

Chemometric Classification of Herb – *Orthosiphon stamineus* According to Its Geographical Origin Using Virtual Chemical Sensor Based Upon Fast GC*

Chew Oon Sim¹, Mohd Noor Ahmad^{1*}, Zhari Ismail², Abdul Rahman Othman³, Nor Amin Mohd Noor² and Ezrinda Mohd Zaihidee¹

¹School of Chemical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

²School of Pharmaceutical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

³School of Distance Education, Universiti Sains Malaysia, 11800 Penang, Malaysia

Author to whom correspondence should be addressed. amnoor@usm.my

Received: 5 June 2003 / Accepted: 28 August 2003 / Published: 31 October 2003

Abstract: An analytical method using Electronic Nose (*E-nose*) instrument for analysis of volatile organic compound from *Orthosiphon stamineus* raw samples have been developed. This instrument is a new chemical sensor based on Fast Gas Chromatography and Surface Acoustics Wave (SAW) detector. Chromatographic fingerprint obtained from the headspace analysis of *O. stamineus* samples were used as a guideline for optimum selection of an array of sensor. Qualitative analysis was carried out based on the responses of each sensor array in order to distinguish the geographical origin of the cultivated sample. The results of the analysis showed variances of volatile chemical compound of the samples even though it is from the same species. However, similarities of main components from all five samples were observed. Usage of pattern recognition chemometric approaches such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Cluster Analysis (CA) for processing

*The paper is already presented at "The 1st International Meeting On Microsensors & Microsystems, National Cheng Kung University, Tainan, Taiwan, 12-14 January 2003".

instrumental data provided good classification of *O. stamineus* samples according to its geographical origin.

Keywords: Electronic Nose; *Orthosiphon stamineus*; Fast Gas Chromatography; Chemometric; Virtual chemical sensor; Pattern recognition.

Introduction

Herbal medicine is an important part of health care to a majority of the world's population. Medicinal herbs namely Misai Kucing (Malaysia) or Kumis Kucing (Indonesia) have been trusted traditionally as a diuretic and has been used in treating urinary lithiasis, edema, eruptive fever, influenza, rheumatism, hepatitis, jaundice and biliary lithiasis [1-2].

In this research, dried leaves of *O. stamineus* cultivated commercially in different geographical origin have been classified using a virtual chemical sensor based on Fast Gas Chromatography (GC) with Surface Acoustic Wave (SAW) detector namely zNoseTM [3]. The resultant instrumental data was processed with chemometrics using a multivariate statistical analysis method in order to classify the raw sample according to its place of origin.

Scientific development on Orthosiphon stamineus

Scientific research on Misai Kucing especially on species *Orthosiphon stamineus* had begun since 1970. According to Van der Venn [4], *O. stamineus* is applied as a medicinal plant because of the diuretic and bacteriostatic properties due to the presence of potassium, inositol and lipophilic flavones in its leaves. Based on the preliminary report [5], Schmidt and Bos further investigated the composition of the essential oil derived from commercial fresh leaf and stem namely *Orthosiphon folium* DAB 8 using Gas Chromatography-Mass Spectrometry (GC-MS). The results obtained shown that b-caryophyllene, b-elemene, humulene, b-bourbonene, 1-octen-3-ol and caryophyllene oxide were the main volatile organic compound.

In 1992, Masuda [6] examine the constituents of *O. stamineus* leaves chemotaxonomically and for the first time, a new highly oxygenated diterpene, orthosiphon A was successfully isolated by repeated silica gel chromatography from the methylene chloride extract. Since then particular attention was made to isolate different diterpenes from this plant [7-10].

In Malaysia, research institute such as the Institute of Medical Research and the Kuala Lumpur Hospital of the Malaysian Health Ministry are participating actively in conducting clinical study to prove the efficacy of *O. stamineus* in treating kidney stones disease. On the other hand, the School of Pharmaceutical Sciences, Universiti Sains Malaysia contribute to the scientific research and development

of *O. stamineus* by conducting the extraction, quality control, standardization, pharmacological and formulation research.

Electronic nose (E-nose)

Electronic nose is a system that mimics the human olfaction by combining the response of a set of chemical sensors with partial specificity for the measurement of volatiles and develop techniques to recognize patterns for data interpretation [11]. This human mimetic technology was developed since early the 1980's and for the first time, researchers at the University of Warwick, Coventry, England successfully developed the sensor array based upon metal oxide semiconductor and continually discovered the conducting polymer for odor detection based on conductivity changes [12]. E-nose has wide applications ranging from food technology, environmental, automobile, perfume, medical diagnosis and pharmaceutical testing.

