

UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Scienze Agrarie e Alimentari

Dipartimento di Scienze per gli Alimenti, la Nutrizione e l'Ambiente



Scuola di Dottorato in

Innovazione Tecnologica per le Scienze Agro-Alimentari e Ambientali

Ciclo XXVIII

***SPECTROSCOPY, IMAGE ANALYSIS AND
HYPERSPECTRAL IMAGING
FOR FOOD SAFETY AND QUALITY:
A CHEMOMETRIC APPROACH***

Settore disciplinare: AGR 15

Tutor: ***Professor Ernestina CASIRAGHI***

Co-tutor: ***Professor Riccardo GUIDETTI***

PhD coordinator: ***Prof. Roberto PRETOLANI***

PhD thesis of:

Cristina MALEGORI – R10153

Anno Accademico 2014 – 2015

A Giorgia e Roberto

INDEX:

a. ABSTRACT	7
b. RIASSUNTO	9
c. PREFACE	11
1. SPECTROSCOPY	13
1.1. Testing of a Vis-NIR system for the monitoring of long-term apple storage.....	15
1.2. Wavelength selection with a view to a simplified handheld optical system to estimate grape ripeness.....	29
1.3. Selection of optimal wavelengths for decay detection in fresh-cut <i>Valerianella Locusta Laterr</i>	43
1.4. Comparison between FT-NIR and Micro-NIR in the evaluation of Acerola fruit quality, using PLS and SVM regression algorithms.....	55
2. IMAGING	61
2.1. Image texture analysis, a non-conventional technique for early detection of biofilm.....	63
3. HYPERSPECTRAL	77
3.1. Hyperspectral image analysis: a tutorial.....	79
3.2. HIS for quality evaluation of vitamin C content in acerola fruit.....	93
3.3. Selection of NIR wavelengths from hyperspectral imaging data for the quality evaluation of acerola fruit.....	101
4. OVERALL CONCLUSIONS	107
d. COMMUNICATIONS AND POSTERS	109
8.1 Conference participation	
8.2 Newspapers article	
e. COURSES AND WORKSHOPS	111

a. ABSTRACT:

Spectroscopy, image analysis and hyperspectral imaging for food safety and quality: a chemometric approach

This PhD project regards different applications of non-destructive optical techniques to evaluate quality and shelf life of agro-food product as well as the early detection of biofilm on food plants. Spectroscopy, image analysis and hyperspectral imaging could play an important role in the assessment of both quality and safety of foods due to their rapidity and sensitivity especially when using simplified portable devices. Due to the huge amount of collected data, chemometric, a multivariate statistical approach, is required, in order to extract information from the acquired signals, reducing dimensionality of the data while retaining the most useful spectral information.

The thesis is organized in four chapters, one for each technique and a final chapter including the overall conclusion. Each chapter is divided in case studies according to the matrix analysed and the data acquisition and elaboration carried out.

The first chapter is about spectroscopy. The aim of the first study - *Testing of a Vis-NIR system for the monitoring of long-term apple storage* - is to evaluate the applicability of visible and near-infrared (Vis-NIR) spectroscopy to monitor and manage apples during long-term storage in a cold room. The evolution of the apple classes, originally created, was analysed during 7 months of storage by monitoring TSS and firmness. Vis-NIR allows an accurate estimation of chemical-physical parameters of apples allowing a non-destructive classification of apples in homogeneous lots and a better storage management.

The work reported in the second paragraph - *Wavelength selection with a view to a simplified handheld optical system to estimate grape ripeness* - is aimed to identify the three most significant wavelengths able to discriminate grapes ready to be harvested directly in the field. Wavelengths selection was carried out with a view to construct a simplified handheld and low-cost optical device. Standardized regression coefficients of the PLS model were used to select the relevant variables, representing the most useful information of the full spectral region. The same approach was followed to discriminate freshness levels during shelf-life of fresh-cut *Valerianella* leaves - *Selection of optimal wavelengths for decay detection in fresh-cut Valerianella Locusta Laterr.* (third paragraph).

The aim of the work presented in the fourth paragraph of the first chapter - *Comparison between FT-NIR and Micro-NIR in the evaluation of Acerola fruit quality, using PLS and SVM regression algorithms* - is to estimate titratable acidity and ascorbic acid content in acerola fruit, using a MicroNIR, an ultra-compact and low-cost device working between 950 – 1650 nm. The spectral data were modelled using two different regression

algorithms, PLS (partial least square) and SVM (support vector machine). The prediction ability of Micro-NIR appears to be suitable for on field monitoring using non-linear regression modelling (i.e. SVM).

In the second chapter, image analysis was performed. The traditional RGB imaging for the evaluation of image texture, a specific surface characteristic, is presented. The texture of an image is given by differences in the spatial distribution, in the frequency and in the intensity of the values of the grey levels of each pixel of the image. This technique was applied for the early detection of biofilm in its early stages of development, when it is still difficult to observe it by the naked eye, was evaluated (*Image texture analysis, a non-conventional technique for early detection of biofilm*).

In the third paragraph, image and spectroscopy were combined in hyperspectral imaging applications. Data analysis by chemometric was crucial in any stage of my PhD project. Chemometric is a multivariate statistical approach that is applied on chemical data to extract the useful information avoiding noise and redundant data. At the beginning of the third chapter - *Hyperspectral image analysis: a tutorial* - proposes an original approach, developed as a flow sheet for three-dimensional data elaboration. The method was applied, as an example, to the prediction of bread staling during storage.

The first application about hyperspectral on acerola is focused on the vitamin C content - *HIS for quality evaluation of vitamin C content in Acerola fruit*. Ten different acerola fruits picked up according to two different stages of maturity, based on the colour of the peel (5 green and 5 red acerola), were analysed. The spectra of pure vitamin C powder was used as references for computing models with two different correlation techniques: spectral angle mapping and correlation coefficient allowing the construction of a qualitative distribution map of ascorbic acid inside the fruit.

The aim of the last one work presented is to evaluate acerola post-harvest quality - *Selection of NIR wavelengths from hyperspectral imaging data for the quality evaluation of Acerola fruit*. Hyperspectral images of 20 acerolas were acquired for five consecutive days and an investigation of time trends was carried out to highlight the most important three wavelengths that characterized the ripeness/degradation process of the Acerola fruit. The false-colour RGB images, derived from the composition of the three interesting wavelengths selected, data enable early detection of the senescence process in a rapid and non-destructive manner.

In conclusion, the three non-destructive optical techniques applied in this PhD project have proved to be one of the most efficient and advanced tools for safety and quality evaluation in food industry answering the need for accurate, fast and objective food inspection methods to ensure safe production throughout the entire production process. **Keywords: spectroscopy, imaging, hyperspectral, chemometric, food**

b. RIASSUNTO:

La spettroscopia, l'analisi dell'immagine e di immagine iperspettrale per la sicurezza e la qualità degli alimenti: un approccio chemiometrico

Questo progetto di dottorato studia le differenti applicazioni delle tecniche ottiche non distruttive per la valutazione della qualità e della shelf-life di prodotti vegetali così come l'identificazione precoce di sviluppi microbici su superfici industriali. La spettroscopia, l'analisi dell'immagine e l'analisi dell'immagine iperspettrale possono giocare un ruolo importante nella valutazione sia della qualità che della sicurezza degli alimenti grazie alla rapidità e sensibilità della tecnica, specialmente quando si utilizzano strumenti semplificati portatili. Un approccio statistico multivariato (chemiometria) è richiesto al fine di estrarre informazioni dal segnale acquisito, riducendo la dimensionalità dei dati e mantenendo le informazioni spettrali più utili.

Lo scopo del primo studio presentato – *Testing of a Vis-NIR system for the monitoring of long-term apple storage* – è la valutazione dell'applicabilità della spettroscopia nel visibile e vicino infrarosso (Vis-NIR) per il monitoraggio e la gestione delle mele durante lo stoccaggio a basse temperature. Per sette mesi è stata seguita l'evoluzione in termini di grado zuccherino e consistenza delle mele suddivise in classi di maturazione. I risultati hanno indicato che la spettroscopia è una tecnica non-distruttiva che consente una stima accurata dei parametri chimico-fisici per la classificazione delle mele in lotti omogenei.

Il lavoro descritto nel secondo paragrafo - *Wavelength selection with a view to a simplified handheld optical system to estimate grape ripeness* – è finalizzato all'identificazione delle tre lunghezze d'onda più importanti per il riconoscimento, direttamente in campo, dell'uva pronta per essere raccolta al fine della messa a punto di un sistema semplificato e a basso costo. I coefficienti di regressione standardizzati del modello PLS (Partial Least Square) sono stati utilizzati per selezionare le variabili più importanti, che racchiudono l'informazione più utile lungo l'intero spettro. La stessa procedura è stata condotta per determinare la freschezza delle foglie di Valerianella durante la shelf-life - *Selection of optimal wavelengths for decay detection in fresh-cut Valerianella Locusta laterr* (terzo paragrafo).

Lo scopo del lavoro presentato nel quarto paragrafo del primo capitolo - *Comparison between FT-NIR and Micro-NIR in the evaluation of Acerola fruit quality, using PLS and SVM regression algorithms* – è stimare l'acidità titolabile e il contenuto di acido ascorbico all'interno del frutto acerola, utilizzando uno strumento compatto e a basso costo denominato Micro-NIR, che lavora nell'intervallo spettrale 950-1650 nm. I dati spettrali sono stati modellati mediante l'applicazione di due algoritmi PLS e SVM (Support Vector Machine). La capacità predittiva dello strumento semplificato è risultata

interessante per applicazioni di monitoraggio in campo, soprattutto modellizzando i dati in modo non lineare.

Nel secondo capitolo, è presentata l'applicazione di immagini RGB per la valutazione delle superfici - *Image texture analysis, a non-conventional technique for early detection of biofilm*. La texture dell'immagine è definita come una differenza nella distribuzione spaziale, nella frequenza e nell'intensità dei livelli di grigio in ogni pixel dell'immagine. Questo metodo è stato determinante per l'identificazione precoce dello sviluppo microbico su superfici normalmente impiegate nell'industria alimentare.

L'approccio chemiometrico è stato cruciale in ogni fase del progetto di dottorato ed è definito come un approccio statistico multivariato che si applica ai dati chimici per estrarre informazione utile, ridurre il rumore di fondo e l'informazione ridondante. Il lavoro presentato all'inizio del terzo capitolo - *Hyperspectral image analysis: a tutorial* - propone una procedura standard per l'elaborazione di dati tridimensionali, presentando un esempio relativo alla predizione del raffermaimento del pane in cassetta.

Il secondo paragrafo del terzo capitolo, presenta una applicazione dell'immagine iperspettrale su acerola, focalizzata sul contenuto di vitamina C - *HSI for quality evaluation of vitamin C content in Acerola fruit*. In questo lavoro, è stata acquisita l'immagine di dieci acerola, raccolte in funzione del livello di maturazione, definito in base al colore della buccia (cinque acerola verdi e cinque rosse). Lo spettro della polvere di vitamina C pura è stato utilizzato come riferimento per l'applicazione di due algoritmi di correlazione (spectral angle mapping e correlation coefficient), consentendo la costruzione di mappe qualitative di distribuzione dell'acido ascorbico all'interno del frutto.

Lo scopo dell'ultimo lavoro presentato è la valutazione della qualità post raccolta dell'acerola - *Selection of NIR wavelengths from hyperspectral imaging data for the quality evaluation of Acerola fruit*. Le immagini iperspettrali di venti acerola sono state acquisite per cinque giorni consecutivi. La valutazione delle modificazioni spettrali durante il tempo ha consentito la selezione delle tre lunghezze d'onda caratterizzanti il processo di maturazione/degradazione del frutto. L'immagine in falsi colori, derivante dalla composizione delle immagini alle tre lunghezze d'onda di interesse, consente l'identificazione precoce del processo degradativo in maniera rapida e non distruttiva.

Le tre tecniche non distruttive impiegate in questo progetto di dottorato hanno dimostrato efficienza e applicabilità per la valutazione della qualità e della sicurezza degli alimenti, rispondendo alla necessità dell'industria alimentare di tecniche accurate, veloci e obiettive per assicurare produzioni ottimali lungo l'intero processo produttivo.

Parole chiave: spettroscopia, analisi d'immagine, iperspettrale, chemiometria, cibo

c. PREFACE

During my PhD, I focused on three non-destructive optical techniques, starting from spectroscopy and ending up with to the hyperspectral imaging, passing through the image analysis. In fact, in spectroscopy, the information in the spectral range 4000-12500cm⁻¹ is correlated to the chemical composition of the matrixes while image analysis is centred on spatial information and allows to know compound distribution in the product. The hyperspectral imaging, which I applied during my experience in Brazil, combine the different approaches with the aim to evaluate food safety and quality. The keystone of these optical techniques is the use of chemometric, a multivariate statistical approach that is applied on chemical data to extract the useful information avoiding noise and redundant data.

The two general subjects of my work are: 1) food quality, intended as ripeness and shelf-life of agro-food product; 2) plant hygiene, as detection of biofilm on food plants.

1) The development of effective quality inspection systems to ensure safe production of food during processing operations is one of the critical aspects for food processing industry and could be solved by the improvement of non-destructive technologies. In this contest, the quality of a particular fresh or processed fruit or vegetable is defined by a series of external characteristics that make it more or less attractive to the consumer. Quality attribute of the post-harvest product included ripeness, size, weight, shape, colour, condition, or presence/absence of defects, stems or seeds, as well as a series of internal properties like sweetness, acidity, texture, hardness, among others. Most of these factors have traditionally been assessed by visual inspection or destructive sampling performed by trained operators.

2) Food safety is normally described as a discipline aiming to ensure that food is safe enough “from-farm-to-fork” for consumers so that outbreaks of food-borne illness can be reduced. The ability of forming biofilms by food microorganisms increases bacterial resistance to antimicrobial agents along the production chain. Biofilm presence in the food industry is frequent and it is a serious problem, as it causes malfunction of plant and microbial contamination of foods. The bacteria involved may deteriorate and impair the food and, if pathogenic, the microorganisms pose a risk to the consumer health. It is therefore important to prevent or, alternatively, recognise biofilm formation. Nowadays, the gold standard for cultivation and detection of biofilms is the combination of flow-cell-grown technique and confocal laser scanning microscopy. This optical technique is promising but requires the specimen to be fluorescent. Therefore, the scientific community is looking for new optical approaches to be implemented on-line for the early detection of food microorganisms or contaminants.

The thesis is organized in four chapters, one for each techniques and a final chapter including the overall conclusion, a comprehensive consideration about the work carried out during my PhD.

Each chapter is divided in case studies according to the matrix analysed and the data elaboration carried out. The first chapter, "Spectroscopy", presents four works about the application of near infrared spectroscopy in the post-harvest field, starting with an application on apples for a better management of the warehouse. This work was already published as scientific article in the "Food and Bioprocess Technology". The second and the third paragraphs are focused on the selection of the most important NIR wavelengths for the development of a simplified device based on LED technology. One application is on wine grapes and was published as scientific article on the "American Journal of Enology and Viticulture"; the second one speaks about Valerianella leaves and was published as scientific article in the "Journal of Food Engineering". In the last paragraph, a simplified device available on the market was used, in comparison with a benchtop FT-NIR, on acerola fruit to evaluate the estimation ability of titratable acidity and ascorbic acid content. This work was presented as poster in the 17th International Conference of NIR Spectroscopy and it is reported in this thesis as extended abstract.

The second chapter, "Imaging", regards the application of traditional RGB imaging for the early detection of biofilm. In this work, that was presented as oral communication at the 2nd Conference on Optical Characterization of Materials and is now in revision as scientific article for the publication in the "Journal of Food Engineering", is proposed an innovative elaboration approach allowing the discrimination of microbial development on different surfaces.

The chapter "Hyperspectral" includes three case studies. The first one underlines the importance of the chemometric approach for the data elaboration. In this paragraph, a flow sheet for three-dimensional data elaboration is proposed, carrying out an example for the prediction of bread staling during storage. This work is submitted to the "Analytical Methods" as scientific article. The other two paragraphs present two applications of hyperspectral on acerola fruit: the first one for the evaluation of vitamin C content, developing two possible mapping of the chemical compound inside the fruit images; the last one is focused on the selection of the three most important wavelengths for monitoring the ripeness/degradation process of this tropical fruit. Both the application were presented in the 17th International Conference of NIR Spectroscopy: the first one on vitamin C as poster, while the second one about ripeness as Oral Communication; in this thesis are presented as extended abstract.

The overall conclusion allows a comprehensive evaluation of the project carried out during the PhD and puts some bases for possible future evolution of the work.

The thesis ends with some details about other activities such as the conference participations and the courses attended during the three years.

1. SPECTROSCOPY

Near infrared (NIR) spectroscopy has proved to be one of the most efficient and advanced tools for monitoring and controlling of process and product quality in food industry. Multivariate statistical techniques are required to extract the information about quality attributes, which is buried in the NIR spectrum. Currently available NIR spectroscopy devices are expensive and not suitable for small-scale food industry. In post-harvest field, simple devices are important to support growers in monitoring ripening progress of various fruits and to plan harvest time; in this context, the availability of low cost miniaturised spectrophotometers could opened up the possibility of portable devices, which can be used directly on field (Costa et al. 2011).

Nowadays, an important challenge in postharvest technology is non-destructive characterization of fruit in order to maximize the quality and reduce waste (Zude et al. 2006, Jha et al. 2010, Bobelyn et al. 2010, Beghi et al. 2012). For the correct management of lots, it is necessary to consider fruit heterogeneity before assessing their prolonged storage. Moreover, performing fruit classification during conferment can allow optimization of the cell opening sequence to reduce waste caused by over-ripening during long-term storage and provides the market with the best available product all year round. In light of these considerations, the aim of the first study presented in this chapter is to evaluate the applicability of visible and near-infrared (Vis-NIR) spectroscopy to monitor and manage apples during long-term storage in a cold room. The evolution of the originally created classes was analysed during 7 months of storage by monitoring TSS and firmness (peak force and penetration energy). The results indicate that the Vis-NIR allows an accurate estimation of chemical-physical parameters for non-destructive classification of apples in homogeneous lots.

In recent years, interest in non-destructive technology has shifted toward the development of portable vis/NIR systems for use in the field, aiming a simplification of the device and a cost saving. However, the spectroscopic approach rely on wide spectral ranges and thus require multivariate techniques for data processing (Cen and He 2007, Williams and Norris 2001). Chemometric can be used to select a few relevant variables that represent the most useful information of the full spectral region (Xiaobo et al. 2010), reducing the size of the required measurement data and the costs for the light source. For these reasons, the work in the second paragraph is aimed to identify the three most significant wavelengths able to discriminate in the field grapes ready to be harvested with a view to a simplified handheld and low-cost optical device. Standardized regression coefficients of the PLS model were used to select the relevant variables, representing the most useful information of the full spectral region. The same work was carried out to discriminate freshness levels during shelf life of fresh-cut *Valerianella Locusta L.* and it is presented in the third paragraph. The shelf life of *Valerianella* leaves

was monitored using a portable commercial vis/NIR spectrophotometer and by traditional analyses (pH, moisture and total phenols content).

Within the last few years, real handheld near-infrared scanning spectrometers become commercially available (Sorak, 2012). A survey of scientific papers published in the last decade shows a steady increase in the number of research and development studies being conducted using these types of portable spectrometers (Teixeira dos Santos, 2013). Accordingly, the aim of the work presented in the fourth paragraph, is to estimate titratable acidity and ascorbic acid content in acerola fruit, using a MicroNIR, an ultra-compact and low-cost device working between 950 – 1650 nm. The spectral data were modelled using two different regression algorithms, PLS (partial least square) and SVM (support vector machine). The prediction ability of Micro-NIR appears to be suitable for on field monitoring using non-linear regression modelling is required.

- Beghi, R., Spinardi, A., Bodria, L., Mignani, I., & Guidetti, R. (2012). Apples nutraceutical properties evaluation through a visible and nearinfrared portable system. *Food and Bioprocess technology*. doi:10.1007/s11947-012-0824-7.
- Bobelyn, E., Serban, A. S., Nicu, M., Lammertyn, J., Nicolai, B. M., & Saeys, W. (2010). Postharvest quality of apple predicted by NIRspectroscopy: study of the effect of biological variability on spectra and model performance. *Postharvest Biology and Technology*, 55, 133–143.
- Cen, H., and Y. He. 2007. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.* 18:72-83.
- Costa G., E. Bonora, G. Fiori, M. Noferini Innovative non-destructive device for fruit quality assessment *Acta Horticulturae*, 913 (2011), pp. 575–581.
- Jha, S. N., & Ruchi, G. (2010). Non-destructive prediction of quality of intact apple using near infrared spectroscopy. *Journal of Food Science and Technology*, 47(2), 207–213.
- Sorak, D., Herberholz, L., Iwascek, S., Altinpinar, S., Pfeifer, F., & Siesler, H. W. (2012). New developments and applications of handheld Raman, mid-infrared, and near-infrared spectrometers. *Applied Spectroscopy Reviews*, 47(2), 83-115.
- Teixeira dos Santos, C. A. T., Lopo, M., Páscoa, R. N., & Lopes, J. A. (2013). A review on the applications of portable near-infrared spectrometers in the agro-food industry. *Applied spectroscopy*, 67(11), 1215-1233.
- Williams, P., and K. Norris (eds.). 2001. *Near-Infrared Technology in the Agricultural and Food Industries*. 2d ed. Am. Association of Cereal Chemists, St. Paul, MN.
- Xiaobo, Z., Z. Jiewen, M.J.W. Povey, M. Holmes, and M. Hanpin. 2010. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* 667:14-32.
- Zude, M., Herold, B., Roger, J. M., Bellon-Maurel, V., & Landahl, S. (2006). Non-destructive tests on the prediction of apple fruit flesh firmness and soluble solids content on tree and in shelf life. *Journal of Food Engineering*, 77(2), 254–260.

