

# Structured Laser Illumination Planar Imaging Based Classification of Ground Coffee Using Multivariate Chemometric Analysis

Olivier K. Bagui<sup>1</sup>, Kenneth A. Kaduki<sup>2</sup>, Edouard Berrocal<sup>3</sup> & Jeremie T. Zoueu<sup>1</sup>

<sup>1</sup>Laboratoire d'Instrumentation Image et Spectroscopie, Institut National Polytechnique Felix Houphouet-Boigny, BP 1093 Yamoussoukro, Cote d'Ivoire

<sup>2</sup>Laser Physics and Spectroscopy Group, Department of Physics, University of Nairobi, P. O. Box 30197-00100, Nairobi, Kenya

<sup>3</sup>Department of Physics, Division of Combustion Physics, Lund Institute of Technology, Box 118, Lund 221 00, Sweden

Correspondence: J. T. Zoueu, Laboratoire d'Instrumentation Image et Spectroscopie, Institut National Polytechnique Felix Houphouet-Boigny, BP 1093 Yamoussoukro, Cote d'Ivoire. Tel: 225-064-6684. E-mail:Jeremie.zoueu@nphb.edu.ci

Received: March 4, 2016 Accepted: March 17, 2016 Online Published: April 19, 2016

doi:10.5539/apr.v8n3p32

URL: <http://dx.doi.org/10.5539/apr.v8n3p32>

## Abstract

Most commercially available ground coffees are processed from Robusta or Arabica coffee beans. In this work, we report on the potential of Structured Laser Illumination Planar Imaging (SLIPI) technique for the classification of five types of Robusta and Arabica commercial ground coffee samples (Familial, Belier, Brazil, Colombia and Malaga). This classification is made, here, from the measurement of the extinction coefficient  $\mu_e$  and of the optical depth OD by means of SLIPI. The proposed technique offers the advantage of eliminating the light intensity from photons which have been multiply scattered in the coffee solution, leading to an accurate and reliable measurement of  $\mu_e$ . Data analysis uses the chemometric techniques of Principal Component Analysis (PCA) for variable selection and Hierarchical Cluster Analysis (HCA) for classification. The chemometric model demonstrates the potential of this approach for practical assessment of coffee grades by correctly classifying the coffee samples according to their species.

**Keywords:** extinction coefficient, optical density, structured illumination, SLIPI, PCA, HCA

## 1. Introduction

Coffee is a popular beverage used throughout the world (Oder, 2015). Though not widely consumed on the continent, African countries make a significant contribution to world coffee trade. Ethiopia was the third largest coffee producer in the world in 2015 with Cote d'Ivoire being number twelve (International Coffee organization, 2015). The quality of coffee depends on many factors, such as the growth environment and processing techniques (Carelli et al., 2006; Clifford & Willson, 1985). Arabica and Robusta varieties represent over 90% of the world production of coffee. Arabica coffee, originally from Ethiopia, is now widely cultivated in South America. Robusta, which is a variety of the *canephora* species, is grown in Africa (mainly in Cote d'Ivoire) and in the Far East (Vietnam in particular). Robusta coffee is richer in caffeine than arabica (from 2% to 3% against 1.3%), (National Coffee Association of USA, 2015). The arabica variety fetches higher prices on the world market.

In order to distinguish between Robusta and Arabica varieties on the market, non-specialists rely on information provided on the packaging. As it is the case with other high value agricultural products, there have been increased incidences of counterfeit coffee on sale in the market. Chemical and genetic procedures for identification of the origin of ground coffee exist but they are time consuming. Another method is sensory evaluation but it is not appropriate for accurate and repeatable classification. Development of measurement techniques and technologies that can objectively discriminate between Robusta and Arabica ground coffee are therefore highly desirable. Nowadays, the use of optical laser techniques for various quantitative measurements is commonly applied for a large number of applications. Here, we aim at measuring the extinction coefficient (e.g. Berrocal et al., 2012; Kristensson et al., 2011, Kristensson et al., 2012) in different coffee solutions and analyze the results by means of chemometrics.

Chemometrics is the use of mathematical and statistical methods to extract chemical information and to correlate quality parameters or physical properties from an experimental data-set. The basic process involves modeling of patterns in the data. The models are then routinely applied to future data in order to predict quality parameters. The chemometrics approach has been gaining interest in assessing product quality. The only requirements are the extraction of reliable measurements and adequate software to interpret the patterns in the data.

In this article, the extinction coefficient of various coffee solutions is measured using a recent approach called single-phase Structured Laser Illumination Planar Imaging (SLIPI) (Berrocal et al., 2012). While transmission measurement records the light intensity of a single beam crossing the sample of interest (where the initial and final intensities are recorder) SLIPI is based on imaging a spatially modulated light sheet from the side (at 90° angle). The main advantage of SLIPI over conventional transmission measurements is its efficient capability in rejecting the light intensity from multiply scattered photons, allowing more accurate measurement of the extinction coefficient in turbid media (such as coffee solutions). We aim, then, at combining SLIPI measurements with a data analysis based on chemometrics to make a reliable classification of various types of coffee.

## 2. Methods

### 2.1 Sample Preparation

The coffee samples were prepared following the same procedure. For each type of coffee, the solutions were prepared by weighting 4g of coffee using a Satorius VIC-303 balance with 1 mg resolution and dissolving it in 100 mL of boiled distilled water. We stirred the water and coffee mixture for 15 seconds to get a homogeneous suspension, filtered the suspension into a sealed glass flask and left it to cool to 20°C before starting measurements.

### 2.2 Experimental Setup

The SLIPI technique was first created and applied in 2008 for imaging spray systems typically used in combustion engines (Berrocal et al., 2008). A description of it, in its various configurations, can be found in the doctoral thesis of E. Kristensson (Kristensson et al., 2012). We employ, here, the single-phase SLIPI approach for the measurements of the extinction coefficients and optical depths. The method has been presented and fully described in (Berrocal et al., 2012). Figure 1 shows a schematic of our experimental setup.

