**Brigitta Jesztl[1], Eszter Luca Benes[1], Marietta Fodor[1]**

Received: February 2018 – Accepted: January 2019

# FT-NIR origin identification of coffee samples

**Keywords:** coffe shrub, *Coffea arabica*, ground coffee, origin identification, spectroscopy, FT-NIR, chemometry, rapid analytics, growing area, roasting method, whole coffee grist, coffee drink.

## SUMMARY

**Two basic variants of the coffee plant are known: arabica (Coffea arabica) and robusta (Coffea canephora). Arabica, with a higher sensitivity to its growing environment, tastes better, while the caffeine content of robusta is about 1.5 times higher than that of arabica. Robusta is less sensitive to its growing conditions, so it can be produced at a lower cost, yet arabica varieties account for two thirds of the world's coffee production.**

**Commercially available cheaper coffees are primarily made from low quality robusta, the cheaper they are, the lower the quality of the coffee is. In the case of the ground versions, we cannot even be sure what else there is in the blend besides coffee.**

**For our measurements, Coffea arabica raw coffee was used as the starting material. Samples from different growing areas, roasted using different methods, and ground under identical conditions were analyzed by Fourier transform near infrared spectroscopy (FT-NIR), and spectral results were evaluated using chemometric methods.**

**Our results demonstrate that a rapid analytical method, requiring no sample preparation and not polluting the environment with chemicals, was successfully used to identify ground coffees by roasting method and growing area.**

## Literature review

### Coffee growing and processing

Ideal environmental conditions for the coffee plant are found in the coffee growing belt between the Tropic of Cancer and the Tropic of Capricorn. Cultivation techniques are adapted to the environmental conditions: shielded, partially shielded or without shielding. The coffee plant requires heavy, clay soil from volcanic ash rich in nitrogen and phosphorus. The seedling developing from the green coffee bean starts producing after 3 or 4 years. Maintenance of plantations that are up to 50 or 60 years old also requires fertilization, mulching, pruning and spraying.

The coffee plant blooms in bunches, resembling jasmine in appearance and smell. Blooming lasts only a few days and 9 to 10 months after pollination the flowers develop into yellow and red cherries that can be picked. There may be mature and immature fruit on the same branch, therefore, beans of the same, adequate maturity are harvested manually. Depending on the growing area and the variety, one or two major harvests, and possibly some minor ones may take place. Typically, the harvest period lasts 4 to 5 months **[1]**.

By removing the flesh of the coffee cherry, raw green coffee beans are obtained. During the processing of the coffee, depending on the growing area and the producer, wet, dry and semi-dry processes are distinguished. The final step in the processing of coffee beans is dry grinding in a mill, during which the last outer layers are removed from the beans, followed by removal of faulty beans, packaging and shipping **[2]**.

Flavoring substances characteristic of coffee drinks obtain their final form during roasting. A green coffee bean is a tasteless seed with a straw-like smell that turns into a product rich in flavors during roasting,

[1] Szent István University, Faculty of Food Science, Department of Applied Chemistry

which can be characterized by thousands of volatile components, depending on the degree of roasting. The degree of roasting is determined by the color of the roasted coffee. During the roasting process, the color of the bean becomes deeper as the temperature rises. ***Lightly roasted coffee*** (***Viennese***) has a light brown color, with no oil on its surface, dominated by a characteristic acid flavor, ***medium roasted coffee*** (***French***) is darker in color, with no oil on the surface of the beans either and a more balanced ration of sour and bitter taste, while the surface of ***heavily roasted coffee*** (***Italian***) is oily, it can be characterized by the dominance of bitter taste.

The most important changes during roasting are loss of weight and moisture, volume growth, carbon dioxide formation, detachment of the parchment membrane surrounding the bean, formation of fragrances and aromas, as well as color change. When carbohydrates are burned, sugars are broken down and united into new molecules by chain reactions and condensations. Maillard reactions take place between sugars and amino groups, resulting in a brown color.

