrices in PLS components. After, the construction of the regression model $\boldsymbol{U = TB}$ ($\boldsymbol{X}$ and $\boldsymbol{Y}$ matrices are represented by their components) is done.

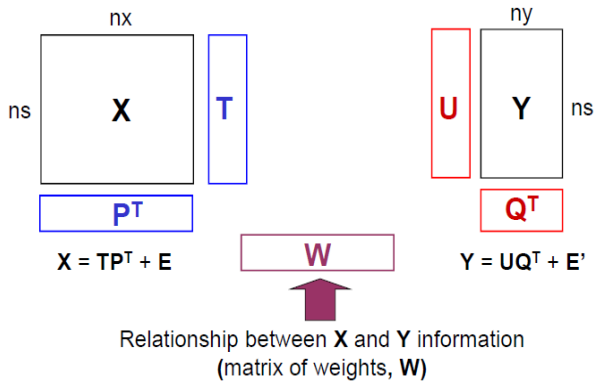


Figure 6. Decomposition of $\boldsymbol{X}$ and $\boldsymbol{Y}$ matrices in PLS components **[55]**

In the calibration step, PLS components and regression vectors are calculated sequentially. There is the decomposition of $\boldsymbol{X}$ and $\boldsymbol{Y}$ in components and the calculation of the regression coefficient.

In the prediction step, first there is the Calculation of scores related to the response of new samples, $\boldsymbol{X_{new}}$:

$$\boldsymbol{T_{new} = X_{new}P}$$

Equation 3

After the prediction of **Y** scores for the new samples ($\boldsymbol{U_{new}}$):

$$\boldsymbol{U_{new} = T_{new}B_{PLS}}$$

Equation 4

And finally, the calculation of the properties of interest ($\boldsymbol{Y_{new}}$) for the new samples using the calculated $\boldsymbol{Y}$ scores:

$$\boldsymbol{Y_{new} = U_{new}Q^T}$$

Equation 5

In summary, building a PLS model includes several steps such as: Preprocessing data sets if required ($\boldsymbol{X}$ and $\boldsymbol{Y}$): in the calibration set or in the validation set and in the new unknown samples; Selection of the size of the calibration model (number of components); Exploration of $\boldsymbol{X}$ and $\boldsymbol{Y}$ data sets and their relationship: study of variance explained by the model and outlier detection and elimination from the

calibration set; Qualitative interpretation of the model; Model validation and finally Prediction of new samples [55].

Partial Least Squares Discriminant Analysis (PLS-DA) is a classical PLS regression, with a regression mode, where the response variable indicates the classes (or categories) of the samples. PLS-DA has often been used for supervised classification, i.e., classification and discrimination problems. The response vector is qualitative and is recoded as a dummy block matrix where each of the response categories is coded with an indicator variable. After this, PLS-DA is performed as if the response vector was a continuous matrix [54].

Classification problems in fingerprints data analysis are complex due to the many variables and few samples/objects issue. This makes that many solutions can be found to separate the classes. The PLS-DA score plots as showed in most classification applications present an overoptimistic view of the separation between the classes.

Even using PLS-DA to discriminate a random data set into two groups does almost always give a PLS score plot with perfect separation between the two arbitrary classes.

The permutation testing and cross model validation are used to assess the validation of classification models. Permutation tests show that when cross validation is not applied appropriately, it leads also to overoptimistic results [56].

Selectivity ratio can be used to detect marker candidates and can be defined for each variable i as:

$$SR_i = v_{expl,i}/v_{res,i} \qquad i = 1,2,3,..$$

Equation 6

where v_{expl} is the explained variance and v_{res} the residual variance. Based on an F-test, this is a valuable property for variable selection especially when the ratio of the number of variables to the number of objects is high [57].

PCA and PLS diverge in the optimization problem they solve to find a projection matrix but they are both linear decomposition techniques and they can be combined with various functions.

The statistical measure of the multivariate distance of each observation from the center of the data set is named Hotelling's T-squared statistic. To calculate the T-squared statistic, PCA uses the main principal components and it is used for the detection of outliers [58].

1.1.5 Information theory applied to chromatographic fingerprint of HM

Information theory is used to evaluate the chromatographic fingerprints. Since the chromatographic fingerprint obtained is deeply dependent on the chromatographic separation degree and concentration distribution of each chemical component, based on the information content it is possible to select the chromatographic fingerprint with the best separation degree and the most uniform distribution of the chemical compounds.

A chromatographic fingerprint may be considered as a continuous signal determined by its shape and according to Ref. [59], the information content of a continuous signal can be defined as:

$$\Phi = -\int p_x log p_x \, dx$$

Equation 7

where p_x is the chromatographic response of all chemical components present in the fingerprint under investigation.

The evaluation of the quality of the HM is done based on similarities and/or differences of the chromatographic shapes and based on the separation degree of each chemical component between the fingerprints obtained for the different HM under study.

Therefore, first a chromatographic fingerprint is normalized with its overall peak area equal to one and after its information content is obtained based on **Equation 8** [60]:

$$\Phi = -\int p_x/[sum(p_x)] log p_x/[sum(p_x)]\, dx$$

Equation 8

Two advantages of calculating the information content according to **Equation 8** is that the whole chromatogram is taken into account and also that the noise should have a small influence in this calculation.

Figure 7 shows four simulated chromatographic fingerprints with different separation degrees (a, b, c and d). The concentration distributions of the four peaks are the same. The values of the chromatographic resolution (R_S) are displayed in **Table 1**. The results suggest that the further chromatographic separation from **Fig. 7a** to **Fig. 7d** (R_s from 1.50 to 2.00), which can not cause any addition to the information content Φ, is unnecessary. However, the serious overlapping situation in **Fig. 7b** and **Fig.7c** (R_s=0.63, 0.31) causes a loss of the information content [60].

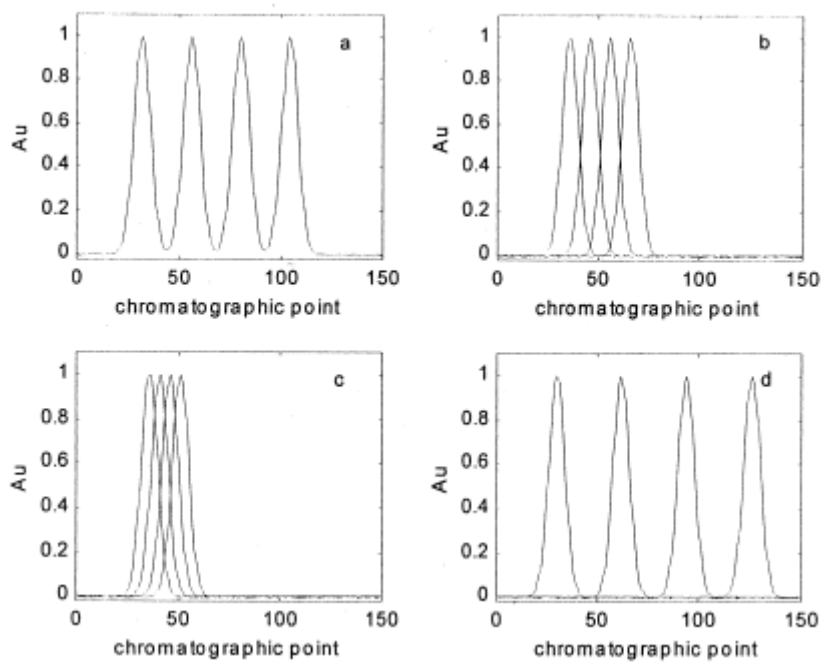


Figure 7. Chromatographic fingerprints simulated with different separation degrees **[60]**

Table 1 - Information content of simulated chromatographic fingerprints with different separation degrees represented in **Figure 7 [60]**

Data	a	b	c	d
R_s	1.50	0.63	0.31	2.00
Φ	6.04	5.50	4.83	6.04

1.1.6 Orthogonal (Taguchi) "L" Array Design

Genichi Taguchi (Japan) developed a method for designing experiments to investigate how different parameters affect the mean and variance of a process performance characteristic that defines how well the process is functioning. This experimental design involves the use of orthogonal arrays to organize the parameters affecting the process and the levels at which they should be varied.

The Taguchi method tests pairs of combinations instead of testing all possible combinations like in factorial design. With this method it is possible to determine which factors most affect product quality with a minimum amount of experimentation, consequently saving resources and time. The method works better for an intermediate number of variables (3 to 50), few interactions between variables and when only few variables contribute significantly.

There are five general steps involved in the Taguchi Method Design of Experiments:

- Define a target value to measure the performance of the process;
- Determine the parameters that affect the process;
- Construct the orthogonal array for the parameter design showing each experiment number and conditions;
- Carry on the experiments indicated in the completed array to collect the data;
- Data analysis in order to check the effect of each parameter on the performance measure.

In **Table 2** it is represented a 3^{4-2} Fractional Factorial Design 4, with Factors at three Levels (9 runs), where P1, P2, P3 and P4 are the parameters that can affect the process.

Table 2 - Orthogonal (Taguchi) L_9 Array Design

Experiment	P1	P2	P3	P4	IC
1	1	1	1	1	IC1
2	1	2	2	2	IC2
3	1	3	3	3	IC3
4	2	1	2	3	IC4
5	2	2	3	1	IC5
6	2	3	1	2	IC6
7	3	1	3	2	IC7
8	3	2	1	3	IC8
9	3	3	2	1	IC9

After calculating the information content for each experiment, the average information content value (K) is calculated for each factor and level. This is done according to **Equation 9**, where i is the level number (1, 2 or 3) and j is the parameter number (1, 2, 3 or 4).

$$K_{i,P_j} = \frac{\sum_i IC}{3}$$

Equation 9

The range R ($R = high\ K - lowK$) of the K for each parameter is calculated and the larger R value for a parameter means a larger effect of the variable on the process [61, 62].

1.1.6.1 Advantages and Disadvantages of "L" Array Design

The advantage of the Taguchi method is that emphasizes a mean performance characteristic value close to the target values and it allows the analysis of many different parameters without a high amount of experimentation. It obtains a lot of information about the main effects in a relatively few number of runs. It allows the identification of key parameters that have the most effect on the performance characteristic value so that further experimentation on these parameters can be performed and the parameters that have little effect can be ignored.

The main disadvantage of this method is that the results obtained are only relative. Also, since the orthogonal arrays do not test all variable combinations, this method should not be used when all relationships between all variables are needed. The Taguchi method provides limited information about interactions between parameters [61, 62].

1.1.7 Pre-processing of data

Prior to chemometric analysis of the chromatographic results using MATLAB and Sirius, it is necessary to pre-treat the chromatograms obtained. There are some pre-processing steps that seem particularly important for the further analysis of the HPLC-DAD data. First, the pre-processing of data done using the Changde program included three processes; namely data smoothing and differentiation and baseline correction. After this, alignment of the chromatographic data is also needed since no internal standard is used during the experiments. And finally, the data obtained was also normalized.

After the pre-processing of fingerprints it is possible to proceed to the chemometric analysis of the data obtained.

1.1.7.1 Data smoothing and differentiation

The aim of data smoothing and differentiation is to remove the random errors from the quantitative information. Disregarding the source of these errors, they are usually described as noise and it is very important to remove as much as possible this noise without losing the basic information.

For this purpose, the method of least squares is used, where the set of points is fitted to some curve and it is assumed that all the error is in the ordinate (y) and not in the abscissa (x). The least squares minimize the sum of squared residuals, where a residual is the difference between an observed value and the fitted value provided by a model.

Using averaging prior to smoothing, it is possible to reduce the noise nearly as the square root of the number of points that were used.

Thereby, this function known as Savitzky-Golay filter for smoothing and differentiation will act as a filter to smooth noise fluctuations and avoid distortions into the dataset [63].

1.1.7.2 Baseline correction

Signals of analytical instruments like chromatography essentially contain chemical information, baseline and random noise.

For the baseline correction an algorithm named adaptive iteratively reweighted Penalized Least Squares (airPLS) was used. This method iteratively changes weights of sum squares errors (SSE) between the fitted baseline and the observed signals. The weights of the SSE are adaptively obtained using the difference between the previously fitted baseline and the observed signals [64].

The airPLS leads to a balance between fidelity to the observed data and the roughness of the fitted data.

In **Equation 10**, x is the vector of the analytical signal and z is the fitted vector and m is the length of both of them. The fidelity of z to x can be expressed as the sum square errors between them.

$$F = \sum_{i=1}^{m} (x_i - z_i)^2$$

Equation 10

Balance between fidelity and smoothness is measured as the fidelity plus penalties on the roughness. It can be given by **Equation 11**, where D is the derivative of the identity matrix such that $Dz = \Delta z$.

$$Q = F + \lambda R = \|x - z\|^2 + \lambda \|Dz\|^2$$

Equation 11

Although adaptive iteratively reweighted procedure is similar to the weighted least squares and to iteratively reweighted least squares [65, 66, 67], it calculates the weights in different ways and adds a penalty item to control the smoothness of the fitted baseline. Each step of this proposed procedure involves calculating a weighted penalized least squares according to **Equation 12**, where w is the weight vector and t represents each iterative step.

$$Q^t = \sum_{i=1}^{m} w_i^t |x_i - z_i^t|^2 + \lambda \sum_{j=2}^{m} \left| z_j^t - z_{j-1}^t \right|^2$$

Equation 12

So, airPLs algorithm can be applied to chromatograms since it gives extremely fast and accurate baseline corrected signals for both fitted and observed signals [64].

1.1.7.3 Automated alignment of chromatographic data

The datasets must be preprocessed before PCA and PLS-DA analysis in a way that the elements in the matrix for individual samples describe the same phenomena. For peak alignment in chromatographic data, several types of approaches have been developed. In some approaches, the retention time shifts have been corrected by making internal standards added or making marker peaks coincide in all chromatograms under study [68-74].

For the peak alignment in this work, automated alignment of chromatographic data was used, where the data is preprocessed in order to correct unwanted time-shifts. This approach includes the selection of a reference sample to warp towards. This selection procedure is used when there are no internal standards available or normalization to correct the signal. This method is used for datasets obtained from quite homogeneous samples with very similar chromatographic profiles [1, 75].

1.1.7.3.1 Correlation optimized warping (COW)

The COW algorithm, introduced by Nielsen *et al.* [76] is a method to correct shifts in discrete data signals. This algorithm, that it is assumed to preserve the properties of peak shape and area, aligns a sample chromatogram (digitized vector) towards a reference chromatogram (reference sample vector). This reference sample is used to align the entire data set.

The COW algorithm requires two user input parameters that are typically selected on a trial and error basis by visual inspection of the chromatographic profiles: the segment length and the slack size (flexibility).

A slightly modified version of COW algorithm developed by Tomasi *et al.* [77] is used here and the main change regards the sharing of the boundaries between adjacent segments. In this new version of COW, the correlation coefficient between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of length N is calculated as:

$$r(\boldsymbol{x},\boldsymbol{y}) = \frac{cov(\boldsymbol{x},\boldsymbol{y})}{\sqrt{var(\boldsymbol{x})var(\boldsymbol{y})}}$$

$$= \frac{[(\boldsymbol{I}_N - \boldsymbol{1}\boldsymbol{1}^T N^{-1})\boldsymbol{x}]^T(\boldsymbol{I}_N - \boldsymbol{1}\boldsymbol{1}^T N^{-1})\boldsymbol{y}}{\|\tilde{\boldsymbol{x}}\|_2(\|\boldsymbol{y}\|_2^2 - N\bar{y})^{1/2}} = \frac{\tilde{\boldsymbol{x}}^T\boldsymbol{y}}{\|\tilde{\boldsymbol{x}}\|_2(\|\boldsymbol{y}\|^2 - N\bar{y})^{1/2}}$$

Equation 13

Where $\tilde{x}$ represents the centered $\boldsymbol{x}$, $\bar{y}$ is the mean of $\boldsymbol{y}$ and the centering matrix $\boldsymbol{I_N} - \boldsymbol{1}\boldsymbol{1}^T N^{-1}$ is symmetric and idempotent [75].

1.1.7.3.2 Reference chromatogram selection

The reference chromatogram (sample) should be as representative as possible for all phenomena of interest in the data set.

This method is based on the product of the correlation coefficients between all individual chromatograms.

For a given chromatogram $\boldsymbol{x}_t$, the similarity index $(0 < similarity\ index \leq 1)$ can be calculated as follows:

$$Similarity\ index = \prod_{i=1}^{I} |r(\boldsymbol{x}_t, \boldsymbol{x}_i)|$$

Equation 14

where $r(\boldsymbol{x}_t, \boldsymbol{x}_i)$ is the conventional correlation coefficient between two chromatograms in the dataset calculated as shown in **Equation 13**.

So, the chromatogram with the highest similarity index is selected to be the reference chromatogram to use with the given dataset [75].

1.1.7.3.3 Simplicity value

The simplicity value is used to measure the alignment of a set of chromatograms towards the reference chromatogram. Its principle is related to the properties of singular value decomposition (SVD), where the size of the squared singular values is directly associated to the sum of squares of the data matrix. Any data matrix, **X** (uncentered) can be decomposed as $\boldsymbol{X} = \boldsymbol{USV}^T$, where $\boldsymbol{S}$ is a diagonal matrix containing the singular values equal to the square roots of the eigenvalues of $\boldsymbol{X}^T\boldsymbol{X}$. $\boldsymbol{U}$ and $\boldsymbol{V}$ are both orthogonal matrices, where the columns in $\boldsymbol{U}$ are the eigenvectors of $\boldsymbol{XX}^T$ and the columns of $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$.

The sum of the first R squared singular values determines how much of the variation is explained by the corresponding R components:

$$Explained\ variance = \sum_{r=1}^{R} \left(SVD \left(\frac{\boldsymbol{X}}{\sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} x(i,j)^2}} \right) \right)^2$$

Equation 15

where SVD (**M**) indicates the single value for a given component *r* and where the data is scaled to a total sum of squares of one.

Though, to find the optimal combination of segment and slack size as the simplicity value ($0 \leq \text{simplicity} \leq 1$), the principle of simplicity is adapted from Henrion & Andersson [78], Christensen *et al.* [79] and Johnson *et al.* [80]:

$$Simplicity = \sum_{r=1}^{R} \left(SVD \left(\frac{\boldsymbol{X}}{\sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} x(i,j)^2}} \right) \right)^4$$

Equation 16

It is possible to achieve high simplicity values in COW alignment with some combinations of segment and slack parameters. This is shown in **Figure 8 (A)** using simulated data.

This method is focused on preserving total area of all peaks in the chromatographic profiles and any change introduced by the alignment procedure is not desired. So, a second criterion that takes into account this area effect has to be included [75].

1.1.7.3.4 The peak factor

The data should be quite homogeneous so that peak area and shape can be the same before and after alignment. Even if the reference chromatogram is carefully selected a change in peak area and shape can still occurs. This change can be quantified by the peak factor, which is a number between 0 and 1.

$$Peak\ factor = \frac{\sum_{i=1}^{I} (1 - minimum(c(i), 1)^2)}{I}$$

Equation 17

where, $c(i) = \left| \frac{\|x_w(i)\| - \|x(i)\|}{\|x(i)\|} \right|$ and $\|x(i)\| = \sqrt{\sum_{j=1}^{J} x(i,j)^2}$ is the Euclidean length or norm for $\boldsymbol{x}_i$; $\boldsymbol{x}_i$ is the chromatogram before warping while $\boldsymbol{x}_w(i)$ is the same sample after alignment.

Values of peak factor measure are shown in **Figure 8 (B)** for simulated data. It is possible to see that some combinations of segment length and slack size give high simplicity values but low peak factor values, so should not be considered as suitable alignment parameters.

1.1.7.3.5 The warping effect

The warping effect combines simplicity and peak factor ($0 \leq warping\ effect \leq 2$):

$$Warping\ effect = simplicity + peak\ factor$$

The simplicity factor and the peak factor have the same influence on the warping effect value. The relation between these three measures is shown in **Figure 8**. If the warping effect has a value closer to two means that peaks are both aligned and that the change in the area is minimal [75].

1.1.7.3.6 Optimization

The warping effect values are optimized in the form of a discrete-coordinates simplex-like optimization routine carried out in several steps [81]. The first step is to establish global search space boundaries from the combination of all segment length and slack sizes of interest. In the second step, by default a 5×5 sparse search grid is selected in both the segment and slack direction and then the warping effect for these 25 points is determined (this is done using simulated data as an example with segment length 10–70 and slack size 1–15). The six best (default choice) combinations, providing the highest warping effect scores, are selected and used as starting points in a discrete-coordinates simplex optimization part.

1.1.7.3.7 Defining the optimization space

As shown in **Figure 8**, the search space for segment lengths includes several possible choices as long as the slack size (flexibility) is large enough. So, longer segment lengths require more flexibility to give good alignment. Though, this will always depend on the chromatographic data available such as datasets obtained from quite homogeneous samples with very similar chromatographic profiles as explained before.

1.1.7.3.7.1 Segment length and slack size (flexibility)

The rule to select the segment length optimization space is:

$$PW_A \pm \frac{PW_A}{2}$$

Equation 18

where PW_A is the approximate peak width average at the base over all peaks in the reference chromatogram. Using this rule, the segment lengths will contain both entire peaks and peak fragments.

For the slack size search space, the number of data points before and after the first and the last peak, respectively, should be roughly the same as the peak widths (ensuring enough flexibility), then a slack size search space ranging from 1 to 15 [75].

The simplicity and optimization routines were freely downloaded from Reference [82].

