Geographical provenance of palm oil by fatty acid and volatile compound fingerprinting techniques

A. Tres a,∗, C. Ruiz-Samblas a,b, G. van der Veen a, S.M. van Ruth a

a RIKILT, Wageningen University and Research Centre, Akkermaalsbos 2, 6708 WB Wageningen, The Netherlands
b Department of Analytical Chemistry, University of Granada, c/Fuentenueva, s.n. E-18071 Granada, Spain

A R T I C L E  I N F O

Article history:
Received 5 June 2012
Received in revised form 21 September 2012
Accepted 26 September 2012
Available online 24 October 2012

Keywords:
Authenticity
Chemometrics
Fatty acid composition
Geographical provenancing
PTR-MS
Palm oil
Volatile compounds

A B S T R A C T

Analytical methods are required in addition to administrative controls to verify the geographical origin of vegetable oils such as palm oil in an objective manner. In this study the application of fatty acid and volatile organic compound fingerprinting in combination with chemometrics have been applied to verify the geographical origin of crude palm oil (continental scale). For this purpose 94 crude palm oil samples were collected from South East Asia (55), South America (11) and Africa (28). Partial least squares discriminant analysis (PLS-DA) was used to develop a hierarchical classification model by combining two consecutive binary PLS-DA models. First, a PLS-DA model was built to distinguish South East Asian from non-South East Asian palm oil samples. Then a second model was developed, only for the non-Asian samples, to discriminate African from South American crude palm oil. Models were externally validated by using them to predict the identity of new authentic samples. The fatty acid fingerprinting model revealed three misclassified samples. The volatile compound fingerprinting models showed an 88%, 100% and 100% accuracy to predict the identity of new authentic samples. The fatty acid fingerprinting model revealed three misclassified samples. The volatile compound fingerprinting models showed an 88%, 100% and 100% accuracy for the South East Asian, African and American class, respectively. The verification of the geographical origin of crude palm oil is feasible by fatty acid and volatile compound fingerprinting. Further research is required to further validate the approach and to increase its spatial specificity to country/province scale.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Palm oil is widely used in the food sector as well as in the non-food sector where it has applications in the biofuel, soap, detergents and toiletry industries among others (Aini & Miskandar, 2007; Rupilius & Ahmad, 2007; Siew, 2002). The demand for palm oil has increased steeply in recent decades leading to a global increase in its production by 9% per year (Fitzherbert et al., 2008). The main factors prompting this high demand are related to the higher use of palm oil for biofuel production, the higher demand of certain countries for food applications and its use by the food industry as a low trans vegetable fat alternative to hydrogenated vegetable oils (Aini & Miskandar, 2007; Fitzherbert et al., 2008; Obidzinski, Andriani, Komarudin, & Andrianto, 2012; Rupilius & Ahmad, 2007). Palm oil is produced in tropical areas with Indonesia and Malaysia currently being the two main producing countries. The rapid expansion of palm oil crops in equatorial regions has raised concerns about its potential detrimental effects on biodiversity (Fitzherbert et al., 2008), leading to intense debates between environmental non-governmental organisations and the palm oil industry (Fitzherbert et al., 2008). The production of sustainable palm oil has emerged, among other initiatives, to reduce the environmental impact and protect biodiversity in those areas (Basiron, 2007; Fitzherbert et al., 2008). Several certification systems have been implemented in order to inform consumers and to protect the sections of the palm oil industry that apply environmentally friendly practices (Basiron, 2007; Roundtable on Sustainable Palm Oil, 2012). This has driven the development of a premium market for sustainable palm oil products. Indeed, several food companies and governments are aiming at using only certified sustainable palm oil within a few years (Task Force Duurzame Palm Olie, 2010).

Like other commodities with added-value, sustainable palm oil is susceptible to fraud. Currently, the certification and verification process of sustainable palm oil is based on administrative controls and inspections. For objective verification of the true identity of sustainable palm oil, in addition to administrative controls, analytical methods would be very useful. As far as we know, however, no analytical methods have been developed so far, mainly because finding an analytical marker to discriminate sustainable palm oil from non-sustainable palm oil is a challenging task.

Nowadays, the production of sustainable palm oil takes place in a limited number of regions (Roundtable on Sustainable Palm Oil, 2012). Therefore, developing an analytical method for the verification of the geographical origin of palm oil would be a first step in the authentication of sustainable palm oil (Tres, Ruiz-Samblas, 2012 Elsevier Ltd. All rights reserved.)
2.3. Fatty acid composition

(South America (11 samples from Brazil).)

Although literature concerning compositional variation in palm oil is scarce, some studies have indicated that the FA composition might vary within the geographical origin (Clegg, 1973; Rosselli, King, & Downes, 1985; Siew, 2002; Tres, van der Veer, Alewijn, Kok, & van Ruth, 2011a). This, together with previous findings in other vegetable oils, suggests that FA composition and VOC profile might be suitable candidates to develop an analytical tool to verify the geographical origin of palm oil.