There are almost two thousand ingredients present in the coffee beans, they are transformed and degraded into alcohols, aroma compounds, aldehydes, esters, ketones and cyclic nitrogen compounds, contributing to the development of the characteristic flavor. During roasting, at high temperatures, fatty acids are converted into essential oils that are visible on the surface of the coffee beans. In raw coffee, tannic acid is bound to caffeine, and most of it becomes free during roasting, thus its amount increases. In addition, the amount of trigonelline and chlorogenic acid decreases, because they are decomposed under the influence of heat. The protein content of coffee is not affected by roasting **[2, 3]**.

### NIR – near infrared spectroscopy

The 780 to $10^6$ nm range of electromagnetic radiation is referred to as the infrared range, within which near (NIR, 800 – 2,500 nm), mid (MIR, 2,500 – $2.5 \cdot 10^4$ nm) and far (FIR, $2.5 \cdot 10^4$ – $10^6$ nm) infrared ranges are distinguished. Because of their complex matrices, food samples are primarily analyzed in the NIR range.

There are several interaction that might occur between infrared photons and the sample. From an analytical point of view, transmission, diffuse reflection and the combination of the two, transflection are the most widely used measurement methods. In the course of coffee-related research, due to its analytical advantages, the NIR technique is most often used.

The NIR method was successfully used by Huck et al. to determine the three major alkaloid compounds of robusta and arabica roasted coffees, caffeine, theobromine and theophylline. Reference data were obtained by an HPLC-MS coupled technique **[4]**.

Esteban-Díez et al. have successfully applied the NIR technique to distinguish between arabica and robusta coffee varieties and the compositions of blends **[5]**.

The detection of coffee counterfeiting is a high priority quality management objective in the food industry, because of the high price differences between growing areas and varieties. Pizzaro et al. used the NIR method combined with multivariate regression procedures to identify robusta coffees and to investigate counterfeiting **[6]**.

Evaluation of NIR spectra subjected to different data treatment methods (orthogonal signal correction, OSC, direct orthogonal signal correction, DOSC) using partial least squares (PLS) regression made the determination of the ash and lipid content of given roasted coffee samples possible **[7]**.

To distinguish between coffee growing areas, multivariate chemometric methods combining gas chromatography, electronic language and a trained sensory panel were successfully used by Várvölgyi et al. **[8]**.

The objective of our research is to distinguish between *Coffea arabica* ground coffees and coffee drinks from different growing areas by growing area and roasting method using the NIR technique.

## Materials and methods

### Test samples

In the research, arabica (*Coffea arabica*) green coffee samples from different growing areas (Brazil, Guatemala, India and Colombia) were used. Brazilian coffee is characterized by its soft taste and low acid content, while the coffee grown in Guatemala is highly acidic with a full, spicy taste. The coffee grown in India is characterized by a sweet, soft taste and a low acid content. At the higher altitude plantations of Colombia, soft, full-bodied coffee with a low acid content is grown, especially of the arabica variety.

### Sample preparation – the roasting of coffee

Green coffee was roasted using the green coffee bean roasting equipment marketed by the company Gene Café (Genecafe, South Korea). The different roasting parameters are summarized in ***Table 1***.

### Measurement method - FT-NIR

For each sample, spectra were recorded using a Bruker MPA™ FT-NIR/NIT instrument (Bruker Ettlingen, Germany). The instrument operates in the 12,500 – 4,000 $cm^{-1}$ wave number range. Spectra were recorded using the proprietary OPUS 7.2. (Bruker, Ettlingen, Germany) program of the instrument. Three parallel spectra of each sample were taken. Spectra were recorded in the diffuse reflection measurement

mode in the case of solid samples, and with the help of a temperature-regulated flow-through cuvette (*l*=1 mm) in the case of liquid samples.

### Statistical methods

To distinguish between the samples, unsupervised and supervised statistical methods were used. An unsupervised method is principal component analysis (PCA), with which it was examined whether our samples could be considered a set of samples taking into account a 95% confidence interval. Thus, the groups later found during the performance of supervised linear discriminant analysis (LDA) could really be considered the result of a successful pattern recognition and were not due to the different sample matrices. The purpose of LDA is to create discriminant functions that, as a linear combination of independent variables, best distinguish between the categories of dependent variables. As a first step, to reduce the number of variables, principal component analysis was performed, reducing the original number of variables of more than one thousand. For the examination of both coffee bean samples and ground coffee samples, as well as coffee drink samples, 20 factors were developed, with the help of which pattern recognition was carried out. PCA and LDA analyses were performed using the Statistica 8.0 (StatSoft, Tulsa, Oklahoma, USA) software.

