Food Chemistry 135 (2012) 213–218

Contents lists available at SciVerse ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Analytical Methods

Qualitative and quantitative analysis in sandalwood oils using near infrared spectroscopy combined with chemometric techniques

Saji Kuriakose [1], Hubert Joe *

Centre for Molecular and Biophysics, Mar Ivanios College, Thiruvananthapuram, Kerala, India

ARTICLE INFO

Article history:
Received 30 April 2011
Received in revised form 16 May 2011
Accepted 15 April 2012
Available online 21 April 2012

Keywords:
Essential oils
Sandalwood oil
Near infrared spectroscopy
Chemometric techniques
Support vector machine regression
Qualitative
Quantitative differences

ABSTRACT

Sandalwood oil is an essential oil which finds very wide application in the flavor and fragrance, pharmaceutical industry. The objective of this study is to use the potential of near infrared spectroscopy as a rapid analytical technique for the qualitative and quantitative assessment of purity in sandalwood oils. The quality and efficacy of sandalwood oils, even though come from the same species, are somewhat different according to growing conditions (origin) and poor extraction methods. Classification of sandal oils based on their NIR spectra is performed by principal component analysis, hierarchical cluster analysis and self organising map (Kohonen neural network). All these techniques clearly differentiate the oils according to the area from which the sandalwood has been cut. Support vector machine regression (SVM R) is used to predict the purity of the oils.

1. Introduction

Sandalwood oil is an essential oil obtained by the distillation of the heartwood and roots of the plant *Santalum album* (family – Santalaceae). *S.* album is a small hemiparasitic tree of great economic value, growing in Southern India, Sri Lanka, Australia and Indonesia. Its trunk contains resins and essential oils particularly the α and β-santalols, santalenes and many other minor sesquiterpenoids (Jones, Ghisalberti, Plummer, & Barbour, 2006). These sesquiterpenoids are responsible for the unique sandalwood fragrance. Sandalwood oil is used in the food industry as a flavour ingredient. This oil serves as a fixative for many high – end perfumes. A number of aromatic and phenolic compounds have also been identified in the oil *S. album* (Kim et al., 2005). The quantity of oil produced in a tree varies considerably according to location (environmental factors) and age of the tree, even in nearly identical growing conditions (Jones, Plummer, & Barbour, 2007). It should also be noted that santalol composition can vary depending on the method of oil extraction (Piggott, Ghisalberti, & Trengove, 1997). There has been serious decline in the population of santalum in India due to complex cultivation requirements and non-stop harvesting (especially from smuggling) associated with limited regeneration (Fox, 2000; Radomiljac, Ananthapadmanabha, Welbourn, & Rao, 1998).

Sandalwood oil is approved for food and flavour uses by Council of Europe (CoE, 2000), Flavour and Extract Manufactures Association (FEMA) and the United States Food and Drug Administration (FDA). The sandalwood oil specifications have been reported in the Food Chemicals Codex (FCC, 2003). The international standard (ISO 3518, 2002) for sandalwood oil and similar authorities stipulate a minimum of 90% w/w santalol (as free alcohol) in the oil (α-santalol comprising approximately 60% and β-santalol comprising approximately 33% of total santalol) (British Pharmaceutical Codex, 1949; ISO 3518, 2002). Sandalwood oils with santalol level below these specifications are of inferior quality due to poor extraction methods, adulteration with synthetic or semi-synthetic substitutes or some other counterfeit substitution.

Kumar and Maddan (1979) have reported the recommended iodine value as 283–288 to identify adulteration in pure sandalwood oil. Verghese, Sunny, and Balakrishnan (1990) have suggested, using gas chromatography (GC), 40–55% of α-santalol and 17–27% of β-santalol in the total santalol content (w/w). GC analysis of *S.* album oil by ISO 3518 (2002) specifies similar proportions of Z–α-santalol (41–55%) and Z–β-santalol (16–24%) (ISO 3518, 2002). However, these reports do not discuss the potential variation in *S. album* oil composition depending on the origin or age of the tree, nor do they address the evaluation of counterfeit.

The quality and efficacy of the sandalwood oil, even from same species, are somewhat different according to growing conditions

* Corresponding author. Address: Centre for Molecular and Biophysics Research, Department of Physics, Mar Ivanios College, Thiruvananthapuram 695 015, Kerala, India. Tel.: +91 471 2531053; fax: +91 471 2530023.
E-mail addresses: hubertjoe@sancharnet.in, hubertjoe@gmail.com (H. Joe).
[1] Present address: St. Thomas H.S.S., Pala, Kerala, India.

0308-8146/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodchem.2012.04.073

based on geographical origin. This enforces the requirement of a rapid and accurate analytical method for the correct value estimation based on origin and for the prevention of illegal distribution. However the existing analytical tools are not sufficient to determine the geographical origin clearly as they are time consuming, complex and tedious. Since sandalwood oil contains more than 100 major components that are slightly different according to growing conditions viz. geo. origin, we can not select several specific components as essential criteria. In view of the current issues associated with sandalwood oil, we conduct an investigative study to develop an appropriate tool to assess the quality of sandalwood essential oils.