Fingerprinting techniques combined with chemometrics are state-of-the-art analytical techniques in food authentication. These fingerprinting techniques aim to find a specific pattern for the authentic product (i.e., for each geographical origin), that might allow discrimination between different products (i.e., those from different geographical origins) (Berrueta, Alonso-Salces, & Héberger, 2007). This pattern might be based on the known compositional data of the product (for instance the FA composition of the product) (Tres & van Ruth, 2011) or even on the raw signal provided by certain analytical equipment (such as chromatograms, or mass or near infrared spectra) (Araghipour et al., 2008; Bosque-Sendra, Cuadros-Rodriguez, Ruiz-Samblás, & de la Mata, 2012; Woodcock, Downey, & O’Donnell, 2008). The advantage of these fingerprinting techniques is that they are able to detect non-conformities that would be unnoticed by the use of a traditional targeted approach.

The objective of this study was therefore to develop a chemometric classification model to verify the geographical origin of crude palm oil based on its FA and VOC fingerprints.

2. Materials and methods

2.1. Samples

A total of 94 crude palm oil samples was collected from palm oil plantations, mills and traders during 2010–2011. The sample set included crude palm oil samples from three different continents with major palm oil production: South East Asia (55 samples from Malaysia, Indonesia, Papua New Guinea and Salomon Islands), Africa (28 samples from West Africa – Ghana, Guinea, Cote d’Ivoire, Nigeria and Cameroon) and South America (11 samples from Brazil).

2.2. Reagents and standards

Sodium methoxide (0.5 N) was purchased from Sigma–Aldrich (St. Louis, MO). Boron trifluoride methanol complex (35%) was obtained from Merck (Darmstadt, Germany). The FA methyl ester mixture of standards was supplied by Supelco (Supelco 37 Component FAME mix, Supelco, St. Louis, MO). All the other reagents were of ACS quality grade.

2.3. Fatty acid composition

Oil was melted at 45 °C and weighed (100 mg) in a screw-capped glass tube. The FA methyl esters were obtained as previously described (Tres & Van Ruth, 2011) using a two-step methylation procedure with sodium methoxide (0.5 N) and boron trifluoride–methanol complex (35%). FA methyl esters were determined by gas chromatography in a Varian (Palo Alto, CA) CP-3800 model gas chromatograph, fitted with a flame-ionisation detector and split-splitless injector port, set at 280 °C and 250 °C, respectively. The split ratio was 1:30. Chromatographic separation of FA methyl esters were performed on CP-Select CB for FAME capillary column (50 m × 0.25 mm i.d.; Varian, Palo Alto, CA). Helium (18 p.s.i.) was used as carrier gas and the oven was programmed as follows: initial temperature, 100 °C, increased at 5 °C/min to 230 °C and held for 9 min. The sample volume injected was 1 µL. Fatty acids were identified by their retention times according to those found in the FAME standard mixture. Results were expressed as normalised peak areas (%). All palm oil samples were analysed in triplicate. Data used were the FA average value of the three replicates of each palm oil sample.

2.4. Proton transfer reaction-mass spectrometry

The VOC fingerprint was measured by proton transfer reaction-mass spectrometry (PTR-MS), which is a one-dimensional rapid technique to measure the VOC fingerprint of a sample. The sample headspace is continuously introduced into a drift tube where it is mixed with H2O+ ions formed in a hollow cathode ion source. VOC that have proton affinities higher than water (>166.58 kcal/mol) are ionised by proton transfer from H2O+, mass analyzed in a quadrupole mass spectrometer and eventually detected as ion counts/s (cps) by a secondary electron multiplier. One of the advantages of PTR-MS is that by using H2O+ as the proton source, the ionisation of most of the common inorganic constituents of air (N2, O2 or CO2) is avoided since they have proton affinities lower than that of H2O. Furthermore, this soft ionisation avoids excessive fragmentation of ions, which makes multicomponent analyte mass spectra simpler and easier to interpret (Hansel et al., 1995). However, its low resolution (comparatively with other techniques such as proton transfer reaction time of flight mass spectrometry (PTR-TOF-MS)) gives rise to difficulties in obtaining unequivocal compound identification (Heenan et al., 2012). However, unequivocal compound identification is not necessary for a fingerprint approach in which chemometrics is applied on the full mass spectrum (mass-to-charge ratios and their intensities). Therefore, PTR-MS is an interesting as a fingerprint technique as (i) it requires no sample pre-treatment, (ii) it allows rapid measurements (typically less than 1 min for a complete mass spectrum) and (iii) it is an extremely sensitive technique (Buhr, van Ruth, & Delahunty, 2002; Luykx & van Ruth, 2008).