1.1 TESTING OF A VIS-NIR SYSTEM FOR THE MONITORING OF LONG-TERM APPLE STORAGE

Roberto Beghi^{1*}, Gabriella Giovanelli², Cristina Malegori², Valentina Giovenzana¹, Riccardo Guidetti¹

¹ Department of Agricultural and Environmental Sciences – Production, Landscape, Agroenergy, Università degli Studi di Milano, via Celoria 2, 20133 Milan, Italy

² Department of Food, Environmental and Nutritional Sciences, Università degli Studi di Milano, Via Celoria 2, 20133 Milan, Italy

Abstract

The development of diseases during long-term storage of apples is a well-known issue causing loss of product for warehouses. Non-destructive characterization of fruit can be helpful in order to reduce waste and maximize apple quality. The aim of this study was to evaluate the applicability of visible and near-infrared (VIS-NIR) spectroscopy to monitor and manage many apples during long-term storage in a cold room. A bench-top VIS-NIR apparatus (600–1200 nm) was used to classify apples from two different cultivars, Golden Delicious and Red Delicious, based on their total soluble solids content (TSS). The evolution of the originally created classes was analysed during 7 months of storage by monitoring TSS and firmness (peak force and penetration energy), and the estimation ability of the VISNIR device was evaluated. The results indicate that the spectroscopic technique allows for an accurate estimation of chemical-physical parameters for non-destructive classification of apples in homogeneous lots. Regarding the estimation ability of the compact VIS-NIR spectrophotometer, the results show good prediction ability both for total soluble solids content and firmness indices. The use of the instrument for on-line selection and classification of fruits is therefore desirable. This can lead to better management of postharvest storage and the destination of lots, with a consequent reduction in fruit wastage. This approach is important to plan the opening sequence of storage rooms during the winter season, providing the market with the best available products all year round.

Keywords: Apple VIS-NIR spectroscopy Postharvest Storage Classification

Introduction

The technologies developed for the preservation of apples (low temperature, controlled humidity, and controlled atmosphere) allow long-term storage of fruits, fully maintaining their physical, chemical, and sensorial characteristics (Beni et al. 2001). External quality is considered of primary importance in the marketing and sale of fruits. The appearance, i.e., size, shape, colour, and presence of blemishes, influences consumer perceptions and therefore determines the level of acceptability prior to purchase. The consumer also associates desirable internal quality characteristics with a certain external appearance. This learned association of internal quality to external quality affects future purchases (Wills et al. 1998; Brosnan and Sun 2004). Disorders generally develop in more mature fruit; the main contributing factors are high or low air temperatures in the pre-harvest period, poor calcium nutrition, maturity status at harvest, and cropping level (Ferguson et al. 1999; Woolf and Ferguson 2000). Studies concerning the development of diseases during cold storage are numerous in the literature (Snowdon 2010; Johnston et al. 2002). Today, an important challenge in postharvest technology is non-destructive characterization of fruit in order to maximize the quality and reduce waste. The availability on the market of innovative technologies based on near-infrared (NIR) and visible and near-infrared (VIS-NIR) spectroscopies opened up new ways for fruit testing, which are not encumbered by any of the inconveniences of a traditional analysis (Peirs et al. 2001; Cen and He 2007). Novel non-destructive and rapid tools are used both for predicting the optimum harvest time and for quick monitoring of fruit quality during the postharvest period. The use of spectrometers for the evaluation of product quality is simple, fast, non-destructive, and therefore applicable to a great number of samples. Interesting applications have been found since the beginning of the 1990s (Frank and Todeschini 1994; Massart et al. 1997; Massart et al. 1998; Basilevsk 1994; Jackson 1991). The food sector has shown interest in NIR and VIS-NIR instruments, both mobile and on-line; devices based on diode array spectrophotometers and FT-NIR desk systems have been developed and successfully applied in this field (Guidetti et al. 2012). Apple has been widely involved in the experimentation of non-destructive testing methods, in order to determine a number of sensory and nutritional parameters. Over the past decade, VISNIR has been successfully used to measure a range of apple quality attributes such as soluble solids content (Quilitzsch and Hoberg 2003; Zude et al. 2006), titratable acidity (Peirs et al. 2002), starch index (Menesatti et al. 2009), chlorophyll content (Zude-Sasse et al. 2002), and volatile compounds with the combined use of an electronic nose (Saevels et al. 2004). The results achieved have encouraged the inclusion of other specific applications related to the prediction of fruit firmness (Mehinagic et al. 2003; McClure 2003; Nicolai et al. 2007; Lu et al. 2000) and non-destructive analysis of nutraceutical properties (Beghi et al. 2012). For the correct management of lots, it is necessary to consider fruit heterogeneity before assessing their prolonged storage. Performing fruit classification during conferment can allow optimization of the cell opening sequence to reduce waste

caused by over-ripening during long-term storage. Disorders such as browning and loss of texture, in fact, make the fruit unmarketable, especially in the last phases of the conservation period, with a consequent economic loss for warehouses. Optimization in the management of storage cells can lead to a reduction in waste and provides the market with the best available product all year round. The aim of this work was to study the ability of a commercial VIS-NIR instrument to classify Golden Delicious and Red Delicious apples in homogeneous lots. Apple storage life analysis was carried out by splitting fruits into two classes based on total soluble solids content (a parameter traditionally used for the characterization of fruit). The evolution of created classes was analyzed during long-term storage by monitoring total soluble solids content (TSS) and firmness (peak force and penetration energy), and the estimation ability of the VIS-NIR device was evaluated. Storage centers will achieve the full potential of an already consolidated preservation technique, such as a modified atmosphere, by using this method.

Materials and Methods

Sampling

The research was carried out in 2010–2011 in Ponte in Valtellina (Sondrio, northern Italy) on the two most representative local apple varieties: Golden Delicious and Red Delicious, which received a Protected Geographical Indication (PGI) in 2001. The apples were picked at the commercial maturity stage. Data collected in the vintage year 2009–2010 were used to obtain prediction models for TSS, penetration energy, and peak force evaluation (Giovannelli et al. 2011). Using the TSS models previously developed which are the ones with the best prediction ability, a total of 280 apples (140 for each variety) were divided in the two classes, both representative of the 2010–2011 vintage: “lower TSS” and “higher TSS”. In order to define these classes, the TSS values taken as the threshold were obtained in a destructive way by calculating the average refractometric data collected specifically for this purpose from 40 Golden Delicious and 40 Red Delicious apples. The 280 classified fruits were stored for 7 months, from December 2010 to June 2011, and their evolution during storage was followed. Conditions generally used at the warehouse were different for the two cultivars; in particular, the values of O₂, CO₂, relative humidity, and temperature were 1 %, 2.5–3 %, 94–99 %, and 1–1.5 °C for Golden Delicious and 1 %, 2 %, 89–90 %, and 0.5–1 °C for Red Delicious apples, respectively. For both varieties, storage conditions included the use of SmartFresh technology with 1- methylcyclopropene (Watkins 2006). For cold storage life monitoring, a portion of the classified apples belonging to the different classes was analyzed monthly. At each of the seven sampling dates, ten apples belonging to the higher TSS class and ten belonging to the lower TSS class for each cultivar were analyzed by the bench-top VISNIR system; then, the TSS and firmness were determined on the same samples by a refractometer and texture analyzer. The acquisitions were

performed on whole apples at two opposite points along the equator region, for a total of 560 spectra. Destructive determinations were subsequently carried out. Each half apple was analyzed separately, both by spectroscopic and destructive analysis, in order to limit the influence of the variability of different parts of the same fruit. The analytical and spectral data were processed to evaluate VIS-NIR device prediction capabilities to obtain classification models, to be Food Bioprocess Technol (2014) 7:2134–2143 2135 able to follow the evolution of apples during cold storage, and finally to build new regression models based on samples from 2010–2011 for the prediction of physical and chemical parameters. Automatic Bench-Top VIS-NIR System An automatic bench-top VIS-NIR system (QS_200) produced by Unitec SpA (Lugo, Ravenna, Italy) that operates in the wavelength range of 600–1200 nm was used for the spectral acquisition. Two acquisitions were carried out on each fruit, along the equator region on opposite sides; in the same areas, the destructive analyses were performed. The system consists of four elements: a lighting system, a fiber optic probe, a portable spectrophotometer, and a PC for data acquisition and instrument control. In this system, the samples were hit by radiation produced by the lighting system, and the reflected component was measured by the spectrophotometer. Light radiation was shone onto the fruit sample through a bidirectional fiber optic probe. The choice of this optical fiber was based on the need to acquire spectra in diffuse reflectance (Huang et al. 2008; Guidetti et al. 2008; McGlone et al. 2002). The measurements were acquired by dedicated software (Spectra software package, UNITEC SpA, Italy).

Physical-Chemical Analysis

TSS were measured using a portable digital refractometer (model DBX-55, Atago, Tokyo, Japan). The determination was made after spectrophotometer readings and the destructive determination of firmness. The two halves, previously analyzed and separated from the rest of the cored apple, were reduced into a puree. A few drops of the latter were placed on the refractometer sensor, previously calibrated with distilled water. The results are directly expressed in degrees Brix (°Brix). The firmness of fruits was determined using a TA.HD plus Texture Analyser dynamometer (TXT, Stable MicroSystem, UK) supported by the software “Exponent”. A force-distance curve was acquired for each sample. Measurements were carried out on each whole apple at two opposite points along the equator region, corresponding to spectra acquisition. Firmness parameters obtained from the elaboration of the force-distance curve were the following:

- peak force (N): the maximum force registered during penetration. This parameter is related to the entire fruit firmness (skin and flesh)
- penetration energy (N mm): the work required to penetrate the apple pulp from 5 to 8 mm. This parameter is related to flesh firmness.

Figure 1 shows as an example a track for the analysis of firmness parameters.

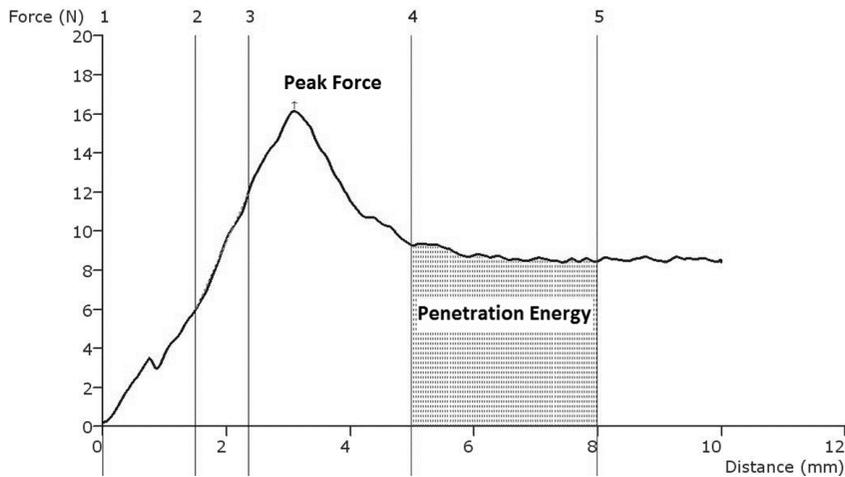


Figure 1: Force-distance curve obtained from the penetration test and

Data Processing

Spectral Data Analysis

Chemometric analyses of 560 spectra collected in 2010–2011 (deriving from 280 apples, 140 apples for each cultivar, two measurements from each fruit) were performed using The Unscrambler software package (version 9.7, CAMO ASA, Oslo, Norway). Different treatments were applied to the VIS-NIR spectra, namely, scatter correction (multiplicative scatter correction (MSC) and standard normal variate (SNV)) and derivatives, before building the calibration models. The first and second derivatives were performed using Savitzky-Golay transformation and smoothing (smoothing window of 15 points and second-order filtering). The resolution of the instrument is 0.3 nm; therefore, a smoothing (moving average) of 15 points correspond to a window of 4.5 nm. Pretreatment parameters were defined, maximizing model accuracy.

VIS-NIR Device Prediction Capabilities

The robustness of the 2009–2010 models was tested using spectral measurements on fruits of 2010–2011 storage season (based on fruits grown under differing conditions) as independent validation set (Magwaza et al. 2012). The differences between the estimated data based on VIS-NIR measurements and the analytical data were calculated, and the limit of acceptability was defined. For TSS, the gap was fixed at 1 °Brix, assuming that a difference in taste is perceptible when apples differed by more than 1 °Brix (P=0.90) (Harker et al. 2002). This difference is equal to 10 % of the average TSS value.

Accordingly, for firmness parameters, a gap of 2 N mm for the penetration energy and 1.5 N for the peak force were identified.

Classification Models

In order to classify the samples by storage time, the VIS-NIR spectra were elaborated by linear discriminant analysis (LDA) (Heberger et al. 2003) using the V-PARVUS package (Forina et al. 2008). The classification technique was carried out after applying the algorithm SELECT to the spectral data as a feature selection technique (Casale et al. 2008). Two different spectral pretreatments were tested (MSC and SNV). Full cross-validation (leave-one-out method) was used, so only one sample at a time was kept out of the calibration. Performance of the LDA classification was evaluated on the basis of predictive ability in the crossvalidation groups.

Regression Models 2010–2011

Regression models for the prediction of physical and chemical parameters were performed to test the accuracy of VIS-NIR models of the 2010– 2011 storage season (based on fruits grown under the same conditions) (Magwaza et al. 2012). The 560 VIS-NIR spectra were correlated with reference indices using the partial least square (PLS) regression algorithm. The PLS method performs particularly well when the various X variables express common information, i.e., when there is a large amount of correlation, or even co-linearity, which is the case for spectral data of intact biological material (Nicolai et al. 2007; Wold et al. 2001). The following statistical parameters were calculated to evaluate model accuracy: the coefficient of determination in calibration (R^2_{cal}), the root mean standard error of calibration (RMSEC), the coefficient of determination in cross-validation (R^2_{cval}), and the root mean standard error of cross-validation (RMSECV) (Sinelli et al. 2008). Percent errors of RMSECV (RMSECV%) were also calculated as equal to RMSECV/ averaged reference values of each parameter. The optimum calibrations were selected based on minimizing the RMSECV. Physical-Chemical Data Analysis In order to highlight statistically significant differences between the samples, one-way ANOVA and the multiple range test were performed using the Statgraphics plus 5.1 package (Graphics Software Systems, Rockville, MD, USA).

Physical-Chemical Data Analysis

In order to highlight statistically significant differences between the samples, one-way ANOVA and the multiple range test were performed using the Statgraphics plus 5.1 package (Graphics Software Systems, Rockville, MD, USA).

Results and Discussion

Traditional Storage Life Analysis

Considering the 40 Golden Delicious apples used for threshold identification, the average refractometric value was 13 °Brix. Therefore, the Golden Delicious apples with a TSS>13 °Brix were classified as “higher TSS”, while the apples with a TSS<13 °Brix were classified as “lower TSS”. Similarly, the average refractometric data for the 40 Red Delicious was 12 °Brix: “higher TSS” for TSS>12 °Brix and “lower TSS” for TSS<12 °Brix. A total of 70 apples from each cultivar, picked in the vintage year 2010–2011, were classified as “lower TSS” and 70 as “higher TSS”. Regarding the evolution of quality indices analyzed during storage, the results showed significant changes in the TSS and firmness parameters at some sampling times.

Table 1: Analytical parameters (mean value±standard deviation, n=10) of Golden Delicious apples during storage

Time (day)	Soluble solids (°Brix)		Texture			
	Lower TSS	Higher TSS	Penetration energy (N mm)		Peak force (N)	
			Lower TSS	Higher TSS	Lower TSS	Higher TSS
0	11.1±0.7 a	13.3±0.8 b	16.9±1.6 d	17.1±1.8 c	10.2±1.0 c	9.7±1.2 b
49	11.3±1.0 abc	13.1±1.1 ab	14.8±1.8 c	15.3±2.2 b	9.2±1.0 b	8.8±1.5 a
76	11.5±0.5 bc	13.3±0.5 bc	13.7±1.3 b	15.0±1.6 b	8.9±0.9 b	9.1±1.0 ab
103	11.7±0.8 bc	13.5±1.0 bc	13.9±1.5 bc	14.5±1.6 ab	9.0±1.0 b	9.0±1.2 a
138	11.7±0.6 c	13.9±0.5 c	13.5±1.4 b	14.3±1.2 ab	8.6±0.7 ab	8.7±1.1 a
168	11.4±0.7 abc	12.6±0.9 a	13.8±1.4 b	13.6±1.2 a	9.1±0.9 b	8.8±1.1 a
189	11.2±0.7ab	13.1±0.5 ab	12.3±1.9 a	13.8±2.3 a	8.1±1.4 a	8.6±1.2 a
<i>p</i> value	*	**	***	***	***	*

For each column, values corresponding to different letters are significantly different at $p<0.05$

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Table 2: Analytical parameters (mean value±standard deviation, n=10) of Red Delicious apples during storage

Time (day)	Soluble solids (°Brix)		Texture			
	Lower TSS	Higher TSS	Penetration energy (N mm)		Peak force (N)	
			Lower TSS	Higher TSS	Lower TSS	Higher TSS
0	12.1±0.7 ab	13.4±1.2 a	26.1±3.8 e	26.2±5.5 d	15.0±1.9 d	14.8±2.3 d
49	13.0±1.5 cd	13.9±0.9 a	25.4±2.5 de	22.8±4.2 c	14.3±1.6 cd	13.3±2.2 c
76	11.7±0.9 a	14.0±1.1 a	25.9±3.5 de	19.2±3.1 b	15.2±1.8 d	12.3±2.1 bc
103	12.3±0.9 ab	13.7±1.0 a	23.8±3.0 cd	20.3±3.4 bc	13.6±1.8 c	12.4±1.6 bc
138	12.6±0.6 bc	15.0±1.9 b	21.2±5.0 b	20.4±5.9 bc	12.4±2.5 b	12.6±3.5 c
168	13.4±1.1 d	13.6±0.8 a	18.3±4.2 a	17.9±4.1 ab	11.1±1.7 a	11.0±2.2 ab
189	12.1±1.1 ab	13.4±0.6 a	22.2±3.7 bc	15.5±4.2 a	13.3±1.5 bc	9.9±2.4 a
<i>p</i> value	***	***	***	***	***	***

For each column, values corresponding to different letters are significantly different at $p<0.05$

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Tables 1 and 2 report the average values of the quality parameters for Golden Delicious and Red Delicious apples, respectively, for both ripening classes during the 7 months of storage. The data show that during storage in a controlled atmosphere coupled with the use of 1-MCP (Fan et al. 1999), no significant evolution of TSS occurred during the storage period, despite some significant differences in this parameter on some sampling

dates. The TSS ranged between 12 and 15 °Brix, and a clear difference between the two ripening classes was maintained all along the storage period (the higher TSS class of apples showed a TSS of about 2 °Brix higher than the other class) (Fig. 2). During the last sampling dates, non-marketable samples (due to disorders) were eliminated. The number of rejected apples was higher for the higher TSS class compared to the lower TSS class (both for Golden and Red samples). This behavior reconfirms the need for a classification tool only to optimize the management of storage and to reduce fruit wastage. Regarding firmness indices, in response to time progression, the fruits of both classes showed a decrease in firmness indices, especially in the penetration energy. A wider variability for firmness parameters was observed as a result of a significant change in apple texture during storage time ($p < 0.001$). However, there were no differences between the two classes, with clearly overlapping patterns, particularly for the cultivar Golden Delicious (Fig. 2).

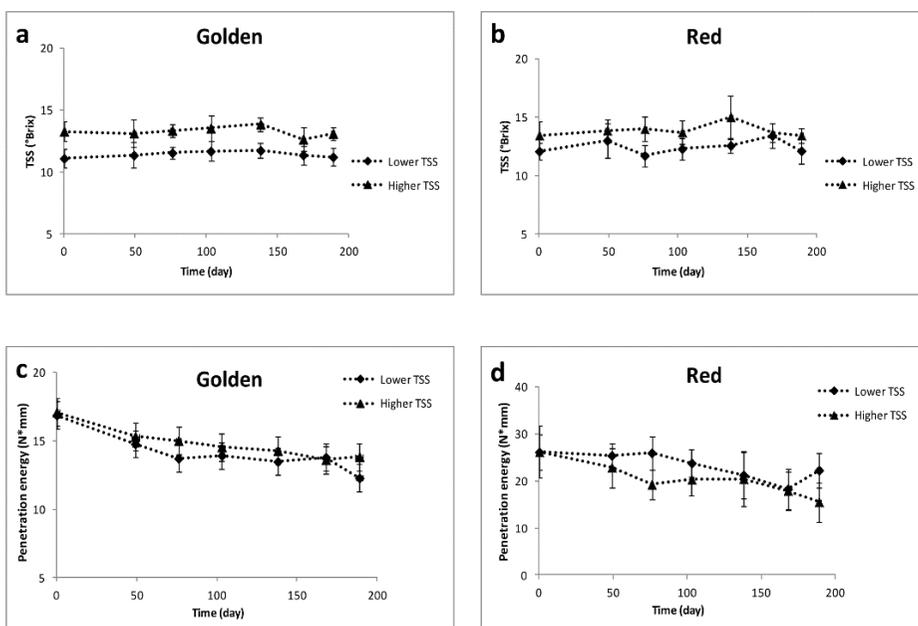


Figure 2: Evolution of TSS and penetration energy during the apple storage life belonging to the two classes of ripeness "higher TSS" and "lower TSS".

VIS-NIR Analysis

Figure 3a shows a particular of the probe of the VIS-NIR bench-top instrument, while Fig. 3b represented the spectra of some samples in the range of 600–1200 nm. The absorption peak at 675 nm (absorption peak of chlorophyll in the visible region) and the peak at about 970 nm (absorption peak relative to the second overtone of the water O–H bond in the near infrared region) can be clearly seen. The variation in the relative absorption, in particular at the chlorophyll peak, is very important because it is directly

related to the degree of ripeness and firmness in fruits (Zude et al. 2006; Merzlyak et al. 2003); the in-depth peak at 970 nm is characteristic of the VISNIR measurements on food very rich in water, like apples (Nicolai et al. 2007).