In the experiment, a coffee solution is illuminated in a cuvette with a spatially modulated laser sheet constructed using by a 5 lp/mm Ronchi grating and shaped by spherical and cylindrical lenses. The incident light is produced by three diode lasers emitting at 450 nm, 532 nm and 638 nm, respectively. A 650 nm high pass filter is positioned in front of the camera to only detect the laser induced fluorescence signal from the coffee.

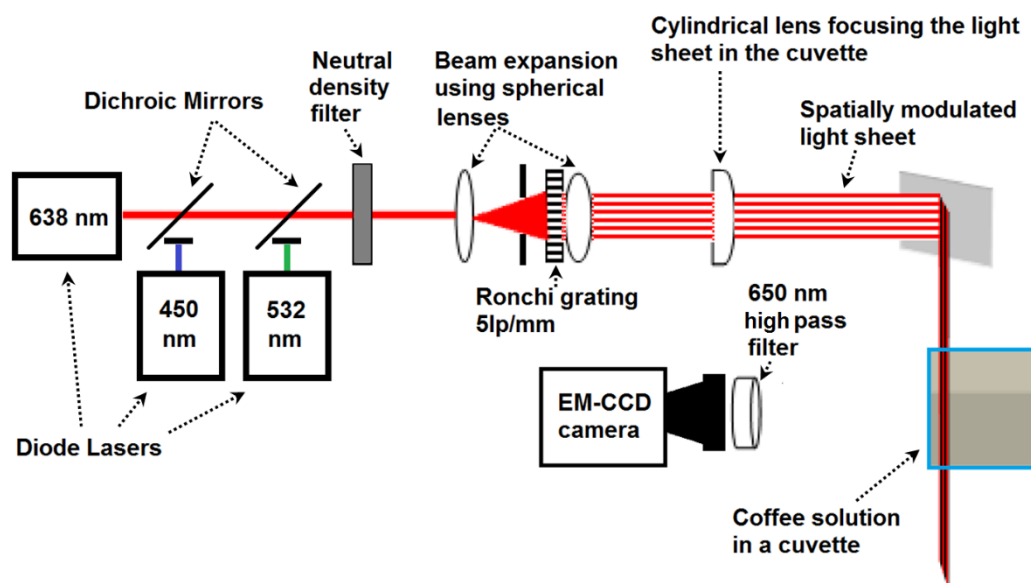


Figure 1. Description of the single-phase SLIPI optical arrangement: A light sheet with a vertically modulated light intensity profile is formed, illuminating the coffee solution. Images of the spatially modulated light sheet are recorded from the side using an EM-CCD camera. By then extracting the amplitude of the modulation, the exponential light extinction through the cuvette can be observed. The measurements are performed sequentially for three different illumination wavelengths corresponding to 450nm, 532 nm and 638 nm respectively

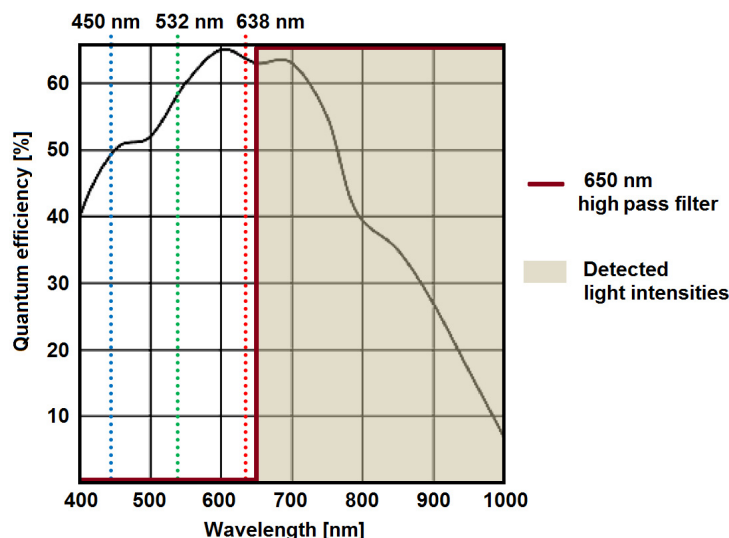


Figure 2. Quantum efficiency curve of the Andor Luca-R EM-CCD camera (data curve from Andor Technology). The wavelengths of corresponding to each illumination scheme and to the low pass filter fixed on the camera objective are also indicated

The images are acquired with an Andor Luca-R Electron Multiplying Charge-Coupled Device (EM-CCD) camera placed at an angle of  $90^\circ$  from the direction of the incident the light sheet. The quantum efficiency of the camera, cooled at  $-20^\circ\text{C}$ , is shown in Figure 2.

### 2.3 Data Analysis

Our clustering algorithm uses the extinction coefficient and the optical depth of each sample for the different laser illumination. Therefore, each sample has six variables making this a multivariate statistical problem. In data sets containing many variables, groups of variables are often inter-related. This can be explained as one variable might be measuring the same underlying principle governing the behavior of the complete system. We can exploit this redundancy by replacing a group of variables with a single new variable. The best way to achieve this is to apply Principal Component Analysis (PCA). PCA generates a new set of variables called principal components. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other, so there is no redundancy of information (Jolliffe, 2002; François Husson, 2014; Besse, 1992).

The first principal component accounts for the most variance and therefore has the most information; the second principal component has the second best variance, and so on.

With this information, one can reduce the original data to represent the significant contrast and trends with only a few variables rather than all contained in the original data by the removal of insignificant variables for the desired contrast. Adding more dimensions do not provide any additional contrast but only increases the noise and reduces the potential contrast of the outcome.