Crude palm oil (5 mL) (melted at 45 °C) was placed in a 250 mL screw cap glass vial. Samples were equilibrated at 45 °C for at least 30 min in a water bath under agitation (60 rpm) in order to release VOC from the sample into the headspace until equilibrium was reached. Measurements were performed using a commercial HS-PTR-MS (high-sensitivity PTR-MS) system (Ionicon GmbH, Innsbruck, Austria). The headspace of the samples was delivered directly to the inlet of the PTR-MS system with a flow rate of 48 mL/min. The temperature of the inlet and drift chamber were both maintained at 60 °C to prevent loss of volatiles along the sampling line. A blank air sample was measured before each sample. Measurements were carried out in the mass full-scan mode and the mass spectra were collected in the range of 20–160 m/z (mass-to-charge ratios). A dwell time of 0.2 s × (m/z)−1 was used, resulting in a cycle time just under 30 s.

Sample analyses were carried out in triplicate. For each replicate, 5 full mass scans were recorded, but the first and last cycles were discarded for calculations. The second, third and fourth scans were averaged and the data were background and transmission corrected, yielding an averaged mass spectrum per replicate. Then, the three averaged mass spectra of the three replicates of each
sample were averaged to obtain a mean mass spectrum per sample. In this manner, a dataset containing mean mass spectra per sample analysed could be compiled. \( m/z \) 32 (\( O_{2}^{+} \)) and \( m/z \) 37 (water cluster ion) that are associated with the PTR-MS ion source were removed from the data set.

2.5. Statistics

Data matrices consisted of as many rows as samples (94) and as many columns as variables (21 for FA composition matrix and 138 for VOC fingerprint matrix). First, Kruskal–Wallis test was conducted with the FA composition data to determine whether the distribution of each FA was similar among the three continents (\( P \leq 0.05 \) was considered significant). Pairwise comparisons were conducted between each two groups as post hoc test. Software used for Kruskal–Wallis test was SPSS v19.0 (IBM Statistics Ink).

For further multivariate modelling and classification, Pirouette 4.0 (Infometrix, Seattle, USA) was used. Principal Component Analysis (PCA) was performed with the 94 samples to screen the multivariate data for outliers and to explore the presence of any natural clustering in the data (Berrueta et al., 2007).

Partial least squares-discriminant analysis (PLS-DA) was conducted to develop the classification model to verify the geographical origin of crude palm oil at a continental scale. Thus, the model was developed to verify whether crude palm oil originated from South East Asia, Africa or South America. PLS-DA performs a variable reduction on the data set by calculating new variables (called latent components or factors) combining the variables in the data set in order to find the maximum correlation between them and the class variable and, thus, the maximum separation among the classes. Then, the model is developed using this reduced variable set (the latent components). PLS-DA might find the maximum separation between by minimising the influence of the common features in the FA profile of different categories, in favour of increasing the influence of the, even subtle, differences between classes.

A hierarchical classification model was developed (Fig. 1), which consisted of two consecutive binary classification models. The first binary model was developed with samples from the three continents and discriminates between Asian and non-Asian crude palm oil. Then, the second binary model was developed with those samples classified by the first binary model as non-Asian. This second binary model therefore discriminates between African and South American palm oil only (Fig. 1). This approach was taken to deal with the imbalance in the data set in terms of sample numbers for the three continents. Two different classification models were then developed for the FA profile of crude palm oil and for the VOC fingerprint.

To develop and validate the PLS-DA classification models, the sample set was divided into a training set (consisting of a random selection of 70% of the samples from each continent) and a validation set (the remaining 30% of samples from each continent). The training set of samples was used to develop the models and to internally validate them by leaving 10% out cross-validation. Once the models were developed, they were externally validated by using each one to predict the continent of origin of the samples in the validation set.

The performance of the model was assessed by its accuracy (i.e., correctly predicted samples divided by the number of samples in the class, as a percentage) in predicting each class correctly. Several data pre-processing techniques were applied to the FA and VOC fingerprinting data sets to find the one offering the highest model’s performance, both during the internal and external validation. The pre-processing techniques assayed were: none, autoscaling (scaling to unit variance), mean centreing, logarithmic transformation, orthogonal signal correction (OSC) and combinations of these. In OSC, the information that is certainly not related to the response variable (or class membership) is largely ignored (Westervuis, De Jong, & Smilde, 2001).

3. Results and discussion

3.1. Fatty acid composition

The main FA composition of crude palm oil was in agreement with previous reported values (FAO/WHO, 1999; Tres et al., 2011a). Following a univariate approach, it was revealed that several FAs significantly varied among the continents of origin of crude palm oil (Table 1). Certain FA, such as C12:0, C16:0 or C20:1n-9 significantly varied among the three continents. Some saturated FA such as C12:0, C14:0, C16:0 were higher in crude palm oil from South East Asia than from the other two continents. African crude palm oil had a higher relative content of C18:3n-3, according to previously reported results (Clegg, 1973). Crude palm oil from South America was richer in monounsaturated FA from the n-9 series, such as C18:1n-9 and C20:1n-9. Higher contents of oleic acid and lower contents of palmitic acid in South American palm oil with respect to palm oil from other locations have been reported in literature (Clegg, 1973; Sambanthamurthi, Sundram, & Tan, 2000).

