SciVerse ScienceDirect

Procedia Food Science 1 (2011) 528 – 535

Procedia Food Science

11th International Congress on Engineering and Food (ICEF11)

# Decomposition of absorption spectra of multi-layered biological materials by spatially-resolved spectroscopy and parallel factor analysis

Mizuki Tsuta*[a,b], Nghia Nguyen Do Trong[a], Els Herremans[a], Josse De Baerdemaeker[a], Wouter Saeys[a]

[a]*BIOSYST-MeBioS K.U.Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium*
[b]*National Food Research Institute, 2-1-12 Kan-nondai, Tsukuba, Ibaraki 305-8642, Japan*

**Abstract**

Visible and near-infrared spectroscopy have been widely used for non-destructive and rapid monitoring of biological materials including agricultural produce and foods. Methods for decomposition of spectra of biological materials into that of each layer are demanded because the target of monitoring is in most cases limited to a single layer (e.g. sweetness of the flesh of a fruit, not the skin). Also, separation of light absorption and scattering properties is considered to lead more robust modeling of biological materials. Therefore, the potential of Parallel factor analysis (PARAFAC) on spatially resolved spectroscopy (SRS) has been investigated in this study. Spatially resolved spectra of 120 'Braeburn' apples with difference in skin color (green and red) and storage condition (before and after 14 days shelf-life storage) were measured in the 400-1100 nm range using 7 detection fibers with different source-detector distance. PARAFAC on the whole data set could decompose the spatially resolved spectra into that of flesh and skin as factor 1 and 2 wavelength loadings, respectively. This result was validated by the facts that the factor 1 loading was almost identical to the 'real' spectra of flesh, which had been acquired by measuring peeled apples, and peaks observed in the factor 2 loading corresponded to absorption bands of pigments contained in apple skin. When PARAFAC was applied to individual sample data sets, difference in skin color and chlorophyll content as well as degradation of chlorophyll in skin during ripening and storage could be modelled as the change in factor 1 wavelength loading. These results support the potential of SRS combined with PARAFAC to decompose the spectra of multi-layered biological materials.


Selection and/or peer-review under responsibility of 11th International Congress on Engineering and Food (ICEF 11) Executive Committee.

*Keywords*: SRS; PARAFAC; Apple; Skin; Flesh

* Corresponding author. Tel.: +32-16-32-85-27
*E-mail address*: mizukit@affrc.co.jp.


doi:10.1016/j.profoo.2011.09.080

## 1. Introduction

Visible and near-infrared spectroscopy have been widely used for non-destructive and rapid monitoring of biological materials including agricultural produce and foods (*1, 2*). However, since many of them are multi-layered materials (e.g. skin and flesh of fruits), measured spectra include mixture of information for each layer, while in many cases only a single layer is the target of monitoring (e.g. sweetness of the flesh, not the skin). Also, traditional spectroscopic methods are often not robust in biological materials, due to the complex interaction of incident light with them involving both absorption and scattering of the light. Therefore, methods for decomposition of spectra into that of each layer, as well as separation of absorption and scattering properties, are highly demanded.

Spatially Resolved Spectroscopy (SRS) combines spectral reflectance measurements at different distances from the incident light beam to separate the light absorbance and scattering properties of a sample (*3, 4*). Spectra measured at different source-detector distances include information at different depth of the sample because the maximum depth of the optical path is deeper for the longer sore-detector distance (*5*). Therefore, separation of absorption and scattering properties, as well as decomposition of the spectrum of each layer of the sample, would be expected by proper analysis of spatially resolved spectra.
SRS measurement yields 3-way data of samples by wavelengths by source-detector distances. Parallel factor analysis (PARAFAC) has been widely applied to such kind of 3-way data analysis because of its robustness, uniqueness and easy interpretation of the results (*6*). Therefore, the potential of PARAFAC on spatially resolved spectra of intact apples to decompose the spectra into that of skin and flesh as well as absorption and scattering coefficients has been investigated in this study.