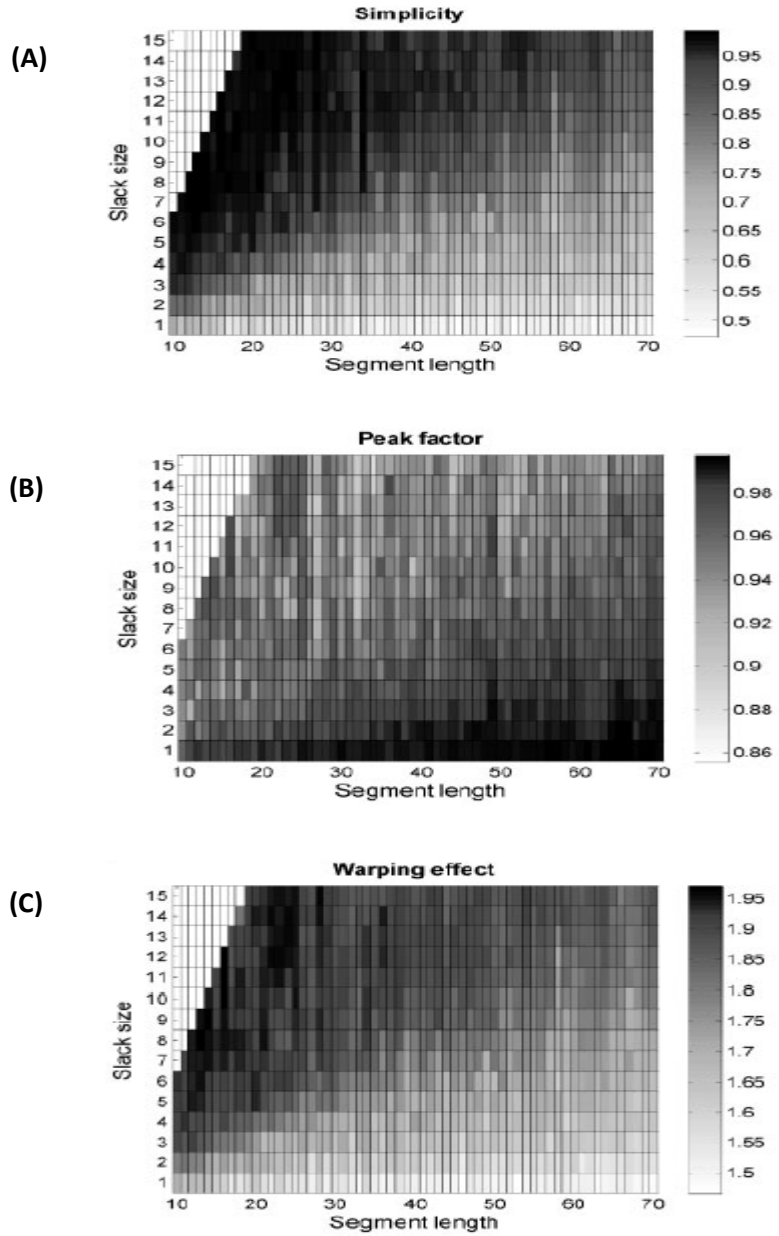


Figure 8. Simplicity **(A)**, peak factor **(B)** and warping effect **(C)** values for all combinations of segment length and slack size using simulated data. For plots (A) and (B) a value close to one indicates that data are well aligned and that the area has changed insignificantly, respectively. For plot (C) a value close to two means that peaks are both aligned and that the change in the area is minimal. The white triangle in the upper left corner contains unfeasible combinations of segment length and slack size in the COW algorithm. **[75]**

1.1.7.4 Data Normalization

Normalization is performed to give objects the same relative or absolute "size". In chromatography, normalization is used to compensate differences in the amount measured at the injector [57].

Normally the variables are divided by one variable and in chromatography this happens when one analyte is added and used as an internal standard. An internal standard is a substance very similar but not identical to the chemical substances of interest present in the sample, whose peak is well resolved relative to the peaks of other substances present in the sample. The ratio of analyte signal to the internal standard signal is not affected by small variations in the injected volume and in the chromatographic conditions [46].

When an internal standard is not available, block normalize can be done and it consists on dividing all selected variables in each object with their sum to obtain the relative distribution of the variables in each object. This procedure is also known as normalizing to constant sum. Normalizing to constant sum corresponds to the transformation:

$$\frac{X_{ki}}{100 * \sum X_{ki}} \rightarrow X_{ki}$$

Equation 19

where the index k runs over the objects and index i over the variables [57].

1.2 Aims of the study

The main objective of the present study is to investigate the chemical differences between the two officinal Chinese pharmacopoeia species *Dendrobii Caulis* (Shihu) and *Dendrobii Officinalis Caulis* (Tiepi Shihu). Therefore, the main steps of the present study are:

- To optimize the extraction method of flavonoids, one kind of active compounds in the traditional Chinese HM *Dendrobii*;

- To optimize the chromatography conditions;

- To establish the fingerprint analysis method and perform HPLC-DAD analysis of *Dendrobii Caulis* (Shihu) and *Dendrobii Officinalis Caulis* (Tiepi Shihu);

- To perform HPLC-MS analysis of *Dendrobii Caulis* (Shihu) and *Dendrobii Officinalis Caulis* (Tiepi Shihu) to identify the main peaks;

- Data analysis of fingerprint based on chemometrics analysis, such as PCA and PLS and find the differences between the two species.

2. EXPERIMENTAL

2.1 Material, reagents and samples

The HM *Dendrobii* samples were reduced to a powder by using a coffee and spice mill Tefal GT30083E (Tefal, China).

Twelve samples of the two species of HM *Dendrobii* were purchased from five Chinese different provinces. In **Table 3** is it possible to see the details of *Dendrobium* samples and the details of *Dedrobium Officinale* samples are described in **Table 4**. In **Figure 9** it is possible to identify the place of origin of each sample.

The hyphenated chromatographic equipment HPLC-UV Dionex Ultimate 3000 LC System (USA) used in Central South University (CSU) in China and in University of Bergen (UiB) in Norway is shown in **Figure 10**. In the LC System it was used a Hypersil ODS (C18) column (reversed phase), with 250mm length×4.6mm internal diameter (ID) and 5µm particle cartridge (Agilent Technologies, USA). The HPLC system consisted of a quaternary pump, a vacuum degasser, an autosampler and the column compartment fixed at 25 ºC was coupled to a variable wavelength diode-array detector (DAD). The injection volume was 10 µL.

For the eluent system with a flow rate of 1mL/min it was used pure Methanol Sigma-Aldrich Chromasolv®, gradient grade for HPLC≥99.9% (lot#SZBC292FV, Germany) and Formic acid Fluka Analytical from Sigma-Aldrich for LC-MS ~98% (lot#BCBG7820V, Germany) with a concentration in water of 0.4%. The water for HPLC analysis was purified by a Milli-Q Millipore water purification system (USA).

Prior to LC analysis, the final samples were filtered using Iso-Disc™ Filters Supelco N-25-4 Nylon 25mm×0.45µm (Germany) and also filters Chromacol 1000×4-SF-45(N) Tecnolab (USA).

All the calculations, plots and fingerprints shown have been performed in Changde 003 Version 1.0, 2008, Central South University, Changsha, PR China, in MATLAB®

version 7.10.0.499 (R2010a), the MathWorks, Inc., USA and in PRS-Sirius Version 8.1, ©Copyright 1987-2009 Pattern Recognition Systems AS, Norway. The simplicity and optimization routines for the automated alignment of chromatographic data were freely downloaded from Quality & Technology Website: http://www.models.kvl.dk, University of Copenhagen, Denmark.

Figure 9. Map of People's Republic of China **[83]**

Table 3 - Description of *Dendrobium* (D) samples

Sample nr	Drugstore	Province of origin (China)	Amount provided (g)
D1	Hunan Qianjin	Guangxi	30
D2	Tianjian	Guangxi	30
D3	Yunxiang	Sichuan	30
D4	Zhilin	Guangxi	30
D5	Yangtianne	Yunnan	30
D6	Hunan Bencaogangmu	Guangxi	30

Table 4 - Description of *Dendrobium Officinale* (DO) samples

Sample nr	Drugstore (Date of production)	Province of origin (China)	Amount provided (g)
DO1	Zhejiang Leqing (20/04/2012)	Zhejiang	30
DO2	Zhejiang Leqing (10/06/2012)	Zhejiang	30
DO3	Zhejiang Leqing (10/07/2012)	Zhejiang	30
DO4	Zhejiang Leqing (13/10/2012)	Zhejiang	15
DO5	Zhejiang Jinhua (15/10/2012)	Zhejiang	10
DO6	Hunan Shaodong (15/09/2012)	Hunan	30

Figure 10. HPLC-DAD Dionex Ultimate 3000 LC System used in CSU and in UiB **[84]**

In HPLC, the most usual chromatographic peak shape distortion is tailing peaks. There can be several mechanisms of analyte retention and in the case of reversed-phase chromatography there are non-specific hydrophobic interactions with the stationary phase. Though, polar interactions with some ionized residual silanol groups on the silica surface are common and are the cause of tailing peaks.

So, to obtain good peak shapes this kind of interactions need to be minimized.

To avoid peak tailing, the chromatographic separation should be performed at a lower pH in order to minimize secondary interactions of the acidic silanol groups because in this way, it is possible to assure that these ionizable residual groups are fully protonated.

So, an aqueous solution of formic acid in a low concentration was used as one of the eluents to improve the tailing peak shapes of the weak acidic phenols and also to help the whole separation because of its acidity, due to its ion pair effect [85-87].

2.2 Optimization of the extraction process

The methods of extraction and sample preparation are very important to obtain good fingerprints of herbal medicines [1].

To achieve this, the optimization of the extraction process of flavonoids was done together with the optimization of the chromatography conditions. Several wavelengths were also tested.

The best approach to perform quality control of complex HM is to perform chromatographic fingerprints especially using hyphenated chromatographic techniques. Any HM sample can contain hundreds of complex phytochemical compounds, thus it is very hard or even impossible to identify all of them by using the common approaches.

In this work, information theory was used to evaluate the chromatographic fingerprints.

For the extraction of flavonoids process, there were used four variables: A–time in minutes, B–the percentage of methanol used, C–volume of methanol used in mL and D–temperature in ºC. For each variable there were three levels. Levels 1, 2, 3 and are respectively for variable A: 30, 40, 50 minutes of ultrasonic extraction; variable B: 50, 70, 100% of methanol; variable C: 30, 40, 50 mL of methanol and variable D: 25, 35, 45 ºC.

So, having these factors at three levels in the extraction process, it was possible to set up a 3^{4-2} Fractional Factorial Design 4 (9 runs) also called (Taguchi) Orthogonal L_9 design.

The results obtained for the information content according to **Equation 8** are shown in **Table 5**.

Table 5 - 3^{4-2} Fractional Factorial Design 4 and respective results obtained for information content relative to the spectra obtained at 254 nm

Experiment No.	A t (min)	B % MeOH	C V (mL)	D T (ºC)	Information content
1	1	1	1	1	11,57
2	1	2	2	2	12,02
3	1	3	3	3	11,96
4	2	1	2	3	11,52
5	2	2	3	1	11,35
6	2	3	1	2	12,24
7	3	1	3	2	11,47
8	3	2	1	3	12,10
9	3	3	2	1	11,48
K1	11,85	11,52	11,97	11,47	
K2	11,70	11,82	11,67	11,91	
K3	11,68	11,90	11,59	11,86	
R	**0,17**	**0,38**	**0,38**	**0,44**	

Although the differences in the information content values are not very significant, a higher information content value means that there is a better chromatographic separation and lower information content value means that might be an overlapping situation. According to this, choosing the higher K values, the best result according to the calculations for the process of extraction of flavonoids is A1B3C1D2 meaning t=30 min, 100% methanol, V=30 mL and T=35 ^{o}C.

Since temperature is the parameter with the highest R value, this should be the variable with the larger effect on the process.

2.2.1 Sample preparation

All the samples were dried and pulverized before use and approximately 3 g of each sample was weighted and extracted with 30 mL of pure methanol in an ultrasonic bath for 30 minutes at 35 ºC. Afterwards, vacuum filtration was done and the extract was evaporated to dryness in a water-bath at 90 ºC. The residue was dissolved with pure methanol to avoid the dissolution of saccharides. The solution was prepared in a 5 mL volumetric flask. Before the HPLC analysis, the sample was filtrated through a 0.45 µm membrane into a LC vial.

During the extraction process, the *Dendrobium* samples showed a yellow-brownish color whereas *Dendrobium Officinalis* samples displayed a greenish color. Parallel samples were prepared for *Dendrobium Officinalis* in both China and Norway.

2.3 Optimization of chromatography conditions and fingerprinting

The optimization of the chromatography conditions was done taking into account the optimization of the extraction process.

A good experimental design for the optimal chromatographic separation is necessary [1]. The chromatographic conditions for *Dendrobii* species were described in **2.1** Material, reagents and samples. After several attempts, discussions and research to find the gradient elution for the chromatographic separation of flavonoids in *Dendrobii* species it was possible to find a chromatogram with good resolution. The chromatographic separation of flavonoids was carried out using a gradient elution of solvent A: 0.4% formic acid in water and solvent B: 100% methanol at a flow rate of 1 mL/min as follows:

t (min)	0	10	25	40	50	70	90	95
%B	5	15	40	55	55	100	100	5

The DAD detector was set at four different wavelengths: 254, 280, 310 and 335 nm.

The data obtained was normalized because there was no standard solution available for this specific analysis.

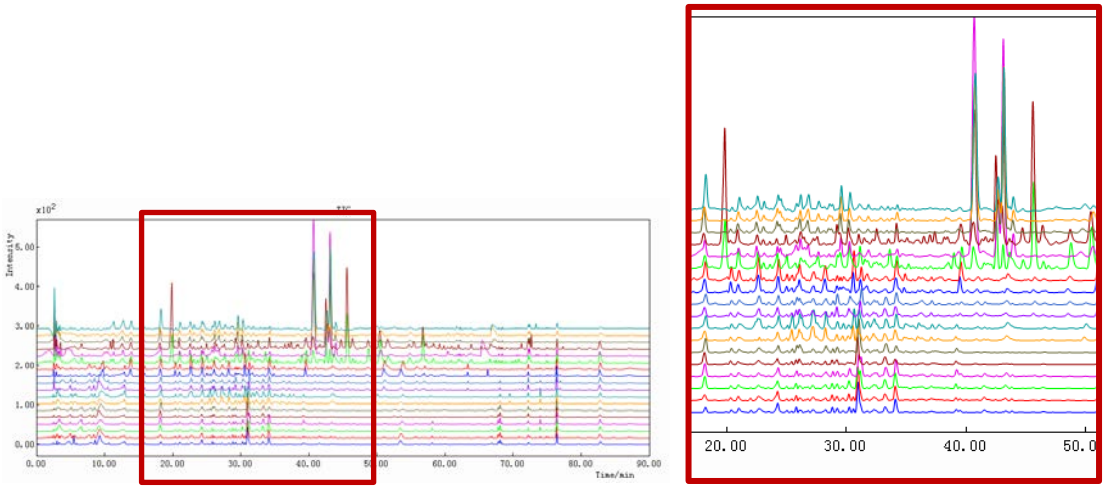
The data was not standardized. Standardization multiplies each variable with the inverse of its standard deviation. Thus, every variable has variance equal to one after this pre-processing. Because data seem to be noisy standardizing it would emphasize peaks with small area (noisy) and decrease the importance of peaks with larger areas.

Scanning the sample solution through the diode array detector, it was found that the baseline is steady at 254 nm and more peaks can be detected. It is possible to see this in **Figure 11**, **Figure 12** and **Figure 13**.

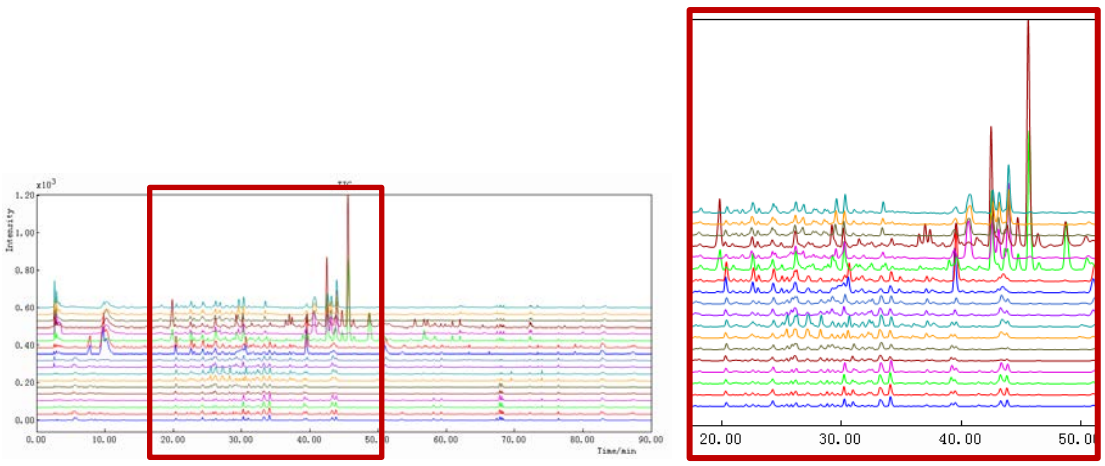
When the wavelength was above 254 nm, the baseline became unsteady gradually, noise increased and the spectrogram became in disorder. Fewer peaks were detected and the values for most peaks remarkably were dropped. So the detection wavelength was set at 254 nm and this was also described in some references of previous works where flavonoids in *Dendrobii* species were studied [1, 50, 53, 88].

To check the stability of the samples prepared, several runs were done for some of the samples. These runs were done on different days. It was verified, by comparing different chromatograms made on different days that the samples were stable at least during the time in which the experiments occurred. The acronyms used to describe the samples are D for *Dendrobii* samples, DO for *Dendrobii Officinalis* samples, 1 and 2 represent the replicates, N represents the samples analyzed in Norway (UiB) and C the samples analyzed in China (CSU).

(A)



(B)



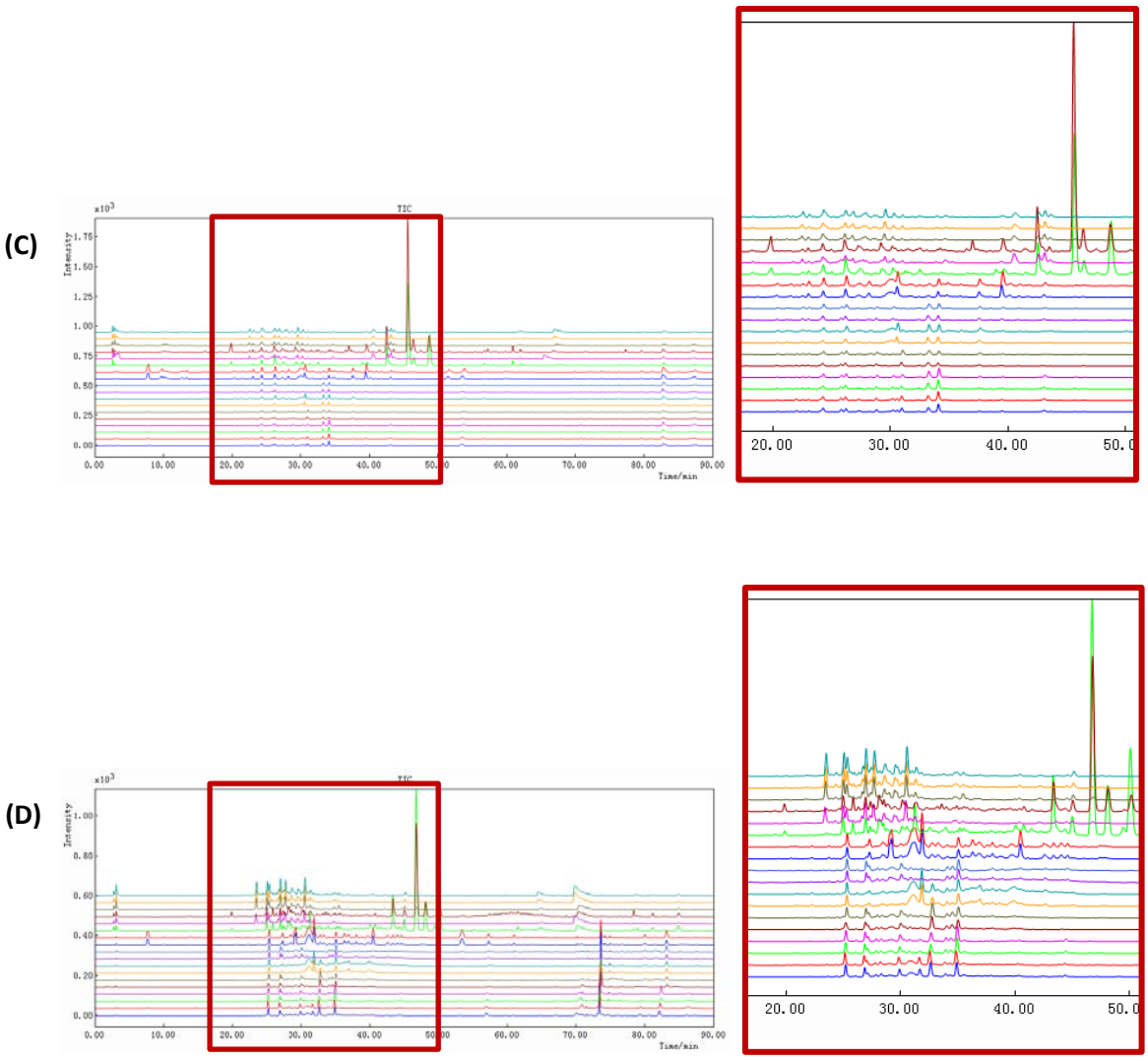
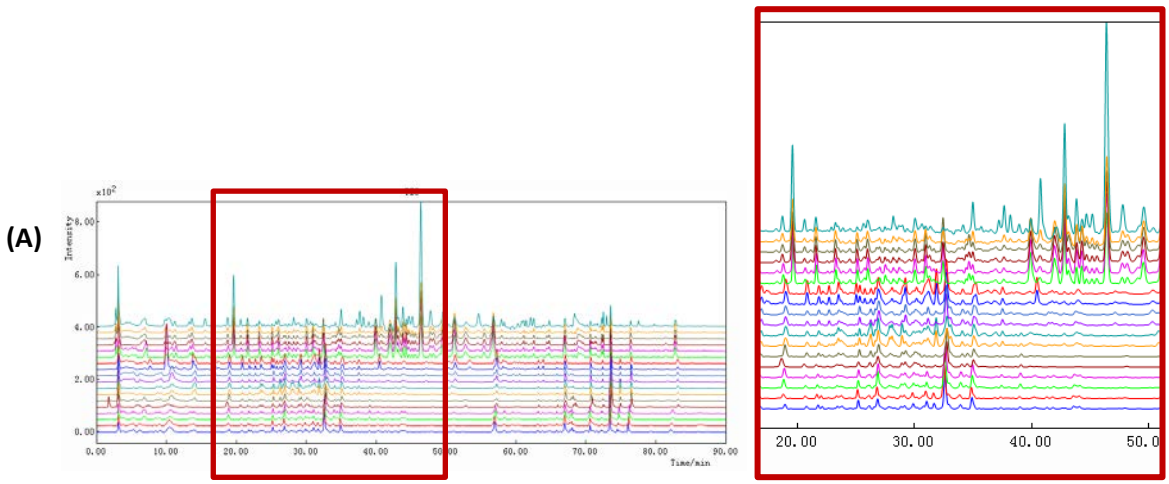


Figure 11. Results obtained in Norway: from 1-12 DO samples and from 13-18 D samples. Results were obtained at (A) 254 nm, (B) 280 nm, (C) 310 nm and (D) 335 nm