3.1.1. Principal component analysis

Food authentication using a univariate approach is not always a robust strategy. For instance, even if we found significant differences in the content of oleic acid among continents, basing the verification of palm oil provenance on only this FA would not be robust enough and the chances of misclassifications would be very high. More powerful and robust methods might be developed by using the full compositional data (multivariate approach) instead of looking at the individual FA data (univariate approach). The main reason for this is that by using the full composition, more information is considered and, thus, the chances of wrong identifications decrease (Ulberth & Buchgraber, 2000). Thus, multivariate statistics was used as a tool to reveal whether there was a specific FA pattern for oils from each continent.

PCA was conducted on the crude palm oil FA data of the 94 samples. The data matrix consisted of 94 rows (samples) and 21 variables (FA). The PCA scores plot revealed one outlying sample, which originated from South America. At room temperature, this sample was distinctly more liquid that all other crude palm oil samples. This led to the decision to exclude this sample from further model development.

PCA on the FA data revealed the presence of some natural clustering (Fig. 2). The first and second PCA factors together explained 63% of the variance in the data set. This clustering was more or less in agreement with the continent of origin of the crude palm oil. South East Asian samples tended to form a narrower cluster, while African and South American samples showed more variability. This is in agreement with previous studies reporting a narrow compositional range in palm oils from various locations in Malaysia (Clegg, 2011a). Following a univariate approach, it was revealed that several FAs significantly varied among the continents of origin of crude palm oil (Table 1). Certain FA, such as C12:0, C16:0 or C20:1n-9 significantly varied among the three continents. Some saturated FA such as C12:0, C14:0, C16:0 were higher in crude palm oil from South East Asia than from the other two continents. African crude palm oil had a higher relative content of C18:3n-3, according to previously reported results (Clegg, 1973). Crude palm oil from South America was richer in monounsaturated FA from the n-9 series, such as C18:1n-9 and C20:1n-9. Higher contents of oleic acid and lower contents of palmitic acid in South American palm oil with respect to palm oil from other locations have been reported in literature (Clegg, 1973; Sambanthamurthi, Sundram, & Tan, 2000).

3.1.1. Principal component analysis

Food authentication using a univariate approach is not always a robust strategy. For instance, even if we found significant differences in the content of oleic acid among continents, basing the verification of palm oil provenance on only this FA would not be robust enough and the chances of misclassifications would be very high. More powerful and robust methods might be developed by using the full compositional data (multivariate approach) instead of looking at the individual FA data (univariate approach). The main reason for this is that by using the full composition, more information is considered and, thus, the chances of wrong identifications decrease (Ulberth & Buchgraber, 2000). Thus, multivariate statistics was used as a tool to reveal whether there was a specific FA pattern for oils from each continent.

PCA was conducted on the crude palm oil FA data of the 94 samples. The data matrix consisted of 94 rows (samples) and 21 variables (FA). The PCA scores plot revealed one outlying sample, which originated from South America. At room temperature, this sample was distinctly more liquid that all other crude palm oil samples. This led to the decision to exclude this sample from further model development.

PCA on the FA data revealed the presence of some natural clustering (Fig. 2). The first and second PCA factors together explained 63% of the variance in the data set. This clustering was more or less in agreement with the continent of origin of the crude palm oil. South East Asian samples tended to form a narrower cluster, while African and South American samples showed more variability. This is in agreement with previous studies reporting a narrow compositional range in palm oils from various locations in Malaysia (Clegg, 2011a). Following a univariate approach, it was revealed that several FAs significantly varied among the continents of origin of crude palm oil (Table 1). Certain FA, such as C12:0, C16:0 or C20:1n-9 significantly varied among the three continents. Some saturated FA such as C12:0, C14:0, C16:0 were higher in crude palm oil from South East Asia than from the other two continents. African crude palm oil had a higher relative content of C18:3n-3, according to previously reported results (Clegg, 1973). Crude palm oil from South America was richer in monounsaturated FA from the n-9 series, such as C18:1n-9 and C20:1n-9. Higher contents of oleic acid and lower contents of palmitic acid in South American palm oil with respect to palm oil from other locations have been reported in literature (Clegg, 1973; Sambanthamurthi, Sundram, & Tan, 2000).
Table 1
Fatty acid composition of crude palm oils from different geographical origins, expressed as peak area normalisation as a percentage.