## Results and evaluation

### Evaluation of FT-NIR spectra

The FT-NIR spectra of both the coffee bean samples and the ground coffee samples, as well as the brewed versions of the coffee samples from the four growing areas (Colombia, Guatemala, India and Brazil) were recorded. ***Figure 1*** shows the spectra of the coffee bean samples from the different areas as a function of the roasting method.

Characteristic peaks are marked with numbers in ***Figure 1***, the qualitative evaluation of which is summarized in ***Table 2***. For the identification of the peaks, literature data were used **[9]**.

Comparing the spectra, it can be seen that there is a significant baseline shift. This is primarily due to the different particle size, but it is worth noting that the order of the curves is the same for all four samples, with the spectrum of the Italian roast being the highest. This phenomenon is not only related to the particle size, but also to the fact that, in the case of this roasting method, there is a significant oil layer on the surface of the particles, which results in light scattering.

When comparing the roasting methods ***(Figure 2)***, significant differences can be seen in the characteristics of the 8,500 – 8,200 $cm^{-1}$ and 7,200 – 7,100 $cm^{-1}$ ranges, which can be explained by the different presence of aliphatic hydrocarbons. Viennese roasting is the most gentle process, the least damaging to organic components, and it is supposed that this is the cause of the characteristic difference. The peaks in the 4,700 – 4,000 $cm^{-1}$ range are related to the protein, fat/oil, carbohydrate, fiber and cellulose content of the sample. It is exactly this area where there is a clear difference between the roasting methods. In Viennese roasted samples, primarily a difference in height could be observed, which is not only related to particle size but also to the concentration. Quantitative conclusions cannot be drawn from the spectral image, but it can be concluded that the spectra of the Colombian and Guatemalan samples coincide, and the same is true for the Indian and Brazilian samples, so it is assumed that these samples are characterized by similar protein and fat content.

In the case of French roasted coffees, the peak belonging to the 8,300-8,100 $cm^{-1}$ wave number flattens out, so it can be concluded that the aliphatic hydrocarbons present are less detectable. Samples are subjected to the most intense heat during Italian roasting. There is no change in temperature, but there is a change in roasting time. In this case, the signal of aliphatic hydrocarbons almost blends into the spectrum. As an effect of heavier roasting methods, significant qualitative and quantitative changes take place in the protein, fat/oil and carbohydrate content. This characteristic change can also be seen in the spectrum in the 4,700-4,000 $cm^{-1}$ wave number range.

FT-NIR spectra of coffee bean samples and ground coffee samples were also compared ***(Figure 3)***. The characteristics of the spectra are similar, as expected. The spectra of the ground coffee samples are typically lower, which is clearly related to the particle size.

During the FT-NIR analysis of coffee drinks, for beverages brewed from samples of different origin and of different roasting methods using the same technological procedure, samples of characteristic spectra in the transmission measurement mode are shown by roasting method and growing area ***(Figures 4. and 5)***.

As expected, virtually all other information is hidden by the characteristic water peaks (7,000 – 6,800 and 5200 – 5150 $cm^{-1}$).

Based on traditional transmission spectra, there is no difference between the spectra of samples of different places of origin roasted using the same method, or between samples of the same place of origin but using different roasting methods. In this case, instead of the traditional spectra, it is worth examining their derivatives. In terms of growing areas, there was no difference between the first derivatives either, but differences could be observed for the different roasting methods ***(Figure 6)***. The difference between

the mild Viennese roasting and the more vigorous French and Italian roasting in the 4,900 – 5,190 $cm^{-1}$ range is related to water-soluble polysaccharides, as well as protein and fat components.

Statistical evaluation of FT-NIR data

**Principal component analysis – PCA**

During the principal component analysis of coffee bean samples it was determined that the variance of the variables is explained by two principal components (PC1=86.96%, PC2=12.2%). The ellipse represents the 95% confidence interval. All samples are within the confidence interval, which means that the coffee bean samples, regardless of place of origin or roasting method, can be considered as a single sample matrix ***(Figure 7)***.