In this study, the use of E-nose technology is expanded to the field of natural and herbal product research. The integration of this chemical sensor technology will hopefully be useful as a stand-alone analytical technique for quantitative and qualitative analysis of chemical constituent in the raw and final herbal product.

Operating principle of GC / SAW electronic nose

The operating principle of this version of electronic nose has many similarities comparative to the human perception as shown in Figure 1. The Gas Chromatography / Surface Acoustic Wave (GC/SAW) electronic nose system is based upon Fast Gas Chromatography (Fast GC) that has a number of advantages compare to normal GC such as (i) low operational costs per sample, (ii) shorter "time-to-result" and (iii) allows several replicated analyses of a sample.

The GC separates the components of a mixture by preferential adsorption in an ascending molecular-weight sequence onto a solid adsorbent material applied as a coating to the interior of the chromatography column. Each gas is identified by its unique retention time at which the center of a symmetrical peak appears on the chromatogram.

Conventional E-nose incorporates sensor arrays with a different adsorbent coating material on each sensor. In this version of E-nose, the gases were first separated in a small capillary loop trap filled with Tenax[®], an adsorbing compound that captures condensable vapors. The gases next pass through the chromatography column and then to a single, uncoated SAW sensor for analysis. The added mass of an analyte condensing on the crystal's surface lowers the vibrational frequency in direct proportion to the amount of condensate.

System component of GC/SAW electronic system

The system consists of a six-port, two-position valve; the loop trap; a sampling pump for pulling vapors into the loop trap; a source of clean helium for use as a carrier gas; the GC column, which is a short section of glass or metal capillary tubing ~0.25 mm in diameter; and the temperature controlled SAW vapor sensor [13].

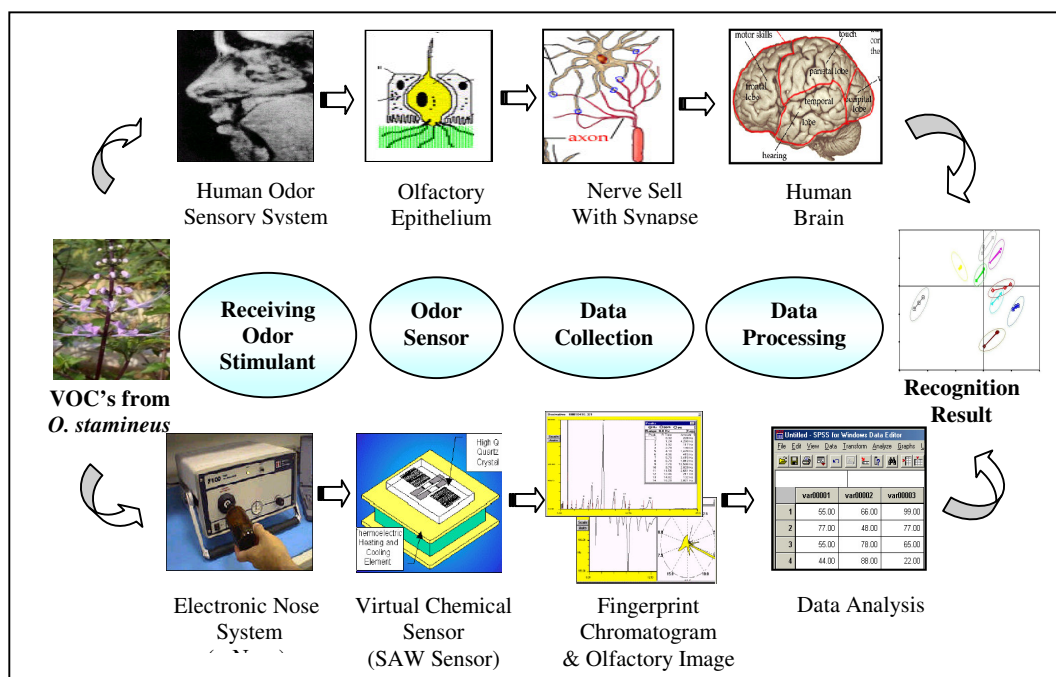


Figure 1. Comparison between human olfaction and GC/SAW electronic nose system.