<table>
<thead>
<tr>
<th></th>
<th>SEAsia (n = 55)</th>
<th></th>
<th>Africa (n = 28)</th>
<th></th>
<th>SAmerica (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Min&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Max&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Average&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Min&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>C8:0</td>
<td>0.01 y</td>
<td>0.01</td>
<td>0.02</td>
<td>&lt;0.01 x</td>
<td>0.00</td>
</tr>
<tr>
<td>C10:0</td>
<td>0.01 y</td>
<td>0.01</td>
<td>0.02</td>
<td>&lt;0.01 x</td>
<td>0.00</td>
</tr>
<tr>
<td>C12:0</td>
<td>0.15 y</td>
<td>0.08</td>
<td>0.24</td>
<td>0.04 x</td>
<td>0.02</td>
</tr>
<tr>
<td>C14:0</td>
<td>0.95 y</td>
<td>0.80</td>
<td>1.08</td>
<td>0.82 x</td>
<td>0.54</td>
</tr>
<tr>
<td>C15:0</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>C16:0</td>
<td>43.44 z</td>
<td>41.50</td>
<td>45.34</td>
<td>42.23 y</td>
<td>39.12</td>
</tr>
<tr>
<td>C17:0</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>C18:0</td>
<td>4.21 x</td>
<td>3.98</td>
<td>4.74</td>
<td>5.22 y</td>
<td>4.21</td>
</tr>
<tr>
<td>C20:0</td>
<td>0.38 x</td>
<td>0.34</td>
<td>0.40</td>
<td>0.43 y</td>
<td>0.37</td>
</tr>
<tr>
<td>C22:0</td>
<td>0.07 x</td>
<td>0.06</td>
<td>0.08</td>
<td>0.08 y</td>
<td>0.07</td>
</tr>
<tr>
<td>C24:0</td>
<td>0.08 x</td>
<td>0.07</td>
<td>0.09</td>
<td>0.10 y</td>
<td>0.08</td>
</tr>
<tr>
<td>trans 18:1</td>
<td>0.02</td>
<td>0.01</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>C16:1 n-9</td>
<td>&lt;0.03 x</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03 y</td>
<td>0.03</td>
</tr>
<tr>
<td>C18:1 n-9</td>
<td>39.32 x</td>
<td>37.52</td>
<td>40.97</td>
<td>39.94 x</td>
<td>37.60</td>
</tr>
<tr>
<td>C20:1 n-9</td>
<td>0.11 x</td>
<td>0.12</td>
<td>0.15</td>
<td>0.14 y</td>
<td>0.12</td>
</tr>
<tr>
<td>C16:1 n-7</td>
<td>0.14 y</td>
<td>0.12</td>
<td>0.15</td>
<td>0.12 x</td>
<td>0.09</td>
</tr>
<tr>
<td>C17:1 n-7</td>
<td>0.02 y</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02 x</td>
<td>0.02</td>
</tr>
<tr>
<td>C18:1 n-7</td>
<td>0.62 y</td>
<td>0.58</td>
<td>0.65</td>
<td>0.53 x</td>
<td>0.40</td>
</tr>
<tr>
<td>C20:2 n-6</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>C18:5 n-3</td>
<td>0.23 x</td>
<td>0.23</td>
<td>0.29</td>
<td>0.30 z</td>
<td>0.23</td>
</tr>
</tbody>
</table>

<sup>a</sup> Average values shown correspond to the average of all samples within each region (one sample was excluded for African set according to results found in principal component analysis).

<sup>b</sup> Min, lowest value found in each class; Max, highest value found in each class.

<sup>c</sup> Significance value obtained by Kruskal–Wallis test for group comparisons *P < 0.05.

Fig. 2. First two factors of the PCA scores plot on the fatty acid fingerprint (auto-scaled data) of crude palm oils of three different continents.

3.1.2. Partial least squares-discriminant analysis

As observed in the PCA scores plot (Fig. 2), crude palm oil samples tend to cluster according to the continent of origin. Thus, a classification model could be developed by applying a supervised classification technique such as PLS-DA. The data matrix used to develop the PLS-DA model consisted of 64 rows (samples in the training set) and 21 columns (FA data). As described in Section 2, a hierarchical classification model was built by combining sequentially two binary classification models (Fig. 1). The first binary classification model (South East Asian vs non-South East Asian samples), based on the auto-scaled FA profile data, provided 100% accuracy, both during internal and external validation. It only required one PLS-DA factor to discriminate South East Asian crude palm oil. Only two samples not from South East Asia were misclassified as being South East Asian (Table 2).

The second binary classification model (African vs South American palm oil) was developed with only the non-South East Asian samples in the training set. As with the first PLS-DA model, auto-scaling provided the most accurate results (Table 2). External validation also reached a 100% correct classification of the African samples, and only one South American sample (out of 3) was not correctly classified. Although some misclassifications were observed, the output of the model was very satisfactory, especially for the South East Asian and African classes. However, it should be taken into account that the sample set was quite unbalanced, especially with regard to the South American samples. The production of palm oil in South America during 2010 and 2011 was rather low, especially when compared with the production in South East Asia and Africa. Several plantations were growing, but production had not started or had just been started. For this reason, it was quite difficult to obtain authentic samples from different producers in South America. This caused the number of South American samples that were available to optimise and to validate the model to be very low (i.e., only 3 samples for external validation). One advantage of classification models such as PLS-DA is that they might be constantly updated when new samples are obtained. This permits the inclusion of new information and variability in the model. In this case, our model was very promising for verifying the origin of palm oil, but it would be very interesting to update it with new samples in the future, especially from South America. That would keep the model updated and would ensure that natural variability of samples is taken into account in the model to avoid wrong identification of test samples.