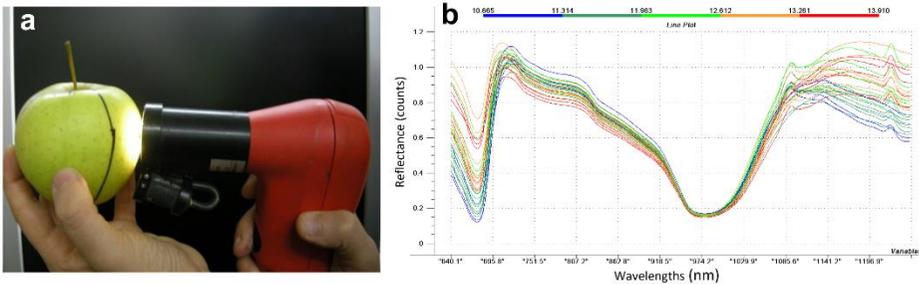


Figure 3: **a** Detail of the probe of the automatic bench-top VIS-NIR system and **b** example of spectra in the 600–1200 nm range (different colors correspond to different TSS values)

VIS-NIR Device Prediction Capabilities

Table 3: Percentage of acceptable and not acceptable prediction of unknown samples for the analyzed parameters

	Acceptable (%)	Not acceptable (%)
TSS (°Brix)		
Golden	83	17
Red	56	44
Penetration energy (N mm)		
Golden	68	32
Red	27	72
Peak force (N)		
Golden	52	48
Red	63	37

Prediction is acceptable if prediction error <10 %; not acceptable if prediction error >10 %

The results highlight the good percentage of acceptable classification for TSS prediction (Table 3). In particular, very good results were obtained for Golden Delicious, with 83 % of samples correctly classified. Regarding firmness parameters, better results were achieved for Golden Delicious compared with Red Delicious apples, especially for penetration energy. The application of NIR spectroscopy for the analysis of firmness parameters often encounters considerable difficulties, highlighted by some published studies (Nicolai et al. 2008). Our data also show that it is difficult to obtain high percentages of correct classifications in both cultivars for these parameters. TSS and firmness parameter prediction on Golden Delicious is better than those on Red Delicious. This is due to the higher variability of the data of Golden Delicious variety, which allowed the development of more robust models. Moreover, the spectral data are influenced by skin color. In Golden Delicious, the skin color evolution is more pronounced and strictly correlated to internal quality. In contrast with the Golden apples, in which the color of the skin changes with the maturation, for the red, this does

not happen. Therefore, the red color of the Red Delicious skin does not change, and this does not reflect the modifications of the internal quality parameters of the apple. This limits the differences in VISNIR spectra at different maturation stages for this variety.

Quantitative VIS-NIR Analysis

At the end of the storage period 2010–2011, all the available data (VIS-NIR spectra and reference destructive analysis) were used for the elaboration of PLS models, based on a large number of samples from the same storage season. Figure 4 shows, as an example, the regression lines of PLS models for penetration energy in Golden Delicious and for TSS in Red Delicious. Descriptive statistics and the statistics of the PLS models for TSS and firmness indices (penetration energy and peak force) are shown in Table 4 for Golden Delicious and Red Delicious. Models with good accuracy were obtained in both analyzed cultivars. TSS estimation gave high R^2 values with a low standard error (RMSECV=0.3 and 0.4 °Brix for Golden Delicious and Red Delicious, respectively), indicating a high level of prediction performance. In the literature, the results obtained with various fruits have shown RMSECV values of about 0.6–1 °Brix (Nicolai et al. 2007; Bobelyn et al. 2010; Jha and Ruchi 2010; Liu and Ying 2004). These results are also similar to those published for similar non-destructive applications on other fruits such as apricot (Camps and Christen 2009) and watermelon (Tian et al. 2007). Regression models were obtained with fairly good correlation coefficients and low errors in the cross-validation both regarding the penetration energy (R^2 cval=0.81 and 0.77 for Golden Delicious and Red Delicious, respectively) and peak force (R^2 cval=0.83 and 0.69 for Golden Delicious and Red Delicious, respectively). Generally, the results obtained for the Golden variety were better than those for the Red variety, both in calibration and in crossvalidation (Bobelyn et al. 2010). Recent studies concerning the application of NIR spectroscopy in the range of 700–1100 nm for the evaluation of Fuji apple firmness report prediction models with r values of 0.93 and percentage error of cross validation (%SECV) of about 8 % (Qing et al. 2007); conversely, application of VIS-NIR (380–1690 nm) spectroscopy for firmness prediction of several apple cultivars gave low correlation coefficients ($R^2 < 0.70$), especially concerning Golden Delicious apples ($R^2 = 0.41$). The estimation of the texture parameters proved to be worse due to the inherent variability of the data, as confirmed in the literature (Paz et al. 2009; Zude et al. 2006).

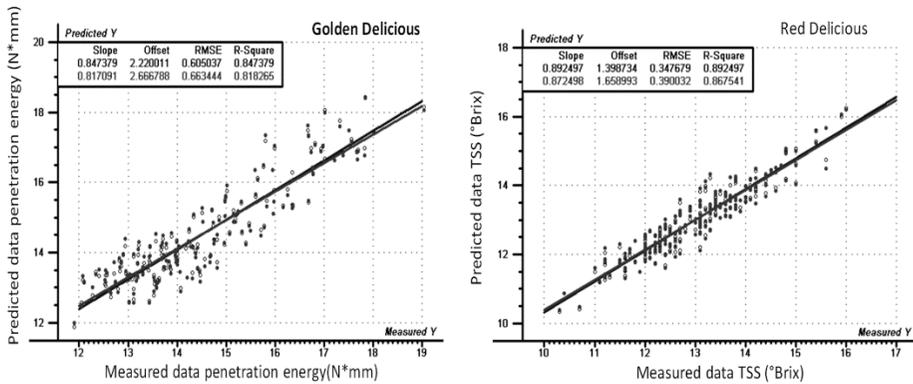


Figure 4: Example of PLS models for Golden Delicious penetration energy and Red Delicious TSS

Table 4: Descriptive statistics and statistics of the PLS models elaborated on Golden Delicious and Red Delicious apples

	Number of samples (fruit)	Mean	S.D.	Calibration			Cross-validation		
				R^2	RMSEC	LV	R^2	RMSECV	RMSECV%
Golden Delicious									
TSS (°Brix)	280 (140)	12.7	1.05	0.94	0.3	7	0.90	0.3	2.4
Penetration energy (N mm)	280 (140)	14.7	2.9	0.85	0.6	6	0.81	0.7	4.8
Peak force (N)	280 (140)	9.1	0.8	0.85	0.4	7	0.83	0.4	4.4
Red Delicious									
TSS (°Brix)	280 (140)	13.8	1.5	0.88	0.4	9	0.86	0.4	2.9
Penetration energy (N mm)	280 (140)	23.0	6.0	0.81	1.6	5	0.77	1.7	7.4
Peak force (N)	280 (140)	13.0	1.7	0.72	0.85	7	0.69	1.0	7.7

R^2 coefficient of determination, *RMSEC* root mean square error of calibration, *RMSECV* root mean square error of cross validation, *LV* latent variables

Qualitative VIS-NIR Analysis

Classification techniques (Linear Discriminant Analysis (LDA)) were also applied to the spectral data in order to discriminate apples on the basis of storage time (each class corresponds approximately to the months of storage, from time zero to 6 months). For the classification, the classes were defined in accordance to the days of storage: class 1 is composed of apples at time 0; classes 2, 3, 4, 5, 6, and 7 are composed of the samples stored for 49, 76, 103, 138, 168, and 189 days, respectively. Table 5 shows the results obtained by LDA classification for Golden Delicious and Red Delicious apples. VIS-NIR spectroscopy gave good classification performances on both apple varieties for both applied pretreatment techniques (SNV and MSC), with average correct classification percentages close to 100 %. This result appears to be very valuable, even when compared with other published works. Kavdir and Guyer (2008) evaluated different pattern recognition techniques for apple sorting, obtaining classification result of 90 % applying LDA. On other products, Craig et al. in 2012 achieved an LDA classification result of 95 % for normalized spectra in prediction, and Uddin et al. (2005) achieved a classification result of 81 %. Conclusions A compact bench-top spectrophotometer,

operating in the range 600–1200 nm, was tested at a refrigerated warehouse for monitoring and optimizing the management of stored apples. The results indicate that the system allows an accurate estimation of chemo-physical parameters (TSS, penetration energy, and peak force). Particularly, the results show excellent prediction ability for TSS, while acceptable performance in the estimation of firmness parameters was obtained. The study of quality characteristics during storage demonstrated that the differences between the two ripening classes, initially selected in a non-destructive way, were maintained during storage in a controlled atmosphere. The instrument proved to be suitable not only for the evaluation of apple quality parameters but also for fruit classification according to the storage time. It can therefore be used as a non-destructive method for apple classification in homogeneous lots with the purpose of a better management of the destination of lots during the months of storage in order to avoid fruit wastage. All these advantages are combined with the well-known ones regarding the use of this optical technology (saving time and analysis of a large number of samples). The use of such tools makes it possible for operators to quickly monitor important characterization parameters attained by the fruit at harvest and during all subsequent storage stages, allowing management optimization and planning of cell opening sequence according to fruit characteristics.

Table 5: Results obtained using the LDA technique Cal calibration, Cv cross-validation

	% correctly classified samples							Average
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	
Golden								
SNV								
Cal	100	100	100	100	100	100	100	100
Cv	100	100	100	100	100	100	100	100
MSC								
Cal	100	100	100	100	100	100	100	100
Cv	100	100	100	100	95	100	100	99.3
Red								
SNV								
Cal	100	100	100	100	100	100	100	100
Cv	100	100	100	100	100	95	100	99.3
MSC								
Cal	100	100	100	100	100	100	100	100
Cv	100	100	100	100	100	100	100	100

Acknowledgments

This study received financial support from Regione Lombardia as “VALORVÌ” research project and from Regione Lombardia and European Social Fund for a Post-Doctoral Research Fellowship (“Progetto Dote Ricerca”).

References

- Basilevsk A. (1994). Statistical factor analysis and related methods: theory and applications. Wiley. ISBN 0-471-57082-6.
- Beghi, R., Spinardi, A., Bodria, L., Mignani, I., & Guidetti, R. (2012). Apples nutraceutical properties evaluation through a visible and nearinfrared portable system. Food and Bioprocess technology. doi:10.1007/s11947-012-0824-7.

- Beni, C., Ianicelli, V., & Di Dio, C. (2001). Il condizionamento dei prodotti ortofruitticoli. Bologna: Edagricole-Edizioni Agricole della Calderini s.r.l.
- Bobelyn, E., Serban, A. S., Nicu, M., Lammertyn, J., Nicolai, B. M., & Saeys, W. (2010). Postharvest quality of apple predicted by NIR spectroscopy: study of the effect of biological variability on spectra and model performance. *Postharvest Biology and Technology*, 55, 133–143.
- Brosnan, T., & Sun, D. W. (2004). Improving quality inspection of food products by computer vision - a review. *Journal of Food Engineering*, 61, 3–16.
- Camps, C., & Christen, D. (2009). Non-destructive assessment of apricot fruit quality by portable visible-near infrared spectroscopy. *Food Science and Technology*, 42(6), 1125–1131.
- Casale, M., Casolino, C., Ferrari, G., & Forina, M. (2008). Near infrared spectroscopy and class modelling techniques for the geographical authentication of Ligurian extra virgin olive oil. *Journal of Near Infrared Spectroscopy*, 16(1), 29–47.
- Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology*, 18, 72–83.
- Craig, A. P., Franca, A. S., & Oliveira, L. S. (2012). Discrimination between defective and non-defective roasted coffees by diffuse reflectance infrared Fourier transform spectroscopy. *LWT – Food Science and Technology*, 47(2), 505–511.
- Fan, X., Blankenship, S. M., & Mattheis, J. P. (1999). 1- methylcyclopropene inhibits apple ripening. *Journal of the American Society for Horticultural Science*, 124(6), 690–695.
- Ferguson, I., Volz, R., & Woolf, A. (1999). Preharvest factors affecting physiological disorders of fruit. *Postharvest Biology and Technology*, 15, 255–262.
- Forina M., Lanteri S., Armanino C., Casolino C., Casale M. & Oliveri P. (2008). V-PARVUS 2008, Dip. Chimica e Tecnologie Farmaceutiche e Alimentari, University of Genova available (free, with manual and examples) from authors or at <http://www.parvus.unige.it>.
- Frank I.E. & Todeschini R. (1994). *The data analysis handbook*. Elsevier, ISBN 0-444-81659-3.
- Giovanelli G., Sinelli N., Beghi R., Guidetti R. & Casiraghi E. (2011). NIR spectroscopy for the optimization of post-harvest apple management. *Proceedings of the CIGR 6th International Technical Symposium "Toward a Sustainable Food Chain"*, Nantes, France - April 18-20, 2011, ISBN 978-2-7466-3203-5.
- Guidetti R., Beghi R., Bodria L., Spinardi A., Mignani I. & Folini L. (2008). Prediction of blueberry (*Vaccinium corymbosum*) ripeness by a portable Vis-NIR device. *Acta Horticulturae*, no. 310, ISBN 978-90-66057-41-8, 877-885.
- Guidetti R., Beghi R. & Giovenzana V. (2012). *Chemometrics in food technology*. In: *Chemometrics in practical applications*, ISBN: 978- 953-51-0438-4, InTech.
- Harker, F. R., Maindonald, J., Murray, S. H., Gunson, F. A., Hallett, I. C., & Walker, S. B. (2002). Sensory interpretation of instrumental measurements. 1: Texture of apple fruit. *Postharvest Biology & Technology*, 24(3), 225–239.
- Heberger, K., Csomos, E., & Simon-Sarkadi, L. (2003). Principal component and linear discriminant analyses of free amino acids and biogenic amines in Hungarian wines. *Journal of Agricultural and Food Chemistry*, 51, 8055–8060.
- Huang, H., Yu, H., Xu, H., & Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review. *Journal of Food Engineering*, 87, 303–313. Jackson J.E. (1991). *A user's guide to principal components*. Wiley. ISBN 0-471-62267-2.
- Jha, S. N., & Ruchi, G. (2010). Non-destructive prediction of quality of intact apple using near infrared spectroscopy. *Journal of Food Science and Technology*, 47(2), 207–213.
- Johnston, J. W., Hewett, E. W., & Hertog, M. L. A. T. M. (2002). Postharvest softening of apple (*Malus domestica*) fruit: a review. *New Zealand Journal of Crop and Horticultural Science*, 30(3), 145–160.
- Kavdir, I., & Guyer, D. E. (2008). Evaluation of different pattern recognition techniques for apple sorting. *Biosystems Engineering*, 99(2), 211–219.
- Liu, Y. D., & Ying, Y. B. (2004). Measurement of sugar content in Fuji apples by FT-NIR spectroscopy. *Journal of Zhejiang University*, 5(6), 651–655.
- Lu, R., Guyer, D. E., & Beaudry, R. M. (2000). Determination of firmness and sugar content of apples using near-infrared diffuse reflectance. *Journal of Texture Studies*, 31(6), 615–630.
- Magwaza, L. S., Opara, U. L., Nieuwoudt, H., Cronje, P. J. R., Saeys, W., & Nicolai, B. (2012). NIR spectroscopy applications for internal and external quality analysis of citrus fruit—a review. *Food and Bioprocess Technology*, 5, 425–444.
- Massart D.L., Vandeginste B.G.M., Buydens L.M.C., De Jong S., Lewi P.J. & Smeyers-Verbek J. (1997). *Handbook of chemometrics and qualimetrics: part A*. Elsevier, ISBN:0-444-89724-0.
- Massart D.L., Buydens L.M.C., De Jong S., Lewi P.J. & Smeyers-Verbek J. (1998). *Handbook of chemometrics and qualimetrics: part B*. Edited by B.G.M. ISBN: 978-0-444-82853-8.

- McClure, W. F. (2003). 204 years of near infrared technology: 1800–2003. *Journal of Near Infrared Spectroscopy*, 11, 487–518.
- McGlone, V. A., Jordan, R. B., & Martinsen, P. J. (2002). VIS-NIR estimation at harvest of pre- and post-storage quality indices for Royal Gala apple. *Postharvest Biology and Technology*, 25, 135–144.
- Mehinagic, E., Royer, G., Bertrand, D., Symoneaux, R., Laurens, F., & Jourjon, F. (2003). Relationship between sensory analysis, penetrometry and visible-NIR spectroscopy of apples belonging to different cultivars. *Food Quality and Preference*, 14, 473–484.
- Menesatti, P., Zanella, A., D'Andrea, S., Costa, C., Paglia, G., & Pallottino, F. (2009). Supervised multivariate analysis of hyperspectral NIR images to evaluate the starch index of apples. *Food and Bioprocess Technology*, 2(3), 308–314.
- Merzlyak, M. N., Solovchenko, A. E., & Gitelson, A. A. (2003). Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. *Postharvest Biology and Technology*, 27, 197–211.
- Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyna, J. (2007). Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biology and Technology*, 46, 99–118.
- Nicolai, B. M., Verlinden, B. E., Desmet, M., Saevels, S., Saeys, W., Theron, K., Cubeddu, R., Pifferi, A., & Torricelli, A. (2008). Timeresolved and continuous wave NIR reflectance spectroscopy to predict soluble solids content and firmness of pear. *Postharvest Biology and Technology*, 47(1), 68–74.
- Paz, P., Sanchez, M. T., Perez, M. D., Guerrero, J. E., & Garrido-Varo, A. (2009). Evaluating NIR instruments for quantitative and qualitative assessment of intact apple quality. *Journal of the Science of Food and Agriculture*, 89, 781–790.
- Peirs, A., Lammertyn, J., Ooms, K., & Nicolai, B. M. (2001). Prediction of the optimal picking date of different apple cultivars by means of VIS-NIR-spectroscopy. *Postharvest Biology & Technology*, 21(2), 189–199.
- Peirs, A., Scheerlinck, N., Touchant, K., & Nicolai, B. M. (2002). Comparison of Fourier transform and dispersive near-infrared reflectance spectroscopy for apple measurements. *Biosystems Engineering*, 81(1), 305–311.
- Qing, Z., Ji, B., & Zude, M. (2007). Predicting soluble solid content and firmness in apple fruit by means of laser light backscattering image analysis. *Journal of Food Engineering*, 82(1), 58–67.
- Quilitzsch, R., & Hoberg, E. (2003). Fast determination of apple quality by spectroscopy in the near infrared. *Journal of Applied Botany and Food Quality*, 77(5–6), 172–176.
- Saevels, S., Lammertyn, J., Berna, A. Z., Veraverbeke, E. A., Di Natale, C., & Nicolai, B. M. (2004). An electronic nose and a mass spectrometry-based electronic nose for assessing apple quality during shelf life. *Postharvest Biology and Technology*, 31, 9–19. *2142 Food Bioprocess Technol* (2014) 7:2134–2143
- Sinelli, N., Casiraghi, E., Tura, D., & Downey, G. (2008). Characterisation and classification of Italian virgin olive oils by near- and mid-infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 16(3), 335–342.
- Snowdon A.L. (2010). *Postharvest diseases and disorders of fruits and vegetables*; vol. 1, chapter 4. Manson; ISBN 9781840765984.
- Tian, H., Ying, Y., Lu, H., Fu, X., & Yu, H. (2007). Measurement of soluble solids content in watermelon by VIS-NIR diffuse transmittance technique. *Journal of Zhejiang University*, 8(2), 105–110.
- Uddin, M., Okazaki, E., Turza, S., Yumiko, Y., Tanaka, M., & Fukuda, Y. (2005). Non-destructive visible/NIR spectroscopy for differentiation of fresh and frozen-thawed fish. *Journal of Food Science*, 70(8), 506–510.
- Watkins, C. B. (2006). The use of 1-methylcyclopropene (1-MCP) on fruits and vegetables. *Biotechnology Advances*, 24, 389–409.
- Wills, R. T., Graham, D., Mc Glasson, W. B., & Joyce, D. (1998). *Postharvest: an introduction to the physiology and handling of fruits, vegetables and ornamentals* (4th ed.). Sydney: UNSW Press Ltd.
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109–130.
- Wolf, A. B., & Ferguson, I. B. (2000). Postharvest responses to high fruit temperatures in the field. *Postharvest Biology and Technology*, 21, 7–20.
- Zude, M., Herold, B., Roger, J. M., Bellon-Maurel, V., & Landahl, S. (2006). Non-destructive tests on the prediction of apple fruit flesh firmness and soluble solids content on tree and in shelf life. *Journal of Food Engineering*, 77(2), 254–260.
- Zude-Sasse, M., Truppel, I., & Herold, B. (2002). An approach to non-destructive apple fruit chlorophyll determination. *Postharvest Biology and Technology*, 25(2), 123–133.

1.2 WAVELENGTH SELECTION WITH A VIEW TO A SIMPLIFIED HANDHELD OPTICAL SYSTEM TO ESTIMATE GRAPE RIPENESS

Valentina Giovenzana¹, Roberto Beghi^{1*}, Cristina Malegori², Raffaele Civelli¹, Riccardo Guidetti¹

¹ Department of Agricultural and Environmental Sciences - Production, Landscape, Agroenergy
Università degli Studi di Milano, via Celoria 2, Milano, 20133 Italy

² Department of Food, Environmental and Nutritional Sciences
Università degli Studi di Milano, via Celoria 2, 20133 Milan, Italy

Abstract

The aim of this work was to identify the three most significant wavelengths able to discriminate in the field those grapes ready to be harvested using a simplified, handheld, and low-cost optical device. Nondestructive analyses were carried out on a total of 68 samples and 1,360 spectral measurements were made using a portable commercial vis/near-infrared spectrophotometer. Chemometric analyses were performed to extract the maximum useful information from spectral data and to select the most significant wavelengths. Correlations between the spectral data matrix and technological (total soluble solids) and phenolic (polyphenols) parameters were carried out using partial least square (PLS) regression. Standardized regression coefficients of the PLS model were used to select the relevant variables, representing the most useful information of the full spectral region. To support the variable selection, a qualitative evaluation of the average spectra and loading plot, derived from principal component analysis, was considered. The three selected wavelengths were 670 nm, corresponding to the chlorophyll absorption peak, 730 nm, equal to the maximum reflectance peak, and 780 nm, representing the third overtone of OH bond stretching. Principal component analysis and multiple linear regression were applied on the three selected wavelengths in order to verify their effectiveness. Simple equations for total soluble solids and polyphenols prediction were calculated. The results demonstrated the feasibility of a simplified handheld device for ripeness assessment in the field.