## 2. Materials and Methods

### *2.1. Spatially-Resolved Spectroscopy Setup (SRS Setup)*

A setup for SRS measurements in the 400-1100 nm range has been built. This setup consists of a contact probe with accurately placed fibers which is linked to a spectrograph for simultaneous measurement of the reflectance at the different distances by a CCD camera. The optical probe has been designed and assembled at the Swiss Federal Institute of Technology (EPFL, Lausanne, Switzerland). The fibers used are Thorlabs multimode silica fibers (FVP-200 PF) with a numerical aperture of 0.22 and a core diameter of 200 μm. The 7 detection fibers are placed at various distances from the illumination fiber, ranging approximately from 0.3 to 1.2 mm with a step of about 0.15 mm. The illumination fiber of the probe is connected to AvaLight-DHc (Avantes, Eerbeek, The Netherlands) halogen lamp through an optical switch. The detection fibers from the SRS probe and a fiber from the optical switch of the light source are aligned in the entrance slit of a CP200 133 g/mm spectrograph (Horiba Jobin-Yvon, New Jersey, USA) which splits the light from each of these fibers into its spectral components and projects these onto a Hamamatsu C7041 CCD camera with a S7041-1008 detector (Hamamatsu, Louvain-La-Neuve, Belgium). Spectra collected by detection fibers are measured at 1044 wavelengths with approximately 0.73 nm wavelength intervals. The signal from this camera is transferred to a computer by means of a PCI MIO-16E-4 data acquisition card. Control of the light source, optical switch and camera is performed in LabView software (National instruments, TX, USA). The measurement setup for spatially resolved spectroscopy is illustrated in Fig. 1.

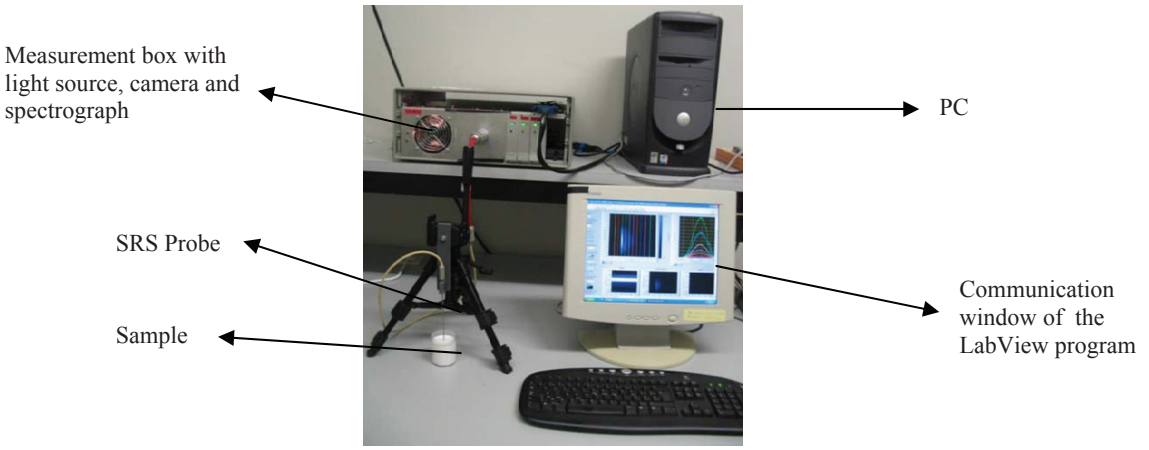


Fig. 1. The setup for spatially-resolved spectroscopy

### *2.2. Conversion of Spatially Resolved Spectra into Optical Density*

The acquired spatially resolved spectra were converted into optical density (OD) by following equation;

$$OD_i(\lambda, f) = -log\left(\frac{I_{sample,i}(\lambda, f) - D(\lambda, f)}{I_{sphere}(\lambda, f) - D(\lambda, f)}\right) \quad (1)$$

where $OD_i(\lambda, f)$ is OD of the $i^{th}$ measurement with $f^{th}$ fiber at wavelength $\lambda$, $I_{sample,i}(\lambda, f)$, $I_{sphere}(\lambda, f)$ and $D(\lambda, f)$ are the intensity values acquired when the sample, integrating sphere and dark noise were measured, respectively. In the case of the whole data set, $i$ indicates the sample number, while replication number is indicated in the individual data sets.