By comparing the PLS-DA scores plots (Fig. 3A and C) with the PLS-DA loadings plots (Fig. 3B and D) of each binary model, the FA most contributing to the separation of each class is revealed. Unsaturated fatty acids from the n-7 series, together with myristic (C14:0) and palmitic acid (C16:0), were highly correlated among them and had a high contribution in discriminating crude palm oils from South East Asia (Fig. 3A and B). Other FA such as C18:0, oleic and linolenic had a higher contribution to the non-South East Asian class. This is in agreement with the univariate approach findings of 1973; Rossell et al., 1985; Siew, 2002), while African oils showed considerably larger variation (Siew, 2002).
that revealed significantly higher C14:0, C16:0 and some n-7 FA, and lower C18:1n-9, C20:1n-9 and C18:3n-3 in crude palm from South East Asia than in palm oil from other continents (Table 1). Also, higher contents of C16:0 and lower contents C18:0 have been reported in palm oil from South East Asian countries compared to those from African countries (Clegg, 1973; Rossell et al., 1985).

In the second PLS-DA binary model, some saturated FA such as C18:0, C20:0 and C24:0 showed a high contribution to the separation of the African crude palm oils from the South American palm oils (Fig. 3C and D). Unsaturated FA from the n-9 series, together with some short chain FA, showed a high contribution for the South American class (Fig. 3C and D). By using the univariate approach, these FA were also higher in the South American palm oil than in the African palm oil (Table 1). This is in agreement with previous reports describing highly unsaturated palm oil coming from South America (Clegg, 1973; Sambanthamurthi et al., 2000; Siew, 2002). These differences in the FA composition between continents might be attributable to differences in the botanical identity (i.e., varieties, breeds), climate conditions, soil conditions and agronomical practices (for instance, fruit ripeness at harvesting), among others (Clegg, 1973; Ekpa, Fubara, & Morah, 1994; Monde et al., 2009; Rossell et al., 1985; Siew, 2002). The results from the multivariate approach are in agreement with findings from the univariate approach, but as stated (Ulberth & Buchgraber, 2000), the multivariate approach offered a validated and robust model to verify crude palm oil provenance. To our knowledge, this is the first study dealing with the development of classification models to verify the geographical origin of crude palm oil from three continents.

3.2. VOC fingerprint

3.2.1. Volatile compound fingerprint

The VOC fingerprint of crude palm oils was determined by PTR-MS analysis. To the authors’ knowledge, this is the first time that PTR-MS has been used to assess the VOC profile of crude palm oil for authentication purposes. In PTR-MS, VOC in the sample head-space are detected without prior chromatographic separation and the outcome is a spectrum of the VOCs (their mass-to-charge ratios and their intensities) that might be used as a fingerprint of
the sample. Fig. 4 presents the VOC fingerprints of three crude palm oil samples from South East Asia, Africa and South America.

The most abundant mass-to-charge ratios found in crude palm oil samples were m/z 33, m/z 43, m/z 45, m/z 59, m/z 61, m/z 73, m/z 75 and m/z 83. As stated above, PTR-MS is not intended to identify the volatile compounds as several compounds might yield the same m/z signal. However, the mass/ion intensities might be tentatively assigned to some compounds, according to previous studies on the oil’s volatile fraction as reported in literature (Araghipour et al., 2008). Fragmentation in PTR-MS is mostly reduced due to the chemical ionisation of the compounds; thus, the molecular structure of most volatile compounds is preserved although some fragmentation might still be found. For instance, it is highly probable that m/z 33 corresponds to a fragment of larger compounds (Aprea et al., 2006). It has also been suggested that m/z 59 could correspond to a fragment of hexenol or to propanal, and m/z 75 to propanoates (Buhr et al., 2002). The m/z 43 and m/z 61 might be related to aliphatic alcohols and acetates, respectively, and m/z 73 might correspond to butanal or butanone (among others) (Buhr et al., 2002). The signal observed at m/z 83 might correspond to a fragment of hexenol or to propanal, and m/z 75 and m/z 83 might correspond to hexenal and hexeno or hexenyl acetate (Araghipour et al., 2008). Some of these compounds, such as hexanal, hexenol or hexenyl acetate, are oxidation products that could be formed by the lipooxygenase pathway (Angerosa et al., 2004).