Special feature on GC/SAW electronic system

There are three special feature of this version of electronic nose compared to the others available in the marketplace. The three special features identified are as follow:

- i) SAW detector response depends on the ability of analyte to absorb onto the cooled surface. Thus it is capable to detect an indefinite number of analyte without regard to analyte polarity or electronegativity. Besides, the SAW detector needs a lower voltage power source because radioactive ionization sources are excluded as compared to the Flame Ionization Detector in the normal GC system.
- ii) Preconcentrator vapor trap which functions as the purge and trap system in the normal GC apparatus. Only a small amount of sample is needed since the volatile compound from the headspace sample can be preconcentrated at a certain period before its was carried to the GC column by carrier gas flow.

iii) Direct-heated GC column with built-in heating element and shorter column (~1m) makes data acquisition within 10 seconds a reality.

With the above mentioned integrated features, electronic nose for the first time can serve as an alternative analytical techniques for herbal analysis that is less time consuming, cost effective and easy to operate compared to conventional analytical techniques such as High Performance Liquid Chromatography (HPLC), Thin Layer Chromatography (TLC) and Gas Chromatography-Mass Spectrometry (GC-MS).

Multivariate statistical analysis

Chemometrics

Chemometrics defined as the chemical discipline that uses mathematical, statistical and other methods employing formal logic to (a) design or select optimal measurement procedures and experiments and (b) to provide maximum relevant chemical information by analyzing chemical data [14]. A widely applied discipline of chemometrics is pattern recognition, which involves the classification and identification of samples. Its purpose is to develop a semiquantitative model that can be applied to the identification of unknown sample patterns [15]. As a conclusion, chemometrics analysis is used to analyze and interpret a cluster of raw data into knowledgeable information using statistic and mathematics model (refer Figure 2).



Figure 2. Transformation order of data to knowledge chemometrically.

Principal Component Analysis (PCA)

Principal Component Analysis is an unsupervised pattern recognition technique. This means that there is no prior knowledge of the classes that the samples will fall into. Originally in statistical analysis, PCA is a technique that reduces the dimensions of the data. Principal components Z_1, Z_2, \dots, Z_n are linear combinations of the original variables X_1, X_2, \dots, X_n [16]. The first two principal components (Z) were defined as follows:

$$Z_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \dots + a_{1n}X_n$$

$$Z_2 = a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + \dots + a_{2n}X_n \text{ etc.}$$

Coefficients a_{11} , a_{12} etc are selected so that the new variables are not correlated to each others. Besides that, first principal component (PC1), Z_1 has largest variance percentage and second principle component has the second largest variance. They were selected to show the data in two dimensions rather than in original n dimension. From a PCA, one therefore obtains two main results, namely (i) a two dimensional representation of the data so that relationships among points may be observed and (ii) the component coefficients a_{ij} and the correlations between the principal components and the original n variables give an indication about their significance in explaining the data structure in simpler terms and the partitioning of the data into clusters [14].

Cluster Analysis (CA)

Cluster analysis (CA) is a method to assign objects to groups. Like PCA, CA is also an unsupervised pattern recognition technique. Most CA techniques are hierarchical that is the resultant classification is in the term of nested classes. The goal in this studies is to identify a smaller number of groups so that elements residing in a particular group are, in some sense, more similar to each other than the elements belonging to the other groups. The construction of the homogeneous subgroups is generally based on the (dis)similarity of the measurement profiles [17].

Linear Discriminant Analysis (LDA)

Compared to PCA and CA, Linear Discriminant Analysis (LDA) is a supervised pattern recognition technique .That is, a learning sample (of known classes) is use to obtain a classification rule. This rule is then use to classify the test sample [14].

The first step in LDA is forming the linear dicriminant function, Y . Y is the linear combination of the origin variable X_1 , X_2 etc.

$$Y_1 = a_1X_1 + a_2X_2 + \dots + a_nX_n$$

The original data, n from the measurements of each object has been combined to one Y value. As a result, data in n dimensions were reduced to one dimension. Based on the Y value the criteria for assigning objects to the respective classes were determined.

$$\text{Classification power} = \frac{\text{No. of correctly classified individuals}}{\text{Sample population}} \times 100$$

The classification power of the analytical data is give by the number of objects correctly predicted to belong to the assigned classed was (expressed as a percentage of the class population) [18].

Experimental

Materials

Samples *O. stamineus* (dried leaves) from five different geographical origins (Figure 3) were collected from the distributors and the samples were named using alphabetical codes as shown in Table 1.

Table 1: The list of *O. stamineus* samples according to its geographical origin.

Code*	Location	State	Country
SRKBPM	Kepala Batas	Pulau Pinang	Malaysia
STJGCM	Jengka	Pahang	Malaysia
ZBPRAM	Parit	Perak	Malaysia
NNPPDM	Pasir Puteh	Kelantan	Malaysia
NHPJI	Pulau Jawa	Jakarta	Indonesia

* Example SRKBPM: SR: distributor; KB: location; P: state; M: Country.