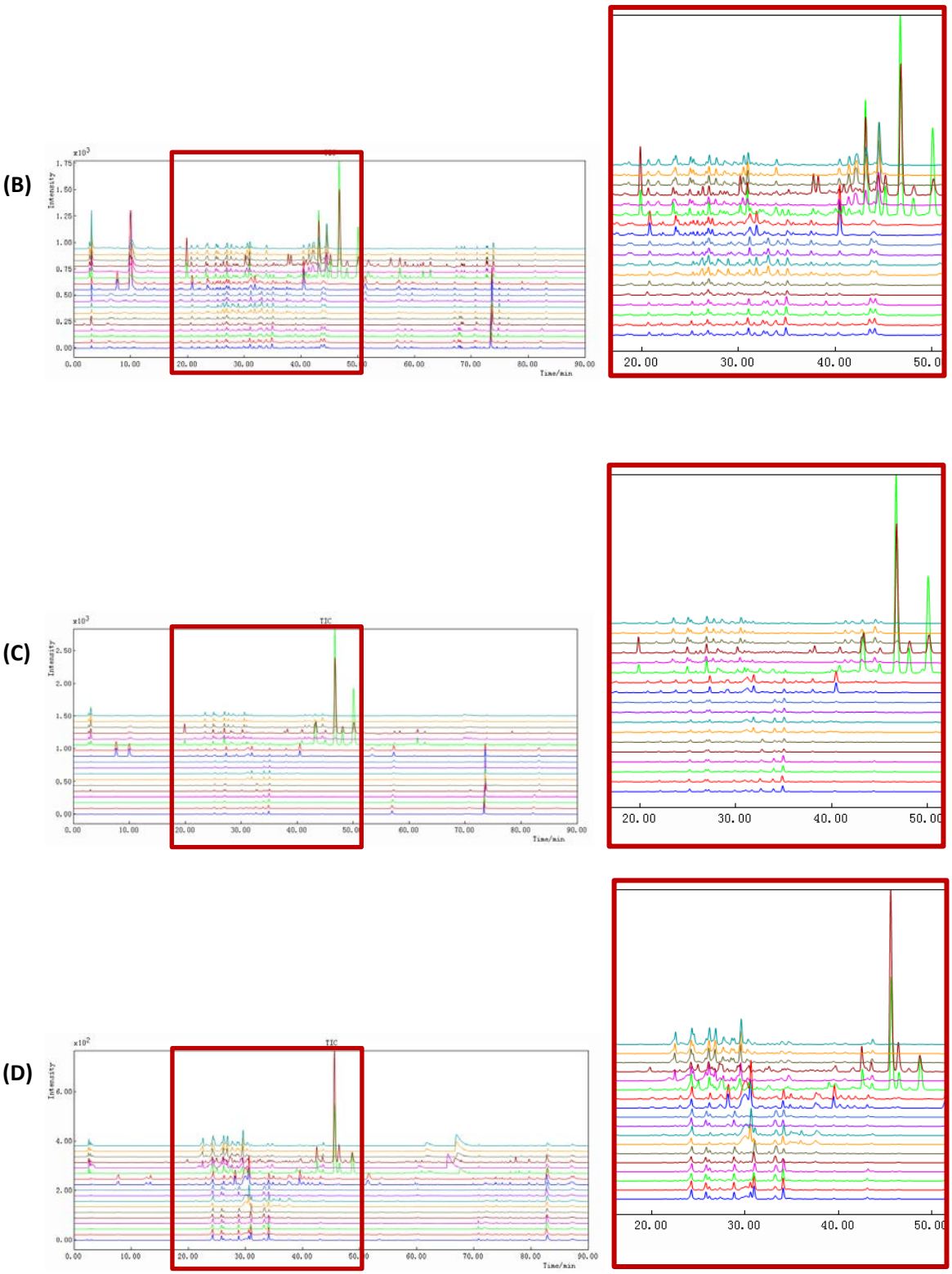
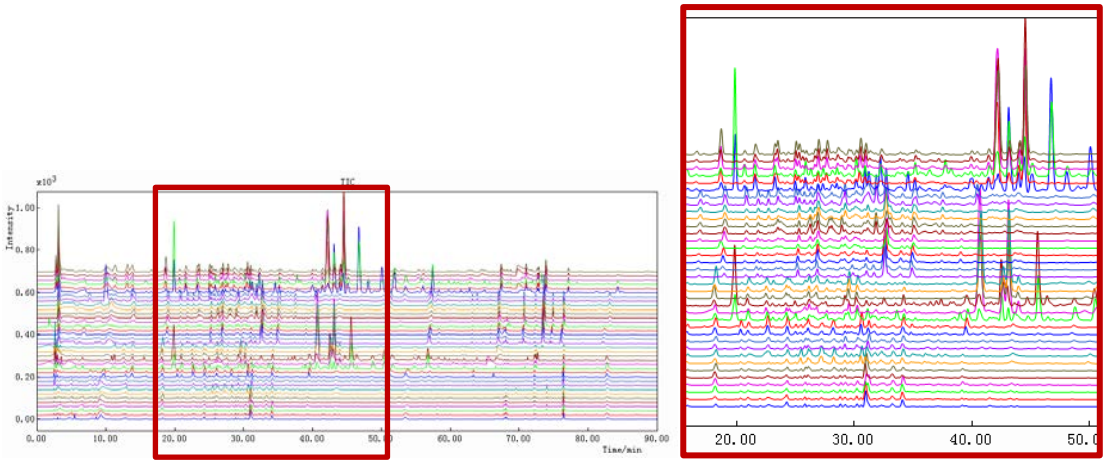
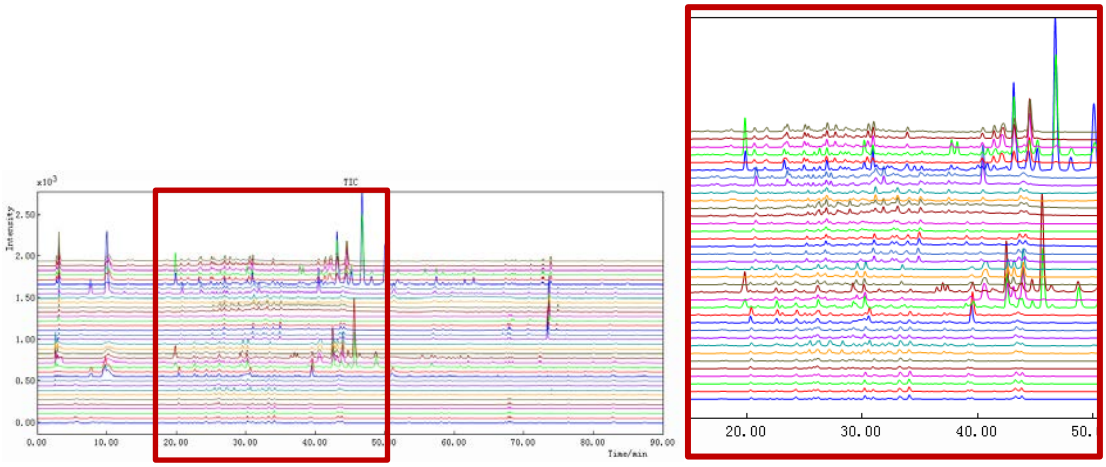


Figure 12. Results obtained in China: from 1-12 DO samples and from 13-18 D samples. Results were obtained at (A) 254 nm, (B) 280 nm, (C) 310 nm and (D) 335 nm

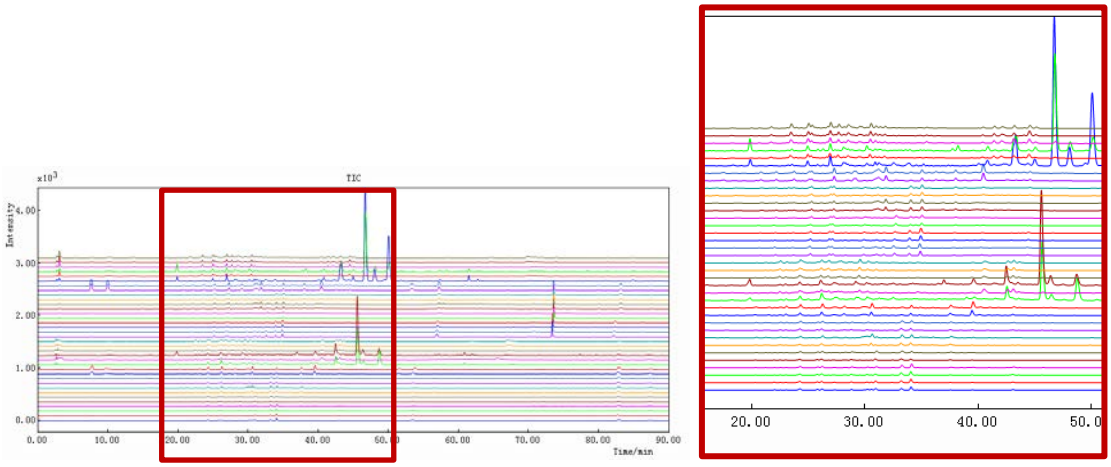
(A)


TIC
Intensity
Time/min
20.00
30.00
40.00
50.0

(B)


TIC
Intensity
Time/min
20.00
30.00
40.00
50.

(C)


TIC
Intensity
Time/min
20.00
30.00
40.00
50.0

(D)

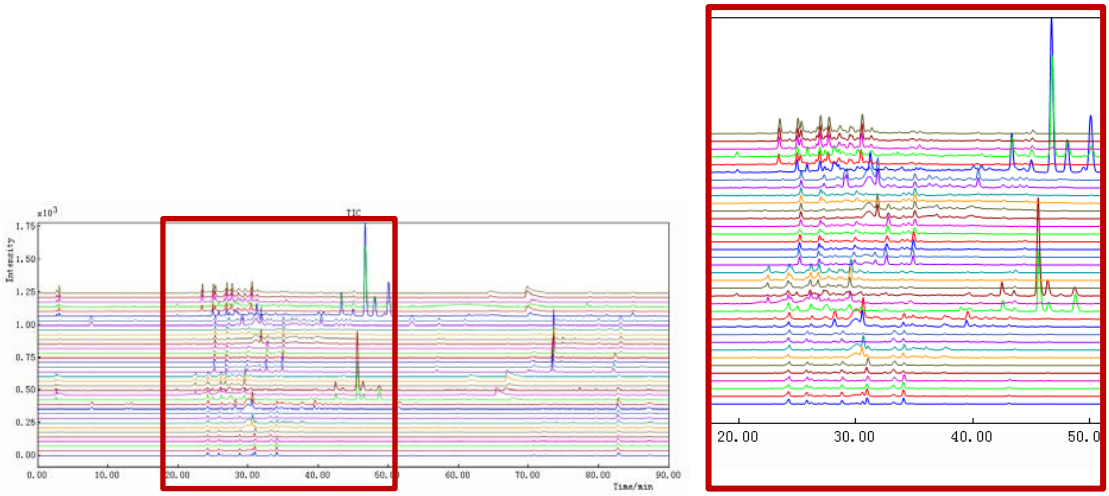


Figure 13. Results obtained in Norway: 1-12 DO samples and 13-18 D samples. Results obtained in China: 19-30 DO samples and 31-36 D samples. All the results were obtained at (A) 254 nm, (B) 280 nm, (C) 310 nm and (D) 335 nm

3. RESULTS AND DISCUSSION

It was necessary to pre-treat the chromatograms obtained prior to the chemometric. First, the pre-processing of data was done using the Changde program and after this, alignment of the chromatographic data was performed in MATLAB since no internal standard was used during the experiments.

After the pre-processing of fingerprints in both programs it was possible to proceed with the chemometric analysis of the data obtained.

3.1 Fingerprint analysis

The fingerprint analysis results are divided into the results obtained at one wavelength – 254 nm – and the sum of the results obtained at four different wavelengths – 254, 280, 310 and 335 nm.

3.1.1 Results using the data obtained at one wavelength: 254 nm

Since more peaks can be detected at the wavelength of 254 nm, looking closer at the fingerprints obtained at this wavelength it is possible to see specific main peaks present in *Dendrobii* spectra that are not present in *Dendrobii Officinalis* spectra.

The following figures show the results obtained at 254 nm before peak alignment (A) and after peak alignment (B). The results obtained in UiB are shown in **Figure 14**, the results obtained in CSU are in **Figure 15** and in **Figure 16** it is possible to see the results obtained in UiB and CSU all together. From (A) to (B) it is possible to see some improvement with respect to the peak alignment.

In the results obtained in UiB there are several main peaks common to all *Dendrobii* spectra with variables between ~4500 and ~5500 that corresponds to retention times between ~40 min and ~50 min. In the results obtained in CSU the main peaks are found in the same range.

This fact might represent that the two species of samples – *Dendrobii* and *Dendrobii Officinalis* – have some different compounds with retention times in this range that could help distinguish them.

In total around 100 characteristic fingerprint peaks were detected in the twelve samples of *Dendrobii* and *Dendrobii Officinalis* collected from five different places in China.

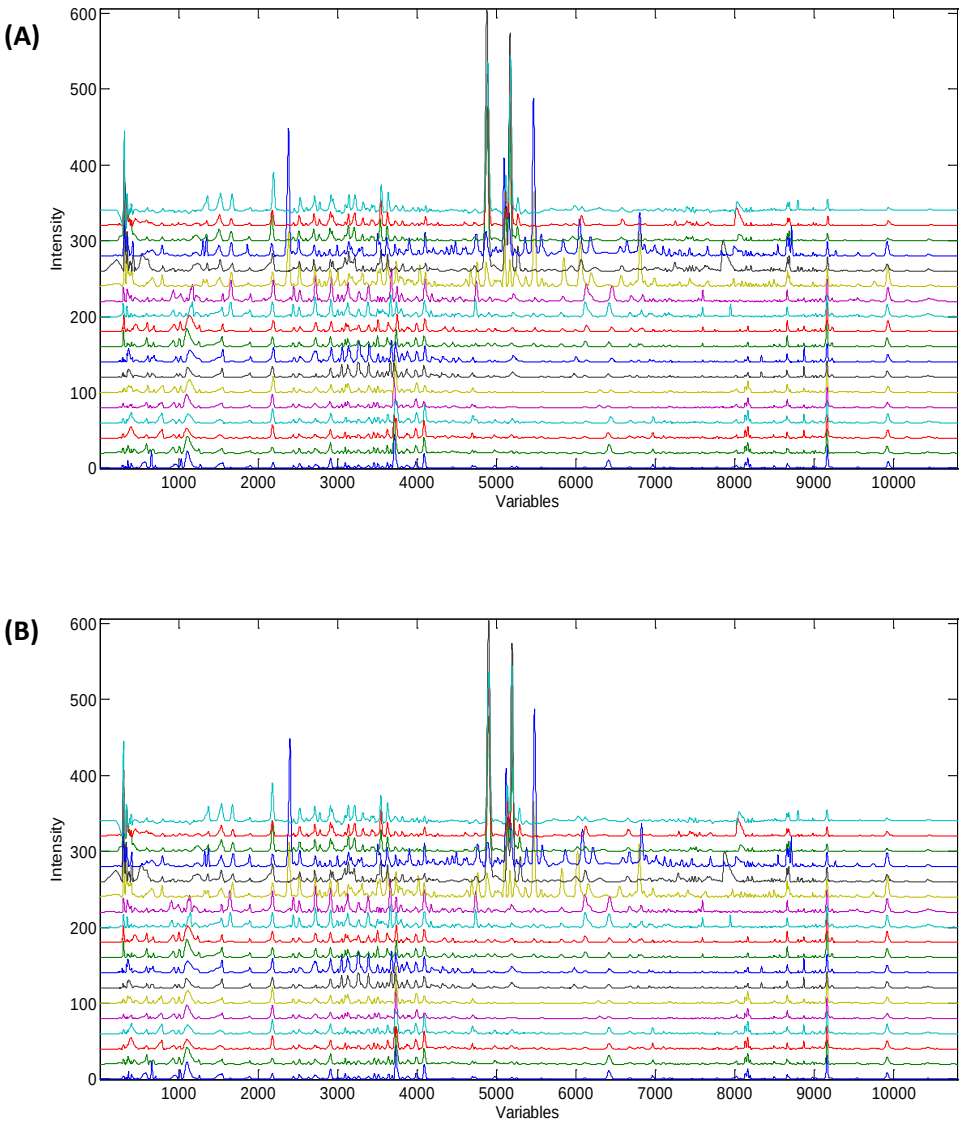


Figure 14. Results obtained in UiB at 254 nm. Spectra 1-12 DO and 13-18 D: (A) before peak alignment and (B) after peak alignment

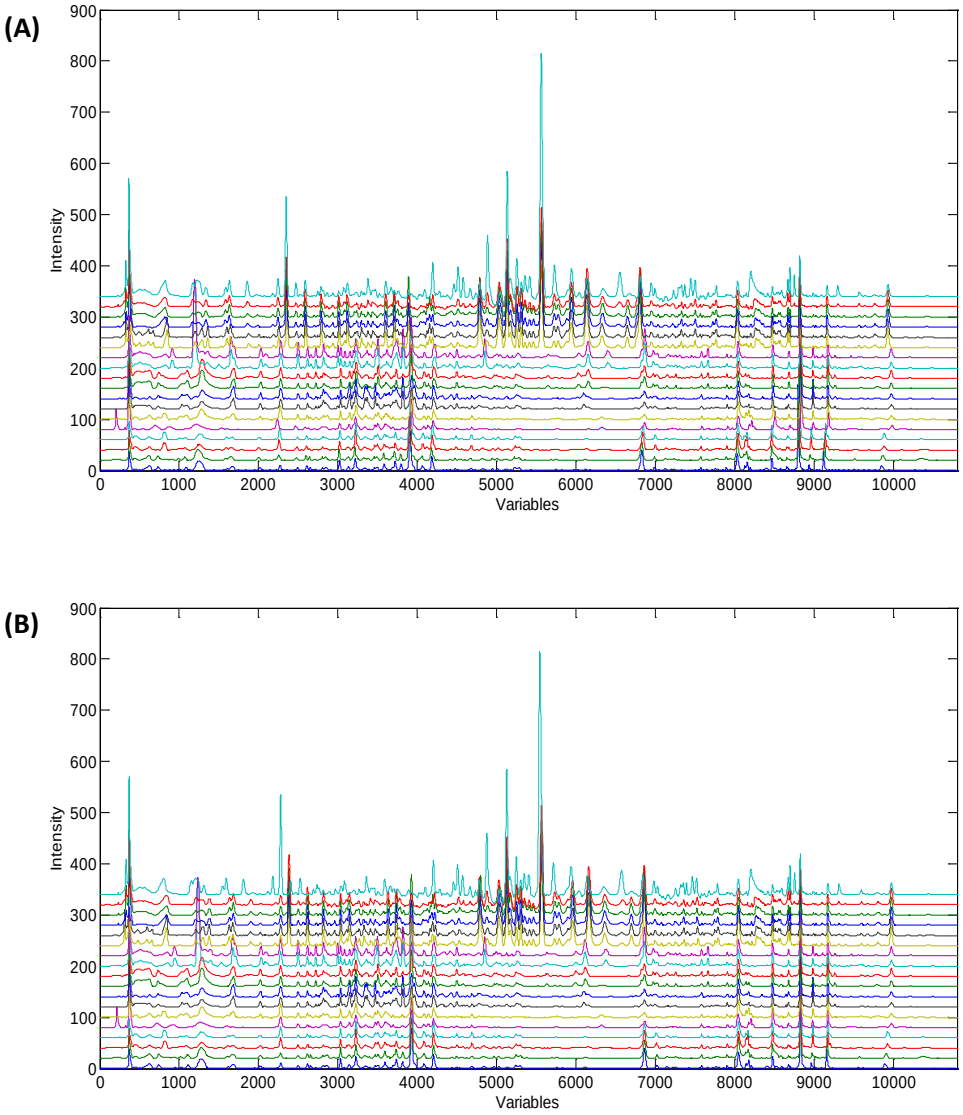


Figure 15. Results obtained in CSU at 254 nm. Spectra 1-12 DO and 13-18 D: (A) before peak alignment and (B) after peak alignment

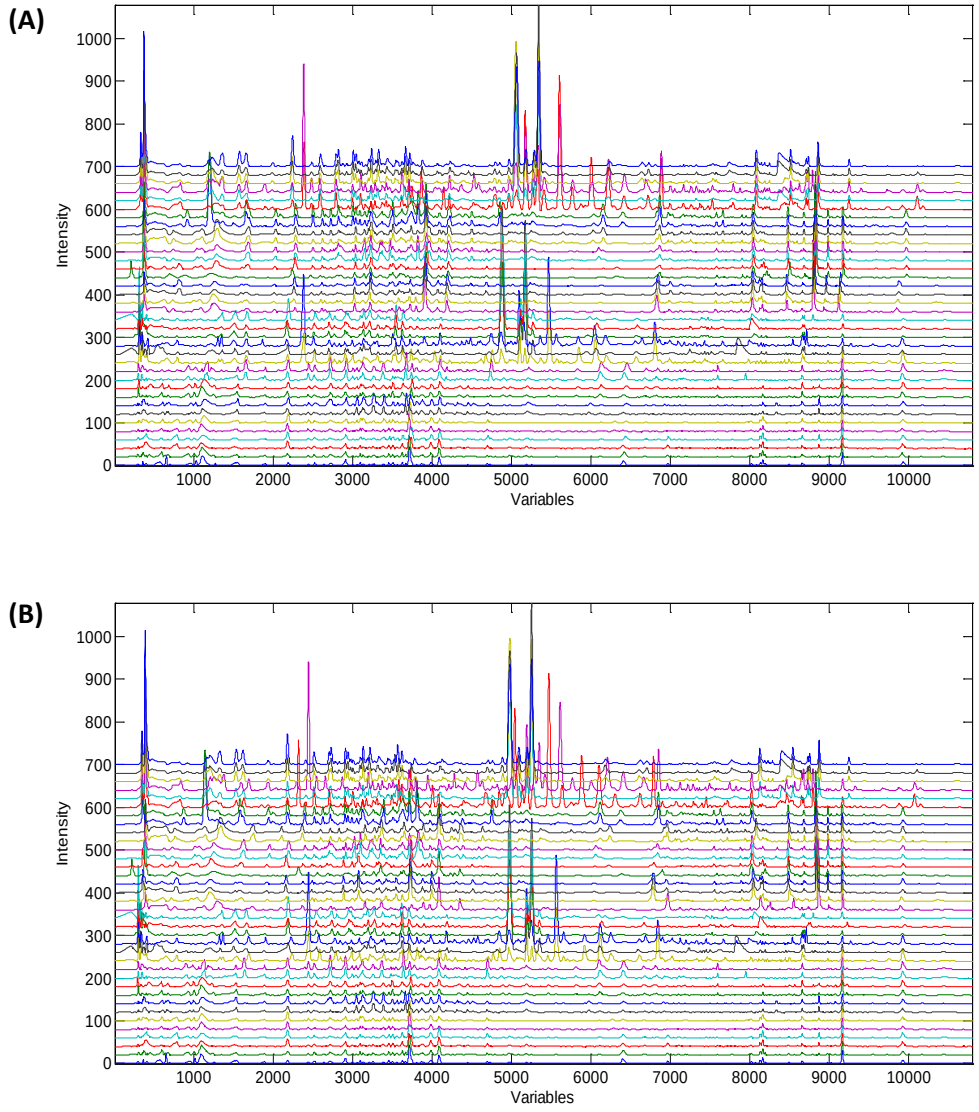


Figure 16. Results obtained in UiB and CSU together at 254 nm. Norway: 1-12 DO, 13-18 D and China: 19-30 DO and 31-36 D. (A) before peak alignment and (B) after peak alignment

3.1.2 Results using the sum of the data obtained at four different wavelengths: 254, 280, 310 and 335 nm

The following figures are shown to illustrate when the results of the four wavelengths are added before peak alignment (A) and after peak alignment (B). In **Figure 17** are shown the UiB results, in **Figure 18** the CSU results and in **Figure 19** the UiB and CSU results all together. From (A) to (B) it is also possible to see if there was some improvement with respect to the alignment. Here, the number of main peaks is reduced compared with the ones obtained when using only one wavelength. Taking a closer look to the fingerprints obtained in UiB and CSU separately it is possible to see that there are still some compounds with retention times in the variable range from ~4500 to ~5500 that corresponds to the range of ~40 to ~50 minutes that are present in *Dendrobii* samples but are not present in *Dendrobii Officinalis* samples. Therefore, it still possible to distinguish between the two species of herbal medicines. When using the sum of the wavelengths, the peak alignment for UiB and CSU together does not work so well as when using only one wavelength. When comparing the fingerprints this can maybe be explained due to the fact that there are fewer peaks when all the wavelengths results are added (**Figure 16** and **Figure 19**). There is also a slight deviation of the main peaks after 50 minutes retention time.