3.2.2. Principal component analysis

As can be seen in the three examples in Fig. 4, the volatile profile of each of the three palm oil samples differed not only in the m/z detected, but also in their intensity. The PTR-MS fingerprint can be investigated by means of chemometric data analysis, which is especially suited to handling large amounts of data. PCA was conducted on the crude palm oil PTR-MS data of the 94 samples to detect outliers and to reveal any natural clustering. The data matrix consisted of 94 samples (rows) and 138 m/z intensities (variables). Among the pre-processing techniques assayed, log10 transformation and auto-scaling led to the clearest tendency of crude palm oil samples to cluster according to their continent of origin (Fig. 5) and to the lowest number of outliers (5 samples). The three first PCA factors explained 61% of the variance.

3.2.3. Partial least-squares discriminant-analysis

As found by PCA, crude palm oil samples tended to cluster according to the continent of origin based on their VOC profiles (Fig. 5). In a similar way to the FA data, a classification model was developed by applying PLS-DA using a two-level hierarchical approach (Fig. 1). The data matrix used to develop the PLS-DA model consisted of 61 rows (samples in the training set, after exclusion of outliers) and 138 columns (m/z intensities).

The first binary model (South East Asia vs non-South East Asia) led to quite successful results both during internal and external validation (Table 3). It provided 97% accuracy for the South East Asian class during internal validation and 100% accuracy for the non-Asian class. In external validation, only one sample was incorrectly predicted as South East Asian, and two South East Asian samples did not match the model. The accuracy obtained by the second binary model was high, especially for the identification of African samples. In the internal validation, South American samples were slightly more difficult to assign to the correct class, similarly to what was found in the FA fingerprinting model. These results show that the geographical origin of crude palm oil samples might be verified by means of their VOC fingerprint (measured by PTR-MS) as well. PTR-MS offers the advantages of being a fast and non-destructive method that could be suitable for in line or on site applications.

Loadings defining both PLS-DA binary classification models showed how the volatile profile led to this classification output (Fig. 6). A few m/z intensities make a high contribution to the first loading for the South East Asian model, especially m/z 80, followed by m/z 97, m/z 98, m/z 107, m/z 108, m/z 117, m/z 120, m/z 131, m/z 132 or m/z 160 (Fig. 6A and B). Loadings defining the African and the South American class are mirror images because they belong to the same binary classification model (Fig. 6C and D). The m/z intensities making a high contribution to the first loading for the African class are for instance m/z 55, m/z 56, m/z 83, m/z 84, m/z 101, m/z 102, m/z 105, m/z 106, m/z 109, m/z 110; while m/z 33, m/z 34, m/z 51, m/z 52, m/z 80, m/z 81, m/z 96, m/z 97, m/z 98 and m/z 120 made a high contribution to the first loading for the South American class.
As noted above, PTR-MS is not intended to identify the VOC compounds of its fingerprint, mainly because it is a low resolution one-dimensional technique (Heenan et al., 2012). However, based on previous literature, some of these intensities might be tentatively assigned to some compounds. For instance, m/z 80, which played a quite important role in the separation of the South East Asian samples from the non-South East Asian samples, and in the separation of the South American from the African samples, might correspond to a fragment of the polyene chain of some carotenoids and their degradation compounds (Caris-Veyrat et al., 2001; Ouyang et al., 1980; Van Breemen et al., 1995). Similarly, other masses such as m/z 96, 131, 132, 107, 120 might also correspond to fragments of carotenoids or their oxidation products (Caris-Veyrat et al., 2001; Ouyang et al., 1980; Van Breemen et al., 1995). Since the main carotenoids in palm oil are β-carotene and α-carotene (Sambanthamurthi et al., 2000; Tres et al., 2011a) which are not very volatile, they would hardly be measured by PTR-MS. However, it is quite probable that they might have degraded during palm oil extraction and storage, producing more volatile degradation compounds, and those degradation compounds could be the ones yielding these fragments. Thus, this indicates that the variation within the carotenoid profile of crude palm oils (Clegg, 1973; Monde et al., 2009), together with the agronomic and technologic practices in each region, would contribute to a different VOC profile between the continents of origin of palm oil. Indeed, Clegg (1973) showed that the carotenoid

Table 3
Validation results of the 2-step classification tree (SE Asia vs non-SE Asia, and Africa vs South America) to identify the continent of origin of crude palm oil samples based on their volatile profile.a

<table>
<thead>
<tr>
<th></th>
<th>Number of samples predicted as</th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>South East Asia</td>
<td>Non South East Asia</td>
</tr>
<tr>
<td>First binary model (with OSC data pre-treatment)</td>
<td>Internal validation</td>
<td>South East Asia</td>
</tr>
<tr>
<td>Origin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South East Asia</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>Non South East Asia</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Second binary model (without OSC data pre-treatment)</td>
<td>Internal validation</td>
<td>South America</td>
</tr>
<tr>
<td>Africa</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>South America</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>External validation</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>

a Volatile profile data were log 10 transformed and auto-scaled.

