## APPLICATION OF FTIR SPECTROSCOPY FOR ASSESSMENT OF GREEN COFFEE BEANS ACCORDING TO THEIR ORIGIN

**S. M. Obeidat [1*], A. Y. Hammoudeh [1], A. A. Alomary [2]**

*[1] Yarmouk University, Irbid, Jordan; e-mail: Safwan@yu.edu.jo, Safobeidat@yahoo.com*
*[2] The Islamic University of Al-Madinah Al-Munawarah, Al-Madinah Al-Munawarah, Saudi Arabia*

*Samples of green coffee beans originating from five different countries were ground and analyzed using FTIR spectra in the region of 600–4000 $cm^{-1}$. Successful discrimination of each coffee type based on their origin was achieved applying a PCA algorithm on the obtained IR spectra for all samples. PCA loading plots show that the IR bands at 2850, 2920, and 1745 $cm^{-1}$ corresponding to the symmetric, and antisymmetric vibrations of $CH_2$ and the stretching vibration of C=O bond in ester, respectively, are the most significant peaks in distinguishing the origin of the above coffee samples.*

***Keywords:*** *green coffee beans, adulteration, FTIR, principal component analysis.*

## ПРИМЕНЕНИЕ ИК-ФУРЬЕ-СПЕКТРОСКОПИИ ДЛЯ ОПРЕДЕЛЕНИЯ ПРОИСХОЖДЕНИЯ СЫРЫХ ЗЕРЕН КОФЕ

**S. M. Obeidat [1*], A. Y. Hammoudeh [1], A. A. Alomary [2]**

*УДК 539.19:663.93/.94*

*[1] Университет Ярмук, Ирбид, Иордания; e-mail: Safwan@yu.edu.jo, Safobeidat@yahoo.com*
*[2] Исламский университет Медины, Медина, Саудовская Аравия*

*(Поступила 19 декабря 2016)*

*С помощью ИК-фурье-спектров в области 600–4000 $см^{-1}$ проведен анализ образцов молотых сырых кофейных зерен, происходящих из пяти различных стран. За счет обработки ИК спектров всех образцов по алгоритму метода главных компонент (PCA) удалось установить отличие каждого типа кофе в зависимости от места происхождения. Загрузочные данные метода главных компонент показывают, что ИК полосы при 2850, 2920 и 1745 $см^{-1}$, относящиеся к симметричным, антисимметричным колебаниям $CH_2$ и валентным колебаниям связи C=O в эфире соответственно, являются наиболее характерными для установления происхождения выбранных образцов кофе.*

***Ключевые слова:*** *сырые (необжаренные) зерна кофе, фальсификация, ИК-фурье-преобразование, метод главных компонент.*

**Introduction.** Coffee is certainly one of the most popular beverages all over the world. For many people coffee drinking is an essential part of their lifestyle and daily habits. Consumption of coffee might be referred to its excellent taste, its stimulatory effects due to the presence of caffeine, and to its health benefits (i.e., antioxidation effect) [1, 2]. In general, coffee can be divided based on quality into two main types: arabica and robusta. Coffee arabica is harvested from the plant of *Coffea arabica*, whereas coffee robusta comes from the *Coffea canephora* plant [3, 4]. The composition of all coffee types is very similar with small differences such as in the caffeine content [5]. Coffee arabica is considered the best in quality and is the most expensive coffee type. The big difference in price between coffee arabica and coffee robusta makes the adulteration of coffee financially alluring, especially since distinguishing between the above types of coffee is a difficult task due to the high similarity in shape and aroma. This makes uncovering coffee adulteration a very tedious procedure. In some countries such as Jordan, the arabica and robusta assessment of coffee is not popular. Alternatively, the classification of coffee according to its origin is much more common. Hence, one