Keywords: grape, ripeness, vis/NIR spectroscopy, wavelength selection, handheld device, chemometrics

Introduction

Near-infrared (NIR) spectroscopy has gained wide acceptance in different fields, particularly in postharvest fruit and vegetable production (Wang and Paliwal 2007). The main advantage of NIR technology is its ability to record spectra nondestructively both for solid (using mainly diffuse reflectance acquisition technique) and liquid (using mainly transmittance mode) samples without any pretreatment, allowing for the rapid analysis of products. Cost savings are often achieved for NIR measurements in terms of improved control and product quality and the technique can provide results faster than traditional laboratory analysis. NIR technology has resulted in a number of publications, including a synthesis of the status of NIR in the agrifood industry (Wang and Paliwal 2007) and a comprehensive overview of NIR spectroscopy for measuring the quality attributes of fruit and vegetables (Nicolai et al. 2007). There have been several NIR applications to estimate the ripeness parameters of different fruit species, especially grapes (Bellincontro et al. 2009, Cozzolino et al. 2006). In recent years, interest has shifted toward the development of portable vis/NIR systems for use in the field and these systems been tested in controlled laboratory conditions and directly in the field. Laboratory applications included the feasibility of using vis/NIR spectroscopic devices in transmission mode to predict the soluble solids content (SSC) and total acidity of fresh-cut KaoNumpung pomelos (Puangsombut et al. 2012), the ability of a portable single-channel vis/ NIR spectrometer to determine the SSC and total acidity of two thick-peel mandarin cultivars (Antonucci et al. 2011), and the use of portable NIR to evaluate apricot quality during postharvest (Camps and Christen 2009). Applications in uncontrolled field conditions have included an optical technique to estimate the ripeness of red-pigmented fruits (Bodria et al. 2004), a portable NIR instrument (640–1300 nm) for determining ripeness in winegrapes (Larrain et al. (2008), a vis/ NIR device in reflectance configuration to predict blueberry ripeness (Guidetti et al. 2008), and portable vis/NIR systems to evaluate grape quality parameters (Guidetti et al. 2010) and to assess the nutraceutical properties of apples (Beghi et al. 2003). All these approaches rely on wide spectral ranges (thousands of wavelengths) and thus require multivariate techniques for data processing to build calibration and prediction models (Cen and He 2007, Williams and Norris 2001). Complex mathematical techniques (chemometrics) are required to explain chemical information encoded in spectral data (Cogdill and Anderson 2005). The most commonly used chemometric techniques are spectral preprocessing, to remove any irrelevant information; principal component analysis (PCA), to perform qualitative data analysis; and partial least squares (PLS) regression, to obtain a quantitative prediction of relevant parameters (Wold et al. 2001, Naes et al. 2002, Nicolai et al. 2007, Cen and He 2007). Chemometrics can be used to select a few relevant variables that represent the most useful information of the full spectral region (Xiaobo et al. 2010). This selection eliminates variables containing mostly redundant information and spectral noise and reduces the cost of the potential miniaturized devices built only with the selected wavelengths. Generally, the selection of these

optimal wavelengths reduces the size of the required measurement data while preserving the most important information contained in the data (Sun 2010). Three effective wavelength selection methods combined with vis/ NIR spectroscopy have been proposed to determine the SSC of beer, including a successive projections algorithm (SPA), regression coefficient analysis (RCA), and independent component analysis (ICA) (Liu et al. 2009). The maximum number of selected wavelengths, by SPA, ranged from 4 to 21 depending on different pretreatments; 10 essential wavelengths were obtained by both RCA and by ICA. There are several simplified nondestructive commercial devices based on a few wavelengths available on the market, including an innovative and simplified NIR system (DAMeter for apples and Kiwi-Meter for kiwifruits) patented by the University of Bologna (Costa et al. 2011). These devices determine the stage of fruit maturity through indices based on absorbance differences between specific wavelengths. These indices are correlated with the main traditional parameters as well as with changes in flesh color. The limited adoption of NIR technology by the enology and viticulture sectors could be attributed to cost, technical limitations, grower resistance, and supply-chain weakness (Magwaza et al. 2012). Thus, the development of simplified handheld devices may encourage adoption. Based on these considerations, the aim of this research was to identify the most significant wavelengths able to discriminate grapes ready to be harvested during the final ripening stages. The specific objectives were (1) to obtain essential wavelength variables based on the PLS-RCA variable selection method; (2) to choose the three most informative wavelengths in a view of a simplified and handheld device; (3) to compare the prediction performance of the calibration models established by PLS on the full vis/NIR spectra and multiple linear regression using the selected wavelengths; and (4) to define simple equations for the estimation of grape ripeness. The possible final application would be a low-cost and user-friendly device that supports small-scale growers in determining optimal harvest date according to ripening degree.

Materials and Methods

Sampling

The experimental plan monitored the grape ripening process just before harvest in the Valtellina viticultural area (Sondrio, northern Italy) using *Vitis vinifera* cv. Nebbiolo (ecotype Chiavennasca), one of the most important red varieties in Italy. Samples were drawn from 17 different zones, throughout the entire viticultural area of the valley, in order to represent environmental variability and monitor the entire production region of DOC (controlled denomination of origin) and DOCG (controlled and guaranteed denomination of origin) wines. A total of 68 samples of grape clusters were collected on four sampling dates (7, 16, and 29 Sept and 12 Oct 2011). For each date, nondestructive analyses were carried out on each sample using a portable commercial vis/ NIR spectrophotometer; destructive chemical analyses were then performed. For each of

the 68 samples, the spectral acquisitions of 10 individual berries were carried out, and for each berry two acquisitions were performed in the equatorial region, for a total of 1,360 spectral measurements. Finally, an average spectrum (20 acquisitions) for each sample was calculated and used, coupled with reference chemical data, for the chemometric analysis.

Portable vis/NIR device

Spectral acquisitions were performed on berries directly in the field using a vis/NIR spectrophotometer (Jaz, Ocean Optics, Dunedin, FL), an optical portable system operating in the wavelength range of 400 to 1000 nm. The Jaz equipment consists of five components: (1) vis/NIR lighting system (halogen lamp), (2) fiber-optic probe for reflection measurement, (3) spectrophotometer, (4) hardware for data acquisition and instrument control, and (5) battery power supply. Spectra were acquired in reflectance mode: light radiation was guided from the light source to the sample through a Y-shaped, bidirectional fiber-optic probe. The Y-shaped fiber guided light from the halogen lamp to illuminate the sample while simultaneously collecting the radiation from the berry and guiding it back to the spectrophotometer. The tip of the optical probe was equipped with a soft plastic cap to ensure contact with the skin of the sample during measurements, while minimizing environmental light interference. The integrated spectrophotometer was equipped with diffractive grating for spectral measurements optimized in the range of 400 to 1000 nm and a charge-coupled device sensor with a 2048 pixel matrix, corresponding to a nominal resolution of 0.3 nm. Spectral measurements were taken in the field on individual berries after artificial illumination with the probe tip. In order to reduce the possible influence of environmental conditions, especially related to diurnal changes in sunlight, spectral acquisitions were consistently taken a few hours in the morning with the help of the plastic cap, ensuring contact between the probe tip and the measured berry. Specific tests conducted at different times of the day confirmed the repeatability of measurements with artificial lighting and the berry/probe contact configuration, which were evidently sufficient to cancel out the possible influence of ambient illumination on samples. Air temperature changes during acquisition were limited to a range from 15 to 25°C. The field data set of spectra were assumed to randomly embed possible environment influencing factors.

Chemical analyses

Total soluble solids (TSS) content was measured using a digital pocket refractometer (model DBX-55; ATAGO, Tokyo, Japan) and grape titratable acidity (TA; g tartaric acid dm⁻³) was measured using an automatic titrator (TitroMatic KF 1S, Crison Instruments, Milan, Italy). Grape phenolic content was estimated according to the Glories method (Glories 1984), in which potential anthocyanins (PA) and extractable anthocyanins (EA),

extracted at pH 1 and pH 3.2, respectively (mg anthocyanins dm⁻³), and total polyphenols (TP) were evaluated. Phenolic compound quantification was based on optical density (OD) measurement at 520 nm and 280 nm for anthocyanins and polyphenols, respectively, using a UV/vis spectrophotometer (model 7800; Jasco, Tokyo, Japan).

Data processing - Spectral data

Chemometric analyses were performed using The Unscrambler (ver. 9.6; CAMO Software, Oslo, Norway) to extract the maximum usable information from the spectral data and to select the most significant wavelengths that could be used in a simplified device. Collected spectra were preprocessed using smoothing (moving average, 15-nm-wide window) and reducing techniques. Principal component analysis (PCA) was performed to explore the possible clustering of sample spectra from the same sampling date. PCA identifies the natural clusters in the data set, with the first principal component (PC) expressing the largest amount of variation, followed by the second PC which conveys the second most important factor of the remaining analysis, and so forth (Di et al. 2009). Loading plots obtained with PCA were subsequently analyzed to confirm the variable selection performed using regression coefficient analysis (RCA), deriving from partial least square (PLS) analysis (Chong and Jun 2005). The correlation between the spectral data matrix and technological (TSS) and phenolic (total polyphenols; TP) parameters were carried out using a PLS regression algorithm, the most widely used regression technique that relates two data matrixes, X and Y, by a linear multivariate model. In this procedure, full cross-validation was used to develop a PLS regression model. To evaluate model accuracy, the statistical parameters used were the coefficient of determination in calibration (R^2_{cal}), the coefficient of determination in crossvalidation (R^2_{cv}), the root mean square error of calibration (RMSEC), and the root mean square error of cross-validation (RMSECV). The best calibrations were selected based on minimizing the RMSECV. Percent errors in cross-validation (RMSECV%) were also calculated as: $RMSECV(\%) = RMSECV / \text{averaged reference values of each parameter}$ (Nicolai et al. 2007, Naes et al. 2002). Moreover, the ratio performance deviation (RPD) value was calculated, which is defined as the ratio between the standard deviation of the response variable and RMSECV (Williams 2001, Fearn 2002). RPD values 2.0, quantitative predictions are possible.

Data processing - Wavelength selection

RCA was carried out for relevant variable selection, representing the most useful information of the full vis/NIR spectral region (Xiaobo et al. 2010, Chong and Jun 2005). Standardized regression coefficients of the PLS model were used for the elaboration. The standardization took into account both the standard deviation of reflectance, for each wavelength considered, and the standard deviation of the reference data (TSS and

TP) (Frank and Todeschini 1994). Regression coefficients obtained by the PLS model were used to calculate the Y variable response value (TSS and TP in grape) from the X variables (grape spectra). The size of the numerical coefficients gave an indication of the impact of different variables on the response (Y). The final aim was to find which variables were important for predicting the Y variable. High absolute values indicate the importance and the significance of the effect on the prediction of Y variable preference. Hence, RCA could be used for essential wavelength selection. Peaks and valleys represented the extreme of the regression coefficient plot, and the three higher peaks (absolute values) were chosen. Finally, qualitative evaluation of the average spectra and loading plot derived from PCA were considered to confirm the effectiveness of RCA variable selection. Spectral reflection intensity measured at selected wavelength ranges were finally used to predict quality parameters (TSS and TP) for the determination of different grape ripening stages. Therefore, the three wavelengths were used as the input data matrix of PCA and for the elaboration of multiple linear regression (MLR) models (Wu et al. 2010, Fernández-Navales et al. 2009, Li et al. 2007). Compared to PLS, MLR allows for the development of models using only few important variables to predict the outcome of a response. MLR is also well-suited when the number of variables is less than the number of samples and is not affected by collinearity (Næs and Mevik 2001). Verification of the prediction ability of the MLR models was performed to study the efficiency of the selected wavelengths.

Results

Table 1: Descriptive statistics of *V. vinifera* cv. Nebbiolo ripening parameters, including standard deviation (SD) and standard error (SE).

Parameter	Unit	N	Range	Mean	SD	SE
Total soluble solids	Brix	68	15.9–23.8	21.1	1.8	0.2
Titrateable acidity	g tartaric acid dm ⁻³	68	5.7–16.9	9.1	2.0	0.3
Potential anthocyanins	mg dm ⁻³	68	38.6–911.6	514.5	139.3	19.0
Extractable anthocyanins	mg dm ⁻³	68	199.9–600.8	357.8	93.7	11.8
Total polyphenols	OD ₂₈₀ nm	68	22.1–51.5	36.2	7.7	1.0

Table 2: Statistics of the PLS models to predict maturity indices and nutraceutical properties of *V. vinifera* cv. Nebbiolo berries

Parameter ^a	Unit ^b	Calibration ^a			Cross-validation ^a					
		LV	R ² _{cal}	RMSEC	R ² _{cv}	Bias	Slope	RMSECV	RMSECV%	RPD
TSS	Brix	7	0.83	0.66	0.77	0.007	0.80	0.78	3.8%	2.26
TA	g tartaric acid dm ⁻³	8	0.76	0.58	0.62	-0.008	0.68	0.75	8.2%	2.66
PA	mg dm ⁻³	7	0.59	76.49	0.41	0.511	0.49	93.48	19.1%	1.61
EA	mg dm ⁻³	7	0.60	49.53	0.39	0.793	0.49	62.18	17.4%	1.51
TP	OD ₂₈₀ nm	7	0.81	3.24	0.74	-0.018	0.77	3.88	10.7%	1.98

^aAbbreviations: TSS, total soluble solids; TA, titrateable acidity; PA, potential anthocyanins; EA, extractable anthocyanins; TP, total polyphenols; OD, optical density; LV, latent variables; R²_{cal}, coefficient of determination in calibration; RMSEC, root mean square error of calibration; R²_{cv}, coefficient of determination in cross-validation; RMSECV, root mean square error of cross-validation; RMSECV%, percent errors in cross-validation; RPD, ratio performance deviation.

Descriptive statistics for berry technological indices (TSS and titrateable acidity) and for nutraceutical parameters (potential anthocyanins, extractable anthocyanins, and total polyphenols) were determined and average data are shown, based on all four sampling

dates (Table 1). Statistics related to the PLS models obtained by vis/NIR spectroscopy for berry qualitative parameters are shown (Table 2). The model developed for TSS presented a good determination coefficient and a low RMSEC value of 0.78. In validation, R^2_{cv} was lower (0.77); RMSECV% was 3.8%. The titratable acidity data had a high coefficient of determination in calibration and a fair RMSECV value (Table 2). In validation, the determination coefficient was slightly lower and RMSECV was slightly higher. The elaboration of a model for PA and EA displayed, in validation, low values of the determination coefficients and RMSECV% values of 19.1% and 17.4%, respectively. Ratio performance deviation (RPD) values for technological parameters (TSS and TA) were >2.0 and for PA and EA phenolic indices were ~ 1.5 . The RPD for total polyphenols (TP) was 1.98. The best correlations were found for TSS and TP. Therefore, the relevant wavelengths describing the features of the spectra for the determination of TSS and TP were selected. RCA was performed, starting from the PLS models for TSS and TP. In particular, the standardized regression coefficients of the PLS model were used to select the relevant variables. The standardization took into account both the standard deviation of reflectance, for each wavelength of the vis/NIR spectra, and the standard deviation of the reference data (TSS and TP). Plots of the standardized regression coefficients, for TSS and TP, versus vis/NIR wavelengths (400–1000 nm) are shown (Figure 1). The trend and shape of the regression coefficient plots were similar for TSS and TP. The three most relevant wavelengths were selected by choosing the higher absolute regression values (Liu et al. 2008, Cen et al. 2006): 670, 730, and 780 nm. The performance of RCA was confirmed using a loading plot derived from PCA and the visual inspection of average spectral curves (Sun 2010). With close observation, the PCA loading plot (Figure 2) indicated some positive and negative peaks at certain wavelengths. Corresponding to the three wavelengths (670, 730, and 780 nm) sorted by RCA, the loading plot showed (1) maximum differences between the PCs (high positive value for PC1 and high negative value for PC2) at 670 nm, (2) the maximum PC1 peak explaining 97% of the total variance at 730 nm; and (3) the maximum positive value of the PC2 couplet with a fairly high positive value for PC1 at 780 nm. The average measured spectra of the four sampling dates are shown (Figure 3). The study of spectral evolution over time highlighted which wavelengths were more sensitive to grape variability during ripening: 670 nm, corresponding to the chlorophyll absorption peak (McGlone et al. 2002); 730 nm, equal to the maximum reflectance peak; and 780 nm, representing the third overtone of OH bond stretching (Clement et al. 2008, Bertrand 2000). Spectral reflection intensity measured at selected wavelengths was finally used to predict the quality parameters (TSS and TP) for the determination of different berry ripening stages. Therefore, the three essential wavelengths were used as the input data matrix of PCA and for the elaboration of simple MLR models. Verification of the prediction ability of the MLR models was done to study the efficiency of the selected wavelengths. The PCA derived from the initial full vis/NIR spectra (Figure 4) and the PCA derived from only the three selected variables (Figure 5) were compared. For the PCA

score plot of the full spectra, the ripening process was well fitted by PCA: spectral data corresponding to different sampling dates were sorted from negative PC1 and positive PC2 values to positive PC1 and negative PC2 values. The same behavior was displayed in the PCA score plot of the PCA arising from the effective wavelengths. In both cases, PC1 accounted for separating berries in terms of ripening stages. According to the variable selection methods stated above, the selected wavelengths were used as the inputs for MLR model elaboration for TSS and TP (Table 3). The selection based on an analysis of the main changes the optical spectra of the berries measured during ripening led to the equation,

$$Y = b_1 \cdot I_{670} + b_2 \cdot I_{730} + b_3 \cdot I_{780} + b_0 \quad \text{Eq. 1}$$

where the parameters b_1 , b_2 , b_3 , and b_0 are computed from a multilinear fit of known pair values (spectral intensities measured at the three wavelengths, 670 nm, 730 nm, and 780 nm, and corresponding chemical data) for the berries using MLR analysis. Equations of MLR models for TSS and TP are reported as:

$$\text{Model}_{\text{TSS}} \quad Y_{\text{TSS}} = 13.15 \cdot I_{670} - 38.55 \cdot I_{730} + 37.81 \cdot I_{780} + 15.69 \quad \text{Eq. 2}$$

$$\text{Model}_{\text{TP}} \quad Y_{\text{TP}} = 60.75 \cdot I_{670} - 254.93 \cdot I_{730} + 382.73 \cdot I_{780} - 162.20 \quad \text{Eq. 3}$$

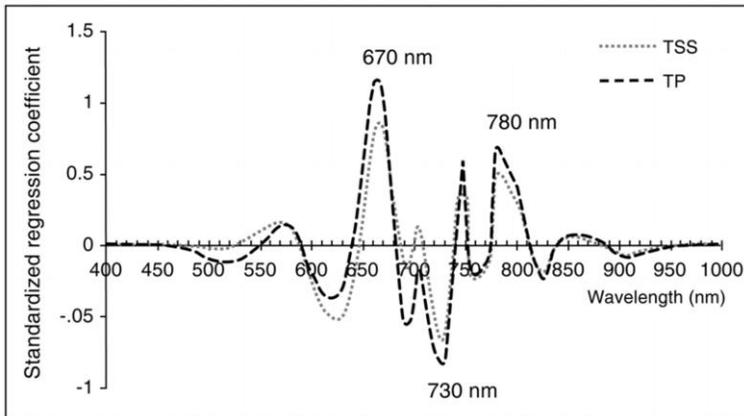


Figure 1: Standardized regression coefficients for total soluble solid

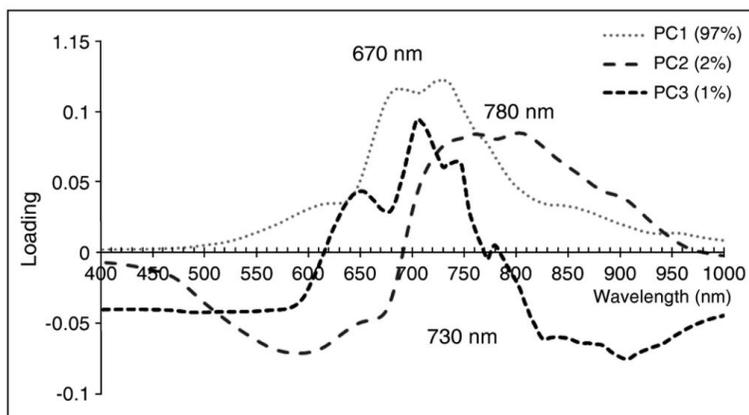


Figure 2: Loading plot of the first three principal components (PC)

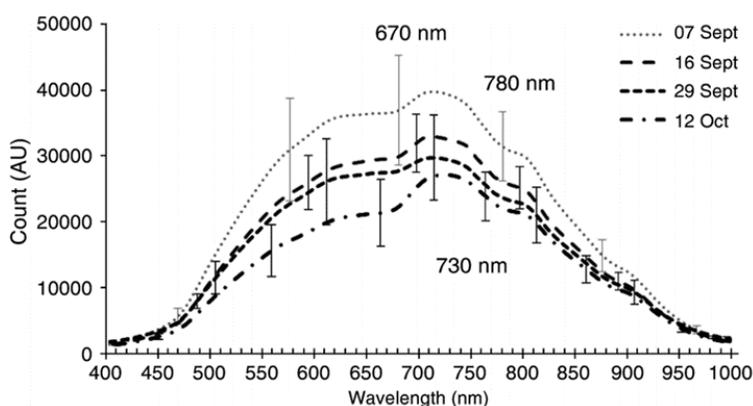


Figure 3: Average raw spectra of 1,360 berries grouped in four sampling dates corresponding to the final stages of ripening.

