

Assessment of Herbal Medicines by Chemometrics – Assisted Interpretation of FTIR Spectra

Chew Oon Sim¹, Mohammad Razak Hamdan¹,
*Zhari Ismail¹ and Mohd Noor Ahmad²

¹School of Pharmaceutical Sciences and
²School of Chemical Sciences,
Universiti Sains Malaysia, 11800 Minden Penang, Malaysia.

*Fax: 604 – 6563443, E-mail: zhari@usm.my

Abstract

Pharmacognosical analysis of medicinal herbs remain challenging issues for analytical chemists, as herbs are a complicated system of mixtures. Analytical separation techniques for example high performance liquid chromatography (HPLC), gas chromatography (GC) and mass spectrometry (MS) were among the most popular methods of choice used for quality control of raw material and finished herbal product. The application of infrared (IR) spectroscopy in herbal analysis is still very limited compared to its applications in other areas (food and beverage industry, microbiology, pharmaceutical etc). This article attempts to expand the use of FTIR spectroscopy and at the same time creating interest among prospective researcher in herbal analysis. A case study was conducted by incorporating appropriate chemometric methods (Principal Components Analysis, PCA and Soft Independent Modelling of Class Analogy, SIMCA) as tools for extracting relevant chemical information from the obtained infrared data. The developed method can be used as a quality control tool for rapid authentication from a wide variety of herbal samples.

Keywords: Herbal analysis; FTIR spectroscopy; Chemometrics; Principal Components Analysis; Soft Independent Modelling of Class Analogy

“Assessment herbal medicines by chemometrics-assisted FTIR spectra ” send to Journal Of Analytica Chimica Acta by January 2004.

1. Introduction

Unlike synthetic drugs, herbal medicine is a complicated system of mixtures. Thus, the methods of choice for identification of 'botanical drug' are mainly intended to obtain a characteristic fingerprint of a specific plant that represent the presence of a particular quality defining chemical constituents. For such purposes, chromatographic techniques such as high performance liquid chromatography (HPLC), gas chromatography (GC), gas chromatography – mass spectrometry (GC-MS) and thin layer chromatography (TLC) were used widely as reported in numerous publications [1-3].

Initially, the use of infrared (IR) spectroscopical method is restricted only for structural elucidation of isolated compounds from the herbal matrices. It is also found useful in phytochemical studies as a 'fingerprinting' devices, for comparing a natural with synthetic sample [4]. Due to the inherent complexity of the IR spectrum, the actual interpretation may be difficult and the operation require much experience. Indeed, slight differences in the spectra within the same plant species may not be obvious and generally not visible to the naked eye. Thus, the application of IR spectroscopy in herbal analysis is still very limited compared to its applications in other areas (food and beverage industry, microbiology, pharmaceutical etc).

With the advance of computer technology, chemometric methods has become a leading tool among the scientific communities towards faster analysis results and shorter product development time [5]. Among others, an unsupervised pattern recognition technique such as Principal Components Analysis (PCA) is the most often used method for handling multivariate data without prior knowledge about the studied samples [6]. While the supervised classification procedure using Soft Independent Modelling of Class Analogy (SIMCA) based on making a PCA model to assign unknown samples into the predefine class model has also been applied to the analysis of infrared spectra [7].

This study aim to extending the used of FTIR transmission spectroscopy associated with the appropriate chemometric methods (PCA and SIMCA) to classify *Orthosiphon stamineus* Benth. (well known as Java Tea for treating infection of the urinary tract, kidney and bladder stones disease [8]) based on its geographical origin and varieties from the obtained characteristic infrared spectrum. Besides, the potential relationship between the complex chemical constituent, geographical origin and varieties of the samples were also studied.

Natural product including plants undergoes natural variation. The quality of samples from different localities and growing conditions based on their geographical origin may vary. Thus, the identification of crude herbs based on geographical origins is crucial in order to ensure authenticity, quality, safety and efficacy of the raw material before it is converted to the final product. Herbal product manufacturer always seeking for a faster and cost effective verification method since the traditional wet chemistry analysis are too laborious and time consuming. The authors undertook this effort to introduce and developing a rapid quality verification method with the integration of statistical and mathematical modeling for extracting relevant information base on the infrared spectroscopy data [9]. To our knowledge, this is the first attempt to incorporate Fourier Transform Infrared (FTIR) Spectroscopy analysis with chemometric techniques for herbal analysis.

2. Experimental

2.1 Materials and Methods

Dried leaves of samples from ten different geographical origins consisting of two varieties (white and purple flowers) were collected from the respective distributors in Malaysia and Indonesia. The samples were named using alphabetical codes as shown in Table 1.

An amount of dried leaf samples were milled until fine powder and were filtrated with sieves 0.071 and 0.500 mm mesh size. KBr of spectroscopy grade were also filtrated with sieves 0.071mm mesh size. 2 mg samples were mixed uniformly with 100 mg KBr (2% w/w) and homogenized by using stir vortex CENCO 34 (Breda, Netherlands).