### *2.3. PARAFAC*

PARAFAC is a decomposition method for multi-way data set which is characterized by three or more sets of variables or 'modes' (*6*). An example of 3-way data set is fluorescence excitation-emission matrix (EEM) with 'sample', 'excitation', and 'emission wavelength' modes. Typical multivariate methods such as principal component analysis decompose data into sets of 'scores' and 'loadings', while a 3-way data set can be decomposed into three sets of loadings corresponds to each mode using PARAFAC. When samples contain $F$ components are measured, the PARAFAC model for the acquired 3-way data set can be described as

$$x_{ijk} = \sum_{f=1}^{F} a_{if} \cdot b_{jf} \cdot c_{kf} + e_{ijk} \quad (2)$$

$$i = 1...I; j = 1...J; k = 1...K$$

where $x_{ijk}$ is the value of the $i^{th}$ sample at the $j^{th}$ variable in the second mode and at the $k^{th}$ variable in the third mode, $a_{if}$ is the loading (or sample score) corresponds to the $f^{th}$ component for the first mode (or sample mode), $b_{jf}$ and $c_{kf}$ are the loadings corresponds to $f^{th}$ component for the second and third mode, and

$e_{ijk}$ is the variation which is not captured by the model (*6, 7*). Therefore, when PARAFAC is applied to spectral data, 'pure' spectrum of each component is expected to be extracted as one loading vector.

A PARAFAC solution can be constrained using the specific knowledge of the data to acquire more interpretable or stable model. For example, based on the knowledge that fluorescence intensity of a pure EEM spectrum should not be negative, non-negativity constraint can be applied to make sure that every element in the loadings correspond to excitation and emission wavelengths is positive (*6, 7*).

### *2.4. Decomposition of Spatially Resolved Spectra by PARAFAC*

Based on Beer-Lambert's law, optical density can be expressed as a product of total attenuation coefficient $\mu_t$ and path length $L$ as follows;

$$OD = \mu_t \cdot L \tag{3}$$

$$\mu_t = \mu_a + \mu_s' \tag{4}$$

where $\mu_a$ is absorption coefficient and $\mu_s'$ is reduced scattering coefficient. In this study, an apple has two layers of skin and flesh and thus acquired OD is the sum of that in each layer;

$$OD_i(\lambda, f) = \mu_{t,skin}(\lambda) \cdot L_{skin,i}(f) + \mu_{t,flesh}(\lambda) \cdot L_{flesh,i}(f) \tag{5}$$

When the mean path length $\overline{L}$ for each layer and a scaling factor $\alpha$, which is the ratio of the path length of each sample to $\overline{L}$, are introduced, equation 5 can be rewritten as

$$OD_i(\lambda, f) = \mu_{t,skin}(\lambda) \cdot \alpha_{skin,i} \cdot \overline{L_{skin}}(f) + \mu_{t,flesh}(\lambda) \cdot \alpha_{flesh,i} \cdot \overline{L_{flesh}}(f) \tag{6}$$

Therefore, PARAFAC was applied to decompose the spatially resolved spectra into $\alpha$, $\mu_t$ and $\overline{L}$ for each layer as sample score, wavelength and fiber loadings, respectively. First PARAFAC was applied to the whole data set with changing the number of factors to investigate appropriateness of the application of PARAFAC to spatially resolved spectra. The core consistency diagnostics (CORCONDIA) was utilized to determine the optimal number of factors with which the core consistency of the PARAFAC model was close to 100% while increasing the number of factors resulted in significant drop of that value (*8*). Then PARAFAC was applied to individual sample data sets to build 120 PARAFAC models. Since $\alpha$, $\mu_t$ and $\overline{L}$ are supposed to be positive, non-negativity constraint for each mode was applied in both cases.