may judge the coffee quality based on the country of origin, which can therefore determine not only the quality but also the price of the coffee. So consumers search usually for coffee of high quality from beans of known origins. Once more, the potential of adulteration is very high, but this time the country of origin of coffee is targeted. This may take place through dishonest sellers who blend coffees from different origins and misinform the consumer about the true origin of the coffee or through the mislabeling of the coffee bags [6, 7]. The authenticity of coffee with respect to its geographical origin (country) is quite important for both producers and consumers. Several studies can be found in the literature addressing the above issue. GC-MS and principal component analysis (PCA) have been applied to distinguish coffee from six different origins [8]. Compared to MS and chromatographic techniques, spectroscopic techniques are in general faster and require little or no sample preparation. Hence, many spectroscopic approaches have been employed to study and classify coffee. $^{13}$C NMR-based metabolomics has been used for the assessment of green coffee beans according to variety and origin [9]. NIR spectroscopy has been used for coffee assessment by quality and origin through preparing calibration curves for the nine main chemical compounds found in coffee such as lipids, proteins, sucrose, caffeine, etc. [10]. NIR spectroscopy has also been used along with partial least squares regression (PLSR) and PCA for studying defective beans among nondefective ones [11]. Classification of coffee mainly by quality has also been achieved using FTIR analysis of the dry extract of coffee originating from various countries [10]. In this work, FTIR spectroscopy and PCA are used for the assessment of green coffee beans according to their origin without any prior chemical treatment of the coffee samples. The entire IR spectrum for all studied samples will be used to maximize the sensitivity.

**Materials and methods.** This study was performed on 48 green coffee bean samples from five different countries: Brazil (B), Colombia (C), Ethiopia (E), Kenya (K), and Yemen (Y). These origins represent the sources of the green coffee beans available in the Jordanian market. An additional five test samples (one from each country) were also collected for the purpose of testing the resulting PCA calibration model. All samples were provided by the exclusive importers in Jordan. Eight to eleven samples of green coffee beans were obtained from each origin (country).

FTIR spectra were collected in the range of 600–4000 cm$^{-1}$ using a Bruker ALPHA spectrometer with an ATR module equipped with a ZnSe crystal.

All samples of liquid-nitrogen frozen green coffee beans were finely ground using a mortar and pestle before the FTIR analysis. The IR spectrum was recorded for each sample in the range of 600–4000 cm$^{-1}$ and saved as an Excel file consisting of 1650 data points. The spectra of the 48 samples were concatenated in a single file of 48×1650 dimensions. The PCA algorithm was then applied to the entire data set. Before analysis, the spectral region of 2750–1775 cm$^{-1}$ was removed from the spectra because it contained no IR absorption except that of $CO_2$ in the atmosphere. The spectra were normalized before applying PCA using a unity-based normalization according to the equation

$$X_{i=0-1} = (X_i - X_{\min})/(X_{\max} - X_{\min}), \quad (1)$$

where $X_{i=0-1}$ is data point $i$ normalized between 0 and 1, $X_i$ is the data point $i$, and $X_{\min}$ and $X_{\max}$ are the minimum and maximum values among all the data points.

Data processing and analysis were executed using MATLAB 7.0.4 and PLS_Toolbox 3.5. The leave-one-out cross validation method was used to validate the resulting PCA model. To test the resulting PCA model, five test samples were treated exactly like the studied samples. Data from these five samples were concatenated in a separate excel file and then applied to the pre-constructed PCA calibration model.

PCA is a pattern recognition algorithm that is being employed to find similarities and differences among the objects (samples) in a particular data matrix. It decomposes the data matrix $X$ as the sum of the outer product of vectors $t_i$ and $p_i$ plus a residual matrix $E$ as

$$X = t_1 p_1^{\mathrm{T}} + t_2 p_2^{\mathrm{T}} + \ldots + t_k p_k^{\mathrm{T}} + E. \quad (2)$$

where the $t_i$ vectors are called scores and contain information about the samples, while the $p_i$ vectors are called loadings and contain information about the variables.

PCA can minimize the data size without losing significant information. Mathematically, PCA represents the eigenvectors for the covariance or correlation matrix of a data matrix. The eigenvector associated with the greatest eigenvalue is known as the first principal component (PC1). The second principal component (PC2) is the eigenvector that is associated with the next greatest eigenvalue and so on. All resulting PCs are mutually orthogonal. PC1 accounts for the maximum variation in the data, while the rest of succeeding components account for as much of the remaining variability. Hence, usually only the first few PCs are used in the analysis because they contain the maximum variation in the data set [12, 13].