The FTIR spectra were recorded in the mid-IR region $4000 - 400 \text{ cm}^{-1}$ at resolution 4 cm^{-1} with 16 scans using Thermo Nicolet FTIR Nexus spectrometer coupled with DTGS (deuterated tri-glycine sulphate) detector. The interferometer and the detector chamber were purged with dry nitrogen to remove spectral interference due to atmospheric carbon dioxide and water vapor. Air background spectrum was recorded before each sample and all experiment were performed in six triplicates (six pellets KBr with three scan each).

2.2 Data Preprocessing

Prior to data analysis, each spectrum was baseline corrected and the absorbance was normalized so that peak absorbance of the most intense band is set to unity. The spectra were

transferred via a JCAMP.DX format into the statistical software program *The Unscrambler*[®] 8.0 (CAMO, Trondheim, Norway).

2.3 Chemometric Data Analysis

2.3.1 Principal Components Analysis (PCA)

PCA algorithms was used in this study for reducing the high-dimensional spectroscopic data by constructing a linear combination of the original variable into a few orthogonal principle components which contain most of the variability of the data set. This projection method allows (1) visualization of the natural clustering in the data, (2) primary evaluation of the between-class similarity and finally (3) finding the reasons behind the observed pattern by making correlation with the chemical or physico-chemical properties of the studied samples.

2.3.2 Soft Independent Modelling of Class Analogy (SIMCA)

In order to test the robustness of this study for classification and discrimination of geographical origin and varieties from *O. stamineus* samples, a popular classification method – SIMCA was used to assign unknown samples to its existing classes. There are five steps involved to obtain the classification model: (1) construction of separate PCA models for each class, (2) determine the optimal number of PCs by validation, (3) fitting the unknown samples to each predefine model, provided that the class are distinct enough (4) deciding whether the samples belong to the corresponding class by referring to the object-to-model distance and leverage (distance of the sample to the model center) and finally (5) validate the classification results with statistical test called significance test.

3. Results and Discussion

Beer's Law assumes that the absorption of light at a given wavelength is due entirely to the absorptivity of the constituents in the sample. Therefore, any variations in the spectral due to spectrometer and sampling error should be eliminated prior to data analysis. In this study, two chemometric pre-processing step namely baseline correction and normalization were used. Baseline effects introduced by the spectrometer (eg. detector drift, changing environmental condition such as temperature, spectrometer purge and sampling accessories) [10] were removed by auto-correcting the spectrum baseline. On the other hand, normalization of the absorbance spectra to the most intense band (hydroxyl group, 3406 cm⁻¹)

deletes the differences between spectra due to different amounts of sample or pathlength variation [11].

3.1 The FTIR Spectra Analysis

3.1.1 Classification of Geographical Origin and Varieties

Referring to Malaysian Herbal Monograph [8], the main chemical constituents of *O. stamineus* are: Flavonoids (mainly methoxylate compounds), Caffeic acid derivatives, Diterpene ester, Triterpene saponins and other minor compounds. Each of these compounds play an indispensable role in the complicated system of mixture contributing to the efficacy and potency due to the synergistic effect contributed from a number of constituents present in the herb. Fig. 1. shows the FTIR spectra ($4000 - 400 \text{ cm}^{-1}$) of ten different samples origin of (a) white flowers and (b) purple flowers varieties.

The sharp absorption peak at $1600 - 1760 \text{ cm}^{-1}$ are assigned to C=O stretching vibration in carbonyl compounds which may be characterized by the presence of high content of terpenoids and flavanoids in the complex mixture of *O. stamineus*. The presence of a narrow and sharp peak at $\sim 2925 \text{ cm}^{-1}$ and $\sim 2853 \text{ cm}^{-1}$ was assigned to C-H and C-H (methoxy compounds) stretching vibration respectively. The present of diterpenes were further proven with the absorption band of hydroxyl ($3500 - 3480 \text{ cm}^{-1}$), ester carbonyl ($1270 - 1150 \text{ cm}^{-1}$) and phenyl ($1600, 1420 \text{ cm}^{-1}$) [12].

By visual recognition, there are no significant difference in the characteristic absorption bands but the intensity of certain wavelength do differ from each others especially at the fingerprint region ($1800 - 800 \text{ cm}^{-1}$). This interprets the similarity of certain main chemical component observed among different *O. stamineus* sample origin. The presentation of spectrum in 3-D [Fig. 2] further enhances the visualization of the variability in the intensity absorption bands between the respective origins as discussed earlier.

Differences between spectra are generally not visible to the naked eyes; it is more practical to incorporate statistical method for the aid of interpreting the obtained measurement results from spectroscopic analysis. Since the discrimination of different geographical origin of herb based on the slight differences among particular absorption bands are too subjective and the results may vary between analysts as reported [13]. Hence, we report a more reliable method for discrimination of samples origin by incorporating chemometric analysis in interpreting the resultant spectra.