## 3. Results and Discussion

### *3.1. Spatially Resolved Spectra*

Mean spatially resolved spectra of all samples are shown in figure 2. Clear peak is observed around 680 nm, which is close to chlorophyll absorption band, in spectra of all fibers. This could be caused by chlorophyll concentrated in apple skin (*9, 10*). On the other hand, the peak around 980 nm, which is close to water absorption band, becomes more prominent for further source-detector distance, which indicates light reaches deeper in the sample and thus absorbed more by water in flesh.

The data below 613 nm and above 1056 nm were removed because of high amount of noise and data at 599 wavelengths remained. Some of the values in the data for fiber 1 showed negative value, which would not be suitable for non-negativity constraint on PARAFAC. Also, unexpected positive and negative peaks were observed around 950 nm in the spectra of fiber 2. Therefore, data for these two fibers were removed.

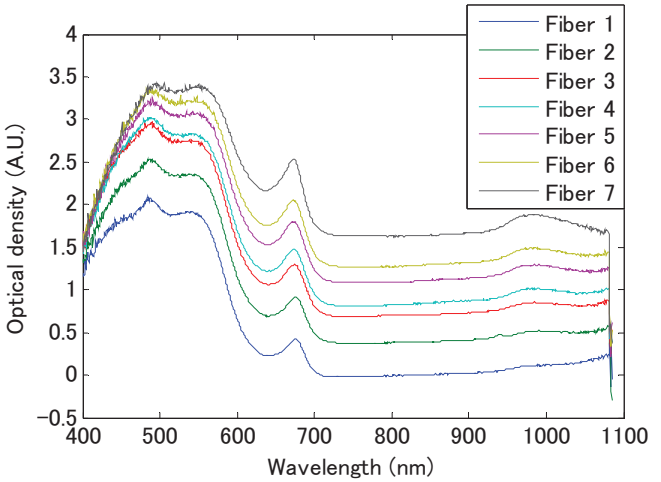


Fig. 2. Mean spatially resolved spectra.

### *3.2. PARAFAC on the Whole Data Set*

The core consistency of the PARAFAC model on the whole data set with 2 factors was 98.2%, while that with 3 factors was -77.4%. Therefore, the optimal number of factors was considered to be 2.

Figure 3 shows the wavelength loadings acquired by PARAFAC on the whole data set. Small peaks are observed in the factor 1 loading around absorption bands for chlorophyll and water. As for the factor 2 loading, there is a prominent peak around chlorophyll absorption band and the loading value increases with decreased wavelength below 650 nm, which indicates strong absorption and/ or scattering in the visible range. Apple skin contains chlorophyll and anthocyanin with light absorption around 678 nm and in the range of 400-625 nm, respectively (*10*). Therefore, it could be considered that the light absorption in the apple skin was modelled as the factor 2 loading.