Figure 3. Location map of *O. stamineus* samples from Malaysia and Indonesia.

Sample preparation

An amount of dried samples were milled until they became fine powder. 0.1g samples were placed in a 2ml headspace vial, which was then closed with a PTFE (Kimble Glass Inc.) septum cap. Fast GC/SAW electronic nose system namely zNose™, Model 7100 (Electronic Sensor Technology, California) was used to analyze volatile organic compound (VOC) from the headspace samples.

Analytical conditions

Samples headspace vial were thermostating 30 minutes at 60°C with heating block (Cole Palmer Inc.). GC/SAW electronic nose operating condition were as follows; injection time 5 second, inlet temperature 180°C, valve temperature 120°C, detector temperature 40°C, ramp 40°C-100°C @ 10°C/sec, capillary column: DB-624, 1m x 0.25mm x 1mm, carrier gas: helium 10 ml/sec, data acquisition time 10 second. Triplicate measurements per vial were carried out for each different geographical origin samples.

Data transformation and data analysis

Data transformation was performed using MS Excel. First, a set of particular GC peak was chosen as “virtual sensor array” based on the corresponding GC profile. The frequency data of each peak (“sensor”) was then calculated as the mean average frequency obtained from the triplicate deviation by using SPSS 9.0. Finally, PCA, CA and LDA were also carried out using SPSS 9.0 in order to classify the sample according to its geographical origin.

Results and discussion

Optimum virtual chemical sensor array selection

The perfect combination of GC direct-heated column and the SAW detector makes a virtual physical sensor. Although the system contains a single physical sensor the compatible system software, namely MicroSense 3.6 can create hundreds of virtual chemical sensors based upon retention time windows. This means that each peak of the GC profile can be identified as a response of a single chemical sensor and at the same time correspond to only one analyte or chemical compound found in the sample. This approach was also reported by Dittmann [19] which assumes that each fragment ion obtained from the mass spectrum characterized certain chemical compound. Typical profile chromatograms of five samples with numerical label representing selected virtual sensor responses are shown in Figure 4.