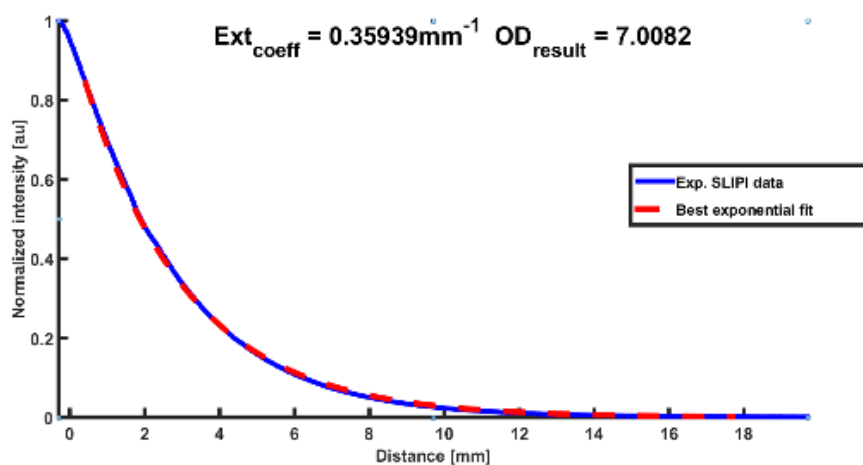
Hierarchical clustering and dendrogram representation (Hastie, Tibshirani, & Friedman, 2009) were applied to summarize the interdistance of the PCA scores to see if there were any discrete clusters of data points in the new coordinate system and how related these were.

## 3. Results

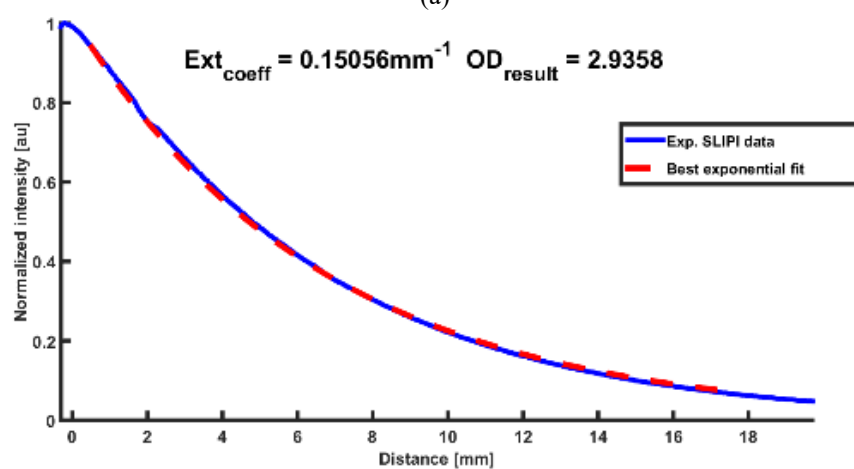
The coffee samples were identified as follows, based on the labels on their respective packages: Belier and Malaga are 100% Robusta while Brasil, Columbia and Familial are 100% Arabica.

### 3.1 SLIPI Results

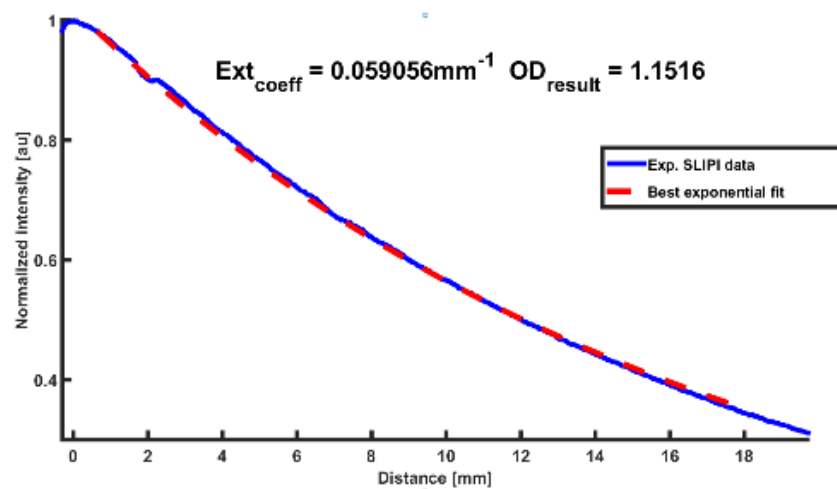
The experimental results from SLIPI measurements are presented in Figure 3 (a,b,c,d,e) for the different ground coffee and extinction coefficient and optical density obtained with the different laser (450nm, 532nm, 638nm) grouped as shown in Tables 1, 2 and 3.



(a)



(b)



(c)

Figure 3. Experimental results of extinction coefficient and optical density from SLIPI measurements with Belier coffee at 450 nm (a), 532 nm (b) and 638 nm (c)