Discussion

Similar results for PLS models were obtained using a portable vis/NIR system (450–980 nm) showing for fresh Nebbiolo berries (same conditions as this study) $R^2_{cv} = 0.72$ and $RMSECV = 0.79$ Brix for TSS, $R^2_{cv} = 0.66$ and $RMSECV = 1.48$ g TA dm^{-3} for titratable acidity, and $R_{cv2} = 0.50$ and 0.46 for the phenolic parameters PA and EA, respectively (Guidetti et al. 2010). Similar results were also found using a portable NIR device (640–1300 nm) on winegrapes ($RMSECV$ from 1.01 to 1.27 Brix) under field conditions (Larrain et al. 2008). Portable vis/NIR devices were also tested for the prediction of ripening indices (TSS and TP) of fresh blueberries (Guidetti et al. 2008), of two apple varieties just before fruit harvest (Beghi et al. 2012), and of fresh apricot under laboratory conditions (Camps and Christen 2009), with similar results to our findings reported here. These results show the possible application of vis/NIR technology for the estimation of many ripening parameters (Nicolai et al. 2007). The application of portable devices under field

conditions is certainly more complex than laboratory experimentation, given the uncontrolled environmental conditions such as ambient light and fluctuating temperatures, as highlighted in several studies (Nicolai et al. 2007, Wang and Paliwal 2007). The explorative PCA conducted on the full vis/NIR spectra resulted in three most significant PCs explaining 100% of the total data variance (PC1, 97%; PC2, 2%; and PC3, 1%). The PCA score plot (Figure 4) shows the evolution of sample ripening as a function of sampling date on PC1. PC loadings (Figure 2) were analyzed in a search for the main wavelength bands contributing to PCs as candidate discriminators for the final stages of the ripening process. As expected, the spectra exhibit significant differences according to sampling date, corresponding to the final stages of ripening. Changes in the spectra obviously reflect modifications in quality parameters during ripening. The observed changes in the visible region spectra between 500 and 700 nm are due to changes in the amount of pigment, especially linked to anthocyanin accumulation, during ripening. This leads to a decrease in reflectance in the visible band associated with the anthocyanin absorption peak centered around 540 nm (Tamura and Yamagami 1994). For TSS and TP prediction, the statistics of the MLR models were equal to $R^2_{cv} = 0.71$ and 0.70 , $RMSECV = 0.83$ Brix and 4.3 OD280, respectively. A similar selection approach was tested on transmittance spectra (325–1075 nm) for the investigation of TSS and pH of grape juice beverages, and $RMSECV$ values, in validation, were 0.360 and 0.054 , respectively (Wu et al. 2010). These results are better than our findings, but the experimental settings under controlled laboratory conditions and the homogeneous matrix of the juice helped provide these excellent outcomes. Good results were also obtained on tea-based soft drinks (Li et al. 2007) and on beer (Liu et al. 2009), using analogous variable selection and validation approaches. The RCA method has been also applied to select the most important wavelengths to determine the TSS and pH of rice vinegars (Liu et al. 2008) and of orange juice (Cen et al. 2006). A comparison of the PLS derived from the full vis/NIR spectra and the MLR arising only from the three wavelengths was carried out for TSS and TP. The overall calibration and prediction results of the MLR models were satisfactory, although the performance of the MLR models was slightly less accurate than the PLS models. The obtained $RMSECV$ values were similar for PLS (0.81 Brix, 3.9 OD280) and MLR (0.71 Brix, 4.3 OD280) models. Similarly, the RPD value for TSS decreased from 2.26 for PLS to 2.13 for MLR and the value for TP decreased from 1.98 for PLS to 1.76 for MLR. PLS and MLR models for both the quality parameters showed very low bias values and almost the same slope (~ 0.8). Thus, only a small loss of information was noticeable between the PLS model calculated using 2048 wavelengths and the MLR model using three effective variables. Moreover, the samples were distributed closely to the regression line, indicating excellent spectral analysis performance of the PLS–RCA–MLR method. In the similar selection approach on transmittance spectra for the investigation of TSS and pH of grape juice beverages (Wu et al. 2010), the authors obtained optimal determination coefficients. PLS models derived from the full spectra (325–1075 nm) achieved, for TSS and pH, R^2 values in

validation ranging from 0.89 to 0.97 and from 0.91 to 0.96, respectively. MLR analysis was applied to verify the results of wavelength selection. For TSS, the authors obtained very good results compared with the PLS models with $R^2_{cv} = 0.97$ and 0.98 for five and nine variables selected, respectively. For pH, they achieved analogous results with determination coefficients equal to 0.96 and 0.97 , respectively. In a study on reducing sugar content in red grape must, there was a slight difference between the PLS model calculated using the full spectra (800–1050 nm) and the MLR model based on four sensitive wavelengths (RMSEPPLS = 12.20 g dm^{-3} , RMSEPMLR = 20.51 g dm^{-3}) (Fernández-Navales et al. 2009). Once the prediction capabilities using simple MLR equations were evaluated, the designed principle of a compact sized, low-cost, and easy-to-use device was studied in this work. A possible functional scheme of this device can be envisioned using LED (light-emitting diodes) technology for sample illumination at the specified wavelengths (670, 730, and 780 nm) and filtered photodiodes for the read-out signal. From the intensity signals sensed at the three wavebands, a microcontroller can compute the TSS predicted value by equation 2 and the TP value by equation 3 and display it to the user. Such a system would support the grapegrower in rapidly estimating the ripening grape parameters and making a decision on harvest time.

Table 3: Statistics of the MLR models based on the three selected wavelengths (670, 730, 780 nm) to predict the ripening parameters

Parameter ^a	Unit ^a	Calibration ^a		Cross-validation ^a					RPD
		R^2_{cal}	RMSEC	R^2_{cv}	Bias	Slope	RMSECV	RMSECV%	
TSS	Brix	0.75	0.77	0.71	-0.006	0.74	0.83	3.9%	2.13
TP	OD ₂₈₀ nm	0.74	3.95	0.70	-0.034	0.73	4.30	11.9%	1.79

^aAbbreviations: TSS, total soluble solids; TP, total polyphenols; OD, optical density; R^2_{cal} , coefficient of determination in calibration; RMSEC, root mean square error of calibration; R^2_{cv} , coefficient of determination in cross-validation; RMSECV, root mean square error of cross-validation; RMSECV%, percent errors in cross-validation; RPD, ratio performance deviation.

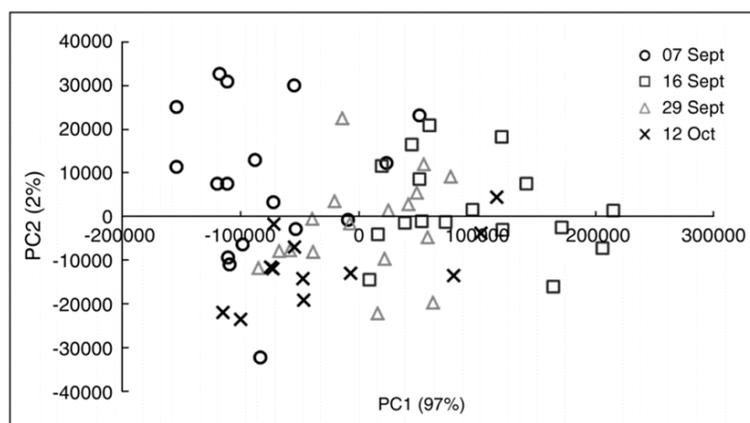


Figure4: Principal component analysis deriving from the initial full vis/ NIR spectra.

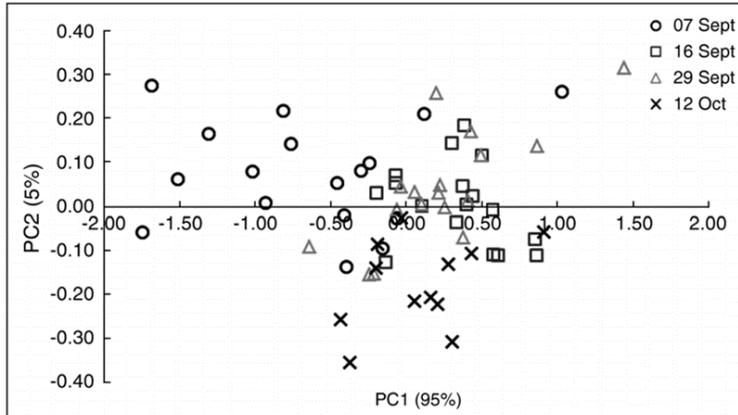


Figure 5: Principal component analysis deriving from the only three normalized spectral ratios.

Conclusion

A variable selection methodology was proposed to determine a reduced set of features that are effective in the detection of berry ripening using vis/NIR spectroscopy. The aim was to identify, using the PLS-RCA variable selection method, the effective wavelengths able to discriminate berries ready to be harvested with a simplified handheld and low-cost optical device. The three selected candidate wavelengths were 670, 730, and 780 nm. PCA and MLR were applied to the wavelengths to validate the prediction ability, compared with the PCA and PLS models using the full spectra, and to verify the effectiveness of selected variables. The overall prediction results of the MLR models, for both TSS and TP, were satisfactory. The obtained determination coefficients and RPD values were similar for PLS ($R^2 = 0.77$ and $RPD = 2.26$ for TSS, $R^2 = 0.74$ and $RPD = 1.98$ for TP) and MLR ($R^2 = 0.71$ and $RPD = 2.13$ for TSS, $R^2 = 0.70$ and $RPD = 1.76$ for TP). Both qualitative (PCA) and quantitative (MLR) analyses highlighted the results, showing good sample separation during ripening and confirming the choice of wavelengths. The potential of vis/NIR to predict TSS and TP was ascertained, and the essential wavelengths, which were strongly related to these indices, were obtained through multivariable analysis. These individual fingerprint wavelengths and the equations for TSS and TP could be used for the design of a simplified, low-cost handheld device which would allow for real-time assessment of berry ripeness in the field. In particular, this device may be based on the measurement and processing of diffuse spectral reflectance at these few appropriately selected wavelengths.

Literature Cited

- Antonucci, F., F. Pallottino, G. Paglia, A. Palma, S. D'Aquino, and P. Menesatti. 2011. Non-destructive estimation of Mandarin maturity status through portable vis-NIR spectrophotometer. *Food Bioprocess Technol.* 4(5):809-813.
- Beghi, R., A. Spinardi, L. Bodria, I. Mignani, and R. Guidetti. 2013. Apples nutraceutical properties evaluation through a visible and nearinfrared portable system. *Food Bioprocess Technol.* 6(9):2547-2554.

- Bellincontro, A., I. Nicoletti, M. Valentini, A. Tomas, D. De Santis, D. Corradini, and F. Mencarelli. 2009. Integration of nondestructive techniques with destructive analyses to study postharvest water stress of winegrapes. *Am. J. Enol. Vitic.* 60:57-65.
- Bertrand, D. 2000. Spectroscopie de l'eau. In *La Spectroscopie Infrarouge et ses Applications Analytiques*. D. Bertrand and E. Dufour (eds.), pp. 93-105. Lavoisier Tec et Doc, Paris.
- Bodria, L., M. Fiala, R. Guidetti, and R. Oberti. 2004. Optical techniques to estimate the ripeness of red-pigmented fruits. *Trans. ASAE* 47(3):815-820.
- Camps, C., and D. Christen. 2009. Non-destructive assessment of apricot fruit quality by portable visible-near infrared spectroscopy. *Food Sci. Technol.* 42(6):1125-1131.
- Cen, H., and Y. He. 2007. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.* 18:72-83.
- Cen, H., Y. He, and M. Huang. 2006. Measurement of soluble solids contents and pH in orange juice using chemometrics and vis-NIRS. *J. Agric. Food Chem.* 54:7437-7443.
- Chong, I.G., and C.H. Jun. 2005. Performance of some variable selection methods when multicollinearity is present. *Chemometr. Intell. Lab.* 78:103-112.
- Clement, A., M. Dorais, and M. Vernon. 2008. Nondestructive measurement of fresh tomato lycopene content and other physicochemical characteristics using visible-NIR spectroscopy. *J. Agric. Food Chem.* 56:9813-9818.
- Cogdill, R.P., and C.A. Anderson. 2005. Efficient spectroscopic calibration using net analyte signal and pure component projection methods. *J. Near Infrared Spec.* 13:119-132.
- Costa, G., E. Bonora, G. Fiori, and M. Noferini. 2011. Innovative nondestructive device for fruit quality assessment. *Acta Hort.* 913:575-581.
- Cozzolino, D., R.G. Damberg, L. Janik, W.U. Cynkar, and M. Gishen. 2006. Analysis of grapes and wine by near infrared spectroscopy. *J. Near Infrared Spec.* 14:279-289.
- Di, C.Z., C.M. Crainiceanu, B.S. Caffo, and N.M. Punjabi. 2009. Multilevel functional principal component analysis. *Ann. Appl. Stat.* 3:458-888. Fearn, T. 2002. Assessing calibrations: SEP, RPD, RER and R2. *NIR News* 13:12-14.
- Fernández-Navales, J., M.I. López, M.T. Sánchez, J. Morales, and V. González-Caballero. 2009. Shortwave-near infrared spectroscopy for determination of reducing sugar content during grape ripening, winemaking, and aging of white and red wines. *Food Res. Int.* 42(2):285-291.
- Frank, I.E., and R. Todeschini. 1994. *The Data Analysis Handbook*. Elsevier, Amsterdam. Glories, Y. 1984. The colour of red wines. *Conn. Vigne Vin* 18:195-217.
- Guidetti, R., R. Beghi, and L. Bodria. 2010. Evaluation of grape quality parameters by a simple vis/NIR system. *Trans. ASABE* 53(2):477-484.
- Guidetti, R., R. Beghi, L. Bodria, A. Spinardi, I. Mignani, and L. Folini. 2008. Prediction of blueberry (*Vaccinium corymbosum*) ripeness by a portable vis-NIR device. *Acta Hort.* 310:877-885.
- Larrain, M., A.R. Guesalaga, and E. Agosin. 2008. A multipurpose portable instrument for determining ripeness in wine grapes using NIR spectroscopy. *IEEE T. Instrum. Meas.* 57(2):294-302.
- Li, X., Y. He, C. Wu, and D.W. Sun. 2007. Nondestructive measurement and fingerprint analysis of soluble solid content of tea soft drink based on vis/NIR spectroscopy. *J. Food Eng.* 82:316-323.
- Liu, F., Y. He, and L. Wang. 2008. Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy. *Anal. Chim. Acta* 610:196-204.
- Liu, F., Y. Jiang, and Y. He. 2009. Variable selection in visible/near infrared spectra for linear and nonlinear calibrations: A case study to determine soluble solids content of beer. *Anal. Chim. Acta* 635:45-52.
- Magwaza, L.S., U.L. Opara, H. Nieuwoudt, P.J.R. Cronje, W. Saeys, and B. Nicolai. 2012. NIR spectroscopy applications for internal and external quality analysis of citrus fruit—A review. *Food Bioprocess Technol.* 5:425-444.
- McGlone, V.A., R.B. Jordan, and P.J. Martinsen. 2002. Vis/NIR estimation at harvest of pre- and post-storage quality indices for 'Royal Gala' apple. *Postharvest Biol. Tech.* 25:135-144.
- Naes, T., and B.H. Mevik. 2001. Regression models with process variables and parallel blocks of raw material measurements. *J. Chemometr.* 15:413-426.
- Naes, T., T. Isaksson, T. Fearn, and T. Davies. 2002. *A User-Friendly Guide to Multivariate Calibration and Classification*. NIR Publications, Chichester, UK.
- Nicolai, B.M., K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron, and J. Lammertyna. 2007. Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol. Tech.* 46:99-118.
- Puangsoambut, A., S. Pathaveerat, A. Terdwongworakul, and K. Puangsoambut. 2012. Evaluation of internal quality of fresh-cut pomelo using vis/NIR transmittance. *J. Texture Stud.* 43:445-452.

- Sun, D.W. 2010. *Hyperspectral Imaging for Food Quality Analysis and Control*. Academic Press/Elsevier, San Diego.
- Tamura, H., and A. Yamagami. 1994. Antioxidative activity of monoacylated anthocyanins isolated from Muscat Bailey A grape. *J. Agric. Food Chem.* 42:1612-1615.
- Wang, W., and J. Paliwal. 2007. Near-infrared spectroscopy and imaging in food quality and safety. *Sens. Instrumen. Food Qual.* 1:193-207.
- Williams, P.C. 2001. Implementation of near-infrared technology. In *Near-Infrared Technology in the Agricultural and Food Industries*. P. Williams and K.H Norris (eds.), pp. 145-169. Am. Association of Cereal Chemists, St. Paul, MN.
- Williams, P., and K. Norris (eds.). 2001. *Near-Infrared Technology in the Agricultural and Food Industries*. 2d ed. Am. Association of Cereal Chemists, St. Paul, MN.
- Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-regression: A basic tool of chemometrics. *Chemometr. Intell. Lab.* 58:109-130.
- Wu, D., Y. He, P. Nie, F. Cao, and Y. Bao. 2010. Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. *Anal. Chim. Acta* 659:229-237.
- Xiaobo, Z., Z. Jiewen, M.J.W. Povey, M. Holmes, and M. Hanpin. 2010. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* 667:14-32.

1.3 SELECTION OF OPTIMAL WAVELENGTHS FOR DECAY DETECTION IN FRESH-CUT *VALERIANELLA LOCUSTA* LATERR.

Valentina Giovenzana¹, Roberto Beghi¹, Raffaele Civelli¹, Cristina Malegori², Riccardo Guidetti¹

¹Department of Agricultural and Environmental Sciences
Università degli Studi di Milano, via Celoria 2, Milano, 20133 Italy

²Department of Food, Environmental and Nutritional Sciences
Università degli Studi di Milano, Via Celoria 2, 20133 Milan, Italy

The aim of this work was to study the feasibility of a simplified handheld and low-cost optical device. This study was focused on identifying the most significant vis/NIR wavelengths able to discriminate freshness levels during shelf-life of fresh-cut *Valerianella locusta* L. The shelf-life of *Valerianella* leaves was monitored using a portable commercial vis/NIR spectrophotometer and by traditional analyses (pH, moisture and total phenols content). The *Valerianella* samples were stored at three temperature: 4 C, 10 C, and 20 C. Through PLS-RCA technique, standardized regression coefficients of PLS models were used to select the relevant variables, representing the most useful information of full spectral region. The four selected wavelengths were 520 nm, 680 nm, 710 nm and 720 nm. Multiple linear regression was applied in order to verify the effectiveness of selected wavelengths. Results demonstrate the feasibility of a simplified device for quickly monitoring the shelf-life of freshcut *Valerianella* leaves.

Keywords: *Valerianella locusta* L., shelf-life, effective wavelength, vis/NIR spectroscopy, low-cost device, chemometrics

Introduction

The international Fresh-cut Produce Association (IFPA) defines fresh-cut products as “any fruit or vegetable or combination thereof that has been physically altered from its original form, but remains in a fresh state”. In recent years, a substantial increase in the consumption of fresh-cut, or minimally processed, fruit and vegetables has been occurred. Fresh-cut production raised in Italy and Europe in recent years (Rico et al., 2007). In Italy, the area of fresh-cut cultivation is approximately 6500 ha and the total harvest is 88,000 t y⁻¹ (Castoldi et al., 2011). 10% of the economic value of the fruit and vegetable market in Italy is covered by fresh-cut product (Baldi and Casati, 2008). The technological treatments extend the shelf-life of the most processed foods. Instead, ready-to-eat products are characterized by a shelf-life shorter than that of the original unprocessed raw material (Guerzoni et al., 1996). In fact the sequence of operations necessary to produce a fresh-cut product (i.e. washing, trimming, peeling and/or cutting) promotes the biochemical and microbial instability of the product itself. These foods are often subjected to rapid loss of colour, organic acids, vitamins and other compounds that determine flavour and nutritional value. Monitoring the quality decay of fresh-cut products is necessary to control the freshness level during the entire production chain and to ensure quality product for the consumer. Hence fresh-cut fruit and vegetable sector could be greatly helped by new analytical methods that are accurate, rapid and could be integrated into the production chain for better managing the shelf-life of minimally processed products and to meet consumer demand. The non-destructive techniques, and in particular the optical analysis in the region of near-infrared (NIR) and visible–near infrared (vis/NIR), have been developed considerably over the last 20 years (Guidetti et al., 2012; Nicolai et al., 2007). NIR and vis/ NIR spectroscopy are based on the study of the interaction of electromagnetic radiation with the structure of the food product. Molecular bonds like OH, CH, CO and NH are subject to vibrational energy changes when irradiated by the radiation. The energy absorption of organic molecules in vis/NIR region occurs when molecules vibrate and this is translated into an absorption spectrum (Cen and He, 2007). These approaches, however, are always related to the analysis of wide spectra (thousands of wavelengths or variables) and, therefore, require multivariate techniques for data processing to build predictive models (Williams and Norris, 2002). In order to explain the chemical information encoded in the spectral data, chemometric analysis is required (Cogdill and Anderson, 2005). For a simplification and greater diffusion of these non-destructive techniques, in recent years, interest has shifted towards the development of portable systems that could be used in pre- and post-harvest (Zude et al., 2006; Temma et al., 2002; Walsh et al., 2000). Chemometrics can be used for the selection of a small number of relevant variables, which represent the most useful information contained in the full spectra (Xiaobo et al., 2010; Sun, 2010). In this way the spectral noise and the variables containing redundant information can be eliminated. Moreover, a reduced cost for potential miniaturized devices, realized to work at only this selected wavelengths, can

be foreseen. Few examples of commercial non-destructive devices based on a small number of wavelengths are already available on the market. These applications are mainly dedicated to fruits. For example, the University of Bologna (Costa et al., 2011) patented innovative and simplified NIRs equipments, namely DA-Meter for apple and Kiwi-Meter for kiwi. These systems are used for the analysis of the ripeness level of the fruit through indices based on differences in absorbance between specific wavelengths. This type of instrument, simple and portable, can be used directly on the fruit on the trees and can help growers in taking decision regarding the best cultural management practices (such as pruning, thinning, and nutrition). In this way the heterogeneity of the product can be reduced and, therefore, can be simplified the management of product lots during post-harvest. Valerianella locusta, also known as lamb's lettuce, is a member of the family Valerianaceae. It is a diploid, autogamous crop, with the chromosome number $2n = 14$ but otherwise little genetic information (Muminovic et al., 2004). Valerianella is actually among the mostly requested baby-leaves commercialised in the Italian market. The aim of this work was to study the feasibility of a simplified handheld and low-cost optical device, based on a few wavelengths appropriately selected, for quality analysis of Valerianella during the production chain or/and to discriminate freshness levels during shelf-life directly at the point of sale. The main objective of this research was the identification of informative wavelengths using the Partial Least Square Regression Coefficients Analysis (PLS-RCA) method of variable selection, correlating the vis/NIR spectra and the Valerianella quality parameters. The prediction performances of partial least square (PLS) models based on the analysis of full vis/NIR spectra were compared with multiple linear regression (MLR) models, created using only the selected wavelengths. Simple equations for the estimation of Valerianella freshness were defined. Finally, a possible functional scheme of a compact-sized LED technology based, low-cost, and easy-to-use device was proposed.