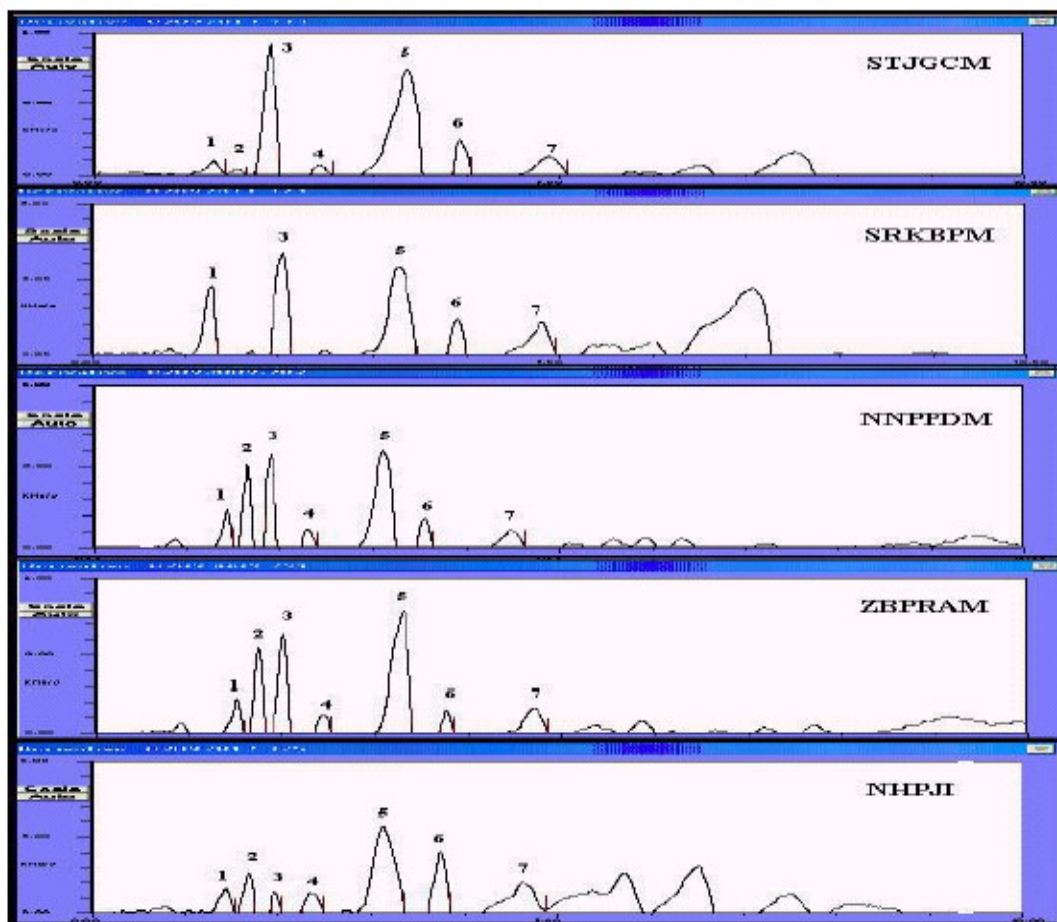


Figure 4. Profile chromatogram representing an array of virtual chemical sensor, which characterized the different geographical origin samples respectively.

Classification by pattern recognition

Image olfactory (VaporPrint™ Image)

According to the inventor of the GC/SAW electronic nose system [20], image olfactory is a high resolution (500 pixel) two-dimensional visually recognizable images, which can also quantify the strength of each chemical within a fragrance. The image is a closed polar plot of the odor amplitude (SAW detector frequency) with radial angles representing sensors. A brief conclusion can be drawn by making comparison among the vapor images shown in Figure 5. Hence, *O. stamineus* from different origins were represented by their own aroma patterns. So the unknown samples can be easily classified according to its origin by making comparison with the vapor image of reference sample. But this approach is not reliable when the vapor images look similar to each other as shown by ZBPRAM and NNPPDM samples. In this situation, high probability of misclassification can happen.

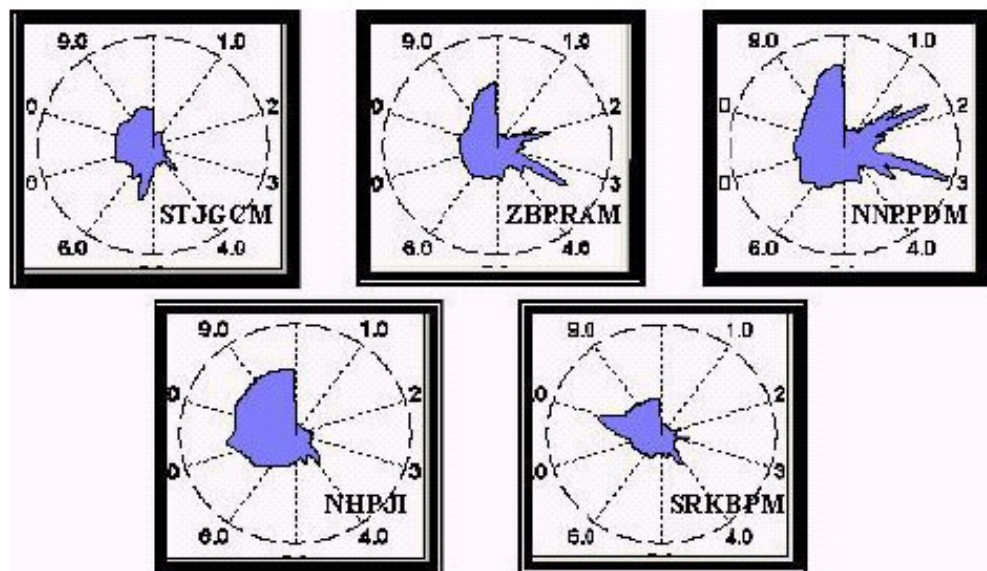


Figure 5. Image olfactory of different geographical origin *O. stamineus* samples.

Due to the stated limitation, the use of chemometric approaches using unsupervised pattern recognition techniques namely PCA and CA were investigated. Besides that, LDA that is a supervised pattern recognition technique is also being applied in this study.

Chemometrics classification

Data pretreatment

According to Massart and coworkers [14], if the distribution of variables as in this study the frequency data are not normal but severely skewed, then the reliable or successful results cannot be obtained from most multivariate statistical analyses. On the other hand, Miller [16] stressed that decision must be made whether raw data or standardized data (min=0 and standard deviation=1) is used for data analysis before PCA and LDA is carried out. Based on the above statement, data frequency from the instrument analysis is standardized to obtain equal weight of all the variables. If no data pretreatment is done, the zero reading in sensor 2 and sensor 4 shown by SRKBPM samples account for such a large variance and false classification may occur.

Principal component analysis (PCA)

Figure 6 shows the scatter plot of the standardized frequency data in two dimensions. Together, first two principal components represent 67.39% of the total variance (PC1 = 41.58% and PC2 = 25.80%). The two principal components are independent. A straight line passing through the data points represents a linear combination of the corresponding variables.