3.2 The chemometric analysis

3.2.1 Classification of geographical origin and varieties

3.2.1.1 Principal Components Analysis (PCA)

PCA was carried out using 520 points (normalized absorbance) between the ranges of the selected spectral region $1800 - 800 \text{ cm}^{-1}$ as discuss earlier. Out of 18 data set from six triplicate measurement of each sample, 12 data set were randomly selected to represent the respective samples origin and varieties as shown in Table 1. Natural grouping of *O. stamineus* from ten different geographical origin were visualize in the two-dimensional scores plot below [Fig. 3]. The first three principal components (PCs) represent only 61% of the total variance (PC1 = 29%, PC2 = 22% and PC3 = 10%). This indicates that these three components were not sufficient to provide effective clustering of the samples origin with clear separation between the groups.

The rule of thumb to make a good classification is to make sure that the variation within different sample must be greater than the variance between individual samples. Thus, spectral reproducibility is important for creating a robust classification model. The systematic variation between replicate spectra due to baseline effect are highlighted and removed after derivatisation. Generally, the use of spectra derivatives with Savitzky-Golay algorithm as a chemometric pre-processing techniques are widely reported in most classification based on FTIR spectroscopy [14 - 16].

The potential relationship and pattern associate between different *O. stamineus* samples source with regards to its complex mixtures were further investigated by computing a PCA based on the first derivative spectral data. The first three PCs represent almost 90% of the total variance in the data set. The first PCs consist 48% of the total variability followed by the second PCs with 34% variance and only 8% variance carried by the third PCs. Overall, each sample was able to formed distinct cluster in the two dimensional plot [Fig. 4]. Examining the space defined by the first and second PCs as shown in Fig. 4 (a), samples from the same origin (state) for example Kelantan (group G and L) and Sarawak (group H, I and N) tends to groups itself together in the first PCs. Samples from Penang origin only observed this relationship when it is projected onto the third PCs, which contain only 8% of the total data variability [Fig. 4 (b)]. From this inherent structure, we can make a sound assumption that the composition of complex chemical constituent do vary according to its origin. But, there are no significant differences in the complex mixture observed between the different plant

varieties as sample with purple flower (rectangular shape) form a close cluster with the white flower samples (oval shape).

This is further shown by taking a closer look at the PCA models in Fig. 5. The first and second PCs which contain almost 60% of the data variances separate the four respective sample origin into four well define space. A good separation between different samples varieties from the sample region, in this case Kelantan and Sarawak only shown in PC3 [Fig. 5 (b)]. Therefore, the slight differences in the chemical composition between plant varieties which is difficult to interpret with our naked eyes, could be easily identified using this chemometric treatment of infrared data. Furthermore, the classification model may be used as a standard references for monitoring the batch-to-batch consistency of raw materials supplied by the respective supplier.

3.2.1.2 Soft Independent Modelling of Class Analogy (SIMCA)

Following the successful application of PCA techniques in discriminating analyzed samples to its geographical origin and varieties accordingly, the ability of the obtained model in performing real situation task such as assigning upcoming unknown samples need to be further proved using a more sophisticated classification technique. SIMCA classification approach was employed for such purpose. The first derivative spectra data sets of each sample group were divided randomly into training and testing set containing two thirds (12 samples) and one third (6 samples) respectively of the total data. PCA model for each class were constructed using the optimum number of PCs determined by leverage validation.

As a result, SIMCA classification based on the predefine PCA model described above yield 100% of correct classification for all classes at 5% significance level. The classification results are well displayed in the Cooman's plot below [Fig. 6]. The plot displays the sample-to-model distance for each sample to two models, which allow us to visualize how different two models are from each other. The newly classified samples (testing set) are displayed in green color, while the training set for the two models are displayed in blue and red. By referring to Fig. 6 (a), sample Sabah_P and BPPM_P with model distance lower than three implies that these two samples are likely to each others in term of their complex chemical mixture. Contrary, the complex chemical constituent of sample USKGSM_P and NNPPDM_P are comparable to each other, as the two sample models were separate far away from one and another.

4. Conclusions

The results from this work give evidence that *O. stamineus* samples from different geographical origin and varieties have varied complex chemical mixture. Sample origin seems to have more dominant effect to the chemical constituent of the plant compared to plant varieties. The good classification model obtained from both PCA and SIMCA further proved that classification of samples from various sourcing and varieties is possible with the incorporation of chemometric techniques and FTIR spectroscopy. Therefore, the obtain classification model may be of great use for quality inspection of raw herbal material on a continuous basis as new batches are produced. Chemometrics analysis of spectra data is rapid and simple since no chemical treatment of samples are required.

Acknowledgements

Financial support from the Ministry of Science, technology and Environmental, Malaysia through IRPA No. 305/Pfarmasi/6123002 for this research is acknowledged. The authors also thank the assistance of Ranjit Viswanathan (Sales Manager, CAMO Software India Pvt. Ltd.) for downloading '**The Unscrambler[®] 8.0**' evaluation version.