Materials and methods

The packages of fresh-cut *V. locusta* L. used for the study were provided by a collaborative producer. The Valerianella leaves were harvested by hand in September 2012. After a minimal process, the leaves were packed in bags made by high-density polyethylene and sealed. The commercial expiration date was fixed by the producer at 4 days from the packaging date. During their transport to the laboratory, the samples were maintained at the temperature of 4 C.

Sampling

Three storage/shelf-life temperature were investigated: 4.0 ± 0.5 C, 10.0 ± 0.5 C and 20.0 ± 0.5 C. The lower temperature, 4 C, represents the optimal storage condition for fresh-cut products (Kader, 2008). The storage at 10 C simulates the realistic supermarket condition (Jacxsens et al., 2001) while 20 C is the extreme storage situation. At the last

temperature the physiological activities of fresh-cut products are accelerated (Toivonen and DeEll, 2002). The duration of the experimentation was different for the different storage temperature. The Valerianella packages preserved at 4 C and 10 C were sampled for 16 days, while samples stored at 20 C were analyzed only for 7 days, due to the rapid degradation of the lettuce at this temperature. The numbers of sampling points during storage monitoring for the fresh-cut leaves were therefore 10, 11 and 6 for 4 C, 10 C and 20 C, respectively (Table 1). The quality decay of samples was evaluated by chemical parameters (pH, moisture and total polyphenols content) and by a nondestructive optical device (vis/NIR spectroscopy).

Chemical analyses

At each temperature and sampling time all the chemical analysis were carried out in triplicate. The pH was measured using a digital pH meter (Ioncheck 45, Radiometer Analytical SAS, Lyon, France) on 20 g of sample blended for 2 min in 40 mL of deionized water. To obtain the moisture content (MC) a thermogravimetric analysis (Sartorius MA150, Bradford, UK) was performed. An amount of 5 g of sample was weighed in the analyzer and heated at 120 C. The difference between the weights of the sample, before and after being heated, represent the MC. The moisture content was calculated as grams of water per 100 g of sample. The sample leaves were crushed and 10 g were weighed in a centrifuge tube and added with 15 mL of methanol. The mixture was stirred for 1 h in the dark and centrifuged at 11,200 rpm for 10 min at 15 C. The solids were extracted two more times using 15 and 10 mL of the methanol for 15 min under shaking in the dark, and centrifuged in the above-described conditions. The gathered extracts were made up to 50 mL with the extraction solvent. The phenolic compounds quantification was based on the Folin–Ciocalteu method (Singleton and Rossi, 1969) and expressed as mg of gallic acid equivalents per 100 g of sample, by comparison with a calibration curve built with 0.5 mg kg⁻¹ , 1 mg kg⁻¹ , 2.5 mg kg⁻¹ , 5 mg kg⁻¹ and 10 mg kg⁻¹ standard solution of gallic acid (Sigma–Aldrich, Italia).

Portable vis/NIR device

Figure 1: Particular of spectral acquisitions on Valerianella leaf with JAZ vis/NIR

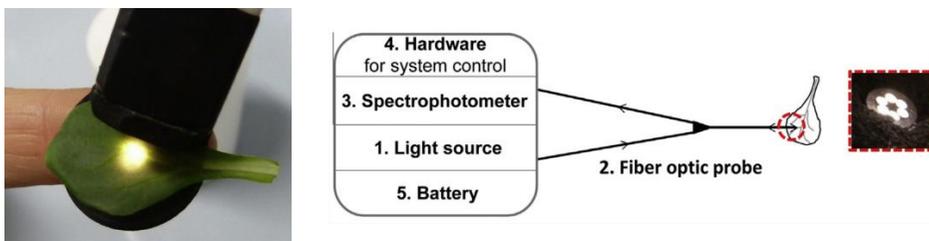


Figure 2: Scheme of the commercial portable vis/NIR spectrophotometer used for spectral acquisitions and particular of the tip of the fiber optic probe.

The optical analyses were carried out with a commercial vis/NIR spectrophotometer (JAZ, OceanOptics, USA, Fig. 1) operating between 400 and 1000 nm. The instrument consists of five components (Fig. 2): (1) a vis/ NIR halogen lamp, (2) a fiber optic probe for reflection measurement, (3) a spectrophotometer, (4) hardware for data acquisition and instrument control, (5) a battery as the power supply. The optic probe is a bidirectional Y-shaped cable (OceanOptics, USA) in which seven fibres are arranged in a 6-around-1 configuration. The six external fibers guide the light from the light source to the sample, while the single central fiber brings it back from the leaf to the spectrophotometer (Fig. 2). The integrated spectrophotometer was equipped with a diffractive grating for spectral measurements optimized in the range of 400–1000 nm and a CCD sensor with a 2048 pixel matrix, corresponding to a spectral resolution of 0.3 nm. Spectra were acquired in reflectance mode, without any sample preparation. Due to the sample thinness, a dark surface was placed on the opposite side of the leaf, in correspondence to the acquisition point. The dark surface was realized through a generic black rubber material, previously and properly tested in order to guarantee the total light absorption along the whole spectral range. In this way the light exceeding the leaf is completely absorbed by the dark surface, without any interference with the measured light reflected from the leaf. Moreover, the tip of the optical probe was equipped with soft plastic cap to ensure contact with the sample's surface during measurements in order to reduce the interference of environmental light (Fig. 1). A total number of 270 Valerianella leaves were analyzed: 100, 110 and 60 samples for 4 C, 10 C and 20 C, respectively. Measurement of 10 leaves were performed for each sampling time. Each sample was obtained by averaging 3 spectral acquisitions made in three different points of the leaf, for a total of 810 spectra. An averaged spectrum on the 10 leaves, representative of each plastic bag, was finally calculated and used for the data analysis.

Data processing

The data processing was carried out using The Unscrambler 9.6 software package (CAMO ASA, Norway). The chemometric approach allows to extract the usable information from the spectral data and to select the most significant wavelengths, in order to develop a simplified device. The collected spectra were pre-processed using smoothing (moving-average, 15 nm wide window) and reducing techniques. The correlations between spectral data matrix and chemical parameters (pH, MC and TP) were carried out using PLS regression algorithm and the variable selection was performed using regression coefficient analysis (RCA), deriving from PLS analysis (Chong and Jun, 2005; Liu et al., 2008). The regression coefficients obtained by PLS model were used to calculate the Y variable response value (pH, moisture and TP) from the X variables (Valerianella leaves spectra). The follow parameters were calculated to evaluate model accuracy (Nicolai et al., 2007; Naes et al., 2002): R² cal, coefficient of determination in calibration. R² cv, coefficient of determination in cross-validation.

RMSEC, root mean square error of calibration. RMSECV, root mean square error of cross-validation. RPD, Ratio Performance Deviation. RPD is defined as the ratio between the standard deviation of the response variable and RMSECV. If RPD is lower than 1.5, the calibration is not useful, on the contrary, if it is higher than 2, the model can perform a quantitative prediction; between 1.5 and 2.0 the algorithm have the possibility to distinguish between high and low values (Williams, 2001; Fearn, 2002). In order to extract the most useful information from the vis/NIR spectra, RCA was carried out (Xiaobo et al., 2010; Chong and Jun, 2005). This approach was already applied in literature by Cen et al. (2006), on orange juice samples for the estimation of TSS an pH as well as by Liu et al. (2008) on rice vinegars samples. Liu et al. (2009) used RCA to determine the TSS content in beer. The regression coefficients deriving from the PLS analysis were standardized considering the standard deviation of reflectance and the standard deviation of the reference data (Frank and Todeschini, 1994). The standardized regression coefficients were used for the wavelength selection. The size of these numerical coefficients gave an indication of the impact of different variables on the response (Y). Large absolute values indicate the importance of the corresponding wavelengths on the prediction of Y parameter. The final aim was to find which variables were important for predicting the Y response. Hence, RCA could be used for essential wavelength selection. As already used by Liu et al. (2008, 2009), the higher peaks (absolute values) of the regression coefficient plot were chosen. Peaks and valleys, in fact, represent the relevance of the corresponding wavelength in predicting the Y-variable. Finally, another regression method, MLR, was applied to test the effectiveness of the selected wavelengths. The effective wavelengths were employed as the input data matrix for the elaboration of MLR models (Wu et al., 2010; Fernández-Navales et al., 2009; Li et al., 2007). This approach, compared with the traditional PLS method, performs quantitative predictions using only a few important variables. MLR works even when the number of variables is less than the number of samples and is not affected by collinearity (Naes and Mevik, 2001). The MLR models were compared to PLS models, allowing to evaluate the efficiency of variable selection. Therefore, reflection intensity measured at the selected wavelengths were used to predict quality parameters, in order to determine the different freshness stages during decay.

Results and discussion

Francois et al. (2008) investigated sensory attributes of different chicory hybrids using vis/NIR spectroscopy; Ferrante and Maggiore, in 2007, evaluated storage time and temperature of Valeriana leafy vegetables using fluorescence technique and Zhang et al. (2012) determined the water content in leaves of potted plants using vis/NIR technique; no specific works are reported in literature regarding vis/NIR spectroscopy on fresh-cut Valerianella leaves. Similar portable vis/NIR device was tested by Guidetti et al. for the prediction of ripening indices (total polyphenols) of fresh blueberries (Guidetti et al., 2008) and of fresh grapes (Guidetti et al., 2010), by Beghi et al. (2012)

on two apple varieties just before fruit harvest and by Camps and Christen (2009) on fresh apricot under laboratory conditions with similar results to our findings. These results show the possible application of vis/NIR technology for the estimation of many ripening parameters, as widely shown in the literature (Nicolai et al., 2007). The plots of the standardized regression coefficients (for pH, MC and TP) vs. vis/NIR wavelengths (400–1000 nm) are shown in Fig. 3. The trend and shape of the regression coefficients plots were different for pH, moisture and TP at specific wavelengths. The final goal was to select wavelengths showing simultaneously high absolute values for all the three parameters considered. This in order to be all used at the same time for the prediction of each parameter in a hypothesis of a simplified system. The most relevant wavelengths were selected by choosing the higher absolute regression values (Liu et al., 2008; Cen et al., 2006). Therefore, the candidate effective wavelengths were: 520 nm, reflection wavelength for the green colour; 680 nm, corresponding to chlorophyll maximum absorption peak; 710 nm and 720 nm, corresponding to bands related to the evolution of chlorophyll spectral forms throughout all stages of leaf development, often not completely resolved (Gitelson et al., 1996). Moreover, the authors stated that peaks at 685–706, 710, 725 and 740 nm were dependent on different degree of leaf age and pigment concentration in the leaves. For MC, the selected wavelengths of 520 and 680 nm show very low values of standardized regression coefficient and their contribute tends to be cancelled out. The real informative wavelengths for this parameter are only 710 and 720 nm (Fig. 3).

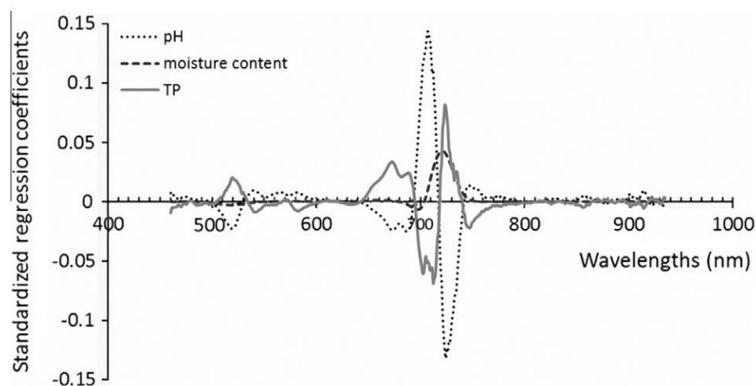


Figure 3: Standardized regression coefficients for pH, moisture content (MC) and total polyphenols (TP)

The performance of RCA was confirmed using the visual inspection of average spectral curves (Sun, 2010) and their correspondence to specific absorption peaks. The average spectra of the different shelf-life classes at storage temperature of 10 C and 20 C in the considered vis/NIR range are shown in Fig. 4. As expected, the spectra exhibit differences, according to sampling date, more evident for 20 C storage (Fig. 4B).

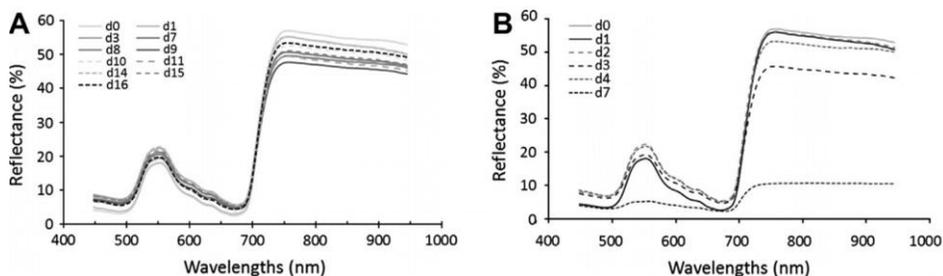


Figure 4: Example of average spectra of *Valerianella* leaf samples grouped for sampling day (from d0 to d16) for samples stored at 10 C (A) and from d0 to d7 for samples stored

Average spectra of leaves stored at 4 C are similar to those stored at 10 C and are not shown. The changes in the spectra obviously reflect modifications in quality parameters during the shelf-life. In particular, the figure shows relevant differences in the reflectance band centred around 540 nm in the visible band associated with green samples reflectance peak. The same differences could be noticed also around 680 nm, associated, in this case, with the chlorophyll absorption peak. In the NIR region a maximum reflectance peak is shown around 740 nm. The spectral reflection intensities measured at the selected wavelengths were finally used to predict the quality parameters for the determination of different freshness levels of the leaves. According to the variable selection methods stated above, the selected wavelengths were employed as the inputs for MLR model elaboration for pH, MC and TP. The results are shown in Table 2. The prediction ability of the MLR models was verified to study the efficiency of the selected wavelengths. For pH, moisture and TP prediction, the statistics of the MLR models were equal to $R^2_{cv} = 0.70, 0.75, 0.80$ and $RPD = 1.83, 2.08, 2.48$, respectively. For MC, an alternative MLR model was carried out using only 710 and 720 nm informative wavelengths. Results demonstrated similar determination coefficients, in calibration and validation, compared to the MLR model with the four effective wavelengths (data not shown). A RPD value between 2 and 2.5 indicates that coarse quantitative predictions are possible (Nicolai et al., 2007). In this work RPD values 3 is considered adequate for analytical purposes in most NIRs applications for agricultural products (Williams, 2001; Fearn, 2002). Nicolai et al. (2007) wrote a review about the application of NIR spectroscopy for measurement of fruit and vegetable quality and indicated that RPD values between 2.5 and 3 or above corresponds to good and excellent prediction accuracy, respectively. A comparison between PLS derived from the full vis/NIR spectra (400–1000 nm) and MLR arising only from the four wavelengths was carried out. The overall calibration and prediction results of the MLR models, for all the parameters, were satisfactory, although the performance of the MLR models was slightly worse than the good PLS models. RPD value for pH decreased from 2.54 for PLS to 1.83 for MLR, regarding the MC, RPD showed a slight decrement from 2.25 to 2.08, and for TP from 3.19 to 2.48 for PLS and MLR respectively (Table 2). This means that only

a small loss of information was noticeable between the PLS models calculated using 2048 wavelengths and the MLR models employing only the four effective variables. Moreover, the samples were distributed closely to the regression line, which shows excellent spectral analysis performance of the PLSRCA-MLR method. Wu et al. (2010) applied a similar selection approach on transmittance spectra for the investigation of TSS and pH of grape juice beverages. The authors obtained optimal values for the coefficients of determination, in validation, for both TSS that for pH ranging from 0.89 to 0.97 and 0.91 to 0.96, respectively, relating to the PLS models derived from the full spectra (325–1075 nm). MLR analysis was applied to verify the results of wavelength selection. Regarding TSS, the authors obtained very good results compared with the PLS models with $R^2_{cv} = 0.97$ and 0.98 for five and nine variables selected, respectively. For pH, they achieved analogous results with coefficients of determination equal to 0.96 and 0.97 , respectively. Good results were also obtained by Li et al. (2007) on tea-based soft drinks and by Liu et al. (2009) on beer, using analogous variable selection and validation approaches. Zhang et al. (2012) proposed (Bipls-SPA) a method to select 25 wavelengths for the estimation of water content in ornamental plant leaves using vis/NIR spectroscopy. PLS model deriving from the full spectrum (200–1100 nm) showed RPD equal to 3.66. After the selection PLS model gave an higher RPD value of 4.86.

Table 2: Statistics of the MLR models, based on the four selected wavelengths (520, 680, 710, 720 nm) to predict the freshness level of Valerianella leaf samples, and of the PLS models

Quality parameters	N ^a	Mean	SD	Calibration MLR			Cross-validation MLR			Cross-validation PLS		
				R ²	RMSEC	RPD	R ²	RMSECV	RPD	R ²	RMSECV	RPD
pH	24	6.45	0.33	0.82	0.13	2.54	0.70	0.18	1.83	0.86	0.13	2.54
TP (mg/100g _{eq} gallic acid)	16	267	40.3	0.88	12.38	3.26	0.80	16.28	2.48	0.89	12.64	3.19
MC (%)	24	93.38	0.27	0.87	0.09	3	0.75	0.13	2.08	0.84	0.12	2.25

SD = standard deviation; LV = latent variables.

The selection based on the analysis of the main changes in leaves' optical spectra led to an equation having the general form,

$$Y = b_1 \cdot I_{520} + b_2 \cdot I_{680} + b_3 \cdot I_{710} + b_4 \cdot I_{720} + b_0 \quad \text{Eq.1}$$

where the parameters b_1 , b_2 , b_3 , b_4 and b_0 are computed from a multi linear fit of known pairs' values (spectral intensities measured at the four wavelengths, I_{520} , I_{680} , I_{710} and I_{720} , and corresponding chemical data) for the Valerianella leaves using MLR analysis. Equations of MLR models for pH, MC and TP are reported hereafter:

$$\text{Model}_{\text{pH}} \quad Y_{\text{pH}} = -1.00 \cdot I_{520} - 16.76 \cdot I_{680} + 9.88 \cdot I_{710} - 10.61 \cdot I_{720} + 6.66 \quad \text{Eq.2}$$

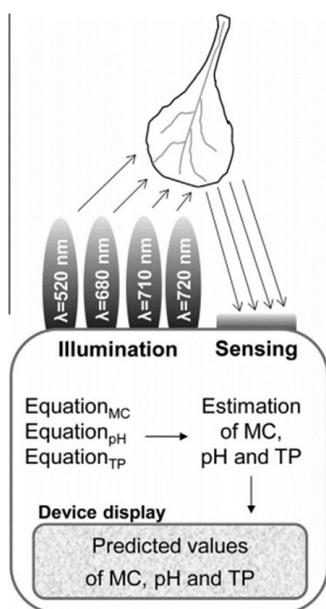
$$\text{Model}_{\text{MC}} \quad Y_{\text{MC}} = 3.74 \cdot I_{520} - 4.67 \cdot I_{680} - 5.80 \cdot I_{710} + 7.37 \cdot I_{720} + 92.31 \quad \text{Eq.3}$$

$$\text{Model}_{\text{TP}} \quad Y_{\text{TP}} = 665.8 \cdot I_{520} + 1835.0 \cdot I_{680} - 724.0 \cdot I_{710} + 660.8 \cdot I_{720} + 242.9 \quad \text{Eq.4}$$

Once the prediction capabilities using simple MLR equations were evaluated, the designed principle of a compact-sized, low-cost, and easy-to-use device was proposed.

In order to highlight the innovative features of a future device, in Fig. 5 a scheme was proposed. A possible solution consists of a 4 LED illumination source (light emitting diodes) at the specific wavelengths (520, 680, 710 and 720 nm), with filtered photodiodes for the readout signal. The elaboration unit, using equations with a general form like [1], assigns specific coefficients to the signal intensity recorded at the four wavelengths coming from the LEDs photodiodes system. In this way, a prediction of pH, MC and TP is realized. Finally, a display shows the estimation to the user. The foreseen system would support the producers to quickly predict the freshness of fresh-cut *V. locusta* L. and, therefore, take decisions regarding the product management directly at the selling point.

Conclusions



In this work, a variable selection methodology has been proposed in order to select a reduced set of wavelengths that are essential in the detection of minimal processed *V. locusta* L. freshness, using vis/NIR spectroscopy. The four selected candidate wavelengths were 520 nm, 680 nm, 710 nm and 720 nm. MLR was applied to the wavelengths in order to validate the prediction ability, compared with the PLS models built using the full spectra, and to verify the effectiveness of selected variables. The overall prediction results of the MLR models, for the three analyzed parameters (pH, MC and TP), were satisfactory. The obtained determination coefficients and RPD values were similar for the PLS and MLR models. These individual fingerprint wavelengths and simply equations, with a general form like [1], could be used for the design of a simplified handheld device

which would allow real-time assessment of *Valerianella* freshness. In particular, this device may be based on the measurement and processing of diffuse spectral reflectance at these few appropriately selected wavelengths. A possible functional scheme of a compact-sized LED technology based, low-cost, and easy-to-use device was proposed. This type of simplified optical tool, avoiding specific chemometric analyses and trained personnel, could support the conventional techniques in the shelf-life assessment of fresh-cut *Valerianella* providing information useful for a better management of the product along the distribution chain. Moreover, the implementation of the device directly at the point of sale should be a guarantee for the consumer.

Acknowledgements

This study received financial support from AGER as “STAYFRESH – Novel strategies meeting the needs of the fresh-cut vegetable sector” research project and from Regione Lombardia and European Social Fund for a Post-doctoral Research Fellowship (“Progetto Dote Ricerca”).