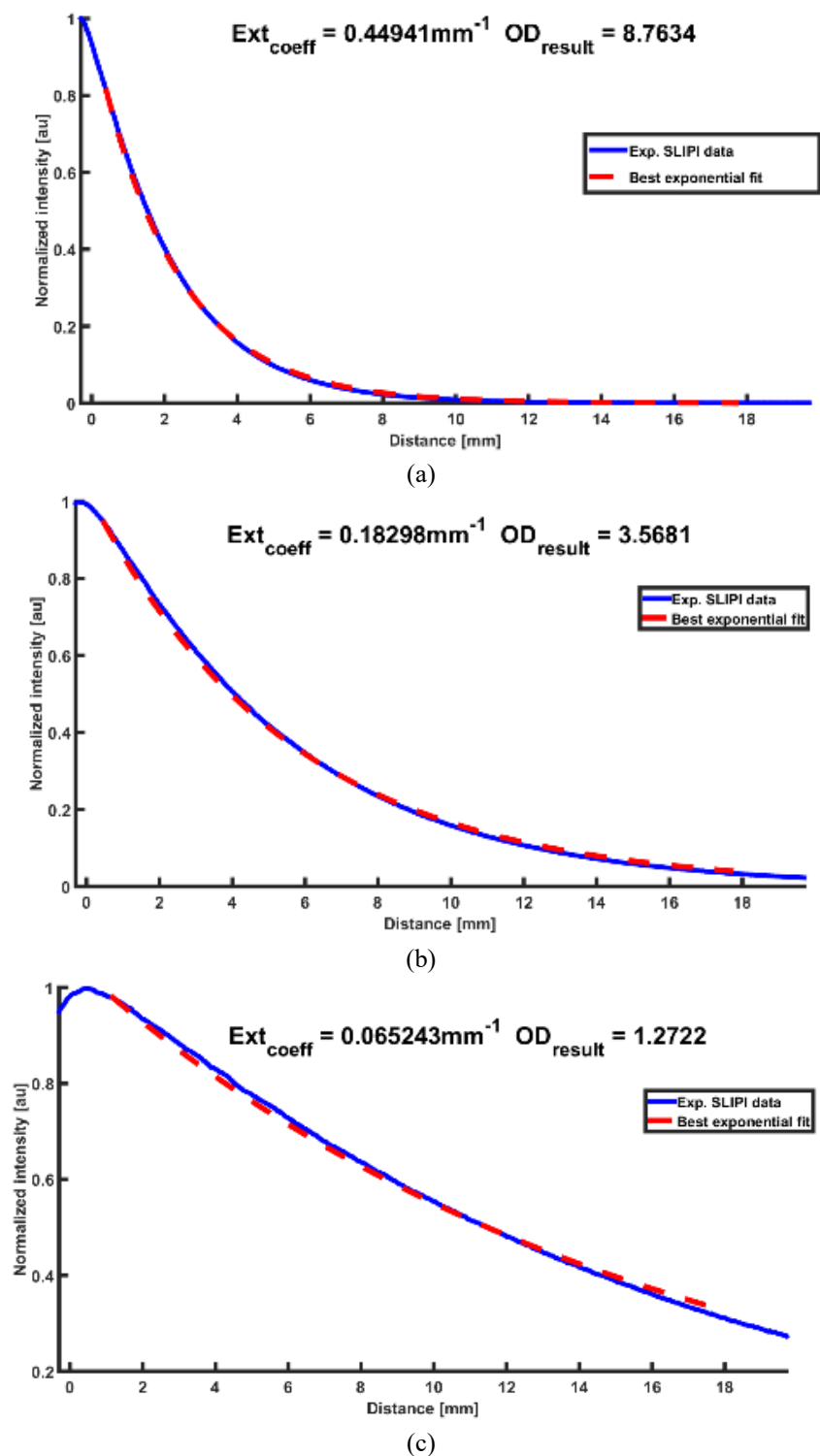


Figure 4. Experimental results of extinction coefficient and optical density from SLIPI measurements with Malaga coffee at 450 nm (a), 532 nm (b) and 638 nm (c)

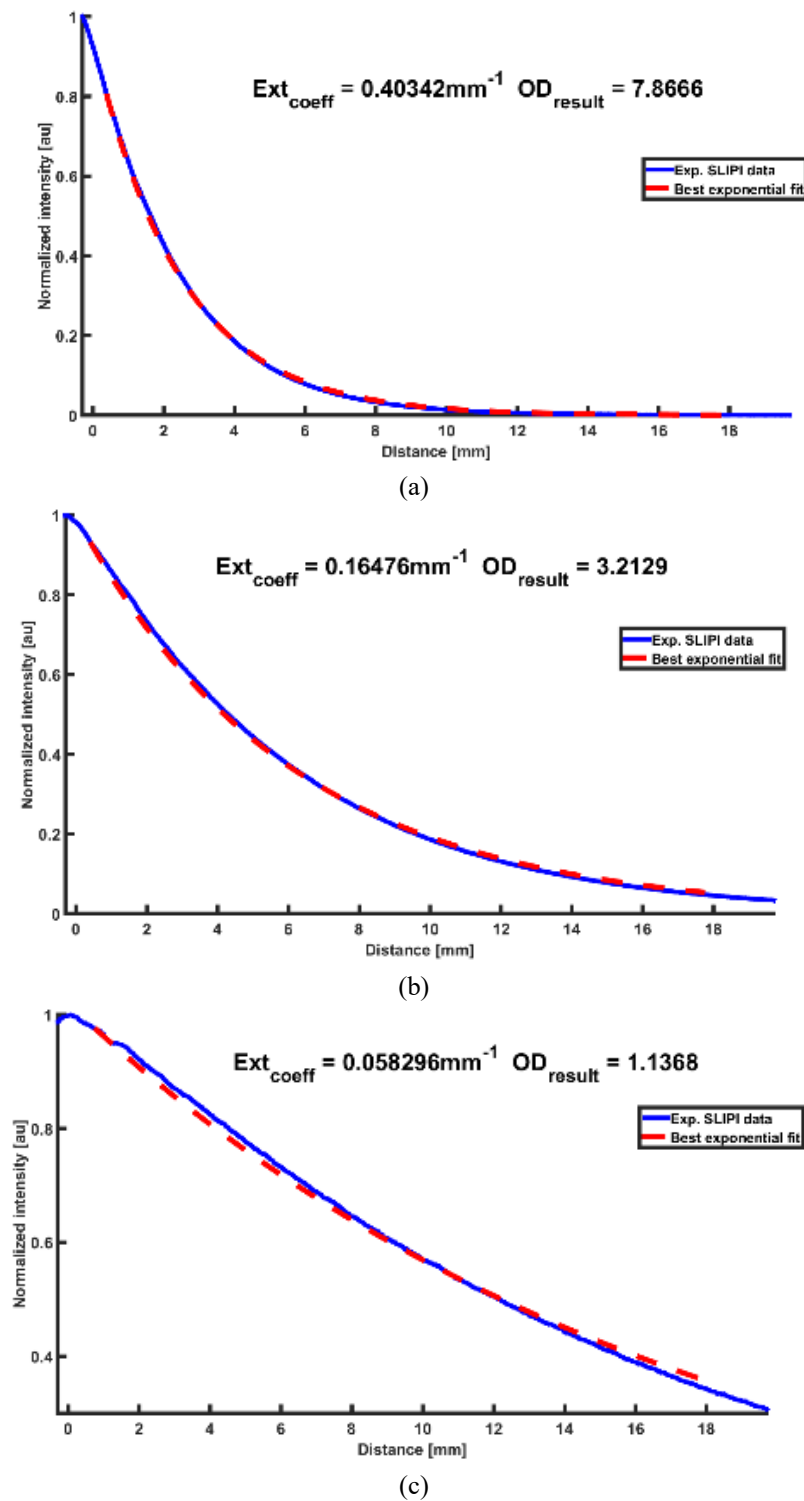


Figure 5. Experimental results of extinction coefficient and optical density from SLIPI measurements with Familial coffee at 450 nm (a), 532 nm (b) and 638 nm (c)