**Results and discussion.** *Visual analysis of the IR spectra.* Figure 1a displays five IR absorption spectra; each spectrum corresponds to a particular coffee sample. A careful inspection for these spectra shows two important regions that can be useful for the visual comparison. Those regions fall between 1775–1500 $cm^{-1}$ (bottom left) and 3030–2750 $cm^{-1}$ (bottom right). These spectra reveal some common and important features such as the bands at 2922 and 2852 $cm^{-1}$ that correspond to the antisymmetric and symmetric vibrations of $CH_2$, respectively. However, in sample C, the band corresponding to the antisymmetric vibration of $CH_2$ is red-shifted by about 5.0 $cm^{-1}$ (centered at 2917 $cm^{-1}$) and it is broader than in the other samples. Another band appears at 3009 $cm^{-1}$ corresponding to the (H–C=) stretching vibration, which can be related to unsaturation in fatty acids, which is obviously greatest in the Brazilian coffee (B). Other features in the spectra appear as shoulders at 2952 and 2870 $cm^{-1}$ corresponding to the antisymmetric and symmetric vibrations of $CH_3$, respectively.

The spectra in Fig. 1 contain also important information about the general composition of the studied coffee samples. The absorption band at ~1744 $cm^{-1}$ belongs to the stretching vibration of the C=O bond in esters and can be attributed to coffee lipids. There are three types of lipids to consider: fatty acid esters of glycerine, sterols, and diterpene alcohols. The broadness of the band at 1740 $cm^{-1}$ is in the order of C > Y > B ≈ K ≈ E. The broadening is attributed to contributions of the C=O stretch at slightly smaller wavenumbers, most obvious in the case of Y as a shoulder at 1735 $cm^{-1}$. In the case of C, this component seems to have a larger contribution so that it merges with that at 1744 $cm^{-1}$, leading to a broad signal centered at 1740 $cm^{-1}$. The lipid composition in C and Y seems to be different than in B, E, and K. Since many of the aroma producing substances are fat soluble, it is believed that this difference in the lipid composition affects the aroma quality of the coffee.

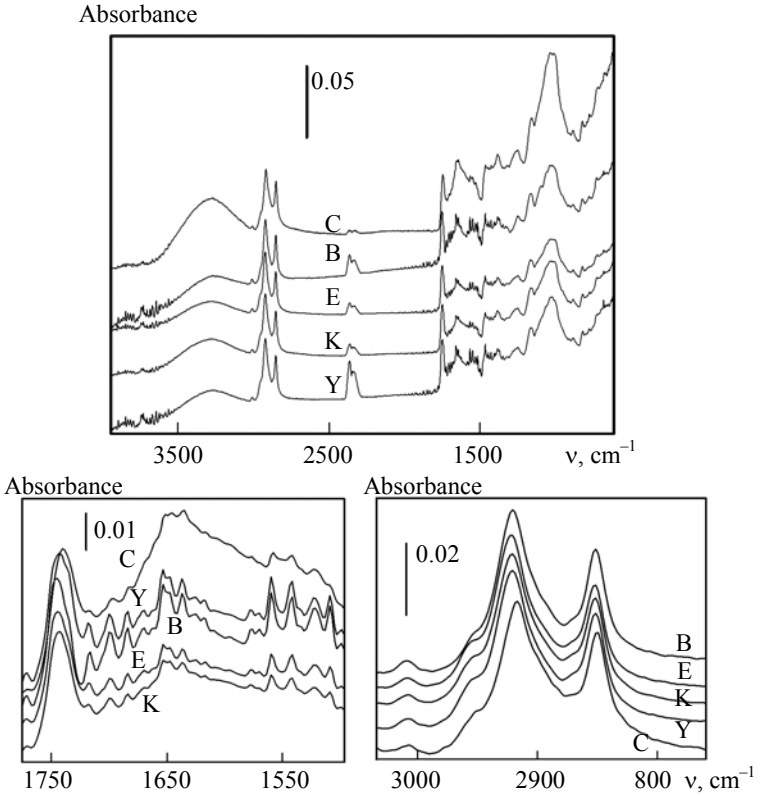


Fig. 1. Full FTIR spectra (a) and zoom in for particular regions from the spectra (b, c) for five green coffee beans from five different countries: Brazil (B), Colombia (C), Ethiopia (E), Kenya (K), and Yemen (Y).

The peptide linkage in proteins contains also a C=O bond; it appears, however, centered around 1650 $cm^{-1}$, the so-called Amide I band. Other structures contributing to the broad absorption in the range of 1650–1625 $cm^{-1}$ include the HO-bending of adsorbed water, free fatty acids, and chlorogenic acids. Caffeine absorbs at 1707 $cm^{-1}$ or below.