References

- [1] J.L. Natalie and P. Peter, *Drug Information Journal*, 32 (1998) 497.
- [2] H.K. Edward, J.V. Kevin, L.W. Jeffery, A.R. Robyn and S.L. Mike, *Drug Information Journal*, 32 (1998) 471.
- [3] B. Rudolf, *Drug Information Journal*, 32 (1998) 101.
- [4] J.B. Harborne, *Phytochemical Methods A Guide to Modern Techniques of Plant Analysis*, 3rd Edition, Chapman & Hall, UK, 1998.
- [5] M.B. Seasholtz, *Chemometrics and Intelligent Laboratory Systems*, 45 (1999) 55.
- [6] J.N. Miller and J.C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, 4th Edition, Prentice Hall, 2000.
- [7] A.R. Cotrim, A. Ferraz, A.R. Gancalves, F.T. Silva and R.E. Bruns, *Bioresource Technology*, 68 (1999) 29.
- [8] Monograph Committee, *Malaysian Herbal Monograph, Volume 2*, Malaysian Monograph Committee, National Pharmaceutical Control Bureau, Ministry of Health Malaysia, 2001.
- [9] O.S. Chew, M.R. Hamdan, Z. Ismail and M.N. Ahmad, 19th Annual Seminar & Workshop of the Malaysia Natural Products Society, Faculty of Science, University of Malaya, Malaysia, 13-16 October 2003.
- [10] R.L. Tranter, *Design and Analysis in Chemical Research*, Sheffield Academic Press, England, 2000.
- [11] J.A. Mike, *Chemometrics in Analytical Spectroscopy*, RSC Analytical Spectroscopy Monograph, The Royal Society of Chemistry, Thomas Graham House, Science Park, Cambridge, 1995.
- [12] A. Suresh, T. Yasuhiro, S. Seikichi, T. Kazuhiko and K. Shigetoshi, *Tetrahedron Letters*, 43 (2002) 1473.
- [13] Z. Wang, S.Q. Sun, Q. Zhou, X.B. Li, Q. Zhou, L. Lin and D.G. Du, *Spectroscopy and Spectral Analysis*, 21(3) (2001) 311.
- [14] L. Mariey, J.P. Signolle, C. Amiel and J. Traver, *Vibrational Spectroscopy*, 26 (2001) 151.
- [15] M. Kansiz, P. Heraud, B. Wood, F. Burden, J. Beardall and D. McNaughton, *Phytochemistry*, 52 (1999) 407.
- [16] N. Dupuy, L. Duponchel, J.P. Huvenne, B. Sombret and P. Legrand, *Food Chemistry*, 57(2) (1996) 245.

Table 1

The list of *O. stamineus* samples according to its geographical origin and varieties.

Code ^a	Location	State	Group ID
BLKBPM	Bumbang Lima	Penang	A
SRKBPM	Kepala Batas	Penang	B
SZBKAM	Bota Kanan	Perak	C
MABTRM	Bohor Temak	Perlis	D
ZARWBM	Rawang	Selangor	E
STJGCM	Jengka	Pahang	F
NNPPDM	Pasir Puteh	Kelantan	G
MSSMSM	Semonggok	Sarawak	H
USKGSM	Kuching	Sarawak	I
NHPJI	Pulau Jawa	Jakarta	J
BPPM_P	Balik Pulau	Penang	K
NNPPDM_P	Pasir Puteh	Kelantan	L
SABAH_P	- ^b	Sabah	M
USKGSM_P	Kuching	Sarawak	N

^a Example BLKBPM: BL: distributor; KB: location; P: state; M: country (M: Malaysia; I: Indonesia); P: purple flowers (white flowers if not indicate)

^b Information not available.