References

- Baldi, L., Casati, D., 2008. The fresh-cut crops sector and his development in Lombardia Region: the economic aspects (Il settore della IV gamma e il suo sviluppo in Lombardia: gli aspetti economici). In: Società Agraria di Lombardia e Camera di Commercio di Bergamo (Ed.), *La IV gamma, Innovazione nell'offerta dei prodotti orticoli*. Bergamo, Italy (in Italian).
- Beghi, R., Spinardi, A., Bodria, L., Mignani, I., Guidetti, R., 2012. Apples nutraceutic properties evaluation through a visible and near-infrared portable system. *Food Bioprocess Technol.* <http://dx.doi.org/10.1007/s11947-012-0824-7>.
- Camps, C., Christen, D., 2009. Non-destructive assessment of apricot fruit quality by portable visible–near infrared spectroscopy. *Food Sci. Technol.* 42 (6), 1125–1131.
- Castoldi, N., Bechini, L., Ferrante, A., 2011. Fossil energy usage for the production of baby leaves. *Energy* 36 (1), 86–93.
- Cen, H., He, Y., 2007. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.* 18, 72–83.
- Cen, H., He, Y., Huang, M., 2006. Measurement of soluble solids contents and pH in orange juice using chemometrics and vis–NIRS. *J. Agric. Food Chem.* 54, 7437–7443.
- Chong, I.G., Jun, C.H., 2005. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* 78, 103–112.
- Cogdill, R.P., Anderson, C.A., 2005. Efficient spectroscopic calibration using net analyte signal and pure component projection methods. *J. Near Infrared Spectrosc.* 13 (3), 119–132.
- Costa, G., Bonora, E., Fiori, G., Noferini, M., 2011. Innovative non-destructive device for fruit quality assessment. *Acta Horticulturae* 913, 575–581.
- Fearn, T., 2002. Assessing calibrations: SEP, RPD, RER and R2. *NIR News* 13, 12–14.
- Fernández-Navales, J., López, M.I., Sánchez, M.T., Morales, J., González-Caballero, V., 2009. Shortwave-near infrared spectroscopy for determination of reducing sugar content during grape ripening, winemaking, and aging of white and red wines. *Food Res. Int.* 42 (2), 285–291.
- Ferrante, A., Maggiore, T., 2007. Chlorophyll a fluorescence measurements to evaluate storage time and temperature of Valeriana leafy vegetables. *Postharvest Biology and Technology* 45 (1), 73–80.
- Francois, I.M., Wins, H., Buysens, S., Godts, C., Van Pee, E., Nicolai, B., De Proft, M., 2008. Predicting sensory attributes of different chicory hybrids using physicochemical measurements and visible/near infrared spectroscopy. *Postharvest Biol. Technol.* 49, 366–373.
- Frank, I.E., Todeschini, R., 1994. *The data analysis handbook. Data Handling in Science and Technology*. Elsevier, Amsterdam, The Netherland, ISBN 0-444- 81659-3.
- Gitelson, A.A., Merzlyak, M.N., Lichtenthaler, H.K., 1996. Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *J. Plant Physiol.* 148, 501–508.
- Guerzoni, M.E., Gianotti, A., Corbo, M.R., Sinigaglia, M., 1996. Shelf life modelling for fresh-cut vegetables. *Postharvest Biol. Technol.* 9, 195–207.
- Guidetti, R., Beghi, R., Bodria, L., Spinardi, A., Mignani, I., Folini, L., 2008. Prediction of blueberry (*Vaccinium corymbosum*) ripeness by a portable vis–NIR device. *Acta Horticulturae* 310, 877–885.
- Guidetti, R., Beghi, R., Bodria, L., 2010. Evaluation of grape quality parameters by a simple vis/NIR system. *Trans. ASABE* 53 (2), 477–484.
- Guidetti, R., Beghi, R., Giovenzana, V., 2012. *Chemometrics in food technology. Chemometrics in Practical Applications*. InTech, ISBN: 978-953-51-0438-4.
- Jacxsens, L., Devlieghere, F., Debevere, J., 2001. Temperature dependence of shelflife as affected by microbial proliferation and sensory quality of equilibrium modified atmosphere packaged fresh produce. *Postharvest Biol. Technol.* 26, 59–73.
- Kader, A.A., 2008. Perspective flavor quality of fruits and vegetables. *J. Sci. Food Agric.* 88, 1863–1868.
- Li, X., He, Y., Wu, C., Sun, D.W., 2007. Nondestructive measurement and fingerprint analysis of soluble solid content of tea soft drink based on vis/NIR spectroscopy. *J. Food Eng.* 82, 316–323.

- Liu, F., He, Y., Wang, L., 2008. Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy. *Anal. Chim. Acta* 610, 196–204.
- Liu, F., Jiang, Y., He, Y., 2009. Variable selection in visible/near infrared spectra for linear and nonlinear calibrations: a case study to determine soluble solids content of beer. *Anal. Chim. Acta* 635, 45–52.
- Muminovic, J., Melchinger, A.E., Lubberstedt, T., 2004. Genetic diversity in cornsalad (*Valerianella locusta*) and related species as determined by AFLP markers. *Plant Breeding* 123, 460–466.
- Naes, T., Mevik, B.H., 2001. Regression models with process variables and parallel blocks of raw material measurements. *J. Chemom.* 15, 413–426.
- Naes, T., Isaksson, T., Fearn, T., Davies, T., 2002. A user-friendly guide to multivariate calibration and classification. NIR Publications, Chichester, UK, ISBN 0- 9528666-2-5.
- Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyna, J., 2007. Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol. Technol.* 46, 99–118.
- Rico, D., Martín-Diana, A.B., Barat, J.M., Barry-Ryan, C., 2007. Extending and measuring the quality of fresh-cut fruit and vegetables: a review. *Trends Food Sci. Technol.* 18 (7), 373–386.
- Singleton, V., Rossi, J.A., 1969. Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. *Am. J. Enology Viticulture* 16, 144–158.
- Sun, D.W., 2010. Hyperspectral imaging for food quality analysis and control. In: Sun, Da-Wen. (Ed.). Elsevier, ISBN 978-0-12-374753-2.
- Temma, T., Hanamatsu, K., Shinoki, F., 2002. Development of a portable near infrared sugar-measuring instrument. *J. Near Infrared Spectrosc.* 10, 77–83.
- Toivonen, P.M.A., DeEll, J.E., 2002. Physiology of fresh-cut fruit and vegetables. In: Lamikanra, I. (Ed.), *Fresh-cut fruits and vegetables*. CRC Press LLC, Boca Raton, Florida, USA, pp. 91–117.
- Walsh, K.B., Guthrie, J.A., Burney, J.W., 2000. Application of commercially available, low-cost, miniaturised NIR spectrometers to the assessment of the sugar content of intact fruit. *Aust. J. Plant Physiol.* 27, 1175–1186.
- Williams, P.C., 2001. Implementation of near-infrared technology. In: Williams, P., Norris, K.H. (Eds.), *Near-Infrared Technology in the Agricultural and Food Industries*. American Association of Cereal Chemist, St. Paul, Minnesota, USA, pp. 145–169.
- Williams, P., Norris, K., 2002. *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd ed. American Association of Cereal Chemist, St. Paul, Minnesota, USA.
- Wu, D., He, Y., Nie, P., Cao, F., Bao, Y., 2010. Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. *Anal. Chim. Acta* 659 (1–2), 229–237.
- Xiaobo, Z., Jiewen, Z., Povey, M.J.W., Holmes, M., Hanpin, M., 2010. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* 667, 14–32.
- Zhang, Q., Li, Q., Zhang, G., 2012. Rapid determination of leaf water content using vis/NIR spectroscopy analysis with wavelength selection. *Hindawi Publishing Corporation Spectrosc.: Int. J.* 27 (2), 93–105. <http://dx.doi.org/10.1155/2012/276795>.
- Zude, M., Herold, B., Roger, J.-M., Bellon-Maurel, V., Landahl, S., 2006. Nondestructive tests on the prediction of apple fruit flesh firmness and soluble solids content on tree and in shelf life. *J. Food Eng.* 77, 254–260.

1.4 COMPARISON BETWEEN FT-NIR AND MICRO-NIR IN THE EVALUATION OF ACEROLA FRUIT QUALITY, USING PLS AND SVM REGRESSION ALGORITHMS

Cristina Malegori ^a, Emanuel Marques ^b, Maria Fernanda Pimentel ^b, Sergio Tonetto de Freitas ^c, Flávio de França Souza ^c, Celio Pasquini ^d, Ernestina Casiraghi ^a

^a DeFENS Department of Food, Environmental and Nutritional Sciences, Università degli Studi di Milano – Milano - ITALY

^b Departamento de Engenharia Química, Universidade Federal de Pernambuco – Recife (PE) – BRAZIL

^c Brazilian Agricultural Research Corporation, Embrapa Semiárido, Petrolina (PE) – BRAZIL

^d Instituto de Química, Universidade Estadual de Campinas, Campinas (SP) – BRAZIL

Abstract

Acerola (*Malpighia emarginata*) is a typical Brazilian super-fruit characterized by a huge amount (from 1.0 to 4.5 mg/100 g) of ascorbic acid. Fresh acerola fruit is mainly transformed with a very high post-harvest waste, up to 40%, due to the high perishability of acerolas. Therefore, the aim of this work is to estimate in a non-destructive manner, titratable acidity and ascorbic acid content of acerolas for a better storage management and waste reduction. The fruit quality was evaluated by NIR spectroscopy comparing the smallest portable device on the market (Micro-NIR 1700) with a bench top FT-NIR.

The variability of chemical parameters follow a non-strictly linear profile. For this reason, two different regression algorithm were compared: PLS and SVM. Regression models obtained with Micro-NIR spectra give better results using SVM algorithm, for both ascorbic acid and titratable acidity estimation. NIR data give comparable results both using SVM and PLS algorithms with lower errors for SVM regression. The prediction ability of the two instrument in prediction is statistically comparable (Duncan test) so Micro-NIR appears to be suitable for on field monitoring.

Introduction

Numerous studies highlight the potential of NIR spectroscopy, applied to fruit and vegetables, and enhance the use of this technique as a rapid and non-destructive analytical tool useful for determining internal and external characteristics, either quantitative or qualitatively (Nicola et al., 2007). Even if, considerable attention has been given to the miniaturization and portability of spectroscopic devices, only within the last few years, real handheld near-infrared scanning spectrometers become commercially available (Sorak et al., 2012). A survey of scientific papers published in the last decade shows a steady increase in the number of research and development studies conducted using these types of portable spectrometers (Teixeira Dos Santos et al., 2013). It is clear that the great advantage of these systems is the possibility of be implemented for new applications on-site and on-line at an industrial level but the potential of these instruments can only be realized if the reduction in size does not compromise the performance of the spectrometer (O'Brien et al., 2012). Although most comparison studies reported that the portable instruments had lower performance scores than the laboratory instruments, the main conclusions were that their flexibility and possibility of field-use were major advantages that made the portable options the best solution (Teixeira Dos Santos et al., 2013). Nowadays most of the scientific studies carried out in the post-harvest with hand-held device are performed under laboratory conditions rather than in field environment (Teixeira Dos Santos et al., 2013). Nevertheless, the fruit and vegetable sector demands simple equipment capable of performing in field analyses to allow producers to evaluate the proper "on-tree ripening" and to establish the most suitable harvest dates, shelf-life and storage conditions consequently ensuring the best possible crop quality levels required by an increasingly more demanding market, while avoiding waste and loss (Teixeira Dos Santos et al., 2013). However, in the last recent years, fruit quality standard was enhanced, although consumers were not always satisfied of the quality of the fruits at the point of sale (Crisosto and Crisosto, 2005).

The elaboration of spectral data obtained with the NIR devices, requires a multivariate statistical approach for extracting useful information from the acquired signals because wavelength-dependent scattering effects, instrumental noise, ambient effects, and other sources of variability may complicate the spectra (Teixeira Dos Santos et al., 2013). To relate the spectra (dependent variables) to the specific key product quality parameters (independent variables) a regression method is required (Thissen 2004). Consequently, many studies are focused on finding the calibration models, testing and comparing several different pre-processing techniques, and optimizing regression methods (Teixeira Dos Santos et al., 2013).

In this work two different regression algorithm are compared: Partial Least Squares (PLS) and Support Vector Machines (SVM). The first one is widely used to make regression models because of its simplicity to use, speed, relative good performance and easy accessibility while a possible large advantage of SVR is its ability to model nonlinear relations (Thissen et al., 2004). In the field of chemistry, and more specifically,

chemometric, only a few applications of SVMs for classification or regression tasks have been published. For detailed in-depth theoretical background on SVM, literature presents interesting works (Suykens and Vandewalle, 1999; Zhang, 2001)

The global aim of this work is to estimate in a non-destructive manner, titratable acidity and ascorbic acid content of acerolas for a better storage management and waste reduction. Reaching this goal, the comparison between a traditional bench-top FT-NIR with the smallest portable NIR device on the market, Micro-NIR from JSDU, was carried out. The estimation ability of the acerola quality parameters using the two instruments was compared using two different regression algorithm: SVM, for non-linear models, and PLS, the classical linear model.

Material and methods

117 acerola fruits characterized by different ripeness stages, depending on the colour of the peel, were analysed during fourteen sampling days.

For the non-destructive acquisitions, two different NIR device were used: a Perkin Elmer Frontier FT-Spectrometer, working in the range of 12000 – 4000 cm^{-1} , equipped with a Reflectance Accessory (NIRA); a MicroNIR 1700 (950 – 1650 nm), an ultra-compact and low-cost device created by JSDU. The MicroNIR weighs less than 60 g, allows measurements from 850 to 2450 nm, and have the option of interchanging different sampling accessories; more instrument details are available in O'Brien et al. 2012.

Regarding calibration parameters, the titratable acidity was obtained by a titration of 1g of acerola's juice, diluted in 50mL of distilled water, with a solution of NaOH 0,1N and the results were expressed in malic acid percentage. The quantification of ascorbic acid was carried out by the titration of 1g of acerola's juice, diluted in 100mL of oxalic acid (0.5%), with a solution of DFI (2,6-dichlorophenol indophenol, 0.02 %). The results were expressed in mg of ascorbic acid /100 g of acerola juice.

The spectral data were modelled using two different regression algorithms, PLS (partial least square) and SVM (support vector machine). The first one algorithm is used because of its simplicity, good performance and easy accessibility while SVM is intended to model non-linear relations with high dimensional input vector. The data were divided in two sets: the first one used for the calibration (77 fruit) and the second one as external test set (40 fruits). The robustness of the models was compared using the RMSE in calibration and in validation, joint with the bias and the R2. The comparison between the estimation ability in prediction of the two instrument was statistically compared using a Duncan test on the ANOVA results.

Results

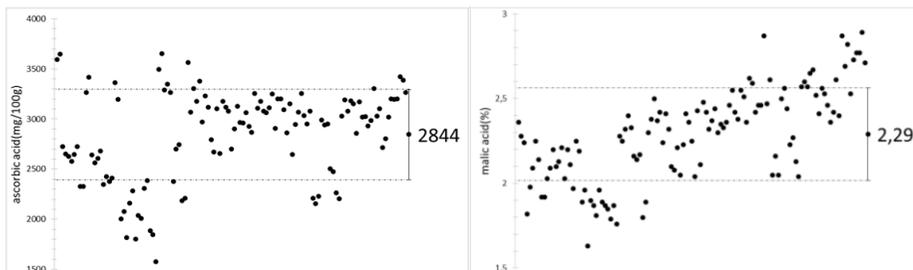


Figure 1: the determinations of the traditional quality parameters, ascorbic acid and titratable acidity, in all the 117 samples.

In figure number 1, the determinations of the traditional quality parameters of the 117 fruits are presented. The mean for the acid ascorbic content is equal to 2844 ± 449 mg/100g (CV%=15,8%) while the titratable acidity is 2.29 ± 0.27 g/100g (CV%=11,9%), expressed as malic acid. To better understand the variability of these data, a repeatability test was carried out for both the chemical analysis, carrying out ten trials of the same sample. The following results were obtained: 1526 ± 27 (CV%=1,8%) mg of ascorbic acid/100g and $1.93\% \pm 0.04$ (CV%=2,1%) of malic acid. In light of these results, the standard deviation of the destructive data, although it is not so wide, is ascribable to the differences among the fruits. These consideration laid the foundations for the correct interpretation of the regression models implemented with the NIR data.

The models parameter are presented in the following two tables; table 1 using FT-NIR and table 2 using Micro-NIR. Inside each table, it is possible to compare the two algorithms applied on both the ascorbic acid and titratable acidity determinations. The models were developed in calibration with 77 samples verifying the prediction ability with the remaining 40 samples.

The PLS regression on FT-NIR data gives similar results for both titratable acidity and ascorbic acid content. The coefficient of determination and the root mean square error are similar in calibration and in prediction, underlining the robustness of the models. The SVM, that well fits non linear models, gives better results for both the chemical parameters with higher R2 and lower errors in calibration. Also regarding the Micro-NIR models, presented in figure number 3, the models fitted using the SVM algorithms gives better results both in calibration and in prediction obtaining very robust models for both the quality parameters.

FT-NIR

PLS	number of samples calibration / prediction	Mean	S.D.%	Calibration			Prediction		
				R ²	RMSEC	bias	R ²	RMSEP	bias
Tritatable Acidity (% malic acid)	77/40	2.3	11.8	0.63	0.17	0	0.74	0.14	-0.02
Ascorbic acid content (mg/100g)	77/40	2844	15.8	0.65	284.92	4.55*10 ⁻¹³	0.49	342.45	-29.06

SVM	number of samples calibration (prediction)	Mean	S.D.%	Calibration			Prediction		
				R ²	RMSEC	bias	R ²	RMSEP	bias
Tritatable Acidity (% malic acid)	77/40	2.3	11.8	0.75	0.14	-0.006	0.69	0.16	-0.02
Ascorbic acid content (mg/100g)	77/40	2844	15.8	0.91	145.49	4.75	0.49	359.69	-46.68

Table 1: calibration and prediction models for FT-NIR using two algorithms: SVM and PLS

MicroNIR

PLS	number of samples calibration / prediction	Mean	S.D.%	Calibration			Prediction		
				R ²	RMSEC	bias	R ²	RMSEP	bias
Tritatable Acidity (% malic acid)	77/40	2.3	11.8	0.40	0.18	4.44*10 ⁻¹⁶	0.66	0.18	-0.004
Ascorbic acid content (mg/100g)	77/40	2844	15.8	0.42	310.75	9.09*10 ⁻¹³	0.40	404.37	-24.53

SVM	number of samples calibration (prediction)	Mean	S.D.%	Calibration			Prediction		
				R ²	RMSEC	bias	R ²	RMSEP	bias
Tritatable Acidity (% malic acid)	77/40	2.3	11.8	0.78	0.11	0.008	0.72	0.16	0.02
Ascorbic acid content (mg/100g)	77/40	2844	15.8	0.71	220.89	5.19	0.65	318.19	-1.64

Table 2: calibration and prediction models for MicroNIR using two algorithms: SVM and PLS

Regarding the comparison between the two devices, the benchtop FT-NIR and the portable Micro-NIR, a Duncan test on the ANOVA results was carried out. At a first glance, models developed using the benchtop FT-spectrometer give better results than those obtained by the portable device. Therefore, as it is shown in figure 4, the mean and standard deviation of the data obtained with the two instruments were compared not only among them but also with the chemical determinations. It is interesting that no statistical differences were highlighted by Duncan test, allowing to conclude that the goodness of the spectroscopic data is comparable with the chemical data. Moreover, no differences were found between the two devices.

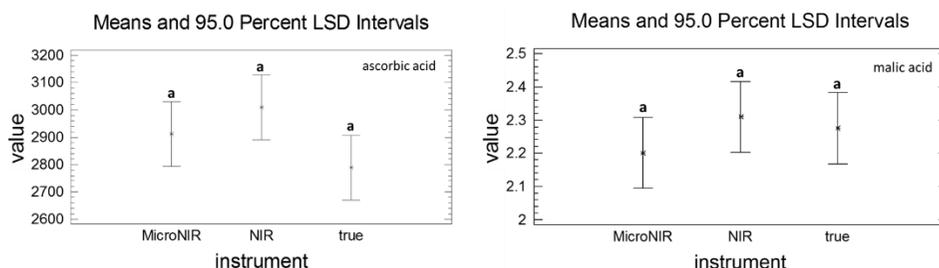


Figure 4: Duncan test on the ANOVA results between the two instruments and the chemical data

Discussion and conclusion

The Support Vector Machine algorithm gives better results for the modelling of quality parameters in acerola, laying the basis for the application of this non-linear approach in other postharvest application. In particular, the application of SVM make the difference in the construction of Micro-NIR model. Regarding the comparison between the two devices, the prediction ability of the two instrument is statistically comparable and reliable if compared with the analytical results. Ones again, the ability of spectroscopy for the non-destructive, rapid and easy determination of chemical features in fruit are demonstrated. The simple construction of the Micro-NIR, the fewer number of wavelengths and the portability make the device very interesting and reliable in field use.

References

- Crisosto, C.H., Crisosto, G.M., 2005. Relationship between ripe soluble solids concentration (RSSC) and consumer acceptance of high and low acid melting flesh peach and nectarine (*Prunus persica* (L.) Batsch) cultivars. *Postharvest Biol. Technol.* 38, 239–246. doi:10.1016/j.postharvbio.2005.07.007
- Nicola, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy : A review. *Postharvest Biol. Technol.* 46, 99–118. doi:10.1016/j.postharvbio.2007.06.024
- O'Brien, N.A., Hulse, C.A., Friedrich, D.M., Van Milligen, F.J., von Gunten, M.K., Pfeifer, F., Siesler, H.W., 2012. Miniature near-infrared (NIR) spectrometer engine for handheld applications, in: *Next-Generation Spectroscopic Technologies V*. pp. 3–9. doi:10.1117/12.917983
- Sorak, D., Herberholz, L., Iwascek, S., Altinpinar, S., Pfeifer, F., Siesler, H.W., 2012. New Developments and Applications of Handheld Raman, Mid-Infrared, and Near-Infrared Spectrometers. *Appl. Spectrosc. Rev.* 47, 83–115. doi:10.1080/05704928.2011.625748
- Suykens, J.A.K., Vandewalle, J., 1999. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* 9, 293–300. doi:10.1023/A:1018628609742
- Teixeira Dos Santos, C. a., Lopo, M., Páscoa, R.N.M.J., Lopes, J. a., 2013. A review on the applications of portable near-infrared spectrometers in the agro-food industry. *Appl. Spectrosc.* 67, 1215–1233. doi:10.1366/13-07228
- Thissen, U., Pepers, M., Üstün, B., Melssen, W.J., Buydens, L.M.C., 2004. Comparing support vector machines to PLS for spectral regression applications. *Chemom. Intell. Lab. Syst.* 73, 169–179. doi:10.1016/j.chemolab.2004.01.002
- Zhang, T., 2001. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. *AI Mag.* 22, 103. doi:10.1609/aimag.v22i2.1566.