As noted above, PTR-MS is not intended to identify the VOC compounds of its fingerprint, mainly because it is a low resolution one-dimensional technique (Heenan et al., 2012). However, based on previous literature, some of these m/z intensities might be tentatively assigned to some compounds. For instance, m/z 80, which played a quite important role in the separation of the South East Asian samples from the non-South East Asian samples, and in the separation of the South American from the African samples, might correspond to a fragment of the polyene chain of some carotenoids and their degradation compounds (Caris-Veyrat et al., 2001; Ouyang et al., 1980; Van Breemen et al., 1995). Since the main carotenoids in palm oil are β-carotene and α-carotene (Sambanthamurthi et al., 2000; Tres et al., 2011a) which are not very volatile, they would hardly be measured by PTR-MS. However, it is quite probable that they might have degraded during palm oil extraction and storage, producing more volatile degradation compounds, and those degradation compounds could be the ones yielding these fragments. Thus, this indicates that the variation within the carotenoid profile of crude palm oils (Clegg, 1973; Monde et al., 2009), together with the agronomic and technologic practices in each region, would contribute to a different VOC profile between the continents of origin of palm oil. Indeed, Clegg (1973) showed that the carotenoid

Fig. 6. (A) First PLS-DA loading for the South East Asian class (South East Asian vs non-South East Asian model), (B) First PLS-DA loading for the African class (African vs South American model) and (C) First PLS-DA loading for the South American class (African vs South American model).
content of palm oil varied between countries. This author attributed a large part of this variation to variations in the ripeness standards for harvesting in each location. But according to Clegg (1973), genetic variation, climate and other geographical factors could also be related to this variation.

The $m/z$ 97 was also quite high in the first loading of the binary model South East Asian vs non-South East Asian samples (Fig. 6A and B). It could be attributed to furfural (Loi, Boo, Mohamed, & Ariffin, 2010) or to a fragment of heptanal (Buhr et al., 2002), among others. Furfural is not naturally present in palm oil, but, according to Loi et al. (2010), it might originate from the sterilisation process in the mill. On the other hand, heptanal is one of the oxidation compounds formed from FA of the n-7 family. As described above (Fig. 3B), the FA of the n-7 series had a high contribution for the separation of the South East Asian class vs the non-South East Asian class in the model based on the FA fingerprinting. Thus different technological practices during palm oil extraction might explain the contribution of $m/z$ 97 in the discrimination of crude palm oil origin. As mentioned, however, further studies would be needed to confirm the identity of the compounds contributing to these $m/z$ intensities.

4. Conclusion

The verification of the geographical origin of crude palm oil by means of both its FA and its VOC fingerprinting combined with chemometrics seems feasible, at least on a continental scale. Our chemometric approach was based on a hierarchical classification approach based on two binary PLS-DA classification models consecutively applied (1st model: South East Asian vs non-South East Asian crude palm oil; 2nd model: African vs South American crude palm oil). Both FA and VOC models were successful in verifying the origin of palm oil at a continental scale; however, more outliers were found in the VOC data set (i.e., prior to the development of the classification model). Since PTR-MS is a fast, sensitive and non-destructive technique that could be suitable for in site applications, it could be proposed as a first screening technique, while FA fingerprinting could be used for further confirmation.

Development of analytical methods for geographical provenance of palm oil, in addition to the effect of administrative controls currently applied, will have positive implications for fair trading of sustainable palm oil. With this study we have taken a first, and to our knowledge unique, step in this direction. Yet, the sample set used for modelling should be extended, especially for the South American class, and this aspect is currently being worked on. Inclusion of new samples in the future would keep the models updated according to the temporal variability of the oil. Furthermore, a higher number of samples would also improve geographical resolution by down-scaling the models to smaller geographical areas (e.g., to country or even province level).

Acknowledgments

The authors are very grateful to the palm oil suppliers: Agropalma (Brazil), Eugenia Vargas from Lanagro-MG/CGAL/SDA/MAPA, Brasil, P. Korst (Unimills B.V.), N. van Oorschot (Loders Croklaan Europe), R. Bennink (Saybolt) and M. Bruggeman (Nofalab). The authors would also like to acknowledge E. Cuijpers, J. van der Roest and E. Pertwii for their help in sample collection and A. Koot for his help in PTR-MS analysis.

This work was partially funded by the by the European Union 7th Framework Programme (FP7/2007–2013) under a grant agreement PIEF-GA-2009-251972 (Marie Curie IEF) awarded to A. Tres. The authors are also grateful to the Fundación Alfonso Martín Escudero (Madrid, Spain) for the research grant awarded to A. Tres, and to the Andalusia Regional Government (Consejería de Innovación, Ciencia y Empresa) (Spain) for the personal postgraduate grant awarded to C. Ruiz-Sambulas. Funding was also received from the Dutch Ministry of Economic affairs, Agriculture and Innovation.

References