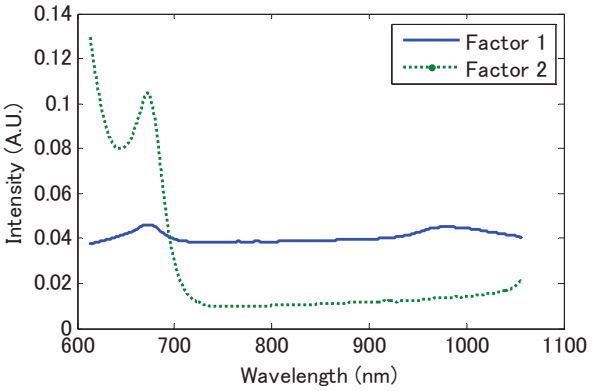


Fig. 3. Wavelength loadings

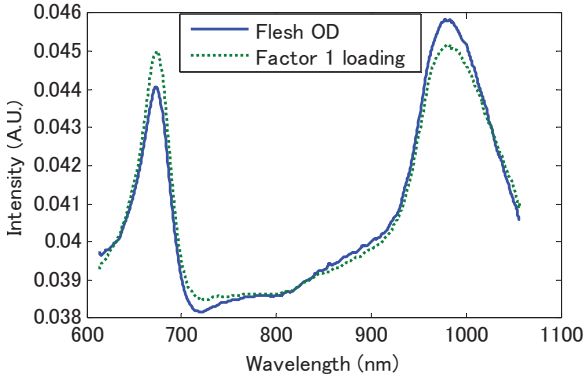


Fig. 4. Spatially resolved spectra of flesh and factor 1 wavelength loading

Normalized spatially resolved spectra of flesh without skin and the factor 1 loading were almost identical with the correlation coefficient of 0.97 as shown in figure 4, which indicates light absorption in flesh was modelled as the factor 1 loading.

A core consistency above 90% indicates that the data set is 'very trilinear' and thus PARAFAC modelling is appropriate (*8*), thus it can be concluded that spatially resolved spectra was successfully decomposed into that of skin and flesh as two wavelength loadings by PARAFAC. On the other hand, absorption and scattering coefficients could not be separated because PARARFAC models with 3 or more factors found to be not valid by CORCONDIA.

### *3.3. PARAFAC on the Individual Data Sets*

Out of 120 acquired PARAFAC models, 62 of them showed core-consistency lower than 50%, which indicates the existence of variation cannot be handled by PARAFAC and invalidity of the models (*8*). The rest with core-consistency higher than 50% were used for further analyses.

When the loadings of the models were averaged, the shape of the factor 1 wavelength loading was almost identical to that of the factor 2 wavelength loading of the PARAFAC model for the whole data set. Therefore, spatially resolved spectra of the skin were considered to be modelled as the factor 1 loading in the case of individual data sets.

Mean factor 2 wavelength loading had poor correlation coefficient of 0.51 to normalized spatially resolved spectra of flesh without skin, which indicated that spatially resolved spectra of flesh could not be appropriately modelled by PARAFAC. This could be explained by the small size of the data sets with only 7 replications which leads to instable PARAFAC models and failure in separation of factors (*11*).

Figure 5 shows the mean factor 1 loading for data sets of the most green and red part of the sample. The loading for red skin samples was higher than that of green skin samples in the wavelength range below 650 nm, which indicates that their color difference was successfully modelled. On the other hand, the loading for green skin samples was higher than that of red skin samples in the chlorophyll absorption band. This result is in good agreement with the study by Knee (*9*) which showed that apples with greener skin contained more chlorophyll in their skin.