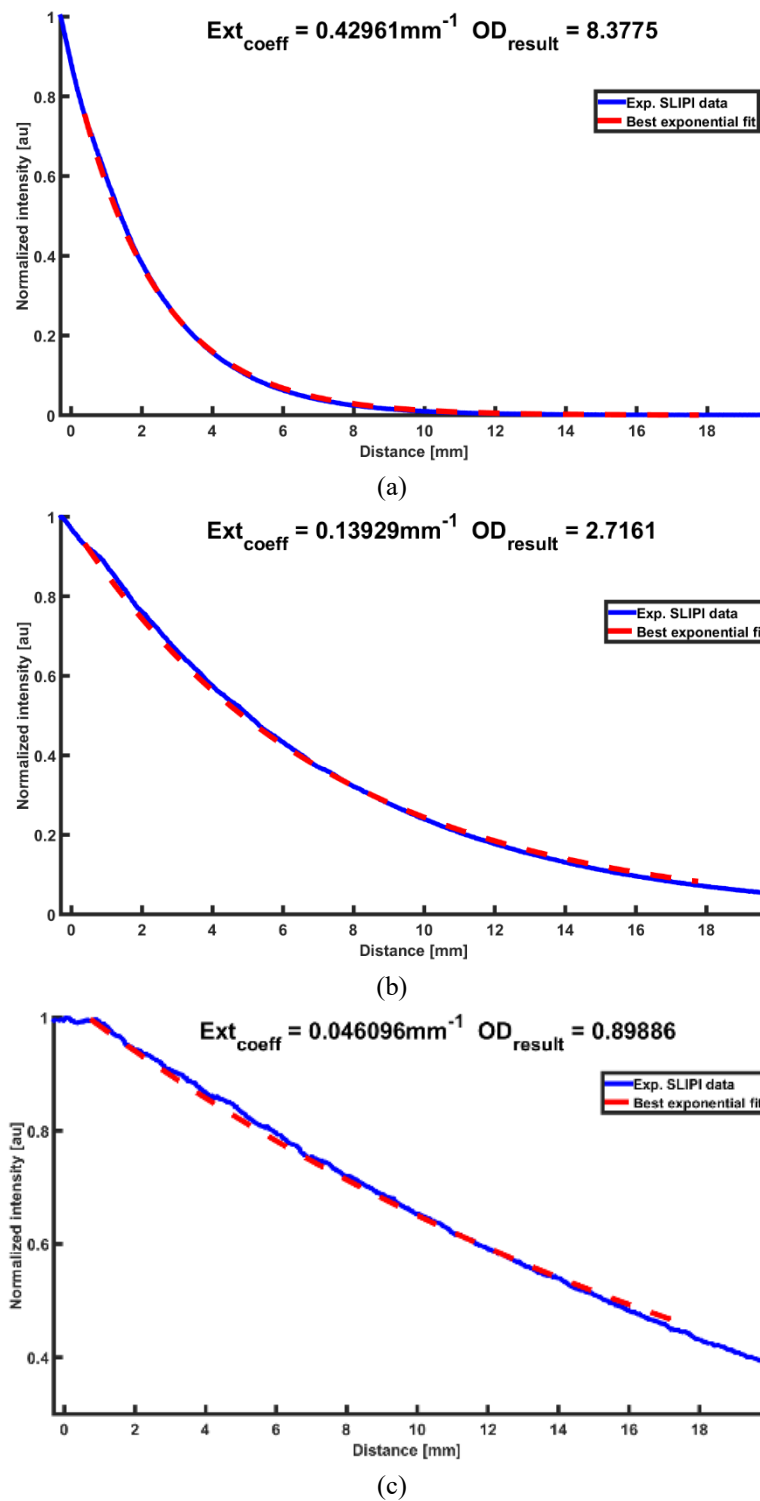


Figure 6. Experimental results of extinction coefficient and optical density from SLIPI measurements with Colombia coffee at 450 nm (a), 532 nm (b) and 638 nm (c)