In summary, the IR spectra of the studied coffee samples look in principle very similar to each other. Even the fine differences in the spectra are very subjective. Therefore, the discrimination among the above coffee types based on visual analysis is extremely difficult. Visual comparison between the samples becomes even more complicated upon comparing a greater number of the spectra. Therefore, a more advanced and reliable method is needed to look at the similarities or differences in the spectra.

*PCA analysis*. For a more efficient and reliable method of distinguishing the similarities and differences among the obtained IR spectra of coffee, PCA was applied to the data matrix of 48×1650 dimensions containing the spectral data of 48 coffee samples. The first few PCs were calculated. Table 1 shows the variance captured in the data set by the first five PCs. The best PCA score model was obtained using the second and the third PCs. This model is displayed in Fig. 2 (without the numbered lozenges) showing the resulting PCA score model that accounts for only less than 5% of the total variation in the data set. This small value of the captured variance of the data reflects the high similarity among the spectra expressed in the first PC as it appears in Table 1. As can be seen in Fig. 2, five well-resolved clusters can be recognized. Each point in this plot represents the IR spectral data of a single coffee sample. In the PCA interpretation, usually every independent cluster contains samples with maximum spectral similarities reflecting the composition similarity. Each cluster in Fig. 2 contains samples from the same country of origin. Thus, successful distinguishing among coffees from different countries was achieved. In this figure, it is also obvious that the sample variation is different from one cluster to another. These variations were sometimes high as in case of the Colombian coffee, and some other times small as in the case of Brazilian coffee. This might be referred to several reasons such as the exact location of the coffee plantation in that specific country and related weather conditions, in addition to the harvesting conditions that might differ slightly from one crop to another. These changes can produce such tiny changes in the coffee composition or aroma. Other reasons could lie in the transportation and storage conditions, which can differ from one shipment of coffee to another [6, 14]. This PCA score model can be used in further studies to identify the country of origin of coffee obtained from the above-mentioned countries. To check out the ability of the presented technique to assess coffee based on its origin, five new coffee beans samples from the studied regions were obtained. Samples were presented by lozenges and numbered from 1 to 5 corresponding to Brazil, Kenya, Ethiopia, Yemen, and Colombia, respectively. These samples were treated and measured by FTIR in the exact manner as described for the control samples. The spectral data for the test samples were assembled in a separate data matrix of the dimensions 5×1650. This data set was then applied to the preconstructed PCA score model (Fig. 2). The result of this test can also be seen in the same figure. As it can be seen in this figure, each test sample represented by a lozenge is found to lie within or very close to a particular cluster, indicating that successful identification of the samples of green coffee beans could be achieved using the current method.

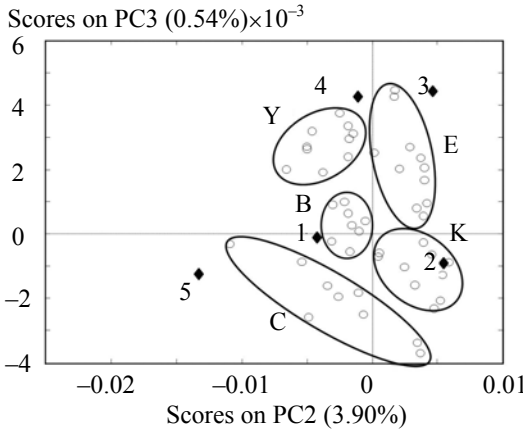


Fig. 2. The result of the application of five test samples (lozenges) of coffee to the PCA scores plot for the five green coffee beans: 1 (Brazil), 2 (Kenya), 3 (Ethiopia), 4 (Yemen), and 5 (Colombia).

For further investigation on the most important variables (spectral regions or wavenumbers) that had the maximum contribution to this assessment, the PCA loadings were displayed. Figure 3 presents the loadings on the second and the third PCs, respectively. In Fig. 3a, it can be seen that the main spectral regions that contribute to the separation in PC2 are around 2920 and 2850 $cm^{-1}$. Again these two peaks are mainly due to the antisymmetric and symmetric vibrations of $CH_2$, respectively. On the other hand, the main spectral

TABLE 1. Percent Variance Captured for the First Five Principal Components of the FTIR spectral Data Matrix of All Coffee Samples

| Principal Component Number | Variance Captured, % |
|---|---|
| 1 | 95.40 |
| 2 | 3.89 |
| 3 | 0.542 |
| 4 | 0.0311 |
| 5 | 0.0223 |