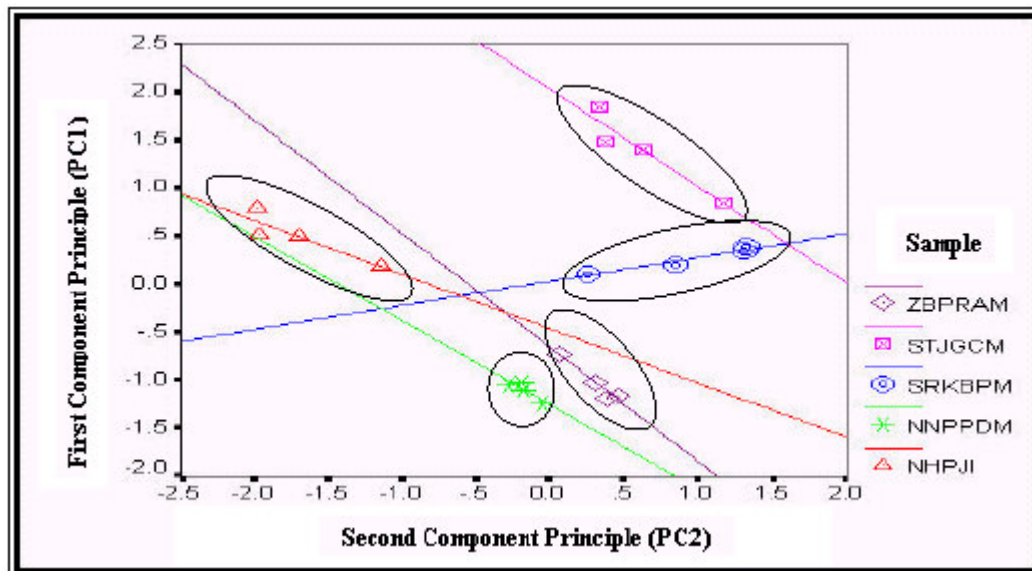


Figure 6. Principal component analysis of the virtual sensor array responses to *O. stamineus* samples of different geographical origin.

A good separation between NHPJI samples from the others samples is obtained. This observation explain the fact that the cultivated area of herbal medicines is a controlling factor in the quality of the herb due to the different growing conditions. Classification of NNPPDM and ZBPRAM samples indicated the volatile composition of the both samples are similar and do not differ enough to make a good separation [21].

As a result, PCA is not so effective for classification of the *O. stamineus* samples according to its geographical origin. Although Togari and coworkers [22] reported that PCA showed effectiveness for classification of tea samples according to its categories (fermented and unfermented) based on the GC profile. Thus, the classification power of LDA a supervised pattern recognition techniques is thus investigated.

Linear discriminant analysis (LDA)

This supervised pattern recognition techniques had been applied widely for the classification purposes. Martin and coworkers [23], have proven that the LDA method shows good classification and prediction capabilities of vegetable oils.

LDA when applied to classify *O. stamineus* samples based on its origin seem to give good classification results as shown in Figure 7. By using LDA, SRKBPM and STJGCM samples separated well on the negative side of the x-axis and the y-axis, which was not clearly classified by PCA. As a conclusion, this study finds that LDA is a more powerful tool compared to PCA in terms of classification.

This is mainly because LDA selects direction, which gives maximum separation between the studied classes [14].

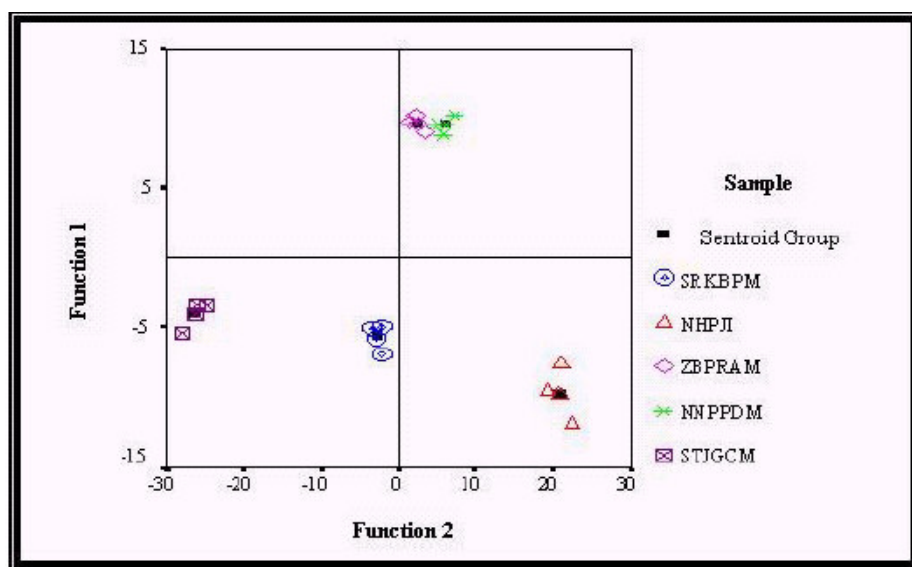


Figure 7. Discriminant plot of *O. stamineus* samples by LDA method.