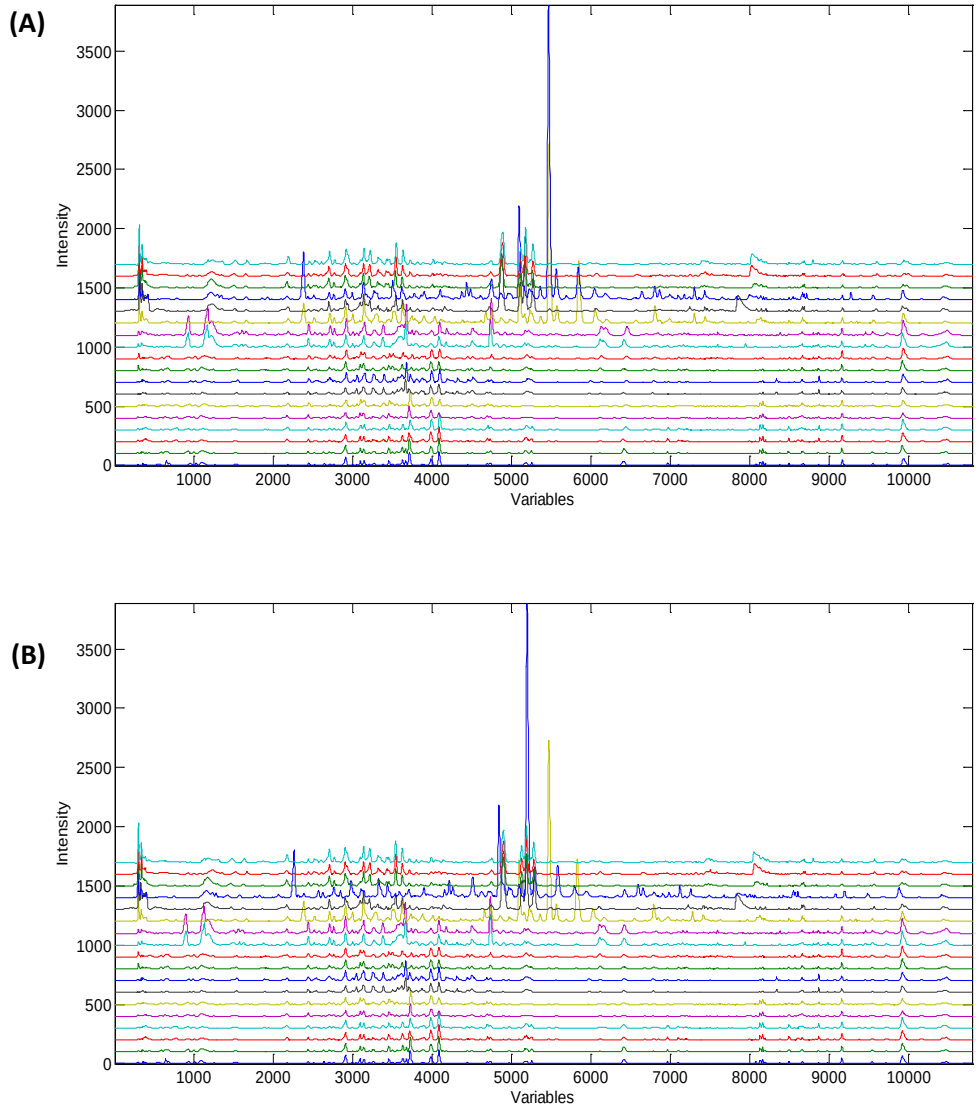


Figure 17. Results obtained in UiB after adding the 4 wavelengths (254, 280, 310 and 335 nm), 1-12 DO and 13-18 D: (A) before peak alignment and (B) after peak alignment

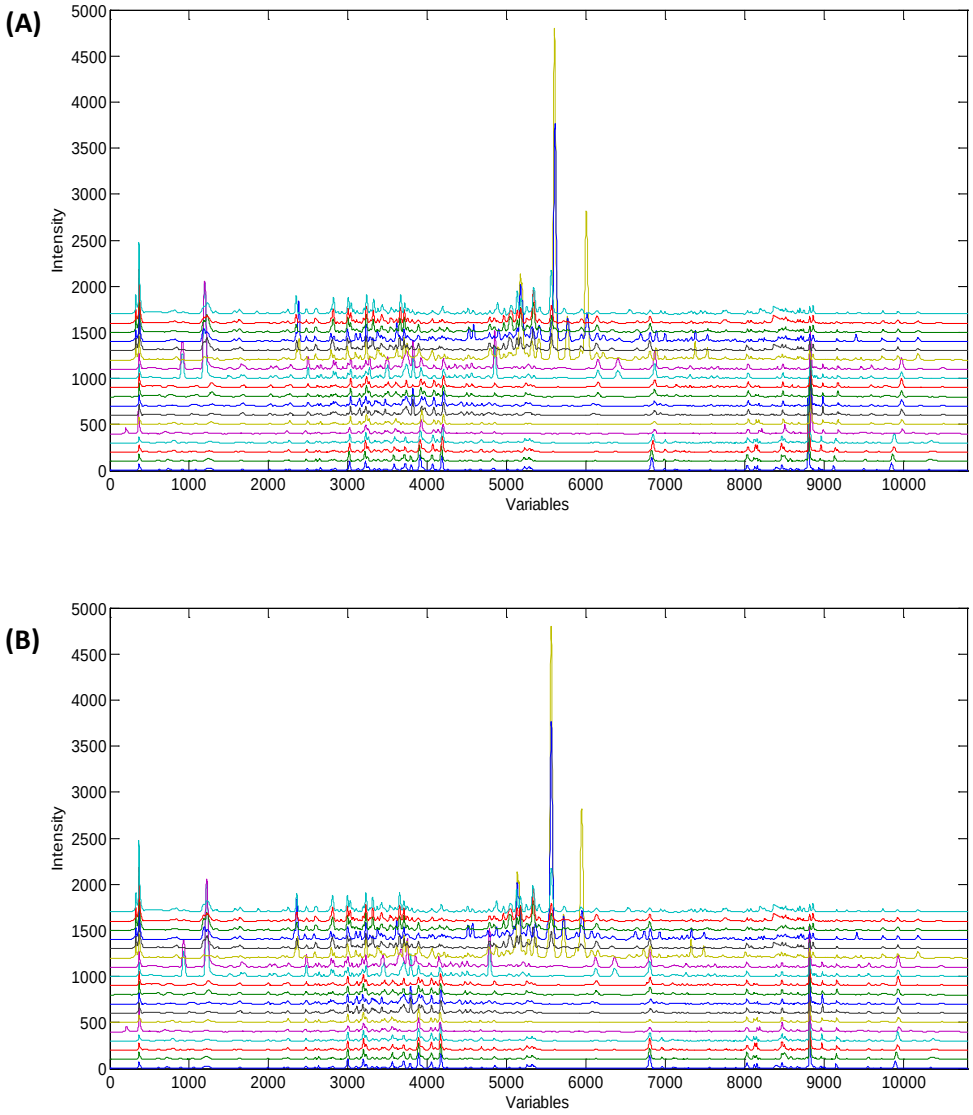


Figure 18. Results obtained in CSU after adding the 4 wavelengths (254, 280, 310, 335 nm), 1-12 DO and 13-18 D: (A) before peak alignment and (B) after peak alignment

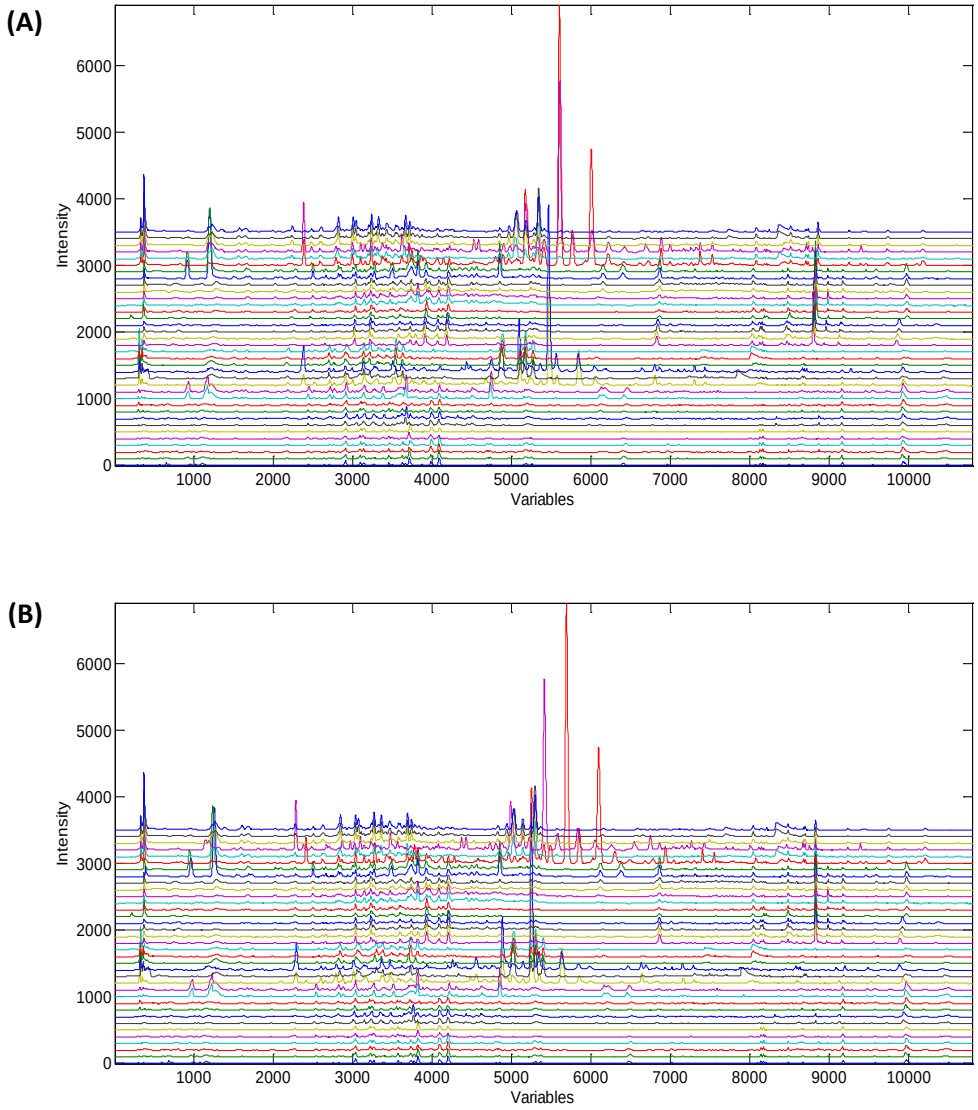


Figure 19. Results obtained in UiB and CSU together after adding the 4 wavelengths (254, 280, 310 and 335 nm). Norway: 1-12 DO, 13-18 D and China: 19-30 DO, 31-36 D. (A) before peak alignment and (B) after peak alignment

3.2 PCA

In order to evaluate the similarity between the two species of *Dendrobii*, PCA was performed on the data set obtained from the HPLC chromatograms. As the previous results, the Principal Component Analysis are also divided into the results obtained at one wavelength – 254 nm – and the sum of the results obtained at four different wavelengths – 254, 280, 310 and 335 nm.

3.2.1 Results using the data obtained at one wavelength: 254 nm

The next figures represent the score plots obtained by PCA where PC1 stands for the scores coordinates of principal component 1 and PC2 for the scores coordinates of principal component 2. The first two components kept a larger percentage variance explained and the third component contributed little to the separation.

The PCA model was built using the data matrix containing all the analyzed samples in each country (18×10801) or in both countries (36×10801).

The loading plots for the first two PC, that represent the object space projected and are used for interpreting relations among variables are not presented in this work because due to the large amount of variables it is not possible to have conclusive results. Therefore, in the figures below there are represented the Loadings vs Variables graphics for the first two PC, where it is possible to identify the variables that contribute most to each PC and then check how is this reflected in the raw data.

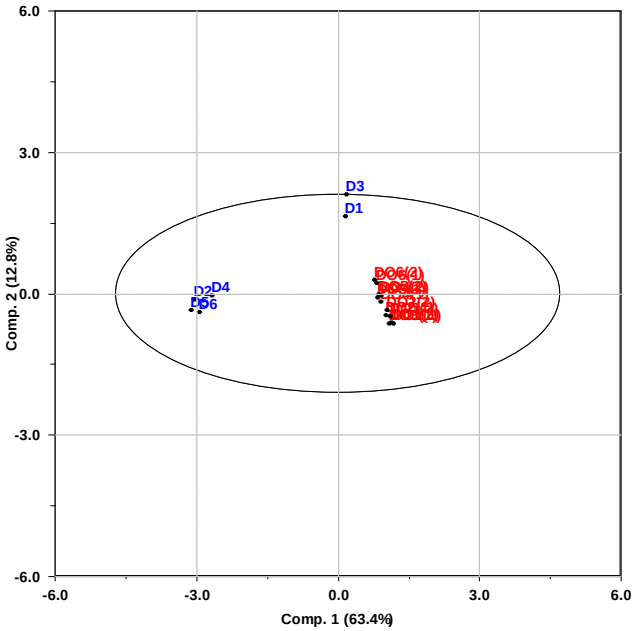
Looking at the results obtained in UiB before and after peak alignment (**Figure 20**) it is possible to see that the results are quite consistent. D and DO samples are mainly separated on the first PC. There are two possible outliers, i.e., samples D1 and D3 but according to Hotelling's T-squared statistic D3 is a clear outlier. The second PC is predominantly describing the difference between these two samples and the rest, where it is possible to see important variables ~5500 and in the raw data that represents a peak that appears in samples D1 and D3. The possible explanation for D3 is that it comes from a province (Sichuan) different from the majority. But here there

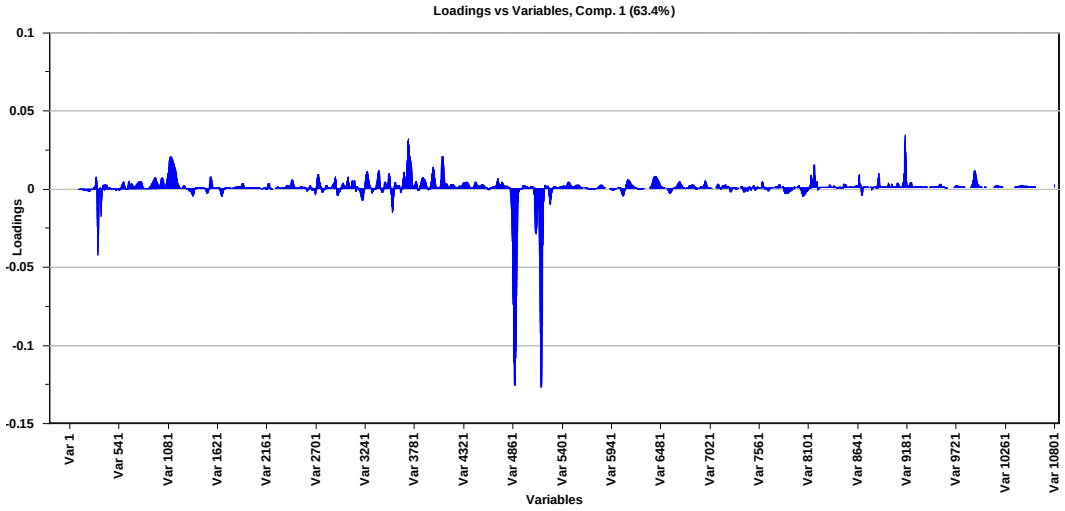
may be many reasons as explained in the introduction of this work such as the harvest season, the drying process or even some contaminations from pesticides or toxins.

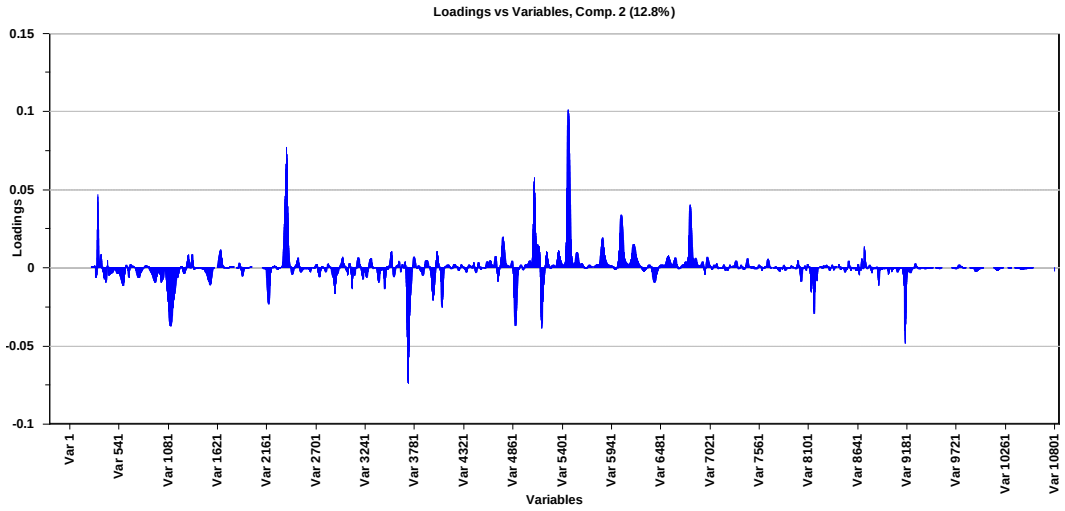
The distribution of the samples in the two groupings before and after peak alignment is almost the same and clearly shows the similarity between the *Dendrobii* samples and the similarity between the *Dendrobii Officinalis* samples. From before to after peak alignment it was also observed a very small improvement of the explained variance.

Looking at the Loadings vs Variables plot for the first component, the most important variables are located in the range from 4861 to 5401 and looking at the raw data in **Figure 14** it is possible to see that this is the main region where some main peaks appear in *Dendrobii* samples but not in *Dendrobii Officinalis* samples. To confirm these results and which compound corresponds to each peak the ideal situation would be to perform LC-MS.

(A)


6.0
3.0
0.0
-3.0
-6.0
Comp. 2 (12.8%)
D3
D1
D2
D4
D5
D6
-6.0
-3.0
0.0
3.0
6.0
Comp. 1 (63.4%)


Loadings vs Variables, Comp. 1 (63.4%)
0.1
0.05
0
-0.05
-0.1
-0.15
Loadings
Var 1
Var 541
Var 1081
Var 1621
Var 2161
Var 2701
Var 3241
Var 3781
Var 4321
Var 4861
Var 5401
Var 5941
Var 6481
Var 7021
Var 7561
Var 8101
Var 8641
Var 9181
Var 9721
Var 10261
Var 10801
Variables


Loadings vs Variables, Comp. 2 (12.8%)
0.15
0.1
0.05
0
-0.05
-0.1
Loadings
Var 1
Var 541
Var 1081
Var 1621
Var 2161
Var 2701
Var 3241
Var 3781
Var 4321
Var 4861
Var 5401
Var 5941
Var 6481
Var 7021
Var 7561
Var 8101
Var 8641
Var 9181
Var 9721
Var 10261
Var 10801
Variables

(B)