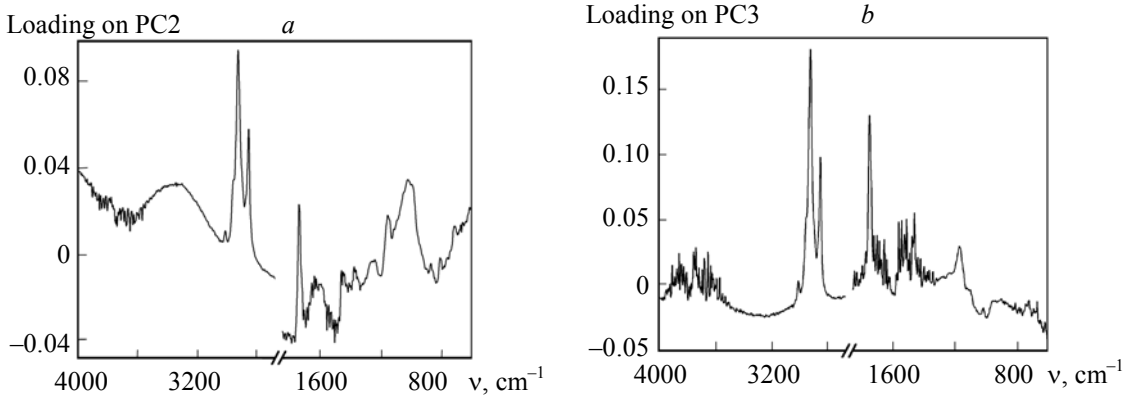


Fig. 3. PCA loadings plot on PC2 (a) and PC3 (b).

region responsible for the separation in PC3 is around 1745 $cm^{-1}$ (Fig. 3b) which is mainly associated with the stretch vibration of the C=O bond in esters. However, since coffee is a complex mixture of organic molecules, it is difficult to identify the molecules responsible for the noticed separation among the coffee types [11].

**Conclusion.** Samples of green coffee beans from five different countries (Brazil, Kenya, Ethiopia, Yemen, and Colombia) were studied. Assessment of coffee based on the geographical origin has been successfully achieved using a combination of FTIR and PCA. The investigation of loadings of the used PCs showed that the symmetric and antisymmetric vibrations of $CH_2$ as well as the stretch vibration of the C=O bond in esters played the major role in distinguishing coffee according to the country of origin. Finally, this paper proves that FTIR spectroscopy combined with PCA for data analysis has a great potential to identify and distinguish green coffee beans from different origins.

## REFERENCES

1. M. Grembecka, E. Malinowska, P. Szefer, *Sci. Total Environ.*, **383**, 59–69 (2007).
2. M. S. Butt, M. T. Sultan, *Crit. Rev. Food Sci. Nutr.*, **51**, 363–373 (2011).
3. P. Pohl, E. Stelmach, M. Welna, A. S. Madeja, *Food Anal. Methods*, **6**, 598–613 (2013).
4. V. R. M. Filho, W. L. Polito, J. A. G. Neto, *J. Braz. Chem. Soc.*, **18**, 47–53 (2007).
5. V. Krivan, P. Barth, A. F. Morales, *Microchim. Acta*, **110**, 217–236 (1993).
6. K. A. Anderson, B. W. Smith, *J. Agric. Food Chem.*, **50**, 2068–2075 (2002).
7. M. J. Martin, F. Pablos, A. G. Gonzalez, *Food Chem.*, **66**, 365–370 (1999).
8. I. Dirinck , I. Van Leuven, P. Dirinck, *Czech. J. Food Sci.*, **18**, 50–51 (2000).
9. F. Wei, K. Furihata, F. Hu, T. Miyakawa, T. Tanokuta, *J. Agric. Food Chem.*, **59**, 9065–9073 (2011).
10. N. Dupuy, J. P. Huvenne, L. Duponche, P. Legrand, *Appl. Spectrosc.*, **49**, 580–585 (1995).
11. J. R. Santos, M. C. Sarraguça, A. O. S. S. Rangel, J. A. Lopes, *Food Chem.*, **135**, 1828–1835 (2012).
12. B. Wise, N. Gallagher, S. Butler, D. White, G. Barna, *J. Chemom.*, **13**, 379–385 (1999).
13. I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York (2002).
14. S. I. Mussatto, E. M. S. Machado, S. Martins, J. A. Teixeira, *Food Bioprocess Technol.*, **4**, 661–672 (2011).