Cluster analysis (CA)

CA was carried out using raw data obtained from the analytical instrument in order to study the capabilities of the selected virtual sensor array for classification of the samples based on geographical origin. This approach is able to assign a group of objects to its respective classes so that similar objects are in the same classes. The resultant dendrogram gives extra information regarding the raw data obtained from instrumental analysis.

This study employs Average Linkage method and Euclidean distance as distance measure between objects. Euclidean distance was used because the distance between any two objects is not affected by the addition of new objects to the analysis, which may be outliers. The dendrogram (Figure 8) horizontal scale (0-25) give pictures of similarity and dissimilarity among samples. Samples from the same origin form individual clusters except SRKBPM sample as indicated by no. 19 form a cluster with ZBPRAM sample. The probable reason for this misclassification was mainly due to the zero reading sensor response shown by SRKBPM samples.

Conclusion

With the special integrated feature in the GC/SAW electronic nose system, for the first time Enose can serve as an alternative analytical technique for herbal analysis that is less time consuming, cost effective and easy to operate compared to conventional analytical techniques such as High Performance Liquid

Chromatography (HPLC), Thin Layer Chromatography (TLC) and Gas Chromatography-Mass Spectrometry (GC-MS). Chemometric pattern recognition applied to the selected optimum virtual sensor data from the GC profile is effective in classifying *O. stamineus* samples according to its geographical origin. The combination of the chemometrics approach and GC/SAW electronic nose shows to be a promising analytical technique for herbal analysis. Further study is needed with some modifications in the analytical procedure that emphasized on quantification of the chemical constituent in *O. stamineus*.

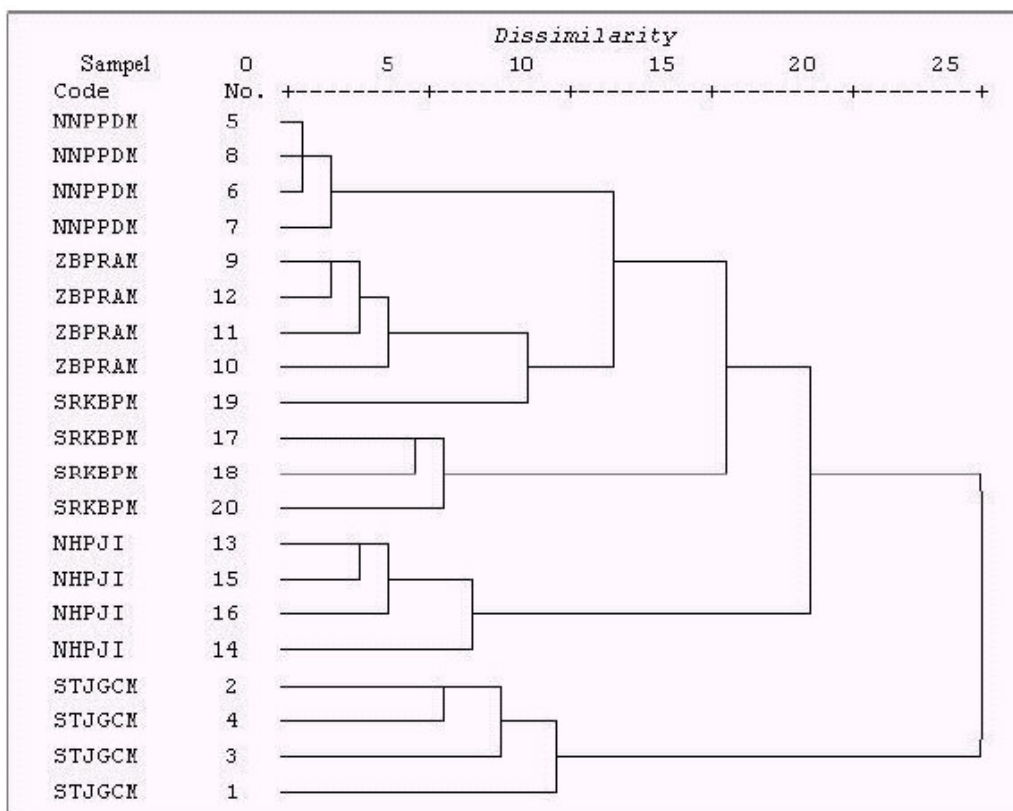


Figure 8. Dendrogram of cluster analysis on different geographical *O. stamineus* samples.