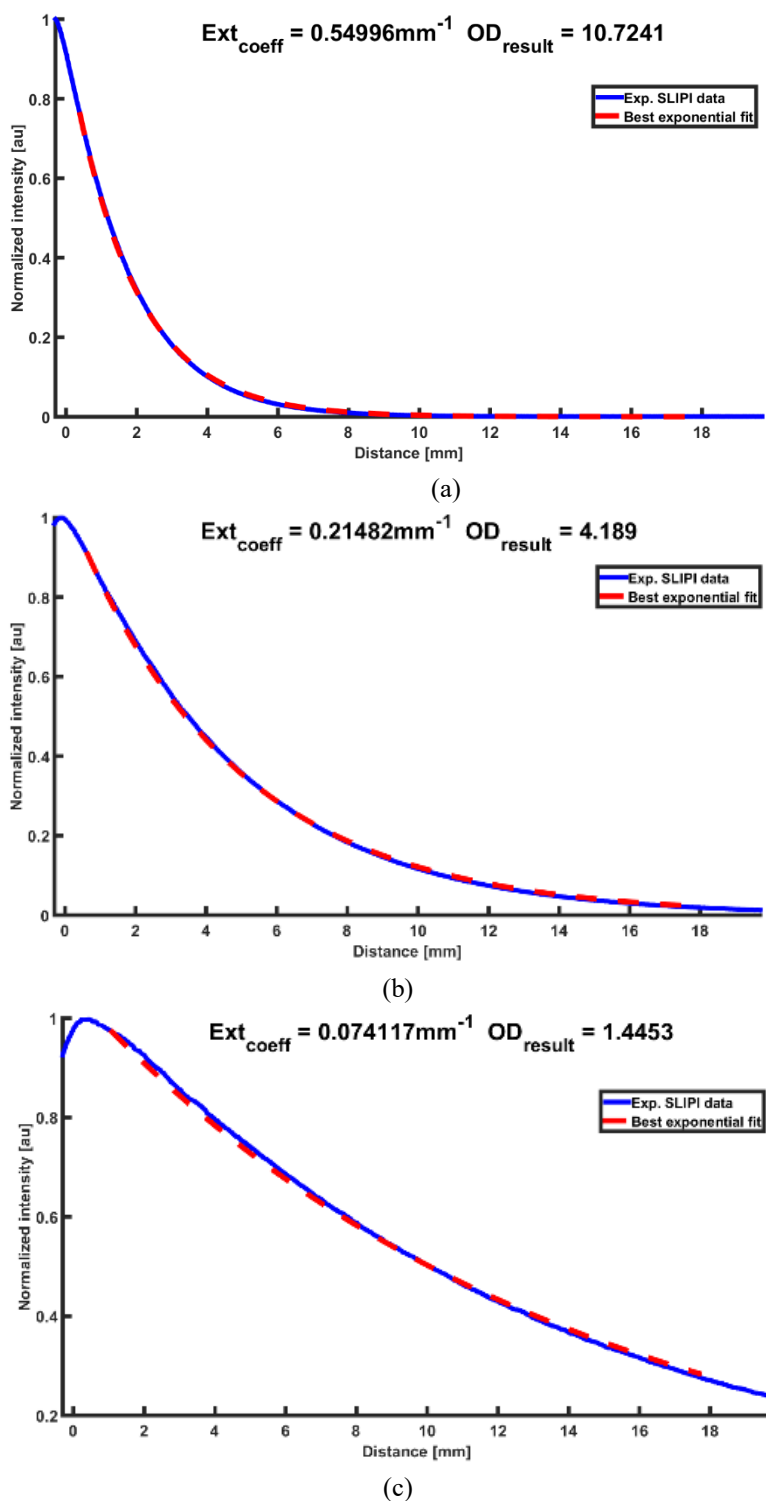


Figure 7. Experimental results of extinction coefficient and optical density from SLIPI measurements with Brazil coffee at 450 nm (a), 532 nm (b) and 638 nm (c)



Table 1. SLIPI results using laser illumination at 450 nm

	450 nm laser illumination	
	Extinction Coefficient ( $\text{mm}^{-1}$ )	Optical Depth (-)
Belier	0.35939	7.0082
Malaga	0.44941	8.7634
Familial	0.40342	7.8666
Colombia	0.42961	2.7161
Brasil	0.54996	10.7241

Table 2. SLIPI results using laser illumination at 532 nm

	532 nm laser illumination	
	Extinction Coefficient ( $\text{mm}^{-1}$ )	Optical Depth (-)
Belier	0.15056	2.9358
Malaga	0.18298	3.5681
Familial	0.16476	3.2129
Colombia	0.13929	0.046096
Brasil	0.21482	4.189

Table 3. SLIPI results using laser illumination at 638 nm

	638 nm laser illumination	
	Extinction Coefficient ( $\text{mm}^{-1}$ )	Optical Depth (-)
Belier	0.059056	1.1516
Malaga	0.065243	1.2722
Familial	0.058296	1.1368
Colombia	0.046096	0.89886
Brasil	0.074117	1.4453

The results in all the tables show a difference between all coffees with regard to the extinction coefficient and optical density parameters. The corresponding plots of the extinction coefficients are given in Figure 8(a) together with the ratio of the extinction coefficients (for each illumination scheme) in Figure 8(b). However, it is very difficult to classify in which type of coffee they are belonging (whether Arabica or Robusta). In order to find a relevant feature to describe the classification it is very important that the variables (extinction coefficient and optical density) are independent with regard to coffee species and laser illumination. To do this we used chemometrics which gives many different ways to solve the discrimination problem in the analysis of data.

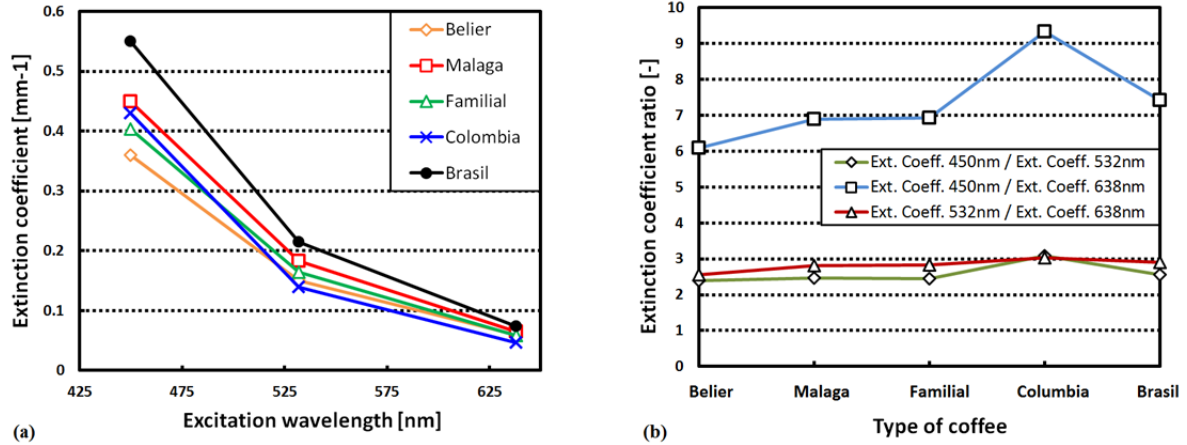


Figure 8. (a) Results of the measurement of the extinction coefficient for each type of coffees for the three illumination wavelengths. (b) Ratio of the extinction coefficients for each type of coffee

### 3.2 Chemometric Results

Before performing the analysis, we checked the correlation between all the variables. The correlation among some variables is as high as 60 % (Figure 8).