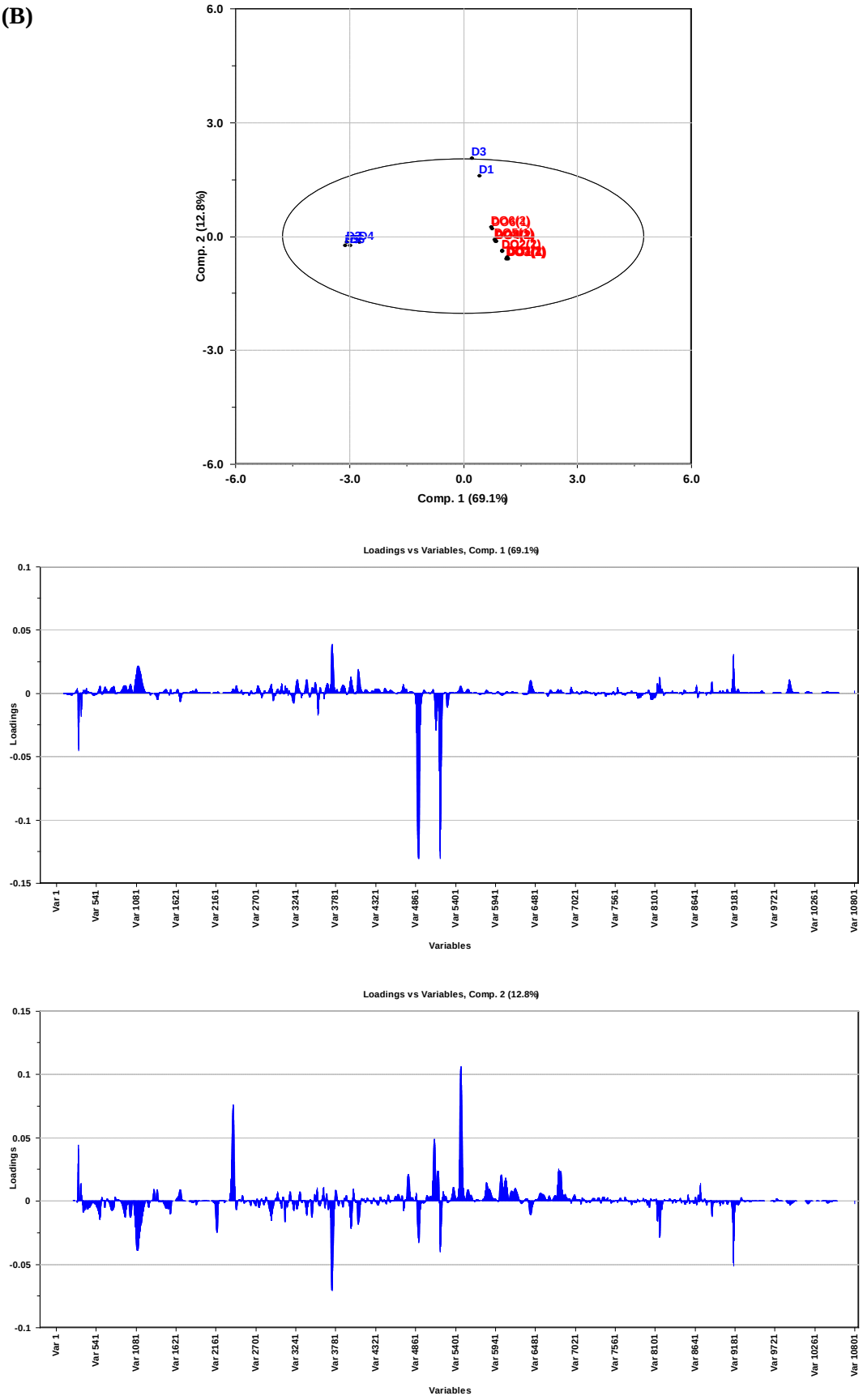
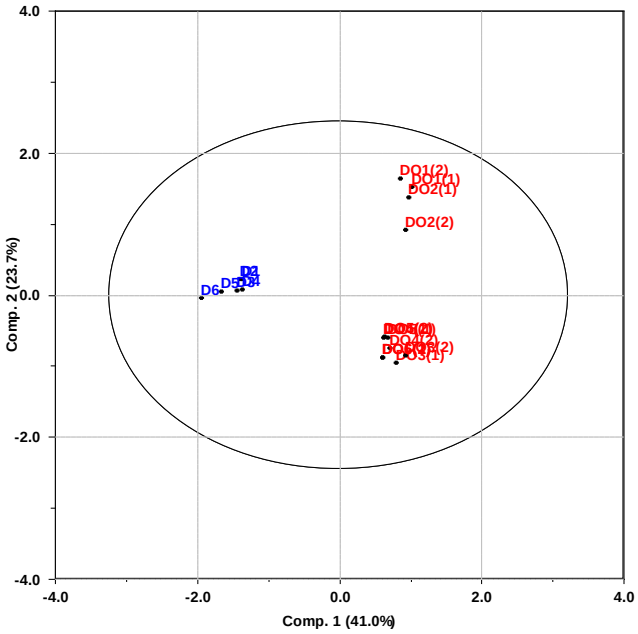
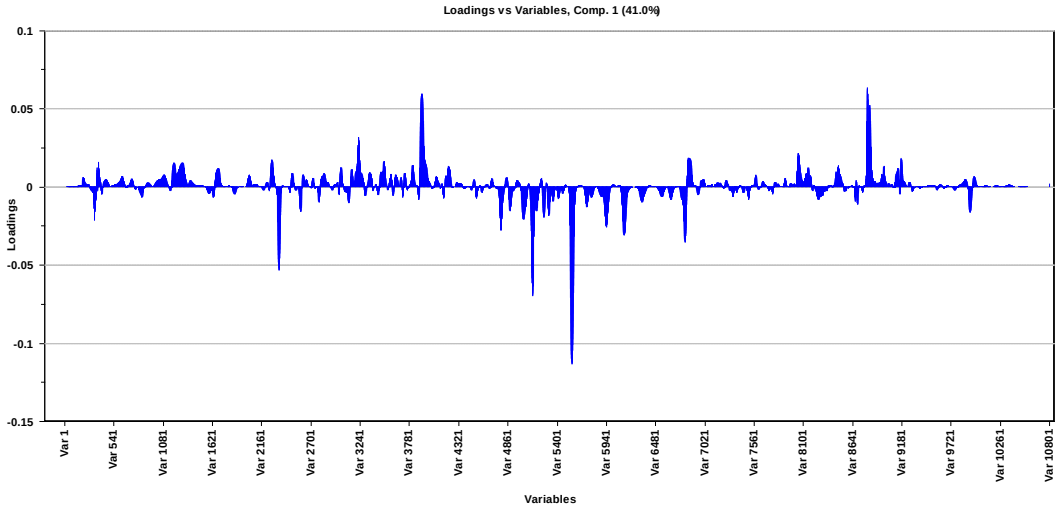


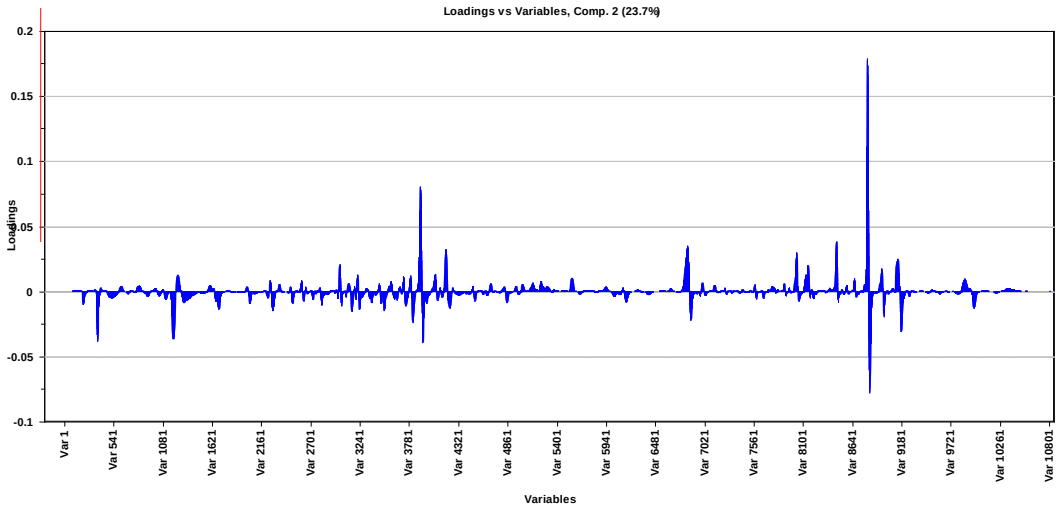
Figure 20. Results obtained in UiB: (A) before peak alignment with PC1–63.4% and PC2–12.8%; (B) after peak alignment, reference: sample DO5(1) PC1–69.1 % and PC2–12.8%. The figures below each score plot represent Loadings vs Variables for Comp. 1 and Comp.2

As expected, in the results obtained in CSU before and after peak alignment (**Figure 21**) it is possible to see two groupings, revealing the similarity between *Dendrobii* samples and between *Dendrobii Officinalis* samples. After peak alignment the similarity between samples remains consistent but most of the replicates are closer to each other and it was also observed a very small improvement of the explained variance. The groupings are mainly separated on the first PC and looking at the Loadings vs Variables, the variables with most significant loadings are located in the range from 4861 to ~5671. And again, looking at the raw data in **Figure 15** it is possible to see that this is the region where some peaks are observed for one kind of samples but not for the other.

(A)


4.0
2.0
0.0
-2.0
-4.0
Comp. 2 (23.7%)
DO1(2)
DO1(1)
DO2(1)
DO2(2)
D6
-4.0
-2.0
0.0
2.0
4.0
Comp. 1 (41.0%)


Loadings vs Variables, Comp. 1 (41.0%)
0.1
0.05
0
-0.05
-0.1
-0.15
Loadings
Var 1
Var 541
Var 1081
Var 1621
Var 2161
Var 2701
Var 3241
Var 3781
Var 4321
Var 4861
Var 5401
Var 5941
Var 6481
Var 7021
Var 7561
Var 8101
Var 8641
Var 9181
Var 9721
Var 10261
Var 10801
Variables


Loadings vs Variables, Comp. 2 (23.7%)
0.2
0.15
0.1
0.05
0
-0.05
-0.1
Loadings
Var 1
Var 541
Var 1081
Var 1621
Var 2161
Var 2701
Var 3241
Var 3781
Var 4321
Var 4861
Var 5401
Var 5941
Var 6481
Var 7021
Var 7561
Var 8101
Var 8641
Var 9181
Var 9721
Var 10261
Var 10801
Variables

(B)

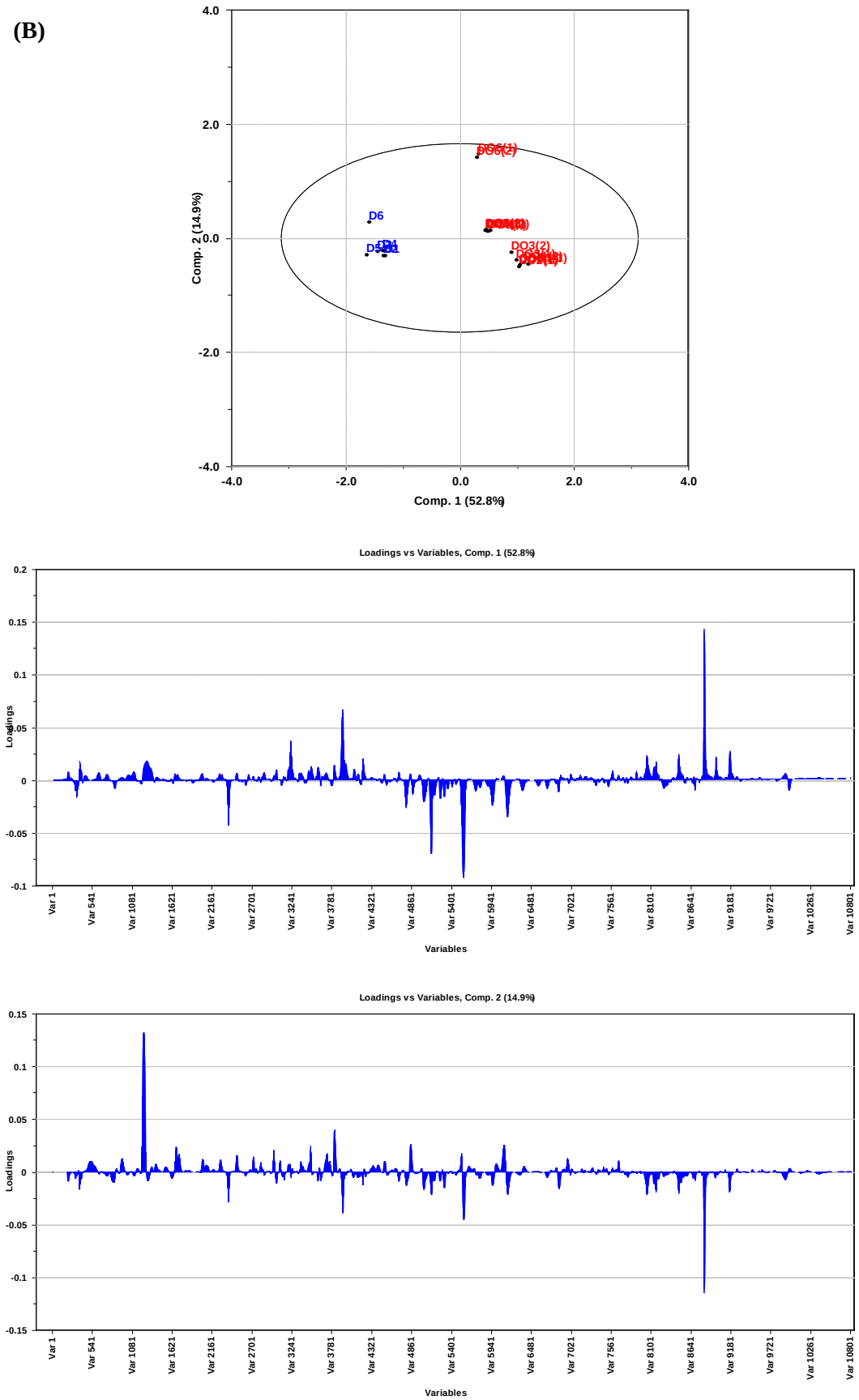


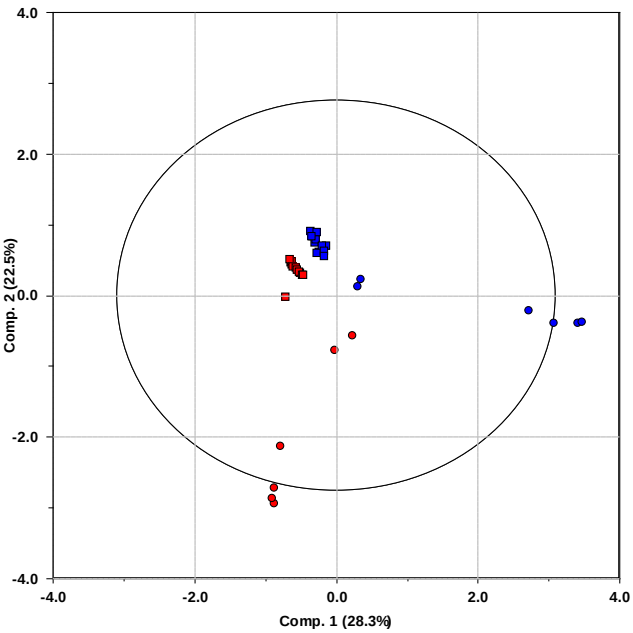
Figure 21. Results obtained in CSU: (A) before peak alignment PC1–41.0% and PC2–23.7%; (B) after peak alignment, reference: sample DO4(1) PC1–52.8% and PC2–14.9%. The figures below each score plot represent Loadings vs Variables for Comp. 1 and Comp.2

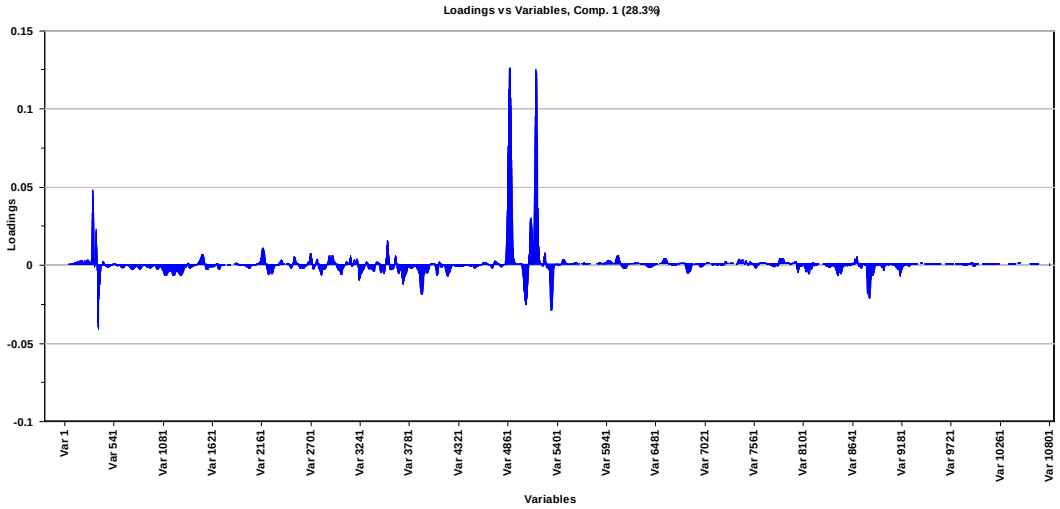
Analyzing the results for UiB and CSU all together (**Figure 22**), before and after peak alignment there is a separation between all the four different groups being investigated. Although it seems that some samples are considered outliers this may be due to the fact that when all the results are mixed together a large part of the variance is therefore unexplained – 49.2% and 42.7% remains in the residual matrix. It is known that models with a larger percentage variance explained are often more trustworthy when exploratory analysis is done. Looking at the Loadings vs Variables, the variables with most significant loadings are located in the range from 4861 to 5401 that corresponds in the raw data (**Figure 16**) to the same region where it is possible to see main peaks that are only present in one type of samples.

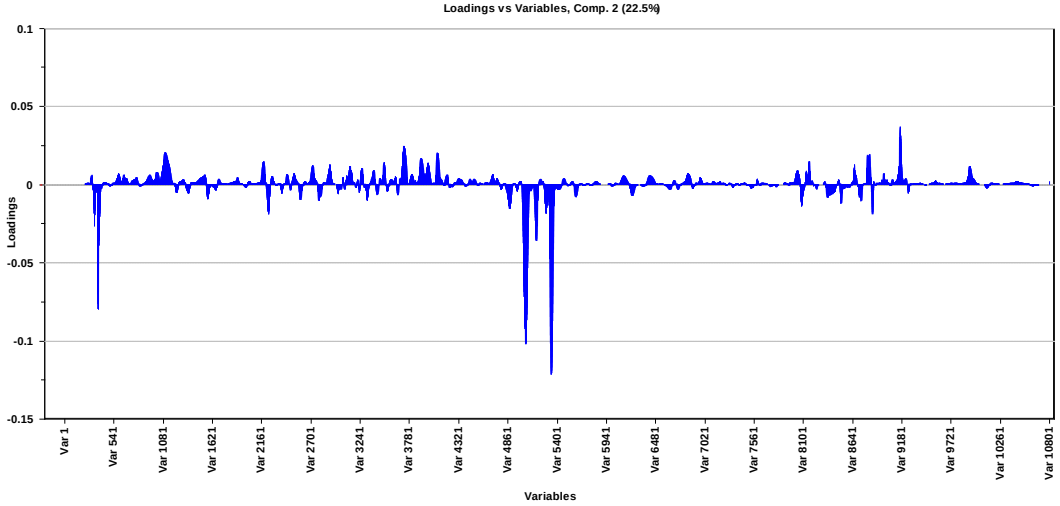
Here, the pre-processing of data was very important because the measurements obtained with different instruments were treated together and also the fact that the data was normalized it reduced the effect of signal strength, which correspond to a different concentration in each sample.

When analyzing the results obtained for UiB, the results for CSU and also the results for UiB and CSU together, there was some improvement for the variance explained by PC1 and PC2 after doing the peak alignment.

(A)


4.0
2.0
0.0
-2.0
-4.0
Comp. 2 (22.5%)
-4.0
-2.0
0.0
2.0
4.0
Comp. 1 (28.3%)


Loadings vs Variables, Comp. 1 (28.3%)
0.15
0.1
0.05
0
-0.05
-0.1
Loadings
Var 1
Var 541
Var 1081
Var 1621
Var 2161
Var 2701
Var 3241
Var 3781
Var 4321
Var 4861
Var 5401
Var 5941
Var 6481
Var 7021
Var 7561
Var 8101
Var 8641
Var 9181
Var 9721
Var 10261
Var 10801
Variables


Loadings vs Variables, Comp. 2 (22.5%)
0.1
0.05
0
-0.05
-0.1
-0.15
Loadings
Var 1
Var 541
Var 1081
Var 1621
Var 2161
Var 2701
Var 3241
Var 3781
Var 4321
Var 4861
Var 5401
Var 5941
Var 6481
Var 7021
Var 7561
Var 8101
Var 8641
Var 9181
Var 9721
Var 10261
Var 10801
Variables

(B)

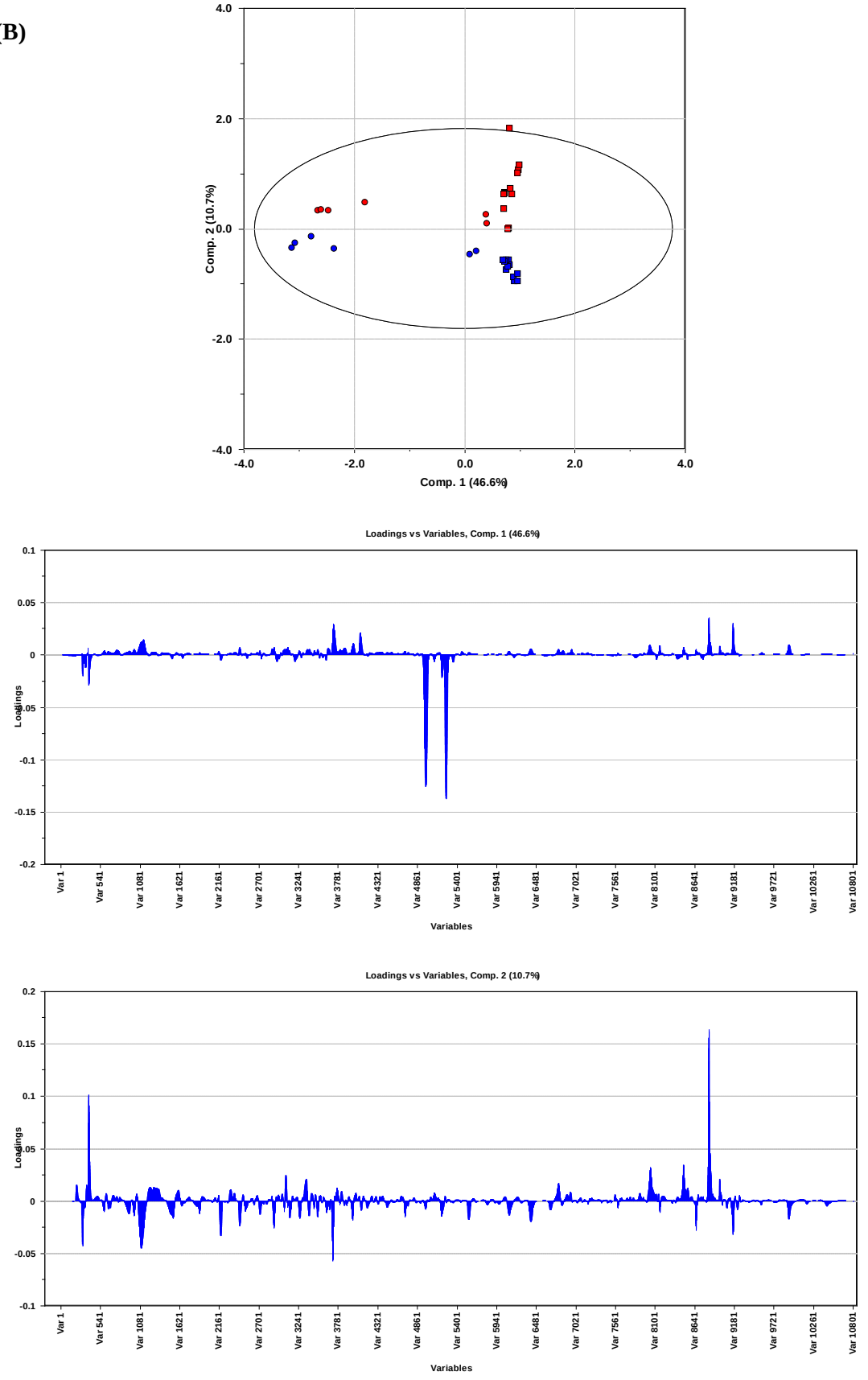


Figure 22. Results obtained for UiB and CSU together: (A) before peak alignment PC1–28.3% and PC2–22.5%; (B) after peak alignment, reference: sample DO2(1)N PC1–46.6% and PC2–10.7%. Blue color represents Norway, red color represents China, the squares represent DO samples and the circles represent D samples. The figures below each score plot represent Loadings vs Variables for Comp. 1 and Comp.2

3.2.2 Results using the sum of the data obtained at four different wavelengths: 254, 280, 310 and 335 nm

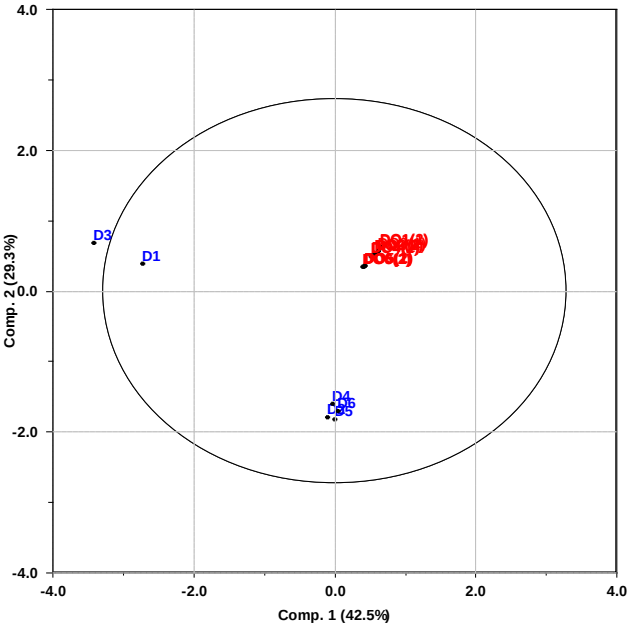
The next figures represent the score plots obtained after performing PCA and again, the first two components kept a larger percentage variance explained and the third component contributed little to the separation.