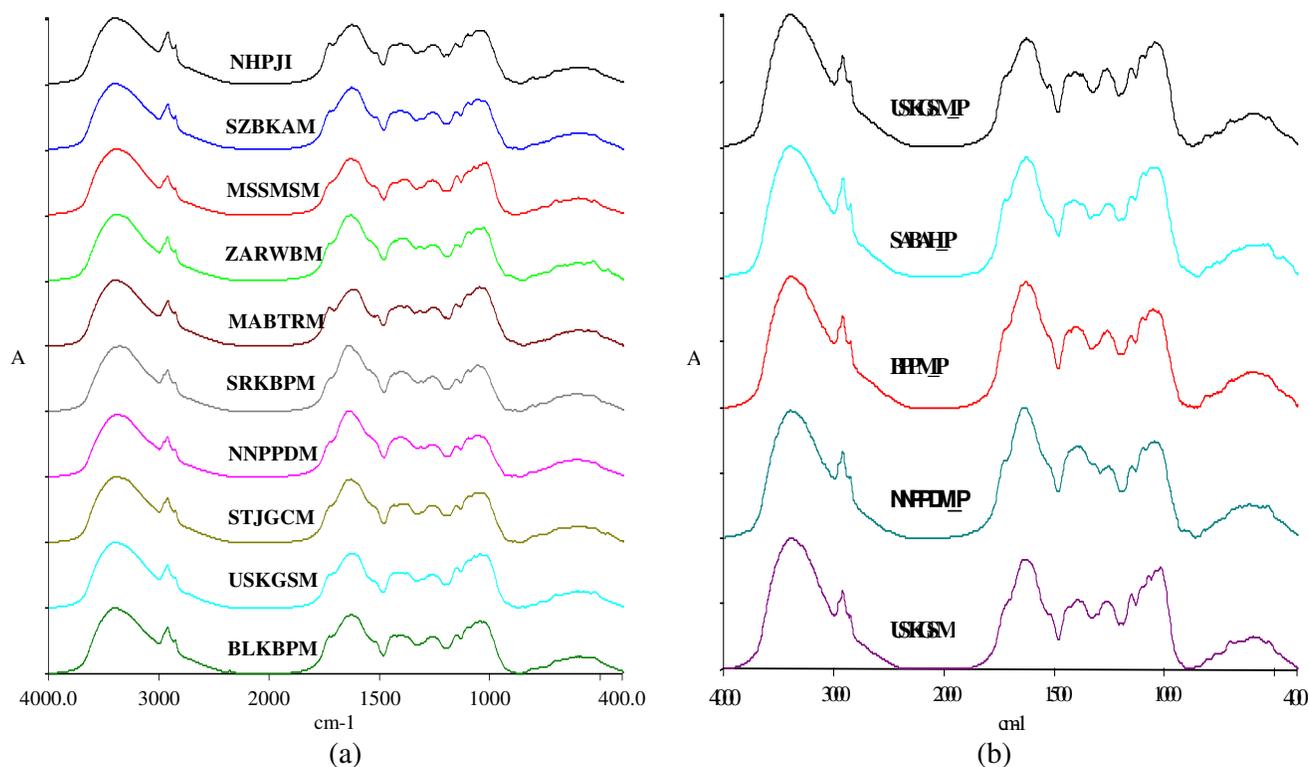


Fig. 1. The characteristic FTIR spectra of *O. stamineus* samples (a) white flowers and (b) purple flowers from different origin in the mid-IR range 4000 – 400 cm⁻¹.

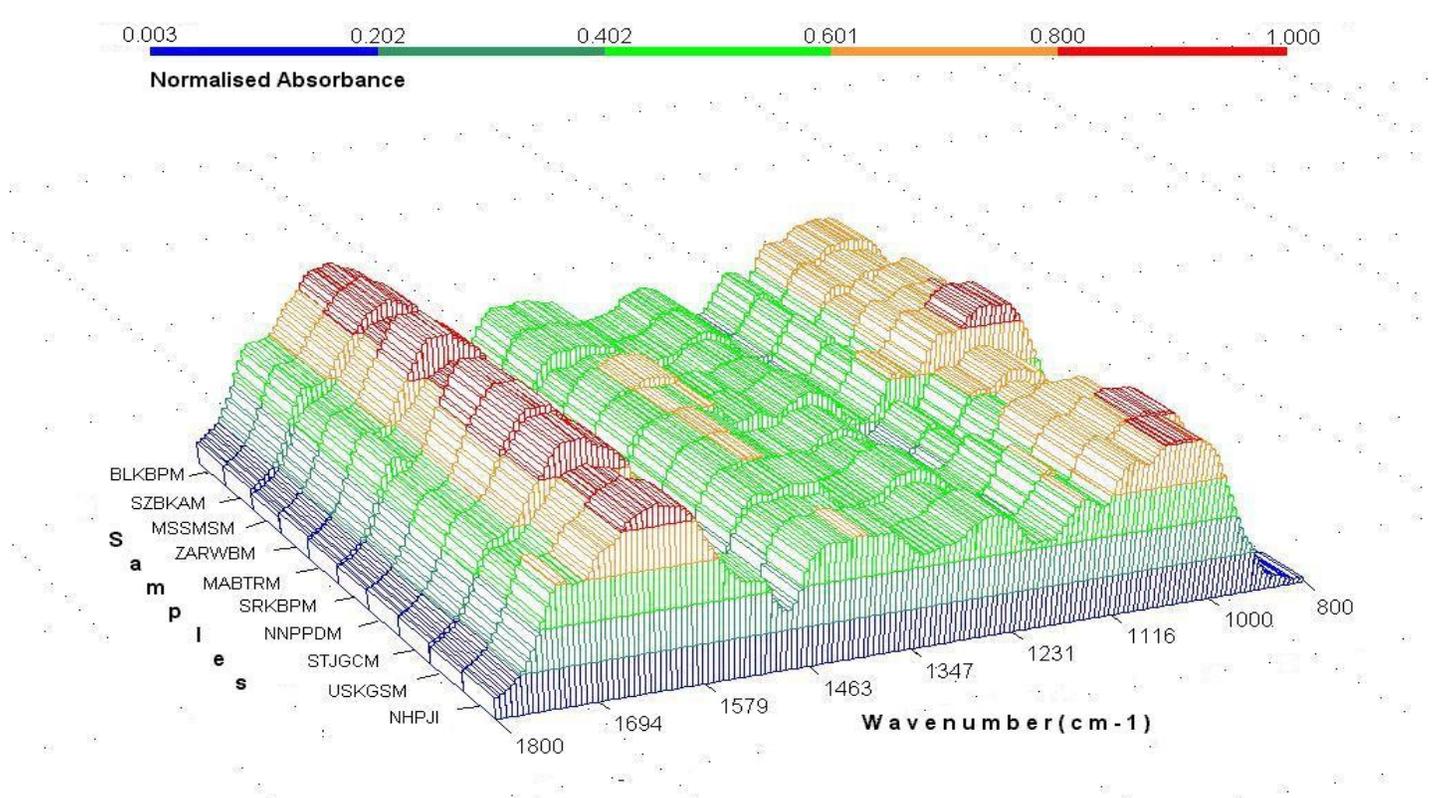


Fig. 2. 3-D absorbance matrix spectra (1800 – 800 cm⁻¹) of different samples origin.