Near Infrared spectroscopy (NIRS) in combination with sophisticated chemometric algorithms can be an excellent tool for quality control purposes and selection of high quality materials. NIR pattern recognition method was applied for the discrimination of roasted coffees (Martin, Pablos, & Gonzalez, 1996) and vegetable oils (Bewig, Clarke, Roberts, & Unklesbay, 1994). These proposed researches were based on the classification of samples with various chemical constituents. In our previous study, we have developed two reliable, accurate and non- destructive models - principal component regression (PCR) and partial least square regression (PLSR) models – to detect and quantify caster oil adulteration in pure sandalwood oil through near infrared spectral data collected from blended sandal oil samples (Kuriakose, Thankappan, Joe, & Venkataraman, 2010).

The objective of the present work is to apply NIRS along with chemometric techniques like principal component analysis (PCA) (Wold, Esbensen, & Geladi, 1987) to ascertain discrimination of sandal oils. Also pattern recognition technique namely hierarchical cluster analysis (HCA) (Armenta, Garrigues, & Guardia, 2007) and self organising map (SOM) (Kohonen, Oja, Simula, Visa, & Kangas, 1996) are applied to classify sandal oils from same species (chemotypes) but different geographical origin. Support vector machine regression (SVM R) (Vapnik, 1995) is used to quantify the % level of counterfeits in the oils.

2. Materials and experimental methods

2.1. Materials

Five sandalwood oil samples are procured from three different geographical origins (three regions of two states, Kerala and Karnataka in India) consisting of the same species. The samples are classified and named into five groups as A–E. Sample A is acquired from Kairali, Arts and Crafts, Kerala Government. Samples B and C are obtained from two different sandalwood factories in Mysore, Karnataka State. Samples D and E are collected from two sandalwood industries in Banglore, Karnataka State. All samples are stored at 4 °C in aluminium bottles and protected from light until they are analysed. None of them are subjected to any treatment as these may change their composition.

2.2. Samples preparation

At least 8 h prior to spectroscopic measurement, the samples are brought to ambient temperature (20 °C). Each sample is diluted to 10% with proper solvent (carbon tetrachloride (v/v)) and homogenised using an electromagnetic stirrer. Stringent protocols have been applied to avoid any sample loss or change. A total of 49 samples are prepared from all the classes. (10 samples each from 4 groups A–D and 9 samples from group E).

2.3. NIR spectra collection

Near infrared spectra of 49 samples are recorded over 800 nm – 2500 nm spectral region, at 1 nm spacing with a NIR spectrophotometer Cary 5000 (SI No. EL 03127331) with a resolution of 0.01 nm. The spectra are collected in 1 nm data intervals. An average spectrum of a number of spectra for each sample is obtained. All the spectra are recorded in absorbance units. The sample spectra are divided into training set ($n = 39$) and test set ($n = 10$), comprising those from each group.

2.4. Chemometrics and data analysis

PCA, HCA, SOM and SVM R are performed using algorithms from PLS Toolbox 6.0.1 supported by Matlab environment (Wise, Gallagher, Bro, & Shaver, 2010; Matlab, 2010a, 2010). (In this study, the unsupervised methods PCA, HCA and SOM are used for model comparison and confirmation of the results obtained). A Proper preprocessing technique namely smoothing (Savizky–Golay filters) coupled with mean centre is used to remove background noise and to increase spectral resolution (Savizky & Golay, 1964). Leave – one-out (LOO) Cross-validation is used to calibrate the model.

3. Results and discussion

3.1. Near infrared absorbance spectra

The near infrared absorption spectra of 5 classes of sandalwood oils (collected from 3 different geographical origins) over the spectral range 800–2500 nm at 1 nm spacing are measured and are converted to ASCII files using Varian software. (Spectra of 49 samples; 10 samples each from classes A–D, 9 samples from Class E).

3.1.1. Spectra investigation

The near infrared region (780–2498 nm) is dominated by overtones and combination bands arising from the unharmonic nature of molecular vibrations. Absorption in the NIR region arises from the vibrational motion of molecules. According to former studies performed on essential oils, the majority of the absorption bands in the near infrared spectra of the analysed oil samples arise from the overtones of hydrogenic stretching vibrations or combination involving stretching and bending modes (Schulz, Prews, & Kruger, 1999). The absorption bands observed at 1135–1215 nm are due to methylene (CH) stretching (2nd (3n) overtone and 2n combination bands). The peaks at 1375–1399 nm are related to methly1 (CH) stretch and bending combination (2nd (3n) overtone and 2n combination bands). The peak around 1690–1695 nm is associated with methyl asymmetric stretching (1st (2n) overtone (CH) stretch) respectively (Westad, Schmidt, & Kemit, 2008). The intense peaks centered near 2308 and 2348 nm are related with combination of CH stretching vibration and deformation tones. (Hourant, Baeten, Morales, Meurens, & Aparicio, 2000).

3.1.2. Wavelength selection

There has been much debate as to the importance of selecting those few wavelengths that contain significant information for optimal model development, thus reducing the number of wavelengths, variables and model complexity. Recently researches are conducting to investigate the importance of combining some wavelengths, (synergic) to those containing problem dependent information (descriptive wavelengths) to improve the model performance. There are considerable differences observed in the spectra of all samples in the wavelengths range 1650–2500 nm. Also water (20 °C) has overtones at 1450, 970 and 760 nm. A combination of OH stretching and bending occurs around 1940 nm. The presence of water content hinders the model performance for this application. Taking into account the above circumstances, the full spectrum range is divided into two regions; 1650–1900 and 2000–2500 nm