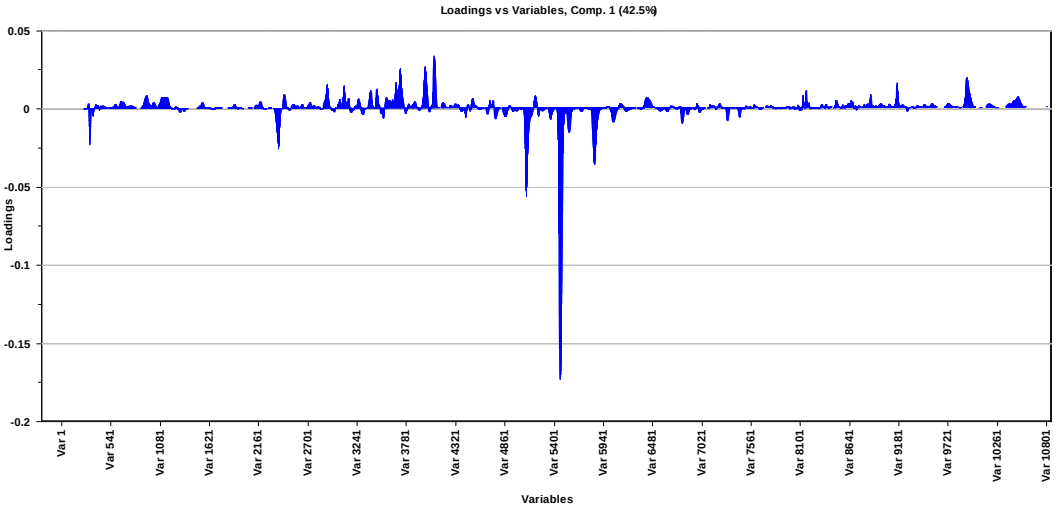
The PCA model was built using the data matrix containing all the analyzed samples in each country (18×10801) or in both countries (36×10801) but this time taking into account all the wavelengths.

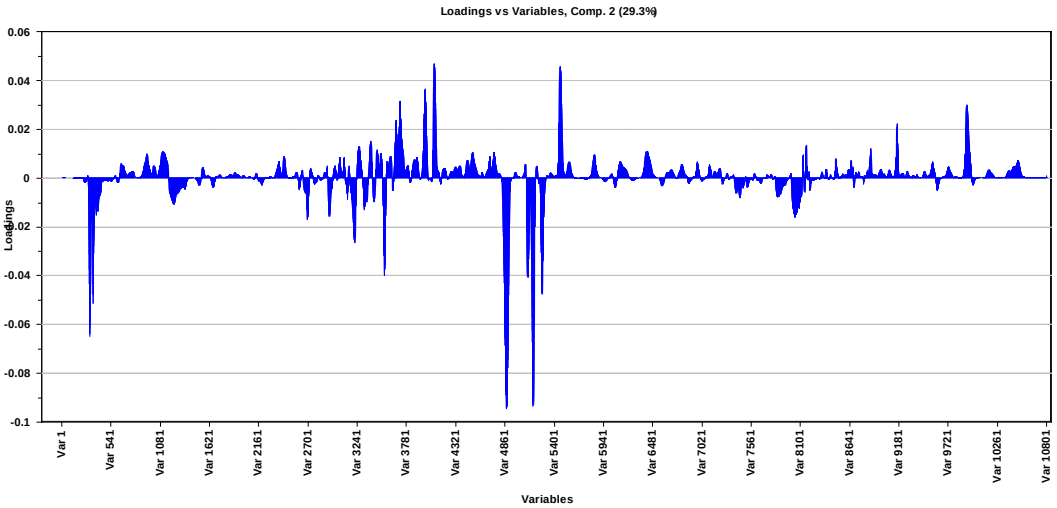
In the figures below there are represented the Loadings vs Variables graphics for the first two PC, where it is possible to identify the variables that contribute most to each PC and then check how is this reflected in the raw data.

From **Figure 23** it is possible to see that for the results obtained in UiB, before and after peak alignment, the *Dendrobii* samples are in one group while *Dendrobii Officinalis* samples are in another group. Therefore, samples D3 and D1 are observed as possible outliers also for the same reasons explained for the results obtained when using one wavelength. It was observed a decrease of the explained variance. The groups are mainly separate on the second PC and again in the variable range from 4861 to 5401 that corresponds in the raw data (**Figure 17**) to the region where some peaks are observed only in one kind of samples. These results are consistent with the ones obtained when using only one wavelength.

(A)


4.0
2.0
0.0
-2.0
-4.0
Comp. 2 (29.3%)
D3
D1
D4
D5
-4.0
-2.0
0.0
2.0
4.0
Comp. 1 (42.5%)


Loadings vs Variables, Comp. 1 (42.5%)
0.05
0
-0.05
-0.1
-0.15
-0.2
Loadings
Var 1
Var 541
Var 1081
Var 1621
Var 2161
Var 2701
Var 3241
Var 3781
Var 4321
Var 4861
Var 5401
Var 5941
Var 6481
Var 7021
Var 7561
Var 8101
Var 8641
Var 9181
Var 9721
Var 10261
Var 10801
Variables


Loadings vs Variables, Comp. 2 (29.3%)
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
Loadings
Var 1
Var 541
Var 1081
Var 1621
Var 2161
Var 2701
Var 3241
Var 3781
Var 4321
Var 4861
Var 5401
Var 5941
Var 6481
Var 7021
Var 7561
Var 8101
Var 8641
Var 9181
Var 9721
Var 10261
Var 10801
Variables

(B)

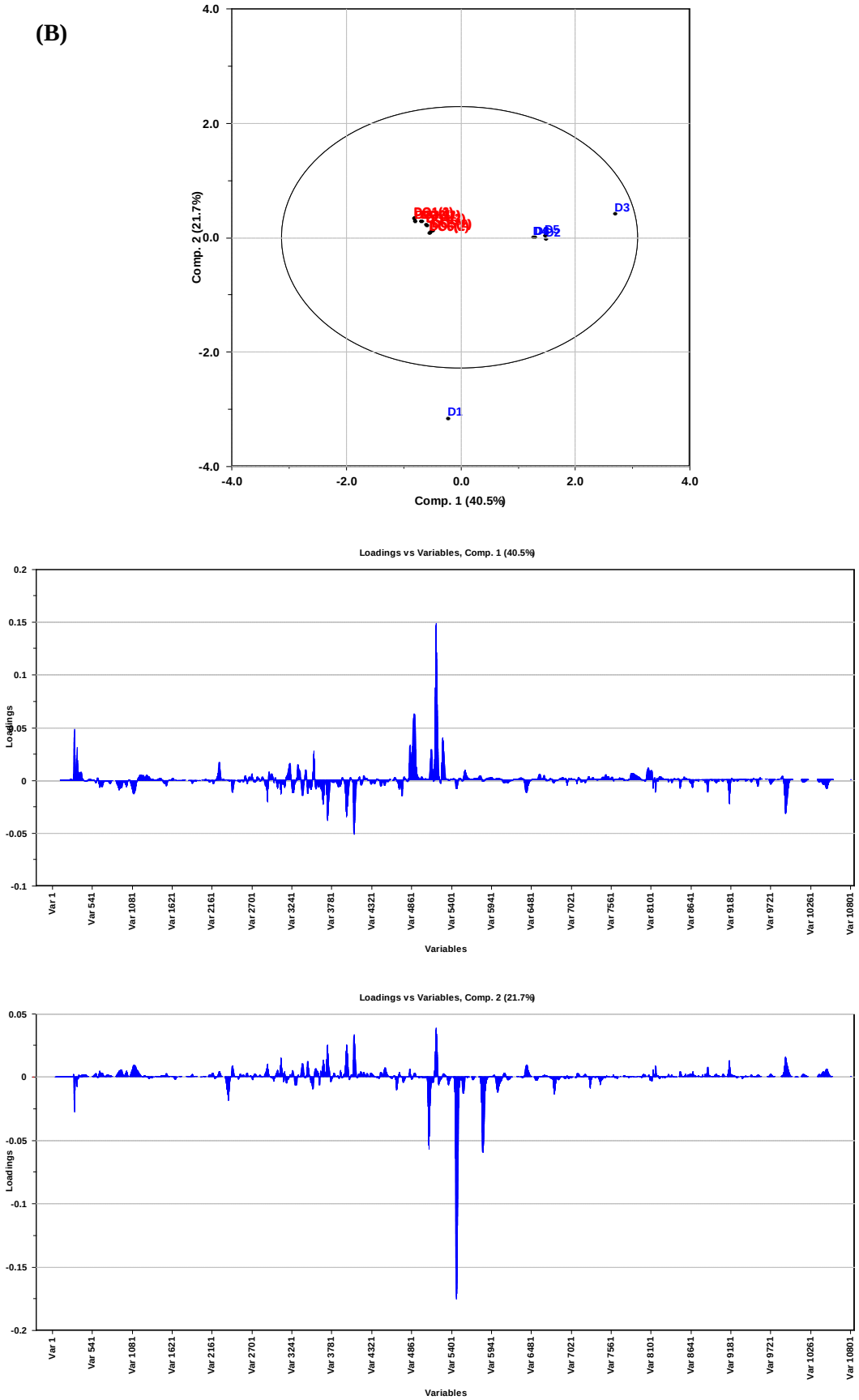
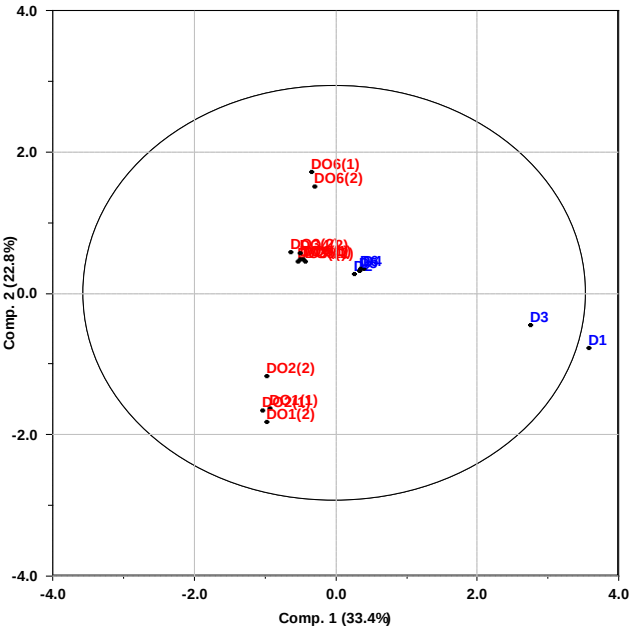


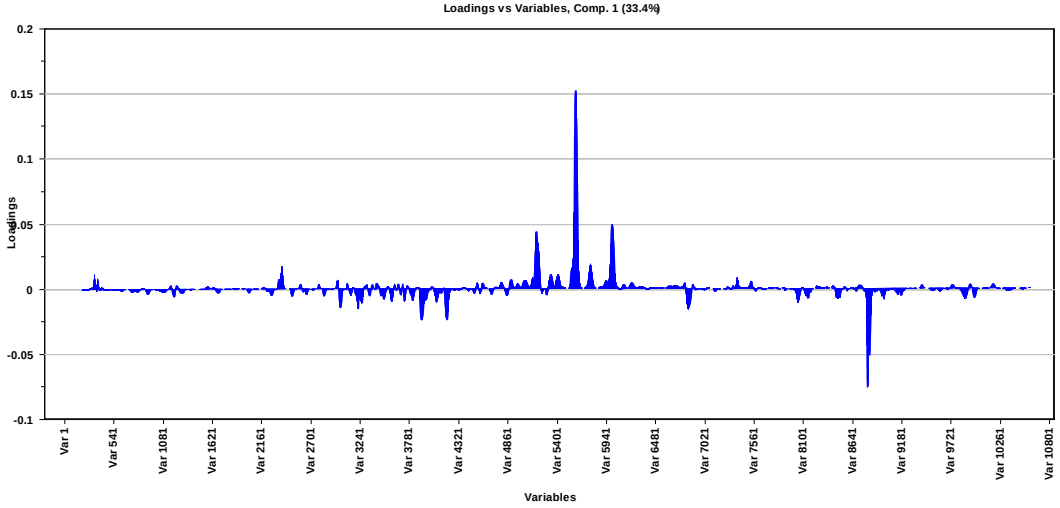
Figure 23. UiB: (A) before peak alignment PC1–42.5% and PC2–29.3%; (B) after peak alignment, reference: sample DO5(1) PC1–40.5% and PC2–21.7%. The figures below each score plot represent Loadings vs Variables for Comp. 1 and Comp.2

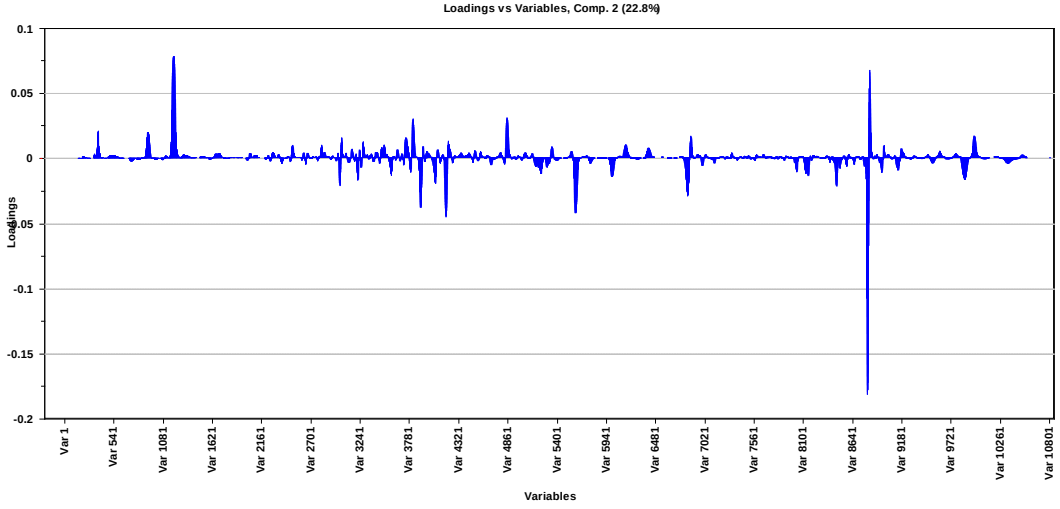
Looking at the results obtained in CSU (**Figure 24**), D1 is observed as an outlier either before or after peak alignment. Here the results are not so consistent with the ones obtained for one wavelength, since before they were not detected outliers. Although, it still possible to observe two groupings with the replicates closer after the peak alignment. From before to after peak alignment there was a slightly increase of the explained variance.

And again, the groupings are mainly separated on the first PC and looking at the Loadings vs Variables, the variables with most significant loadings are located in the range from 4861 to ~5671. Looking at the raw data (**Figure 18**) corresponds to the same range were differences are observed.

(A)


DO6(1)
DO6(2)
D3
D1
DO2(2)
DO1(2)
Comp. 2 (22.8%)
Comp. 1 (33.4%)


Loadings vs Variables, Comp. 1 (33.4%)
Loadings
Variables


Loadings vs Variables, Comp. 2 (22.8%)
Loadings
Variables

(B)

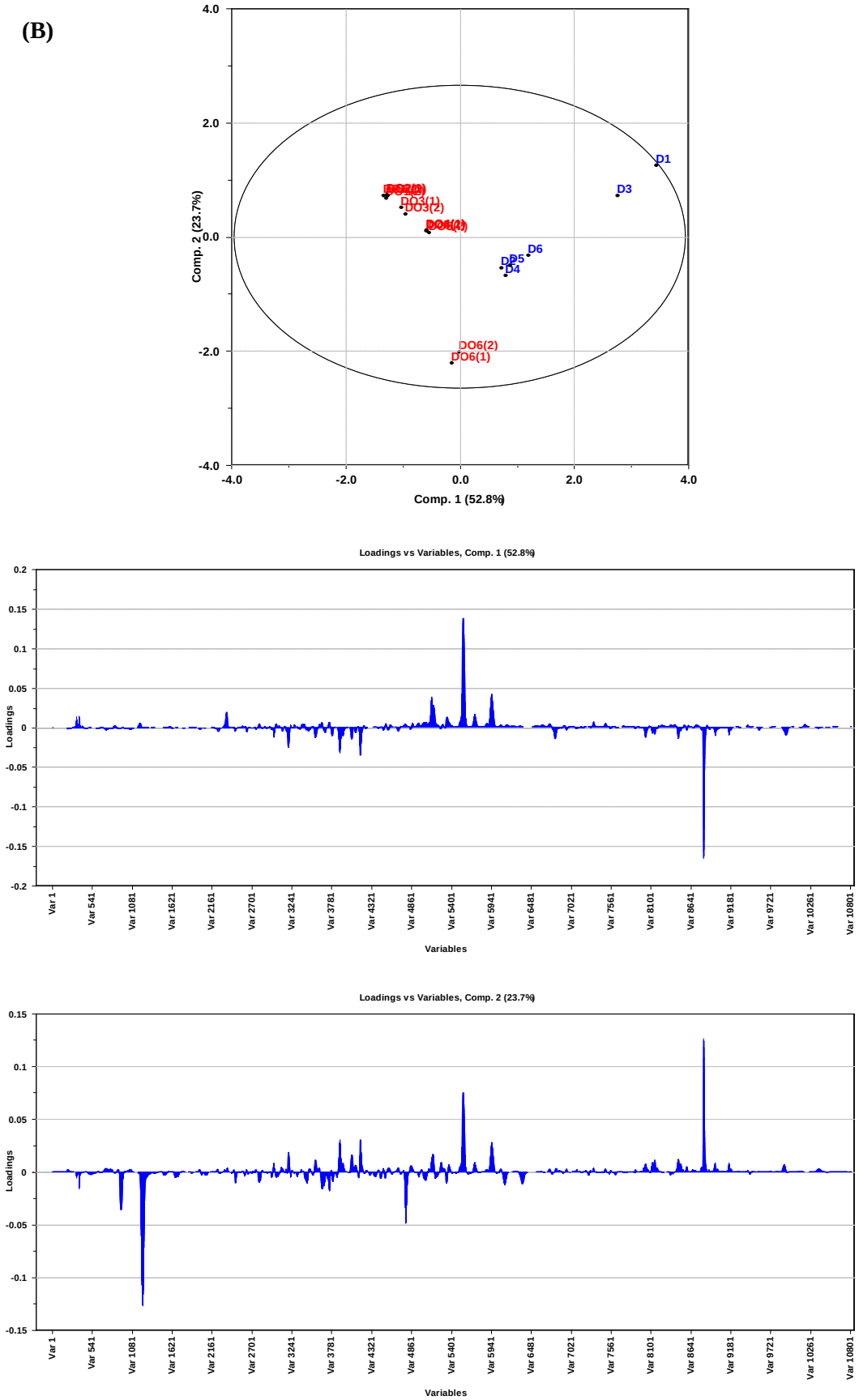
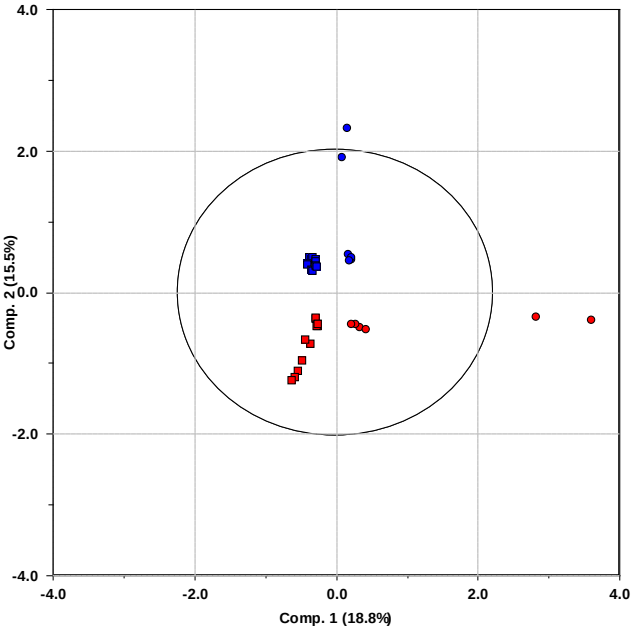


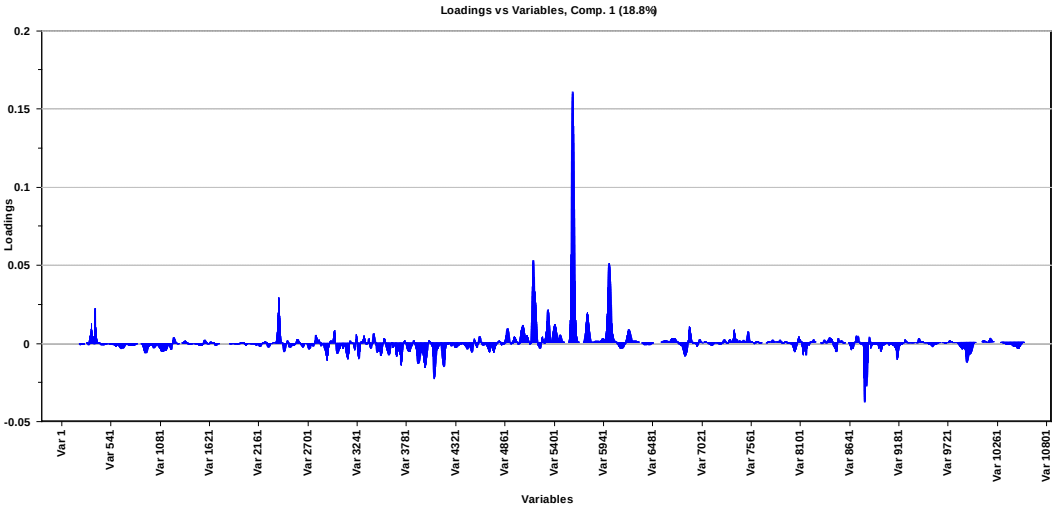
Figure 24. CSU: before peak alignment PC1–33.4% and PC2–22.8%; (B) after peak alignment, reference: sample D2N PC1–52.8% and PC2–23.7%. The figures below each score plot represent Loadings vs Variables for Comp. 1 and Comp.2

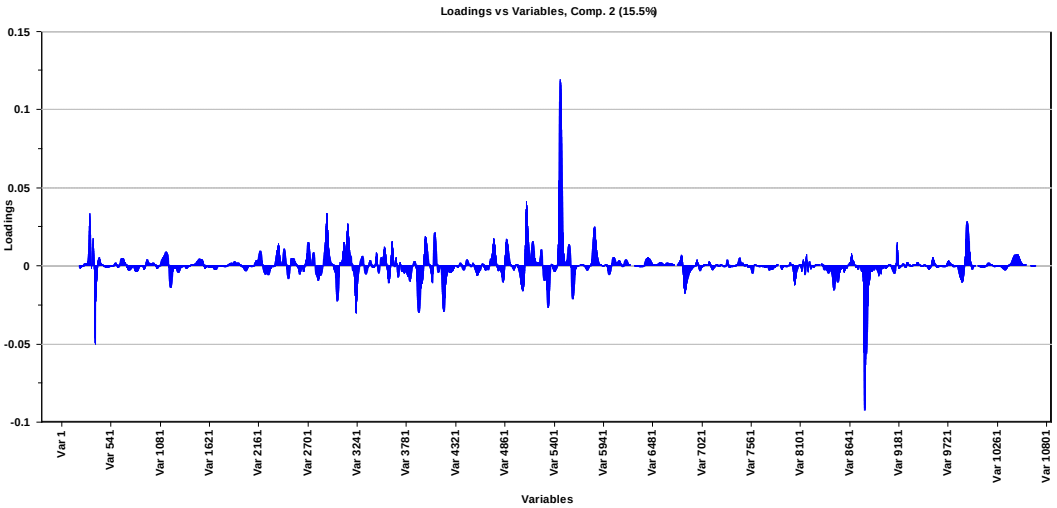
Looking at **Figure 25**, like when only one wavelength was used, the results show that the samples are separated in four groups before peak alignment and after peak alignment. Also here some samples are considered outliers this may be due to the fact that when all the results are mixed together a large part of the variance is therefore unexplained – 65.7% and 59.9% remains in the residual matrix. Looking at the Loadings vs Variables, the variables with most significant loadings are again located in the range from 4861 to 5941 that corresponds in the raw data (**Figure 19**) to the same region where it is possible to see main peaks that are only present in one type of samples.