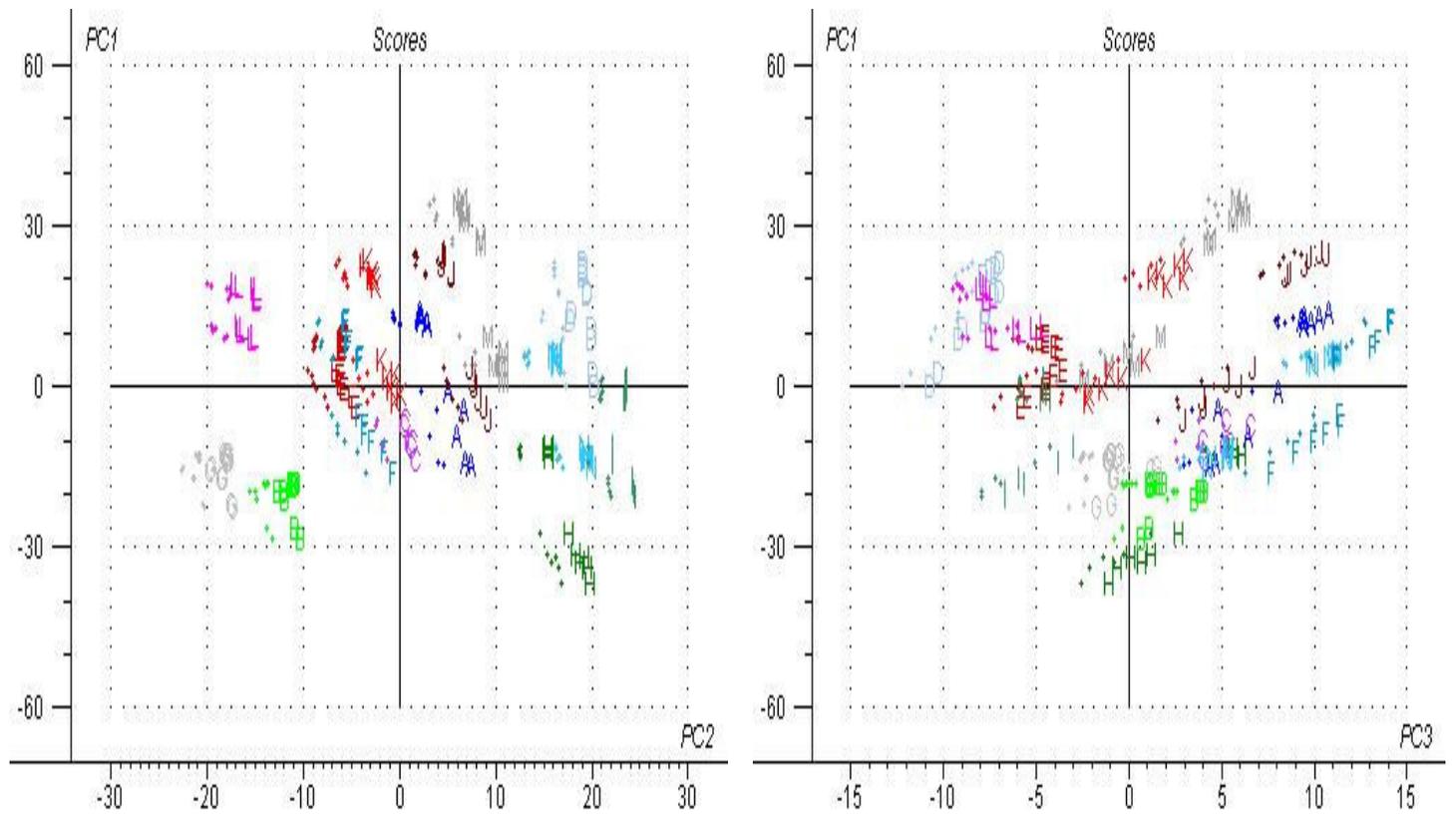


Fig. 3. 2-D scores plot of the first three PCs (a) PC1 = 29% vs PC2 = 22% and (b) PC1 = 29% vs PC3 = 10% obtained from PCA of FTIR spectra.