For ground samples, three principal components were determined (PC1=60.3%, PC2=28.2%, PC3=6.2%), and it can also be stated that all samples are within the ellipse representing the 95% confidence interval. It is also true in this case that the ground samples, regardless of place of origin or roasting method, can be considered a single sample matrix ***(Figure 8)***. In the case of ground samples, the PC1 – PC2 relationship is shown as an example.

**Grouping of coffee bean samples – LDA**

Linear discriminant analysis (LDA) was performed for all places of origin and each roasting method. ***Figure 9*** shows the LDA correlations of coffee bean analysis for the different roasting methods. It can be concluded that the three roasting methods are perfectly separated (for clarity, the ellipse representing the 95% confidence interval is not shown here). LDA tests always need to be checked, and the easiest way to do this is to run a reclassification on a random grouping. If random classification results in jumbled points, then it can be stated that the original grouping is not due to chance. This check was carried out in all cases and based on the results it can be stated that pattern recognition in the case of coffee bean samples was successful for the analysis of roasting methods.

The LDA test was also performed for places of origin. The results of ***Figure 10*** show that Colombian and Guatemalan samples can be distinguished from each other at the 95% confidence level, but the Indian and Brazilian sample groups overlap. This means that certain Indian samples were incorrectly classified as Brazilian by the method. In this case, LDA was also checked using the above-mentioned random classification, as a result of which it was found that classification by place of origin had been successful, with the exception of the two samples in question.

For separation by roasting method, further LDA tests were carried out on the coffee bean samples, which resulted in the perfect separation of the samples based on the roasting level ***(Figure 11)***. Successful pattern recognition was again supported in this case by the LDA check.

**Grouping of ground coffee samples – LDA**

In the case of ground samples, grouping according to the roasting method ***(Figure 12)*** and place of origin ***(Figure 13)*** was also carried out by the LDA method. Based on the roasting method, similarly to the coffee bean samples, each sample was separated, thereby confirming that the FT-NIR measurement was not affected by the particle size of roasted coffee ***(Figure 12)***. Since similar results were obtained for each growing area, only the results of the Guatemalan ground coffee samples are shown in ***Figure 12***.

In the case of ground samples, LDA testing performed according to the growing area showed that the individual growing areas could be completely separated from each other at the 95% confidence level, there was no faulty classification ***(Figure 13)***.

**Grouping of coffee drinks – LDA**

After brewing the coffee, the different roasting methods cannot be distinguished at the 95% confidence level. However, based on the LDA ***(Figure 14)*** it can be seen here as well that the measurement points of French and Italian roasting are closer to each other, they almost overlap. In contrast, samples of Viennese roasting are more separated.

Test results of brews of coffees obtained from different growing areas show that the confidence ellipses of the Indian and Guatemalan samples completely overlap, i.e., these groups cannot be distinguished from each other ***(Figure 15)***. From this it can be concluded that there is a great similarity between the two growing areas, which can be caused by the same soil composition and similar processing methods.

## Conclusions

Based on the results, it can be stated that the FT-NIR method is suitable for distinguishing between *Coffea arabica* coffee bean samples, ground coffee samples and coffee drinks from different growing areas by place of origin and roasting method. Linear discriminant analysis models fitted to the spectra distinguished between the roasting methods perfectly and between the growing areas with a small error in the case of coffee bean samples and ground coffee samples. However, during the analysis of coffee drinks, it was only possible to distinguish between the two extreme roasting levels, and there were several overlaps when running the test on the basis of growing areas. It can be assumed that the more processed the given product is, the more difficult it is for the technique used in our research to work, thus

requiring further tests. In the case of coffee drinks, it should also be taken into account that the sensitivity of FT-NIR to water may significantly influence measurement results.

For accurate analysis of coffee drinks, it is advisable to use more complex analytical techniques, e.g., GC-MS **[10, 11, 12]**. Sensory qualification of coffee drinks may provide additional information using either trained **[13, 14]**, or consumer sensory panels **[15, 16]**, or by the combined evaluation of instrumental and sensory data **[17]**.