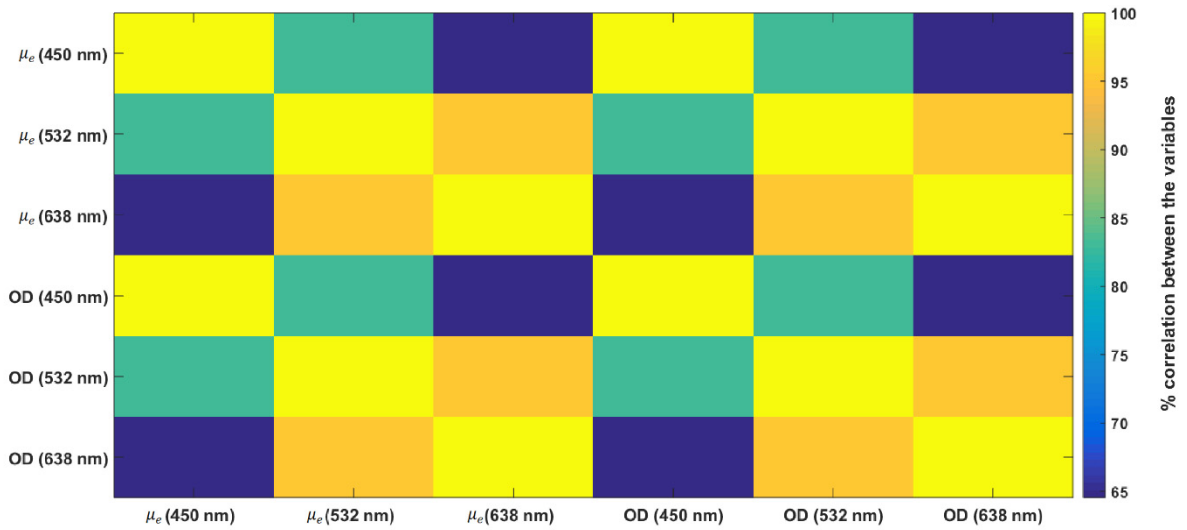


Figure 8. correlation between the variables

Note that there are a large correlation between extinction coefficient and optical density at the same wavelength. This high correlation can be explained by equation 1.

$$OD = l\mu_e \quad (1)$$

where  $\mu_e$  is the extinction coefficient. The optical depth (OD) is an approximation of the mean number of scattering events occurring through a scattering medium of length  $l$ . The extinction coefficient is equal to the sum of the scattering coefficient and the absorption coefficient (Equation 2):

$$\mu_e = \mu_a + \mu_s \quad (2)$$

PCA was then applied to construct independent new variables which are linear combinations of the original variables. The variables do not have the same units so we apply PCA using the inverse variances of the data as weights. To determine which components have high variance and must be retained to describe the data, we made a scree plot of the percent variability explained by each principal component (Figure 9). The scree plot only

shows the first two (instead of the total seven) components that explain 99.7% of the total variance. Thus only the first and the second principal component can be retained.

We then applied Hierarchical Clustering and Euclidean distance as a metric using these two new variables. Hierarchical Clustering groups data over a variety of scales by creating a cluster tree. We used the silhouette criteria to determine where to truncate the cluster. The silhouette value for each point is a measure of how similar that point is to points in its own cluster, when compared to points in other clusters. The silhouette value for the  $i^{\text{th}}$  point ( $S_i$ ) is defined as:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (3)$$

where  $a_i$  is the average distance from the  $i^{\text{th}}$  point to the other points in the same cluster as  $i$ , and  $b_i$  is the minimum average distance from the  $i^{\text{th}}$  point to points in a different cluster, minimized over clusters. A high silhouette value indicate that it is well-matched to its own cluster, and poorly matched to neighboring clusters (Kaufman, L., & Rousseeuw, P. J., 2009). Figure 6 show the silhouette criterion values for the number of clusters tested. The plot shows that the highest silhouette value occurs at five clusters, suggesting that the optimal number of clusters is five. After this we grouped data over a variety of scales by creating a cluster tree using HCA (Figure 10).

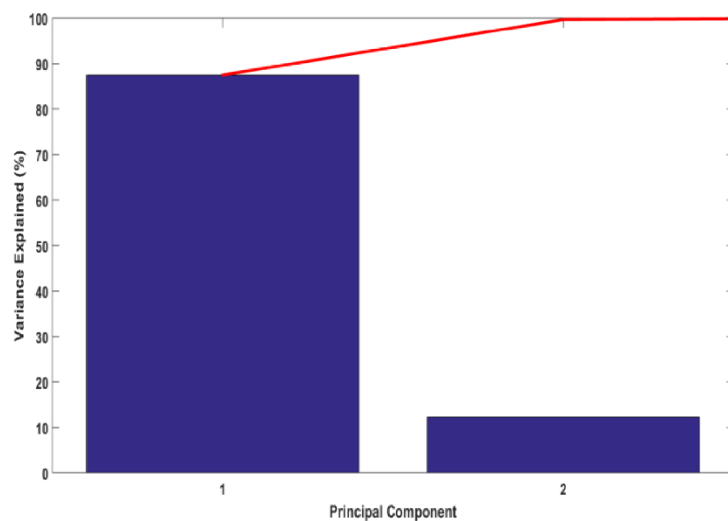


Figure 9. Scree plot of the percent variability explained by the first and second principal component