Acknowledgement

The authors gratefully acknowledge the financial support from the Ministry of Science, Technology and Environment, Malaysia through IRPA No. 610629 for this research. Dr Sani Mohd Ibrahim was thanked for his suggestion and knowledge sharing in the area of gas chromatography.

References

1. Sumaryono, W.; Proksch, P.; Wray, V.; Witte, L.; Hartmann, T. *Planta Med.* **1991**, *57*, 176-180.
2. WHO Regional Office for the Western Pacific Manila and Institute of Material Hanoi, *In Medicinal Plants in Vietnam*, Science and Technology Publishing House: Hanoi, **1990**; pp 271.

3. Ahmad, M.N.; Zaihidee, E.M.; Chew, O.S.; Othman, A.R.; Ismail, Z.; Hitam, M.S.; Md Shakaff, A.Y., Chemometric Classification of Herb – *Orthosiphon Stamineus* (Cat Whisker's) According to Geographical Origin Using a New Chemical Sensor Based on a Fast GC, The 1st *International Meeting On Microsensors & Microsystems*, National Cheng Kung University, Tainan, Taiwan, 12- 14 January 2003.
4. Van deer Veen, X.; Makubard, Th.M.; Zwazing, J.H. *Pharmaceutisch Weekblad*, **1979**, 114, 965.
5. Schmidt, S., Bos, R. Volatile Of *Orthosiphon stamieus* Benth, *Progress in Essential Oil Research*, Walter de Gruyter & Co., Berlin: New York, **1986**; pp 93-97.
6. Masuda, T.; Masuda, K.; Nobuji, N. Orthosiphol A, A Highly Oxygenated Diterpene From The *Orthosiphon stamineus*, *Tetrahedron Letter*. **1992**, 33(7), 945-946.
7. Paviour Stampoulis; Tezuka, Y.; Arjun, H.B.; Kim, Q.T.; Saiki, I.; Kadata, S. Staminol A, A Novel Diterpene From *Orthosiphon stamineus*, *Tetrahedron Letter*. **1999**, 40, 4239-4242.
8. Suresh Awale; Tezuka, Y.; Banskota, H.B.; Shimoji, S.; Taira, K.; Kodata, S. Secoorthosiphols AC: Three Highly Oxygenated Secoisopimarane-type Diterpenes From *Orthosiphon stamineus*, *Tetrahedron Letter*. **2002**, 43, 1473-1475.
9. Suresh Awale; Tezuka, Y.; Banskota, H.B.; Shimoji, S.; Taira, K.; Kodata, S. Norstamine and Isopimarane-type Diterpene Of *Orthosiphon stamineus* From Okinawa, *Tetrahedron*. **2002**, 58, 5503-5512.
10. Suresh Awale; Tezuka, Y.; Banskota, H.B.; Kodata, S. Siphonols A-E: Novel Nitric Oxide Inhibitors From *Orthosiphon stamineus* Of Indonesia, *Bioorganic & Medicinal Chemistry Letters*. **2003**,13, 31-35.
11. Gardner, J.W., Bartlett, P.N. *Electronic Noses: Principle and Applications*, Oxford University Press: New York, **1999**.
12. Nagle, H.T.; Schiffman, S.; Gutierrez-Osuna, R. The How and Why of Electronic Noses, *IEEE Spectrum*. **September 1998**, 22-34.
13. Staples, E.J.; Matura, T.; Viswanathan, S. Real Time Environmental Screening Of Air, Water And Soil Matrices Using A Novel Field Portable GC/SAW System, *Environmental Strategies for the 21st Century, Asia acific Conference, Singapore*. **8-10 April 1998**.
14. Massart, D.L.; Vandeginste, B.G.M.; Deming, S.N.; Michotte, Y.; Kaufman, L. *Chemometrics: A Textbook*, Vol 2, Elsevier: Amsterdam. **1998**.
15. Aboul-Enein, H.Y.; Stefcin, R.; Baiulescu, G. *Quality & Reliability in Analytical Chemistry*, CRC Press UC, **2000**, 49-53.

Sample Availability: Available from the authors.

© 2003 by MDPI (<http://www.mdpi.org>). Reproduction is permitted for noncommercial purposes.