**Acknowledgement**

The authors would like to express their thanks to János Szongoth and Semiramis Kft. for supplying the green coffee samples and their professional advice.

The publication is supported by the EFOP-3.6.3-VEKOP-16-2017-00005 project. The project is cofinanced by the European Union and the European Social Fund.

**References**

**[1]** Wintgens J.N.: *Coffee: Growing, Processing, Sustainable Production*. Wiley-VCH (2012).

**[2]** Spiller A.M.: „The Coffee Plant and Its Processing" in *Caffeine*. Spiller G.A., Ed., CRC Press, 22–28 (1997).

**[3]** Clarke R., Vitzthum O.G.: *Coffee: Recent Developments*. Wiley-Blackwell (2001).

**[4]** Huck C.W., Guggenbichler W., Bonn G.K.: Analysis of caffeine, theobromine and theophylline in coffee by near infrared spectroscopy (NIRS) compared to high-performance liquid chromatography (HPLC) coupled to mass spectrometry. *Anal. Chim. Acta* **538** (1-2), 195–203 (2005).

**[5]** Esteban-Díez I., González-Sáiz J.M., Sáenz-González C., Pizarro C.: Coffee varietal differentiation based on near infrared spectroscopy. *Talanta* **71** (1), 221–229 (2007).

**[6]** Pizarro C., Esteban-Díez I., González-Sáiz J.M.: Mixture resolution according to the percentage of robusta variety in order to detect adulteration in roasted coffee by near infrared spectroscopy. *Anal. Chim. Acta* **585** (2), 266–276 (2007).

**[7]** Pizarro C., Esteban-Díez I., Nistal A.J., González-Sáiz J.M.: Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Anal. Chim. Acta* **509** (2), 217–227 (2004).

**[8]** Várvölgyi E. *et al.*: Application of Sensory Assessment, Electronic Tongue and GC–MS to Characterize Coffee Samples. *Arabian Journal for Science and Engineering* **40** (1), 125–133 (2014).

**[9]** Workman J., Weyer L.: *Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy*. 2nd ed. CRC Press (2012).

**[10]** Radványi D., Gere A., Jókai Z., Fodor P.: Rapid evaluation technique to differentiate mushroom disease-related moulds by detecting microbial volatile organic compounds using HS-SPME-GC-MS. *Analytical and Bioanalytical Chemistry* **407** (2) 537–545 (2015).

**[11]** Radványi D., Gere A., Sipos L., Kovács S., Jókai Z., Fodor P.: Discrimination of mushroom disease-related mould species based solely on unprocessed chromatograms. *Journal of Chemometrics* (2016).

**[12]** Bernhardt B. *et al.*: Comparison of different *Ocimum basilicum* L. gene bank accessions analyzed by GC–MS and sensory profile. *Industrial Crops and Products* **67** 498–508 (2015).

**[13]** Sipos L., Ladányi M., Kókai Z., Gere A.: Leíró vizsgálatot végző érzékszervi bírálók teljesítményértékelési módszereinek felülvizsgálata. *Élelmiszervizsgálati közlemények – Journal of Food Investigation* **63** (1) 1435–1451 (2017).

**[14]** Sipos L., Ladányi M., Gere A., Kókai Z., Kovács S.: Panel performance monitoring by Poincaré plot: A case study on flavoured bottled waters. *Food Research International* **99** (1), 198–205 (2017).

**[15]** Gere A., Szabó Z., Pásztor-Huszár K., Orbán C., Kókai Z., Sipos L.: Use of JAR-Based Analysis for Improvement of Product Acceptance: A Case Study on Flavored Kefirs. *Journal of Food Science* **82** (5), 1200–1207 (2017).

**[16]** Gere A., Sipos L., Héberger K.: Generalized Pairwise Correlation and method comparison: Impact assessment for JAR attributes on overall liking. *Food Quality and Preference* **43**, 88–96 (2015).

**[17]** Gere A. *et al.*: Applying parallel factor analysis and Tucker-3 methods on sensory and instrumental data to establish preference maps: case study on sweet corn varieties. *Journal of the Science of Food and Agriculture* **94** (15), 3213–3225 (2014).