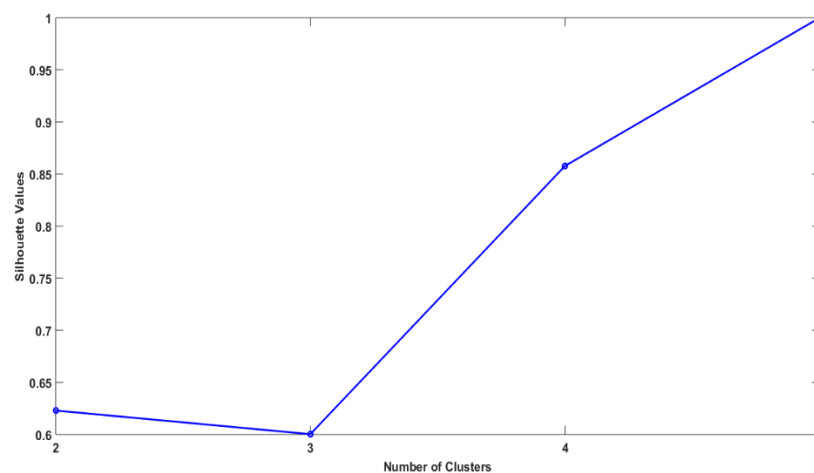


Figure 10. Silhouette values for each clusters tested

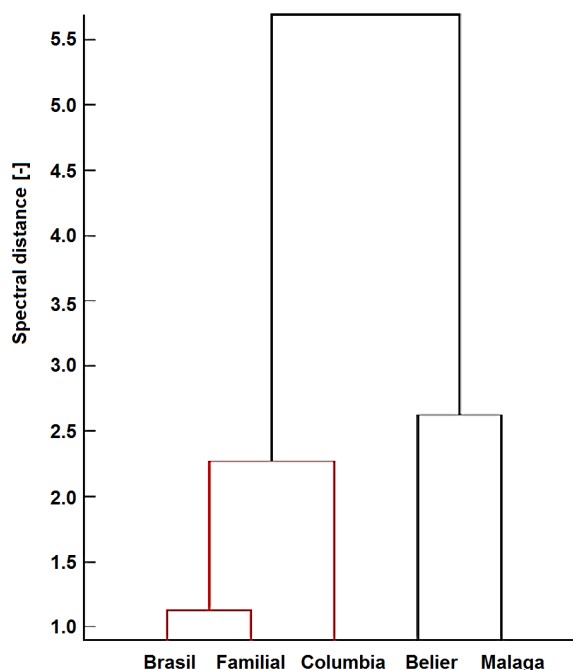


Figure 11. Representation of the hierarchical cluster tree

In the Figure 11, the numbers along the horizontal axis represent the name of the different ground coffee in the original data set. The links between grounds coffees are represented as upside-down U-shaped lines, with the height of the U indicates the distance between the objects.

Based on the silhouette criterion each coffee sample represents an independent group. The links in the dendrogram show that Brasil, Familial and Colombia are near and can form a group. In the same graph, Belier and Malaga are near and can be considered as another group. These results are in agreement with the identity of the coffee as indicated on the packaging labels. Brasil, Familial and Colombia are Arabica types and Belier and Malaga are from the Robusta variety.

#### 4. Discussion

Chemometrics and SLIPI are both powerful techniques for spectroscopic studies; they have been used as complementary methods in this study.

The multivariate approach dealt with the following steps: pre-processing, PCA, variable selection and HCA classification. The data collected with SLIPI technique for each dataset (Belier, Malaga, Brasil, Colombia and Familial ground coffee) were used to show the suitability of this technique to detect similarity between the ground coffee samples.

For every pre-treated dataset, PCA was performed as an explanatory tool in order to get the overspread of data. PCA variance was used to retain the best components to use in describing the variability in the different coffee types and sample groupings.

The HCA plot shows a good grouping of the samples on the basis of the two classes in the space defined by the two first components. This strategy shows that extinction coefficient and optical density measured with SLIPI technique could be useful in the discrimination of coffees species.

#### 5. Conclusion

The strategy showed a clear coffee grouping on the basis of the two classes (Arabica and Robusta). We can conclude that, the SLIPI technique combined with chemometric analysis of coffee samples offer complementary results for the discrimination of products and can be used to accurately classify and evaluate coffee samples.

#### Acknowledgement

The authors wish to thank the International Science Program (ISP) of Uppsala University for equipment and financial support as well as the Lund Laser Center (LLC).

## References

- Berrocal, E., Johnsson, J., Kristensson, E., & Aldén, M., (2012). Single scattering detection in turbid media using single-phase structured illumination filtering. *Journal of the European Optical Society-Rapid Publications*, 7. <http://dx.doi.org/10.2971/jeos.2012.12015>
- Berrocal, E., Kristensson, E., Richter, M., Linne, M., & Alden, M. (2008). Application of structured illumination for multiple scattering suppression in planar laser imaging of dense sprays. *Optics express*, 16(22), 17870-17881. <http://dx.doi.org/10.1364/OE.16.017870>
- Besse, P. (1992). PCA stability and choice of dimensionality. *Statistics & Probability Letters*, 13(5), 405-410. [http://dx.doi.org/10.1016/0167-7152\(92\)90115-L](http://dx.doi.org/10.1016/0167-7152(92)90115-L)
- Carelli, M. L. C., Fahl, J. I., & Ramalho, J. D. C. (2006). Aspects of nitrogen metabolism in coffee plants. *Brazilian Journal of Plant Physiology*, 18(1), 921. <http://dx.doi.org/10.4067/S0718-9516201400500 0018>
- Clifford, M. N., & Willson, K. C. (1985). *Coffee: Botany, Biochemistry, and Production of Beans and Beverage*. Westport, CT: AVI. <http://dx.doi.org/10.1007/978-1-4615-6657-1>
- François HUSSON. (2014). Analyse en composantes principales (ACP) Théorie et pratique.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Hierarchical clustering*. The elements of statistical learning, SPRINGER 2. 520-528.
- International Coffee organization. (2015). *List of countries by coffee production*. Wikipedia. Retrieved January 22, 2016, from [https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_coffee\\_production](https://en.wikipedia.org/wiki/List_of_countries_by_coffee_production)
- Jolliffe, I. T. (2002). Principal component analysis, ser. *Springer Ser. Statist.* (2nd ed.). New York: Springer.
- Kaufman, L., & Rousseeuw, P. J. (2009). Finding groups in data: an introduction to *cluster analysis* (Vol. 344). John Wiley & Sons.
- Kristensson, E. (2012). *Structured Laser Illumination Planar Imaging, SLIPI: Applications for Spray Diagnostics*. Lund University, PhD Thesis.
- Kristensson, E., Berrocal, E. & Aldén, M. (2012). Quantitative 3D imaging of scattering media using structured illumination and computed tomography. *Opt. Express*, 20, 14437-14450.
- Kristensson, E., Berrocal, E., & Aldén, M. (2011). Extinction coefficient imaging of turbid media using dual structured laser illumination planar imaging. *Opt. Lett.*, 36, 1656-1658.
- National Coffee Association of USA. (2015). Retrieved October 30, 2015, from <http://www.ncausa.org/>
- Oder, T. (2015). How coffee changed the world. *Mother Nature Network*. Retrieved October 30, 2015, from <http://www.mnn.com/food/beverages/stories/how-coffee-changed-the-world>

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).