When analyzing the results obtained for CSU and also the results for UiB and CSU together, there was some improvement for the variance explained by PC1 and PC2 after doing the peak alignment. Although, in the results for UiB it was observed a decrease in the variance explained.

(A)


4.0
2.0
0.0
-2.0
-4.0
Comp. 2 (15.5%)
-4.0
-2.0
0.0
2.0
4.0
Comp. 1 (18.8%)


Loadings vs Variables, Comp. 1 (18.8%)
0.2
0.15
0.1
0.05
0
-0.05
Loadings
Var 1
Var 541
Var 1081
Var 1621
Var 2161
Var 2701
Var 3241
Var 3781
Var 4321
Var 4861
Var 5401
Var 5941
Var 6481
Var 7021
Var 7561
Var 8101
Var 8641
Var 9181
Var 9721
Var 10261
Var 10801
Variables


Loadings vs Variables, Comp. 2 (15.5%)
0.15
0.1
0.05
0
-0.05
-0.1
Loadings
Var 1
Var 541
Var 1081
Var 1621
Var 2161
Var 2701
Var 3241
Var 3781
Var 4321
Var 4861
Var 5401
Var 5941
Var 6481
Var 7021
Var 7561
Var 8101
Var 8641
Var 9181
Var 9721
Var 10261
Var 10801
Variables

(B)

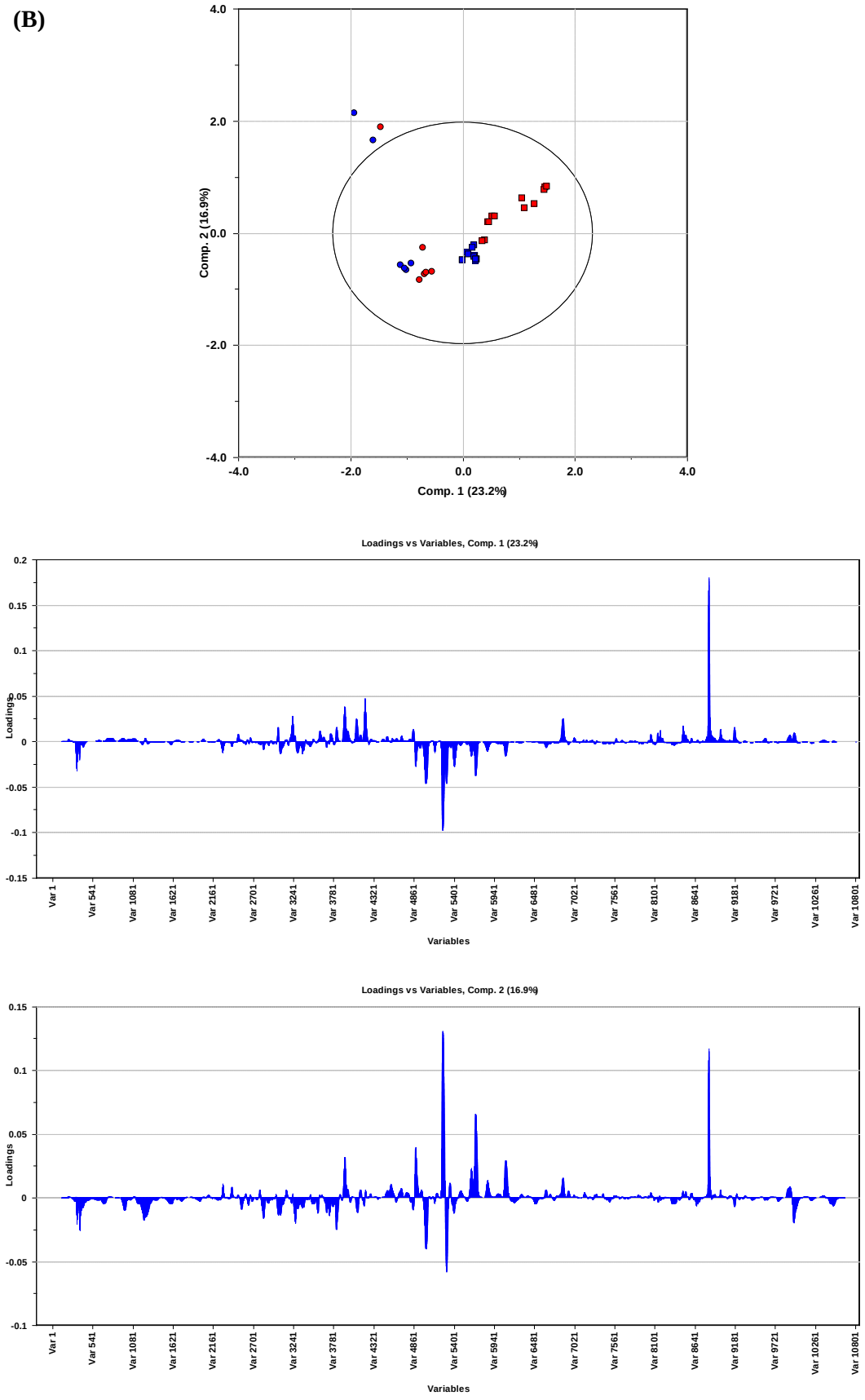


Figure 25. CSU and UiB together: (A) before peak alignment PC1–18.8% and PC2–15.5%; (B) after peak alignment, reference: sample DO4(2)C PC1–23.2% and PC2–16.9%. Blue color represents Norway, red color represents China, the squares represent DO samples and the circles represent D samples. The figures below each score plot represent Loadings vs Variables for Comp. 1 and Comp.2

Looking at **Table 6** and taking into account before and after peak alignment, the only results that had an improvement of the explained variance in both PCs were the results obtained in CSU and UiB-CSU together, whether using one or four wavelengths.

When analyzing the results in UiB, using the sum of four wavelengths, there was a slight decrease in the variance explained by the two PCs.

Table 6 - Summary of PCA results before (*) and after (**) peak alignment

University		**One wavelength** (254 nm)	**Four wavelengths** (254, 280, 310, 335 nm)
UiB*	PC1	63.4%	42.5%
	PC2	12.8%	29.3%
UiB**	Reference	DO5(1)	DO5(1)
	PC1	69.1%	40.5%
	PC2	12.8%	21.7%
CSU*	PC1	41.0%	33.4%
	PC2	23.7%	22.8%
CSU**	Reference	DO4(1)	D2N
	PC1	52.8%	52.8%
	PC2	14.9%	23.7%
UiB-CSU*	PC1	28.3%	18.8%
	PC2	22.5%	15.5%
UiB-CSU**	Reference	DO2(1)N	DO4(2)C
	PC1	46.6%	23.2%
	PC2	10.7%	16.9%

3.3 PLS-DA

The Partial Least Square results are also divided into the results obtained at one wavelength – 254 nm – and the sum of the results obtained at four different wavelengths –254, 280, 310 and 335 nm.

3.3.1 Results using the data obtained at one wavelength: 254 nm

As explained in the theory section, classification problems in fingerprints data analysis are complex due to the many variables and few objects issue. This makes that many solutions can be found to separate the classes. The PLS-DA score plots as showed in most classification applications present an overoptimistic view of the separation between the classes.

To avoid this issue, a new subset was created where the variables with selectivity ratio less than a value obtained by an F-test were removed. The replicate objects and an object from each *Dendrobii* species were also removed from the new subset.

From theory, it is also known that the permutation testing and cross model validation are used to assess the validation of classification models. Permutation tests show that when cross validation is not applied appropriately, it leads also to overoptimistic results. In cross validation, parts of the data are held out and a model is built on the remaining. This process is repeated until all data has been kept out once.

In this case, manual cross validation was used because if automated cross validation was used there might be the risk of removing the objects from the same cluster due to the small amount of objects.

The replicates and an object from each kind of *Dendrobii* species removed from the new subset were used to test the predictive properties of the model. Ideally new sample analysis should be done in the laboratory to later test the models.

After following all the steps to build a PLS model (explained in section **1.1.4.2**), it was possible to obtain some conclusions. As an example, these steps will be explained

and shown in detail for the PLS model for CSU after peak alignment. For the other data sets only final results are shown, as detailed results for all data sets would make the thesis prohibitively long.

The next figures represent the score plots for the performed PLS-DA, the selectivity ratio plot, and finally a comparison between the actual and predicted class membership. In the axis labels of the score plot it is shown the percentage of explained variance by each latent variable for the independent variable (x) and for the dependent variable (y), respectively. PLS-DA was performed to further investigate the separation between the two *Dendrobii* species.

In **Figure 26** it is possible to see a clear separation between the two clusters that represent the different kind of samples. From **Figure 27** it is possible to see that the selectivity ratio corresponds to several specific, continuous regions in the raw data containing peaks. Again, to further investigation these peaks should be identified with another technique such as LC-MS. The predictive models, as represented in **Figure 28**, show good results and are better when SEP is smaller.

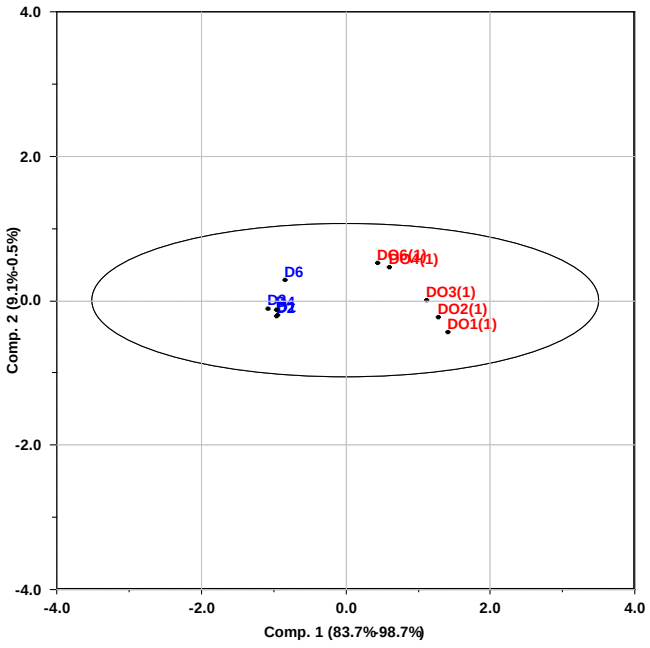


Figure 26. PLS-DA score plots of the first two latent variables for samples tested in CSU after peak alignment. Objects of class -1 (*Dendrobii*) are labeled in blue and objects of class 1 (*Dendrobii Officinalis*) are labeled in red

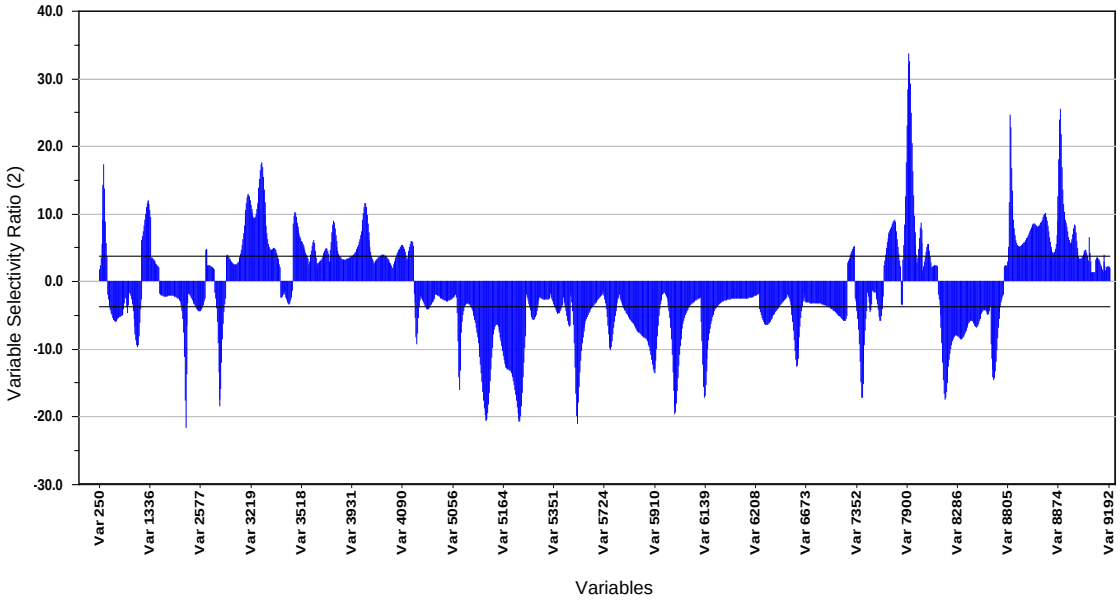


Figure 27. Graphic of variables vs Variable selectivity ratio (3.73 as limit)

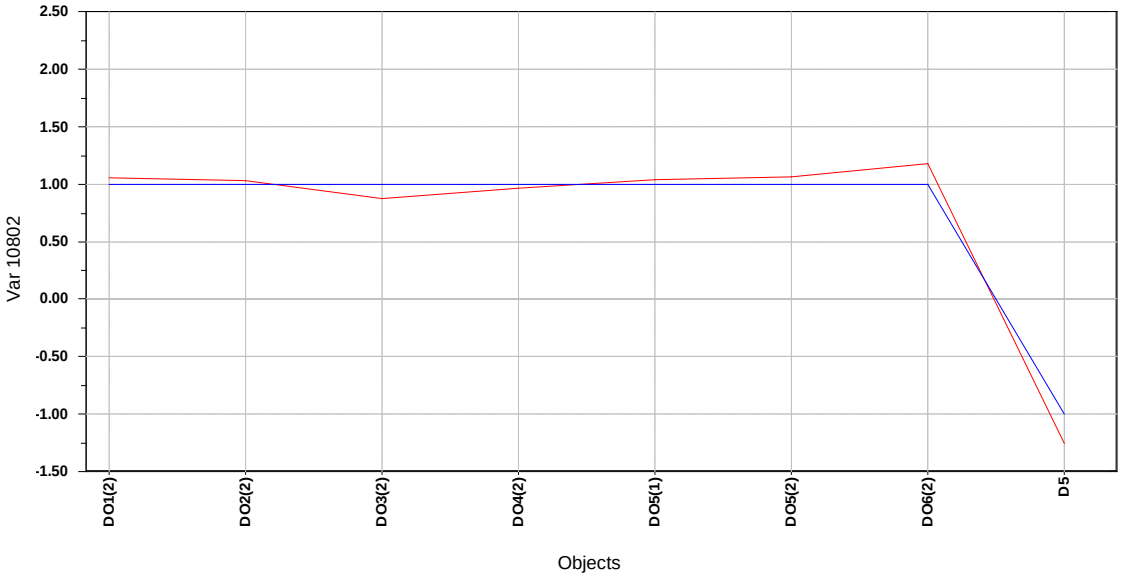


Figure 28. Graphic representation of Predicted (red) and Measured (blue) for Var 10802, SEP = 0.133, Comp. 3

In **Table 7** are shown all the results for UiB and CSU for one wavelength.

The initial subset have always 18 or 36 objects and 10802 variables, after removing the six replicates and one sample from each kind of samples the new subsets have 10

or 20 objects. The number of variables is different according to the value obtained by an F-test for the selectivity ratio.

Also represented in the next table, CsvSD is the ratio between the total prediction error of a model after including a new latent variable, and the total residual standard deviation before this inclusion. The ratio is multiplied by a correction factor to compensate for loss of degrees of freedom when the new latent variable is added. If the resulting number is less than one, the latent variable is significant and included in the model and the procedure continues with the calculation of a further latent variable. This criterion is used in classification and regression/response modeling.

Table 7 - Results obtained for UiB and CSU when one wavelength was used, before (*) and after (**) peak alignment

Analysis	Objects removed	Nr variables in the new subset	LV1 (CsvSD)	LV2 (CsvSD)	Standard Error of Prediction (SEP)
UiB*	Replicates and D1 and DO1 (1)	475	90.2%–97.7% (0.17)	2.4%– 0.9% (1.39)	0.136
UiB**	Replicates and D3 and DO3(1)	615	85.5%–95.4% (0.24)	4.9%– 3.2% (1.09)	0.001
CSU*	Replicates and D2 and DO2(1)	1108	88.2%–90.4% (0.33)	8.7%–6.9% (0.81)	0.108
CSU**	Replicates and D5 and DO5(1)	1384	83.7%–98.7% (0.14)	9.1%–0.5% (0.94)	0.133
UiB-CSU*	Replicates and DO1(1)N, DO1(1)C, D1N and D1C	25	91.7%–83.3% (0.42)	7.3%–4.4% (0.91)	0.631
UiB-CSU**	Replicates and DO2(1)N, DO2(1)C, D2N and D2C	82	79.5%–94.5% (0.25)	7.9%–1.0% (1.15)	0.227

3.3.2 Results using the sum of the data obtained at four different wavelengths: 254, 280, 310 and 335 nm

In the results obtained for the four wavelengths (**Table 8**), they are consistent with the ones obtained for only one wavelength, where the final predictive properties seem to be good due to the low values for SEP. And also here the selectivity ratio shows specific regions in the raw data that could help distinguish between the two species of *Dendrobii*.

Table 8 - Results obtained for UiB and CSU when four wavelengths were used, before (*) and after (**) peak alignment

Analysis	Objects removed	Nr variables in the new subset	LV1 (CsvSD)	LV2 (CsvSD)	Standard Error of Prediction (SEP)
UiB*	Replicates and D2 and DO2 (1)	553	85.0%–97.0% (0.19)	5.8%–1.0% (1.52)	0.163
UiB**	Replicates and D1 and DO(1)	817	87.3%–91.5% (0.37)	3.8%–6.9% (0.80)	0.272
CSU*	Replicates and D2 and DO2(1)	610	81.5%–94.9% (0.26)	8.6%–4.4% (0.72)	0.181
CSU**	Replicates and DO4(1) and D4	756	96.3%–89.3% (0.38)	97.9%–10.0% (0.50)	0.167
UiB-CSU*	Replicates and DO6(1)N, DO6(1)C, D6N and D6C	23	94.8%–74.7% (0.53)	4.9%–1.3% (1.05)	0.947
UiB-CSU**	Replicates and DO3(1)N, DO3(1)C, D3N and D3C	149	78.0%–92.9% (0.29)	13.5%–0.8% (1.12)	0.416

4. CONCLUSIONS

The HPLC method developed during this work can distinguish between *Dendrobii* and *Dendrobii Officinalis,* an identification that can be difficult using only visual inspection. Although, during the extraction process it is possible to observe a small difference in color between the two species, the *Dendrobium* samples showed a yellow-brownish color whereas *Dendrobium Officinalis* samples displayed a greenish color. The extraction process and the sample preparation are relatively simple processes. The main advantages of these processes are low solvent consumption, relatively short extraction time, high extraction efficiency, stability and good repetitiveness.

This HPLC method was developed to separate flavonoids and/or other phenolic components present in these *Dendrobii* species with good resolution. Based on this separation, an efficient chromatographic fingerprint of these species was established.

The pre-processing of data was very important due to the fact that the measurements obtained with different instruments are being treated together and this pre-processing is responsible for removing random errors from quantitative information. When analyzing the fingerprints, the automated alignment of chromatographic data it is also important in order to correct unwanted time-shifts.

In the twelve samples of *Dendrobii* and *Dendrobii Officinalis* collected from five different provinces in China, around 100 characteristic fingerprint peaks were detected and it was found that some compounds with retention times in the range from 40 to 50 min appear in *Dendrobii* fingerprints but not in *Dendrobii Officinalis* fingerprints. This could help to distinguish between the two different species of this herbal medicine. When using the sum of the four wavelengths, the number of main peaks is reduced compared with the ones obtained when using only one wavelength but the general conclusions are quite the same.