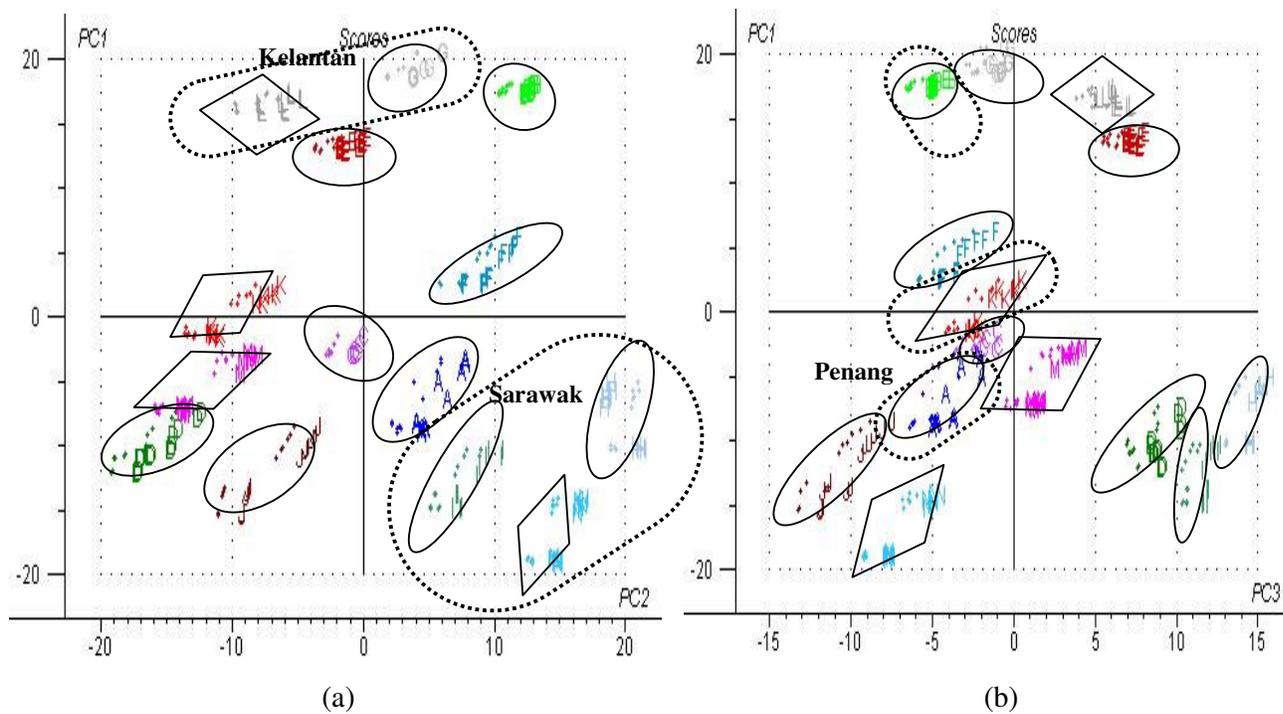


Fig. 4. 2-D scores plot of the first three PCs (a) PC1 = 48% vs PC2 = 34% and (b) PC1 = 48% vs PC3 = 8% obtained from PCA of 2-point first derivative FTIR spectra.