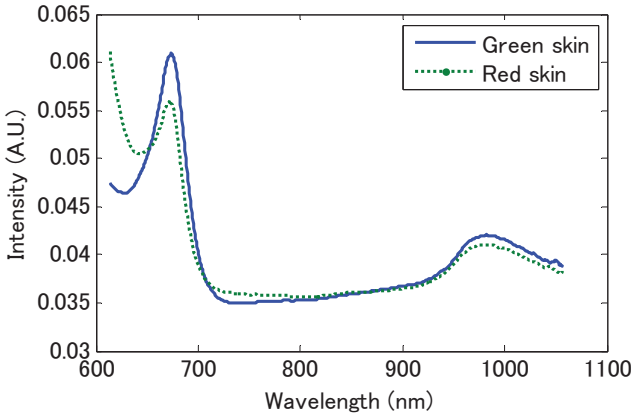


Fig. 5. Mean factor 1 loading of green and red skin data sets

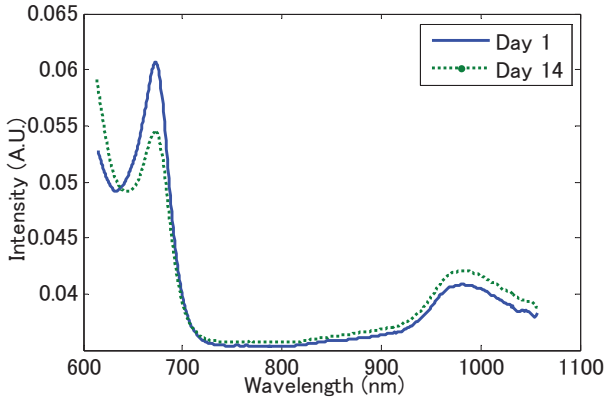


Fig. 6. Mean factor 1 loading of data sets of before and after shelf-life storage

Figure 6 shows the mean factor 1 loading for data sets acquired by measuring before and after the 14 days of shelf-life storage. Decrease in the loading value in the chlorophyll absorption band after the shelf-life storage was observed, which indicates the degradation of chlorophyll in the skin during ripening and storage (*10*).]

## 4. Conclusion

Spatially resolved spectra of intact apples were successfully decomposed into that of skin and flesh by applying PARAFAC on the whole data set. PARAFAC on individual data sets showed that difference in skin color and chlorophyll content as well as degradation of chlorophyll in skin during ripening and storage could be modelled as the change in factor 1 wavelength loading. These results show the potential of SRS combined with PARAFAC to decompose the spectra of multi-layered biological materials.

## Acknowledgements

Mizuki Tsuta would like to thank the Japanese Science Foundation (JSPS) for the mobility grant to stay in K.U. Leuven as a visiting postdoc. Nghia Nguyen Do Trong is a PhD student funded by the Interfaculty Council for Development Cooperation, K.U. Leuven (IRO scholarship).

## References

[1] Pasquini, C., Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society* **2003,** *14*, 198-219.

[2] Karoui, R.; De Baerdemaeker, J., A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. *Food Chemistry* **2007,** *102*, 621-640.

[3] Qin, J.; Lu, R.; Peng, Y., PREDICTION OF APPLE INTERNAL QUALITY USING SPECTRAL ABSORPTION AND SCATTERING PROPERTIES. *Transactions of the Asabe* **2009,** *52*, 499-507.

[4] Farrell, T. J.; Patterson, M. S.; Wilson, B., A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo. *Medical Physics* **1992,** *19*, 879-888.

[5] Tuchin, V. V., Continuous Wave and Time-Resolved Spectrometry In *Tissue optics*, Instrumentation, S. o. P.-O., Ed. SPIE Publications: Bellingham, 2007; pp 451-488.

[6] Bro, R., PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems* **1997,** *38*, 149-171.

[7] Andersen, C. M.; Bro, R., Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *Journal of Chemometrics* **2003,** *17*, 200-215.

[8] Bro, R., A new efficient method for determining the number of components in PARAFAC models. *Journal of chemometrics* **2003,** *17*, 274.

[9] Knee, M., Methods of measuring green colour and chlorophyll content of apple fruit. *International Journal of Food Science & Technology* **1980,** *15*, 493-500.

[10] Merzlyak, M. N.; Solovchenko, A. E.; Gitelson, A. A., Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. *Postharvest Biology and Technology* **2003,** *27*, 197-211.

[11] Christensen, J. H.; Tomasi, G.; Strand, J.; Andersen, O., PARAFAC Modeling of Fluorescence Excitation−Emission Spectra of Fish Bile for Rapid En Route Screening of PAC Exposure. *Environmental Science & Technology* **2009,** *43*, 4439-4445.

Presented at ICEF11 (May 22-26, 2011 – Athens, Greece) as paper EPF223.