Looking at PCA results, the distribution of the samples in the two groupings before and after peak alignment is almost the same and clearly shows the similarity between the *Dendrobii* samples and the similarity between the *Dendrobii Officinalis* samples.

From before to after peak alignment it was also observed a small improvement of the explained variance.

Looking at the Loadings vs Variables plots showed where the most important variables were located looking at the raw data it was possible to see that this was the main region where several main peaks appear in *Dendrobii* samples but not in *Dendrobii Officinalis* samples.

Regarding PLS results, there is a regular relationship between the *Dendrobii* samples and between the *Dendrobii Officinalis* samples. A clear separation between the two clusters was observed. In the results obtained for one wavelength or even four wavelengths, the final predictive properties of the models seem to be quite good due to the low values obtained for SEP. The selectivity ratio shows specific regions in the raw data that could help distinguish between the two species of *Dendrobii*. To further investigate these main differences, some additional analysis should be done such as LC-MS.

So, it can be considered that the results obtained with PLS-DA are consistent with the similarity analysis obtained with PCA.

The fingerprinting quantitative analysis combining similarity evaluation, PCA and PLS cluster analysis is a valid and relatively rapid method for classification of herbal medicine species. Data obtained from this study suggest that the use of HPLC-PCA-PLS-DA can identify and distinguish the two different species of *Dendrobii*. The advantage of using this method is that it is often unnecessary to know the individual components that build the fingerprint. So, this represents a relatively rapid and efficient process for assessment. The fingerprint method established by this study could be applied to other similar *Dendrobii* species for the quality assessment.

5. FURTHER WORK

With the purpose of identifying at least the main peaks detected in *Dendrobii* species but not in *Dendrobii Officinalis* species and also to confirm that the components present in the mixture, that are expected to be flavonoids and/or other phenolic compounds, LC-MS analysis should be performed. Initially, one of the aims of the study (**section 1.2**) was 'to perform HPLC-MS analysis of *Dendrobii Caulis* (Shihu) and *Dendrobii Officinalis Caulis* (Tiepi Shihu) to identify the main peaks' but for technical reasons beyond my control, i.e., the instrument was broken and awaiting repair it was not possible to realize these analysis.

The isolation of the main peaks could also be done to perform UV spectra in order to confirm that some characteristic bands of flavonoids (between 210 and 400 nm) would appear.

Finally, the validation of the analytical method should also be done. This is the process of establishing its performance characteristics and limitations and also the identification of the influences which may change these characteristics and to what extent. The reasons to validate are related with the fact that a new analytical procedure is being developed and also to widening the range of applicability for this procedure, e.g. for determining the same analyte but in a different matrix, the analytical procedure being used in a different laboratory and using different measuring equipment.

Among other steps that could be done for the validation of the method, one should be for example, a reference peak/standard solution should be chosen to calculate the retention times and relative peak areas of the other peaks and with these values estimate instrument and method reproducibility. To check instrument reproducibility, several consecutive replicate analysis of a sample solution should be performed and to check the method reproducibility several replicate samples should be prepared and then each sample solution should be analyzed. The method reproducibility would be an important parameter to be determined since the experiments were done in different countries and consequently different equipment.

The validation is also very important when the measurements are conducted by another person and also when the results of the Quality Control protocol suggest that validation parameters vary in time [89].

6. BIBLIOGRAPHY

[1] Y. Liang, P. Xie and K. Chan, "Review: Quality control of herbal medicines," *Journal of Chromatography B,* vol. 812, p. 53–70, 2004.

[2] X. Liang, Y. Jin, Y. Wang, G. Jin, Q. Fu and Y. Xiao, "Review: Qualitative and quantitative analysis in quality control of traditional Chinese medicines," *Journal of Chromatography A,* vol. 1216, p. 2033–2044, 2009.

[3] WHO, "General Guidelines for Methodologies on Research and Evaluation of Traditional Medicine," *World Health Organization,* 2000.

[4] S. D. A. o. China, "Technical request about fingerprint technology of injection of traditional medicine," *Chinese Traditional Patent Medicine,* vol. 22, p. 671, 2002.

[5] E. Ong, "Chemical assay of glycyrrhizin in medicinal plants," *Journal of Separation Science,* vol. 25, p. 825–831, 2002.

[6] P. Xie, "A Feasible Strategy for Applying Chromatography Fingerprint to Assess Quality of Chinese Herbal Medicine," *Traditional Chinese Drug Research & Clinical Pharmacology,* vol. 12, p. 141, 2001.

[7] W. Welsh, W. Lin, S. Tersigni, E. Collantes, R. Duta and M. Carey, "Pharmaceutical Fingerprinting: Evaluation of Neural Networks and Chemometric Techniques for Distinguishing among Same-Product Manufacturers," *Analytical Chemistry,* vol. 68, no. 19, p. 3473–3482, 1996.

[8] P. Valentao, P. Andrade, F. Areias, F. Ferreres and R. Seabra, "Analysis of vervain flavonoids by HPLC/diode array detector method. Its application to quality control," *Journal of agricultural and food chemistry,* vol. 47, no. 11, pp. 4579-4582, 1999.

[9] R. Bauer, "Quality criteria and standardization of phytopharmaceuticals: Can acceptable drug standards be achieved?," *Drug Information Journal,* vol. 32, no. 1, pp. 101-110, 1998.

[10] N. Lazarowych and P. Pekos, "Use of fingerprinting and marker compounds for identification and standardization of botanical drugs: Strategies for applying pharmaceutical HPLC analysis to herbal products," *Drug Information Journal,* vol. 32, no. 2, pp. 497-512, 1998.

[11] V. E. Tyler, "Phytomedicines: Back to the Future," *Journal of Natural Products,* vol. 62, pp. 1589-1592, 1999.

[12] K. Chan, "Some aspects of toxic contaminants in herbal medicines," *Chemosphere* , vol. 52, no. 9, pp. 1361-1371, 2003.

[13] R. L. Dressler, "The Orchids: Natural History and Classification," *Harvard University Press,* London; 1990.

[14] H. P. Wood, "The Dendrobiums," *AR. G. Gantner Verlag, Ruggell,* Liechtenstein; 2006.

[15] Pharmacopoeia of the People's Republic of China, vol. I, Beijing: People's Medical Publishing House, ISBN 978-7-117-06982-0, 2005.

[16] J. Li, S. Li, D. Huang, X. Zhao and G.Cai, "Advances in the resources, constituents and pharmacological effects of Dendrobium officinale," *Keji Daobao,* vol. 29, no. 18, pp. 74-79, 2011.

[17] L. Chen, Z. Lou, R. Wu, B. Yang, X. Zhang, F. Yang, K. Xia, C. Shen, W. Chen and Z. Wang, "Method for manufacturing medicine from Dendrobium officinale flower for preventing and treating hypertensive apoplexy," *Faming Zhuanli Shenqing,* 2010.

[18] Y. Bai and Y. Zheng, "Chinese medicinal dripping pill containing extracts from Dendrobium and Achyranthes and others for treating cardiovascular and cerebrovascular diseases," *Faming Zhuanli Shenqing,* 2008.

[19] L. Xiang, C. Sze, T. Ng, Y. Tong, P. Shaw, C. Tang and Y. Zhang, "Polysaccharides of Dendrobium officinale inhibit TNF-α-induced apoptosis in A-253 cell line," *Inflammation Research,* vol. 62, no. 3, pp. 313-324, 2013.

[20] L. Jin and C. Liu, "Preventive and therapeutic effect of Dendrobium polysaccharides on rat liver injury induced by cyclosporine A," *Zhongguo Yiyuan Yaoxue Zazhi,* vol. 29, no. 22, pp. 1891-1894, 2009.

[21] Y. Zhang, P. But, Z. Wang and P. Shaw, "Current approaches for the authentication of medicinal Dendrobium species and its products," *Plant Genetic Resources,* vol. 3, no. 2, pp. 144-148, 2005.

[22] "Tjskl.org.cn," *http://www.tjskl.org.cn/products-search/czab2e69/officinal_dendrobium_stem_dendrobium_officinale_kimura_et_migo-pz228af06.html,* [July 2013].

[23] "Age Fotostock," *http://www.agefotostock.com/age/en/Search.aspx?query=Dried%20dendrobium,* [July 2013].

[24] "Pharmacopoeia of the People's Republic of China," *China Medical Science and Technology Press; ISBN 9787506744393,* vol. I, Beijing; 2010.

[25] "State Food and Drug Administration," *www.sfda.gov.cn,* [November 2012].

[26] Q. Cheng, J. Guo and C. Zhang, "Study on the pharmacological effects of flavonoids," *Beihua Daxue Xuebao, Ziran Kexueban,* vol. 12, no. 2, pp. 180-183, 2011.

[27] P. G. Pietta, "Reviews: Flavonoids as Antioxidants," *Journal of Natural Products,* vol. 63, pp. 1035-1042, 2000.

[28] A. D. Agrawal, "Pharmacological Activities of Flavonoids: A Review," *International Journal of Pharmaceutical Sciences and Nanotechnology,* vol. 4, no. 2, 2011.

[29] R. Brouillard and A. Cheminat, "Flavonoids and plant color," *Progress in Clinical and Biological Research ,* vol. 280, pp. 93-106, 1988.

[30] C. Rice-Evans and L. Packer, "Flavonoids in Health and Disease;," *Marcel Dekker, Inc.; ISBN 0-203-91232-2 Master e-book ISBN; 0-8247-4234-6 (Print Edition),* pp. 61-110, New York, USA; 1998.

[31] Ø. Andersen and K. R. Markham, "Flavonoids: Chemistry, Biochemistry and Applications," *CRC Press, Taylor & Francis Group; ISBN-10: 0-8493-2021-6, ISBN-13: 978-0-8493-2021-7,* Boca Raton, US; 2006.

[32] L. Liu, Y. Lu, Q. Shao, Y. Cheng and H. Qu, "Binary chromatographic fingerprinting for quality evaluation of Radix Ophiopogonis by high-performance liquid chromatography coupled with ultraviolet and evaporative light-scattering detectors," *Journal of Separation Science,* vol. 30, no. 16, pp. 2628-2637, 2007.

[33] "Chromatographic Fingerprinting of Flos Chrysanthema Indici Using HPLC," *Application Note 207; Dionex,* 2009.

[34] J. P. L. P. W. X. Q. Zha, "Identification and classification of Dendrobium candidum species by fingerprint technology with capillary electrophoresis," *South African Journal of Botany,* vol. 75, no. 2, pp. 276-282, 2009.

[35] X. Zha, L. Pan, J. Luo, J. Wang, P. Wei and V. Bansal, "Enzymatic fingerprints of polysaccharides of Dendrobium officinale and their application in identification of Dendrobium species;," *Journal of Natural Medicines,* vol. 66, no. 3, pp. 525-534, 2012.

[36] T. B. Ng, J. Liu, J. H. Wong, X. Ye, S. Sze, Y. Tong and K. Y. Zhang, "Review of research on Dendrobium, a prized folk medicine," *Applied Microbiology and Biotechnology* , vol. 93, no. 5, pp. 1795-1803 , 2012.

[37] O. B. A. Halim, H. A. A. Fattah, F. T. Halaweish and A. F. Halim, "Isoflavonoids and alkaloids from Spartidium saharae," *Natural Product Sciences*, vol. 6, no. 4, pp. 189-192, 2000.

[38] A. R. Kuehnle, D. H. Lewis, K. R. Markham, K. A. Mitchell, K. M. Davies and B. R. Jordan, "Floral flavonoids and pH in Dendrobium orchid species and hybrids," *Euphytica* , vol. 95, no. 2, pp. 187-194, 1997.

[39] C. Williams, J. Greenham, J. Harborne, J. M. Kong, L. S. Chi, N. K. Goh, N. Saito, K. Toki and F. Tatsuzawa, "Acylated anthocyanins and flavonols from purple flowers of Dendrobium cv. 'Pompadour'," *Biochemical Systematics and Ecology*, vol. 30, no. 7, p. 667–675, 2002.

[40] T. Phechrmeekha, B. Sritularak, K. Likhitwitayawuid, J. o. A. N. Phechrmeekha, B. Sritularak and K. Likhitwitayawuid, "New phenolic compounds from Dendrobium capillipes and Dendrobium secundum," *Journal of Asian Natural Products Research*, vol. 14, no. 8, pp. 748-754, 2012.

[41] S. S. Whang, W. S. Um, I. Song, P. O. Lim, K. Choi, K. Park, K. Kang, M. S. Choi and J. C. Koo, "Molecular Analysis of Anthocyanin Biosynthetic Genes and Control of Flower Coloration by Flavonoid 3',5'-Hydroxylase (F3'5'H) in Dendrobium moniliforme," *Journal of Plant Biology*, vol. 54, no. 3, pp. 209-218, 2011.

[42] C. Chang, A. F. Ku, Y. Tseng, W. Yang, J. Fang and C. Wong, "6,8-Di-C-glycosyl Flavonoids from Dendrobium huoshanense," *Journal of Natural Products* , vol. 73, no. 2, p. 229–232, 2010.

[43] N. Saito, K. Toki, K. Uesato, A. Shigihara and T. Honda, "An acylated cyaniding glycoside from the red-purple flowers of Dendrobium," *Phytochemistry*, vol. 37, no. 1, pp. 245-248, 1994.

[44] IUPAC, " Compendium of Chemical Terminology-Gold Book," vol. Version 2.3.2, 2012-08-19.

[45] R. E. Ardrey, "Liquid Chromatography–Mass Spectrometry: An Introduction," *John Wiley & Sons, Ltd; ISBNs: 0-471-49799-1 (HB); 0-471-49801-7 (PB),* Cambridge, UK; 2003 .

[46] D. Skoog, F. Holler and S. R. Crouch, "Principles of Instrumental Analysis," *Thomson Brooks/Cole; ISBN-13: 978-0-495-01201-6,* USA, sixth edition,

2007.

[47] "Waters," *http://www.waters.com/waters/pt_PT/How-Does-High-Performance-Liquid-Chromatography-Work%3F/nav.htm?cid=10049055&locale=pt_PT*, [July 2013].

[48] "New Mexico State University," *http://www.chemistry.nmsu.edu/Instrumentation/NMSU_1200HPLC_Procd.html*, [July 2013].

[49] G. Zhou and G. Lu, "Study on HPLC fingerprints of flavone C-glycosides in Dendrobium officinale leaves and determination of index component," *Zhongguo Yaoxue Zazhi; Beijing, China,* vol. 47, no. 11, pp. 889-893, 2012.

[50] H. Ou, J. Cheng, X. Li, R. Zhan, Y. Liang, J. Xu, H. Xu and P. Yan, "HPLC fingerprint of flavonoids and phenols of Dendrobium nobile," *Journal of Chinese medicinal materials,* vol. 32, no. 6, pp. 871-874, 2009.

[51] G. Wei, H. Liu, Y. Huang, X. Xie, L. Yang, X. Liu and D. Liu, "Study on the HPLC characteristic spectrum of fresh Dendrobium officinale," *Zhongchengyao,* vol. 34, no. 9, pp. 1739-1743, 2012.

[52] G. Zhou and G. Lu, "Comparative studies on scavenging dpph free radicals activity of flavone c-glycosides from different parts of dendrobium officinale," *Zhongguo Zhongyao Zazhi,* vol. 37, no. 11, pp. 1536-1540, 2012.

[53] J. Cheng, X. Li, J. Xu, Y. Liang and L. Yi, "Method for determining high performance liquid chromatography (HPLC) fingerprint of Dendrobium," *Faming Zhuanli Shenqing,* vol. CN 101726547 A 20100609, 2010.

[54] R. Brereton, "Chemometrics: Data Analysis for the Laboratory and Chemical Plant," *John Wiley & Sons Ltd; ISBNs: 0-471-48977-8 (HB); 0-471-48978-6 (PB),* University of Bristol, UK, 2003.

[55] A. d. Juan, ""Multivariate Calibration" Lecture, University of Algarve, Faro, Portugal: Erasmus Mundus Master in Quality in Analytical Laboratories," 5-8 June, 2012.

[56] J. A. Westerhuis, H. C. J. Hoefsloot, S. Smit, D. J. Vis, A. K. Smilde, E. J. J. v. Velzen, J. P. M. v. Duijnhoven and F. A. v. Dorsten, "Assessment of PLSDA cross validation," *Metabolomics,* vol. 4, p. 81–89, 2008.

[57] N. ©Copyright 1987- 2009 Pattern Recognition Systems AS, "PRS-Sirius Version 8.1".

[58] "MathWorks," *http://www.mathworks.se/help/stats/pca.html,* [April 2013].

[59] R. Yu, "Information: Theoretical Fundamentals of Modern Analytical Chemistry," *Hunan University Press,* p. 50, Changsha, P.R.C.; 1996.

[60] F. Gong, Y. Liang, P. Xie and F. Chau, "Information theory applied to chromatographic fingerprint of herbal medicine for quality control," *Journal of Chromatography A,* vol. 1002, no. 1-2, p. 25–40, 2003.

[61] "Engineering Statistics Handbook," *http://www.itl.nist.gov/div898/handbook/pri/section3/pri33a.htm,* [July 2013].

[62] "The Michigan Chemical Process Dynamics and Controls Open Textbook," *https://controls.engin.umich.edu/wiki/index.php/Design_of_experiments_via_taguchi_methods:_orthogonal_arrays,* [July 2013].

[63] A. Savitzky and M. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *The Perkin-Elmer Corp., Norwalk, Conn.,* vol. 36, no. 8, pp. 1627-1639, 1964.

[64] Z. Zhang, S. Chen and Y. Liang, "Baseline correction using adaptive iteratively reweighted penalized least," *Analyst,* vol. 135, no. 5, pp. 1138-1146, 2010.

[65] P. W. Holland and R. E. Welsch, " Communications in Statistics: Theory and Methods," vol. 6, pp. 813-827, 1977.

[66] D. B. Rubin, "Iteratively reweighted least squares".

[67] P. J. Green, "Journal of the Royal Statistical Society: Series B (Statistical Methodology)," p. 149–192, 1984.

[68] N. P. V. Nielsen, J. M. Carstensen and J. Smedsgaard, "Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping," *Journal of Chromatography A* , vol. 805, no. 1-2, pp. 17-35, 1998.

[69] J. A. Pino, J. E. McMurry, P. C.Jurs, B. K. Lavine and A. M. Harper, "Application of pyrolysis-gas chromatography pattern-recognitionto the detection of cystic-fibrosis heterozygotes," *Analytical Chemistry,* vol. 57, no. 1, pp. 295-302, 1985.

[70] M. E. Parrish, B. W. Good, F. S. Hsu, F. W. Hatch, D. M. Ennis, D. R. Douglas, J. H. Shelton, D. C. Watson and C. N. Reilley, "Computer-enhanced high-resolution gas-chromatography for the discriminative analysis of

tobacco-smoke," *Analytical Chemistry*, vol. 53, no. 6, pp. 826-831, 1981.

[71] K. J. Johnson, B. W. Wright, K. H. Jarman and R. E. Synovec, "High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis," *Journal of Chromatography A,* vol. 996, no. 1-2, pp. 141-155, 2003.

[72] G. Malmquist and R. Danielsson, "Alignment of chromatographic profiles for principal component analysis - A prerequisite for fingerprinting methods," *Journal of Chromatography A,* vol. 687, no. 1, pp. 71-88, 1994.

[73] D. Bylund, R. Danielsson, G. Malmquist and K. E. Markides, "Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modelling of liquid chromatography-mass spectrometry data," *Journal of Chromatography A,* vol. 961, no. 2, pp. 237-244, 2002.

[74] M. D. Hamalainen, Y. Liang, O. M. Kvalheim and R. Andersson, "Deconvolution in one-dimensional chromatography by heuristic evolving latent projections of whole profiles retention time shifted by simplex optimization of cross-correlation between target peaks," *Analytica Chimica Acta,* vol. 271, no. 1, pp. 101-114, 1993.

[75] T. Skov, F. Berg, G. Tomasi and R. Bro, "Automated alignment of chromatographic data," *Journal of Chemometrics,* vol. 20, pp. 484-497, 2006.

[76] N. Nielsen, J. M. Carstensen and J. Smedsgaard, "Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimized warping," *Journal of Chromatography A,* Vols. 17-35, p. 805, 1998.

[77] G. Tomasi, F. Berg and C. Andersson, "Correlation optimized warping and dynamic time warping as prepocessing methods for chromatographic data," *Journal of Chemometrics,* vol. 18, pp. 231-241, 2004.

[78] R. Henrion and C. Andersson, "A new criterion for simple-structure transformations of core arrays in N-way principal components analysis," *Chemometrics and Intelligent Laboratory Systems,* vol. 47, no. 2, pp. 189-204, 1999.

[79] J. Christensen, G. Tomasi and A. Hansen, "Chemical Fingerprinting of Petroleum Biomarkers Using Time Warping and PCA," *Environmental Science and Technology,* vol. 39, no. 1, pp. 255-260, 2005.

[80] K. Johnson, B. Prazen, D. Young and R. Synovec, "Quantification of naphthalenes in jet fuel with GC × GC/Tri-PLS and windowed rank minimization retention time alignment," *Journal of Separation Science,* vol. 27, no. 5-6, pp. 410-416, 2004.

[81] W. Spendly, G. Hext and F. Himsworth, "Sequential Application of Simplex Designs in Optimization and Evolutionary Operation," *Technometrics,* vol. 4, pp. 441-461, 1962.

[82] "The Quality & Technology Website," *http://www.models.kvl.dk,* [April 2013].

[83] "U China Visa," *http://www.uchinavisa.com/china-provinces-map.html,* [July 2013].

[84] "Dionex," *http://www.dionex.com/,* [July 2013].

[85] L. Yang, Y. Wang, G. Zhang, F. Zhang, Z. Zhang, Z. Wang and L. Xu, "Simultaneous quantitative and qualitative analysis of bioactive phenols in Dendrobium aurantiacum var. denneanum by high-performance liquid chromatography coupled with mass spectrometry and diode array detection," *Biomedical Chromatography,* vol. 21, no. 7, pp. 687-694, 2007.

[86] L. Yang, Z. Wang and L. Xu, "Simultaneous determination of phenols (bibenzyl, phenanthrene, and fluorenone) in Dendrobium species by high-performance liquid chromatography with diode array detection," *Journal of Chromatography A* , vol. 1104, no. 1-2, p. 230–237, 2006.

[87] "Crawford Scientific," *www.crawfordscientific.com,* [May 2013].

[88] X. Fan, Y. Cheng, Z. Ye, R. Lin and Z. Qian, "Multiple chromatographic fingerprinting and its application," *Analytica Chimica Acta,* vol. 555, no. 2, p. 217–224, 2006.

[89] P. Konieczka and J. Namiesnik, "Quality Assurance and Quality Control in the Analytcal Chemical Laboratory - A Practical Approach," *CRC Press; ISBN 13: 978-1-4200-8270-8,* Boca Raton, U.S.A.; 2009.