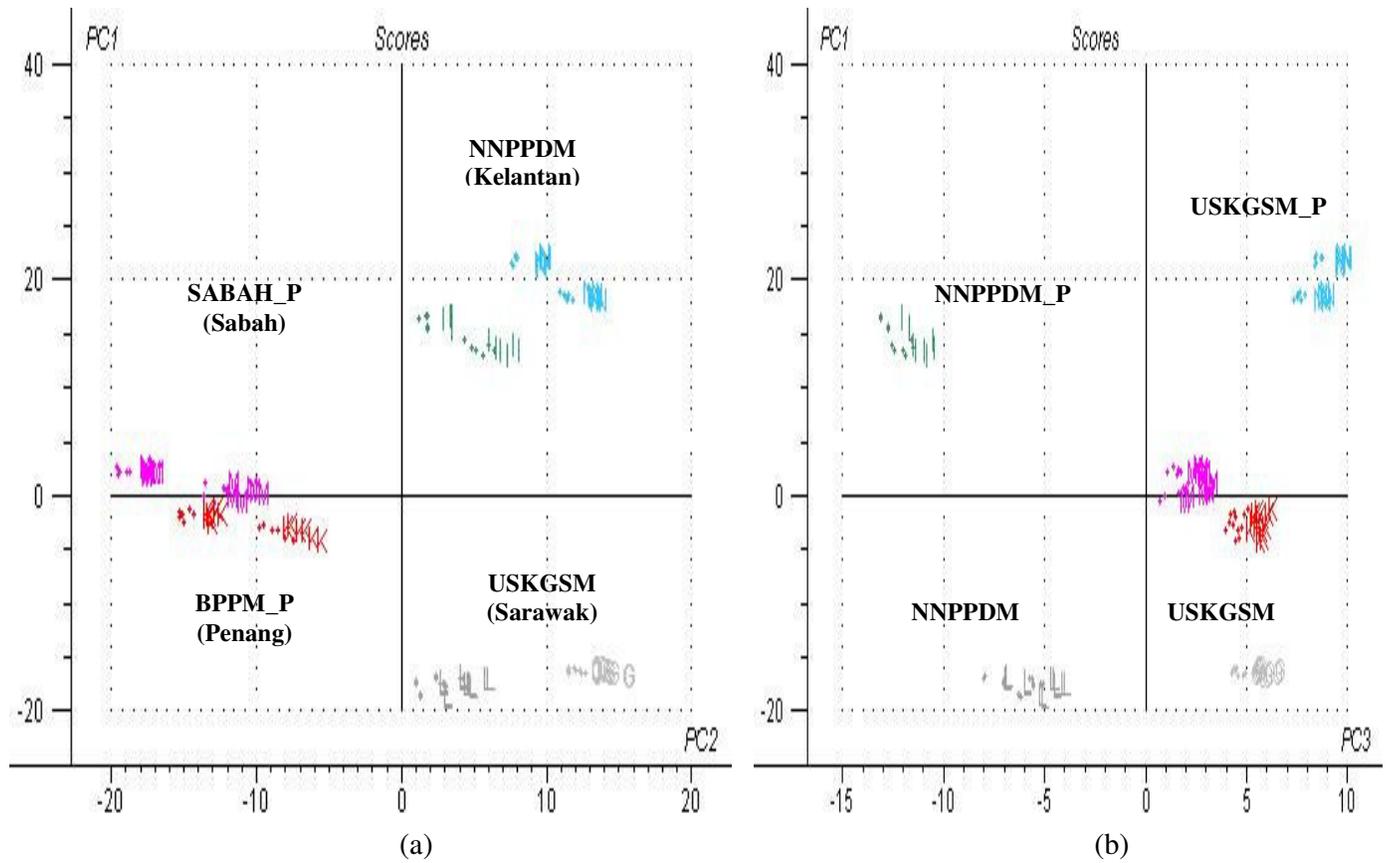
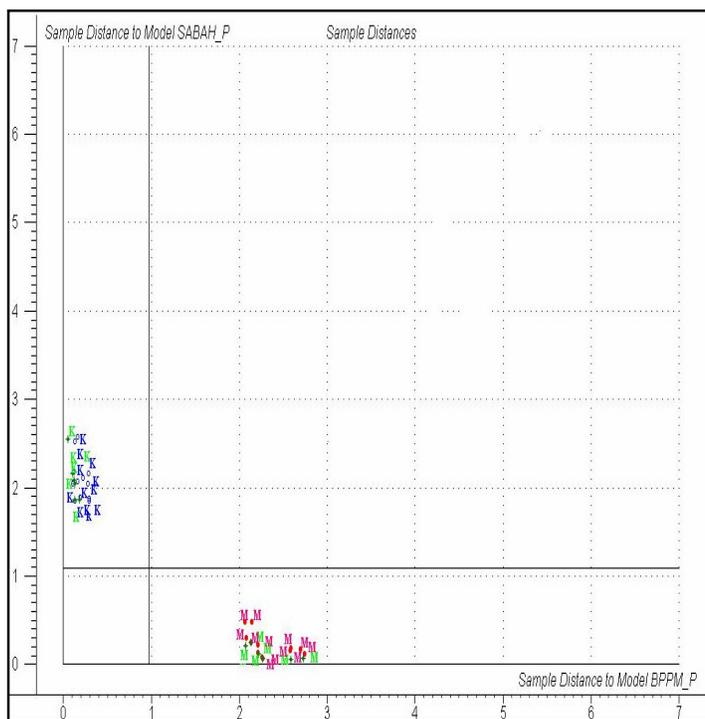
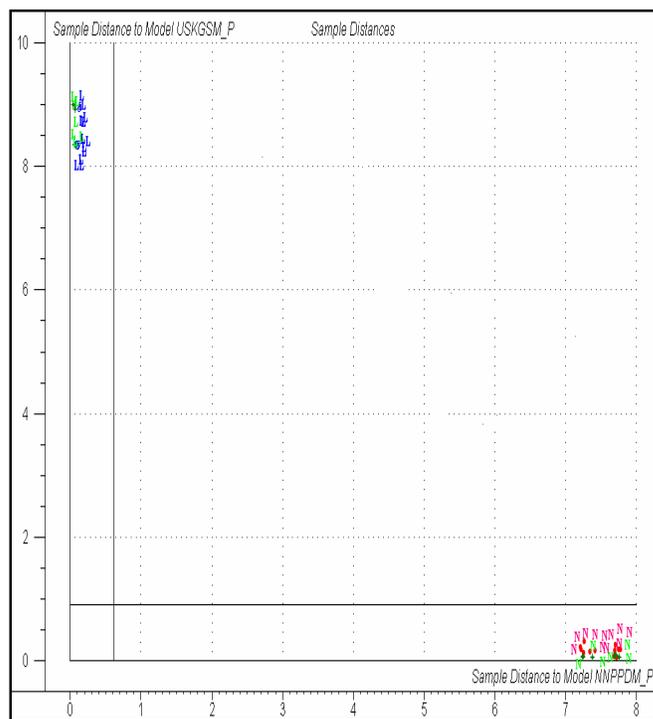


Fig. 5. 2-D Scores plot of the first three PCs (a) PC1 = 40% vs PC2 = 20% and (b) PC1 = 40% vs PC3 = 10% of different samples varieties from the same sourcing.



(a)



(b)

Fig. 6. Cooman's plot of sample-to-model distance for (a) Sabah_P vs BPPM_P samples and (b) USKGSM_P vs NNPPDM_P samples obtained from SIMCA classification.