2. IMAGING

The objective of the traditional application of image analysis is the quantification of geometric characteristic of images, acquired in a form that represents meaningful information (at macro and microscopic level) of object appearance (Diezac 1988). The evolution of these techniques and their implementation in the vision machine, allows a wide flexibility of applications, a high capacity of calculation and a rigorous statistical approach. These reasons suggest why scientific research starts to apply imaging for the evaluation of internal and external quality characteristics of food, valued according to the optical properties of the products. With a suitable light source, it is possible to extract information about colour, shape, size and texture. From these features, it is possible to know many objective aspects of the sample, which are correlated, through statistical analysis, with the food quality parameters (Du & Sun 2006, Zheng et al. 2006).

For the early identification of the biofilm, various methods, such as the total count of viable cells, microscopy, optical density and ATP determination are commonly used (Kumar & Anand, 1998). The present study aims to determine if the traditional RGB imaging could be able to distinguishing clean surfaces from surfaces colonized by a thin layer of biofilm, not jet observable by the naked eye. To reach this aim, the evaluation of surface texture was carried out (Fongaro & Kvaal, 2013). The texture of an image is given by a difference in the spatial distribution, in the frequency and in the intensity of the values of the grey levels of each pixel of the image. This technique was applied for the early detection of biofilm in its early stages of development, when it is still difficult to observe it by the naked eye. The chemometric approach that allow to obtain interesting results from this application is the grey level co-occurrence matrix that describes how each pixel is different from its neighbours. The traditional RGB imaging could be optimally used for a rapid and constant monitoring of the hygienic condition of surfaces in food industry.

- Chen, J., 2007. Surface texture of foods: perception and characterization. *Crit. Rev. Food Sci. Nutr.* 47, 583–598. doi:10.1080/10408390600919031
- Diezak, J.D. (1988). Microscopy and imagine analysis for R&D, Special report. *Food Technol.* pp.110-124
- Du, C.J. & Sun, D.W. (2006). Learning techniques used in computer vision for food quality evaluation: a review. *Journal of food engineering.* Vol.72, N.1, pp. 39–55
- Fongaro, L., Kvaal, K., 2013. Surface texture characterization of an Italian pasta by means of univariate and multivariate feature extraction from their texture images. *Food Res. Int.* 51, 693–705. doi:10.1016/j.foodres.2013.01.044
- Kumar, C.G., Anand, S., 1998. Significance of microbial biofilms in food industry: a review. *Int. J. Food Microbiol.* 42, 9–27. doi:10.1016/S0168-1605(98)00060-9

2.1 IMAGE TEXTURE ANALYSIS, A NON-CONVENTIONAL TECHNIQUE FOR EARLY DETECTION OF BIOFILM

Cristina Malegori ^a, Laura Franzetti ^a, Riccardo Guidetti ^b, Ernestina Casiraghi ^a

^a DeFENS Department of Food, Environmental and Nutritional Sciences
Università degli Studi di Milano – Milano - ITALY

^b Department of Agricultural and Environmental Sciences – Production, Landscape, Agroenergy,
Università degli Studi di Milano - Milano, ITALY

Abstract

Biofilm is a thin layer of microorganisms coated on a surface, linked by an extracellular matrix made by polysaccharides (EPS) synthesized by microorganisms themselves. The bacteria involved may deteriorate and impair the food and, if pathogenic, the microorganisms pose a risk to the consumer health. Biofilm has a particular resistance to detergents and antibiotics, making difficult its removal and surface sanification. In this work the image texture potential, for the detection of biofilm in the early stages of development, was evaluated. The evaluation was carried out on specimens (10x10 cm) of steel, plastic and ceramic for food use. As biofilm forming microorganism *Pseudomonas fluorescens* was chosen. The specimens were placed in a photographic chamber, where images were acquired at regular time intervals, for 7 days. During the timeframe of image acquisition, biofilm formation was monitored by means of classical bacteriology. The images obtained were processed with ImageJ software for image analysis, creating a co-occurrence grey levels matrix (GLCM). The results were then processed with the software MATLAB using PLS Toolbox to perform principal component analysis (PCA). Results shown that image analysis can be a valuable tool for early detection of the biofilm development. The dynamic biofilm formation is strictly related to the samples material. In some cases, a selection of the GLCM *features* used as variables in the PCA was needed. This rapid and non-destructive method could be used for a rapid and constant monitoring of the hygienic condition of surfaces in food industry, also on-line.

Introduction

Biofilm

A biofilm is an aggregation of microorganisms characterized by the secretion of an extracellular matrix adhesive and protective, often containing polysaccharides, characterized by complex biological interactions. The organization of the microorganisms in the biofilm offers important advantages: the microorganisms remain anchored in an optimal position, defended by stressful environmental situations or predation (Kumar and Anand, 1998). Furthermore the microorganisms coexist in a stable set of synergic cooperation between different species and, consequently, they orchestrate the degradation of substrates even complex (Sutherland, 2001).

Among the most common foodborne microorganisms, able to produce biofilm, there is *Pseudomonas*, a common and important altering bacterium of fresh food products. In fact, it is present not only on fruits, vegetables, meats and dairy products slightly acidic, but in the plants of food processing including drains and floors (Gibson et al., 1999). *Pseudomonas* spp., coexists within the biofilm with *Listeria*, *Salmonella* and other pathogens that are capable of survive and form biofilms multi-species, more stable and resistant (Chmielewski and Frank, 2003). The survival of foodborne pathogens or degrading bacteria, due to the insufficient disinfection of surfaces and instruments in contact with food (Fuster, 2006), is the main cause of the contamination of the final product; the consequences are, in addition to the discard of the product with important economic losses, foodborne diseases (Donlan, 2002).

The properties of the contact surface have a role in the formation of the biofilm, which is facilitated the more so the surface is rough and irregular. This is because the cutting forces are reduced and the area is higher the more the surface is rougher. The choice of material is therefore of great relevance in the design of the surfaces that are in contact with foodstuffs and that have to comply with specifications and be officially approved. The materials most commonly used in food plants like stainless steel (Jullien et al., 2003) glass, polypropylene, rubber, aluminium, Teflon and nylon despite the appearance, can still exhibit cracks and irregularities which promote microbial colonization (Allan et al., 2004).

A bacteria suspended in an aqueous medium, near a surface, becomes subject to two forces: the attractive force of Van der Waals, which works at a distance of few hundred nanometres and tends to bring the microscopic particles to the surface; and repulsive forces, with a negative electrostatic charge, acting when the bacterium reaches a distance from the substrate of about 10-20 nm. It has been shown, however, that the hydrophobic interactions between the cell surface and the substrate allow the cell to overcome these repulsive forces and to attach more rapidly to hydrophilic materials such as glass or metal (Fletcher and Loeb, 1979).

Image texture analysis

Although it is possible to associate different image properties such as smoothness, depth, regularity, roughness with the texture (Gonzalez, 2002), does not exist a univocal definition of the concept of image texture (Chen, 2007). The definitions considered the most comprehensive are:

- ✓ a distribution of pixels in an image that creates a texture depending on the composition of the subject being photographed (Gibson, 1978).
- ✓ An index of the arrangement of the pixel grey levels in a region of a digital image (IEEE, 1990).
- ✓ A descriptor of variation in brightness between adjacent pixels in an image (Guidetti et al., n.d.).

Summarizing, the texture of an image is given by a difference in the spatial distribution, in the frequency and in the intensity of the values of the grey levels of each pixel of the image.

Image analysis makes possible the evaluation of the surface texture directly from its digital image through histograms and two-dimensional arrays of co-occurrence. The method most frequently used for the analysis of the texture is based on the calculation of texture features using a co-occurrence matrix of the grey levels, where the number of columns and rows is equal to the number of grey levels in the image (grey level co-occurrence matrix - GLCM) (Haralick et al., 1973). Therefore, the need arises to represent the complexity of a phenomenon composed of a high number of variables. This need is satisfied by the multivariate data analysis and in particular by the Principal Components analysis (PCA). The principal components provide an explanation of the observed variability, with the advantage of describing the phenomenon by two or more orthogonal dimensions ordered in terms of their importance in the explanation of the variance.

For the early identification of the biofilm, various methods, such as the total count of viable cells, microscopy and spectroscopy, determination of ATP are commonly used. The present study aims to determine if the analysis of surface texture, early, rapid and non-destructive, can be able to distinguishing clean surfaces from surfaces colonized by a thin layer of biofilm, not jet observable by the naked eye.

Material and method

Specimens

Samples of different materials among the most commonly used in the food industry, in the form of 10x10 cm (100 cm²) were tested:

- ✓ stainless steel AISI 304 “Scotch Brite” finished (BSSA 2001)(figure1a)
- ✓ Plastic LDPE for food (figure1b)
- ✓ Ceramic (figure1c)

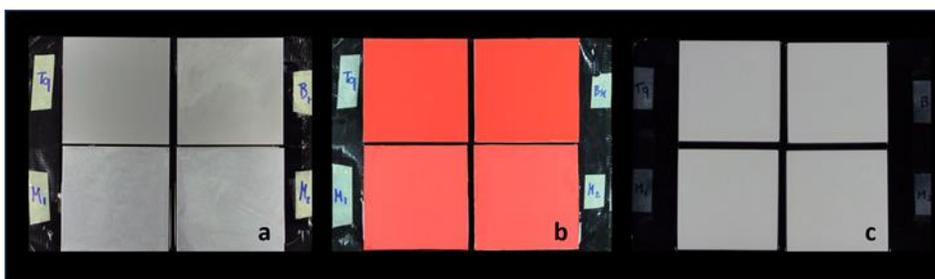


Figure 1: specimens at day 7 – a “steel” b “plastic” c “ceramic” with clean specimens at the top (TQ and Br) and contaminated specimens at the bottom (M1 and M2).

These materials were chosen as examples of work surfaces in contact with food, and industrial environment (floors and walls) (Bower and Daeschel, 1999; Wong, 1998).

Microorganism

As microorganism, the strain of *Pseudomonas fluorescens* DSM 50106 was used (Fletcher, 1988). The choice of this microorganism is due to its high ability to adhere to surfaces and multiply in the form of biofilms. *Pseudomonas fluorescens* is an aerobic microorganism extremely widespread in the food industries (Carpentier and Cerf, 1993; Van Houdt and Michiels, 2010), particularly in water distribution systems. Moreover, it is non-pathogenic microorganism and thus it is safer to use. Its growth was performed on Tryptic Soy Broth (TSB) (Difco-MD, USA), a commonly used soil for aerobic microorganisms, at 30°C for 24 h and was routinely maintained at 4°C. For the count plate Tryptic Soy Agar was used, the same of Tryptic Soy Broth with the addition of 15 g / l agar. For long-term maintenance, stock solution cultures were stored in 20% glycerol (v/v) of appropriate liquid medium at - 80°C.

Biofilm formation

Firstly, to carry out this work was necessary to implement a standardized method able to provide the formation of biofilm on the surface of the samples; so, the method

suggested by Pan, Breidt, & Kathariou, (2006) was followed in order to simulate what occurs in the food industry. The specimens were first washed in ethanol: acetone (9:1) solution, then placed in a Petri plates (120 mm) and submerged in a microbial suspension of the microorganism test with concentration of 10^8 CFU/ml (the microbial concentration has been quantified with bürker chamber method) and incubated at 30°C for 4 hours, to allow for cell attachment. After incubation, the microbial suspension was evacuated by aspiration, and the samples were washed with sterile triptone salt 0.85% to remove loosely attached cells. The materials were then transferred into a new Petri plate and cover with TSB prepared 1:10 and incubate at 22°C for 16 hours to permit biofilm development.

Each sample was then drained for 30 seconds, laid on aseptic paper, and dried for 10 minutes. All operations were carried out aseptically under hood UV.

Capture session

For photographic acquisitions was used a Canon PowerShot Pro90 IS, sensor with 2.6 megapixels. The photos were taken inside a shooting room 90 inches wide, 70 inches deep and 1 m in height. The walls of the room are made of opaque plastic and it was possible to cover the front opening with a dark cloth, to prevent ambient light. Inside the camera, 4 spotlights, with 23-watt fluorescent bulb, were placed. The camera was connected to an external PC from which it is possible to adjust photographic parameters using the software Canon Remote Capture, without intervening directly in the shooting room.

For each capture session were used four specimens of the same material organized as follows (details in paragraph 3):

- 1 background specimen
- 1 specimen dirtied with sterile TSB
- 2 specimens contaminated with biofilm of *P. fluorescens*

Each capture session lasted for 7 days with a frequency of one pictures every 4 hours, for 42 photographs per session. Each complete protocol (capture session) was performed twice.

Data Processing - Grey Level Co-Occurrence Matrix (GLCM)

To proceed in image processing, using the software ImageJ, a region of interest (ROI) (size 400 x 400 pixels) corresponding to the middle section of each specimen was catted and converted to grayscale images. The ROI obtained were organized in stack, three-

dimensional matrixes containing the individual bi-dimensional images for each specimen at various shutting times.

The GLCM creates a square matrix of dimensions equal to the maximum intensity and composed by the frequency of the different intensities of grey within the stack. This processing is strongly influenced by the pixel pitch and the direction. In this work was chosen a step equal to one, which allows considering the more subtle variations of processing. As regards the direction, descriptors calculated for all four angles (0 °, 45 °, 90 °, 135 °) were averaged in order not to take into account any geometrical distributions on the surface of the material. The GLCM calculates up to 14 different features, or descriptors, however those considered in this work are (Fongaro and Kvaal, 2013):

Angular Second Moment, describing the regularity of an image

$$ASM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i, j)\}^2$$

Inverse Difference Moment, evaluating the local homogeneity of the image

$$IDM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i - j)^2} P(i, j)$$

Contrast, considering the local homogeneity of the image

$$CONTRAST = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^G \sum_{j=1}^G P(i, j) \right\}, |i - j| = n$$

Energy, ranging between 0 and 1, where 1 is a constant image

$$ENERGY = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j)^2$$

Entropy, measuring the statistical randomness

$$ENTROPY = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) * \log(P(i, j))$$

Homogeneity, evaluating the homogeneity of the image

$$HOMOGENEITY = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{P(i, j)}{1 + |i - j|}$$

Correlation, considering the relation between a pixel and its neighbours; this correlation can be positive, direct, or negative, indirect

$$CORRELATION = \frac{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{i * j\} * P(i, j) - \{\mu_x * \mu_y\}}{\sigma_x * \sigma_y}$$

A PCA using as variables the features calculated with GLCM, was performed using MATLAB (MathWorks, Inc.) - PLS Toolbox (Eigenvector Research, Inc.) that allows the creation of graphical outputs useful for a qualitative assessment of the phenomenon:

- ✓ the score plot displays the distribution of the samples in an orthogonal space, whose axes are the main components, enabling the interpretation of the differences and similarities;
- ✓ the loadings plot is a graph that rate the weight of each auto-scaled variable in building up of the main components;
- ✓ the average graph represents the specimen variables assessment, in order to evaluate which variables are most useful for differentiating the samples.

Microbial count

Five specimens for each material were located in the shooting room, outside of the camera visual range. A thermometer was placed inside the shooting room to ensure that the temperature was appropriate to the development of *Ps. fluorescens*.

A sample for each material was collected at different times: The day of preparation, 2 days after, 5 days after, 6 days after, 7 days after.

From each specimen biofilm was removed from the 50% of the surface (5 cm x 10 cm = 50 cm²), using a sterile sponge, in a standardized way by wiping the sponge two times. Each sponge was put in sterile bags with 10 ml of triptone salt, then subjected to decimal dilutions and inoculated on Plate Count Agar (Merck Germany). After 48 hours of incubation at 30 °C, the counting of CFU was carried out to perform a quantitative evaluation of the number of cells on the surface of the specimens. The higher the number of microorganisms removed with the sponge, the lower are the remaining microorganisms on the surface. In this way, an indication of the strength of the biofilm of *Ps. fluorescens* on three different materials at different times was provided.

Results expressed in N_s (CFU/cm²) were calculated using the following formula:

$$N_s = \frac{N \times F}{A} \times D$$

N: number of CFU in 1 ml of broth
 F: millilitres of thinner used
 A: sampled surface
 D: dilution factor

Colorimeter technique

On the same samples of the microbial count, following the removal of the 50% of the biofilm, a qualitative assessment of the development of the biofilm through a colorimetric test with 0.1% Safranin was carried out. After waiting for 10 minutes, Safranin excess was removed using a solution of ethanol and acetone (90:10) and the test specimens were dried. In this way, it was possible to make an instant and easy visual assessment of the presence of biofilm at different times. As an example, figure 2 shows the more intense colour of steel specimen as time passes and biofilm grows.

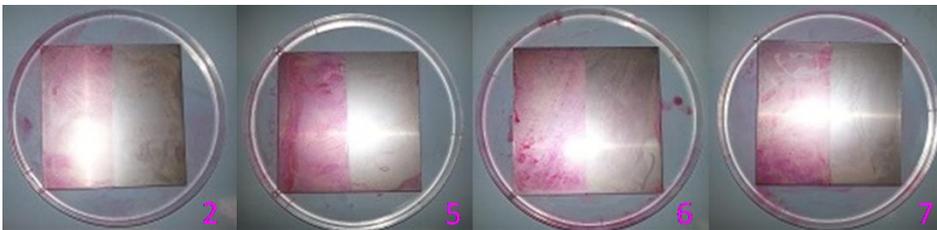


Figure 2: Steel specimens coloured by Safranin in each sampling date, from day 2 to day 7.

Results

For each shooting session four samples were analysed; they were named as follow (figure1):

Br: specimen soiled with sterile culture broth;

M1: specimen contaminated with *Pseudomonas Fluorescens* (10^8 CFU / ml) - trial1;

M2: specimen contaminated with *Pseudomonas Fluorescens* (10^8 CFU / ml) - trial2;

TQ: clean specimen.

Table 1 shows the mean value of the GLCM feature after autoscaling and their loadings on the first Principal Component (PC1). The variability of the samples along time is not presented in the paper because it proves to be irrelevant for the image evaluation.

Table 1: GLCM feature's average trend and PC1 loadings of the four specimens for each material.

		ASM	IDM	Contrast	Energy	Entropy	Homogeneity	Correlation
STEEL	TQ	-0.60	-1.03	1.10	-0.60	0.57	-1.03	-0.16
	Br	-1.09	-0.68	0.55	-1.09	1.10	-0.68	-1.33
	M1	0.95	0.95	-1.09	0.95	-1.04	0.95	0.95
	M2	0.75	0.76	-0.56	0.75	-0.63	0.76	0.55
	PC 1	0.39	0.38	-0.37	0.39	-0.39	0.38	0.35
PLASTIC	TQ	0.19	1.28	-1.15	0.19	-1.39	1.28	-0.57
	Br	-0.55	0.29	-0.43	-0.55	-0.07	0.30	0.68
	M1	1.32	-0.68	0.43	1.32	0.79	-0.66	1.00
	M2	-0.96	-0.90	1.15	-0.96	0.66	-0.91	-1.10
	PC 1	0.04	0.50	-0.49	0.04	-0.49	0.50	-0.01
CERAMIC	TQ	0.91	-0.86	0.88	0.91	-0.86	-0.85	0.55
	Br	-0.99	-0.75	0.81	-0.99	1.05	-0.74	-1.06
	M1	0.82	1.26	-1.11	0.82	-0.85	1.27	1.10
	M2	-0.73	0.35	-0.57	-0.73	0.66	0.32	-0.59
	PC 1	0.39	0.36	-0.33	0.39	-0.40	0.36	0.42

Figure 3 reports for the different specimens, steel, plastic and ceramic, the microbial count to estimate the number of removed microorganism.

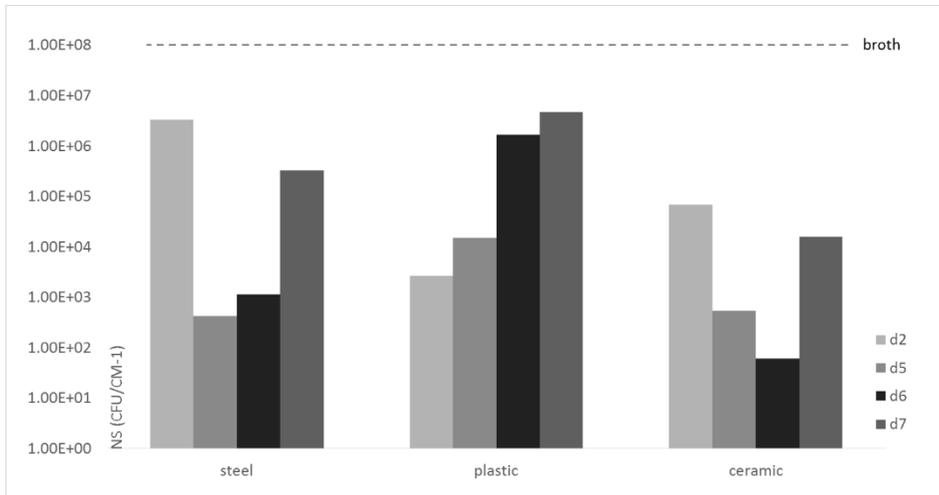


Figure 3: Histogram resuming the microbiological count, for each material at each sampling date.

Steel- Image texture analysis

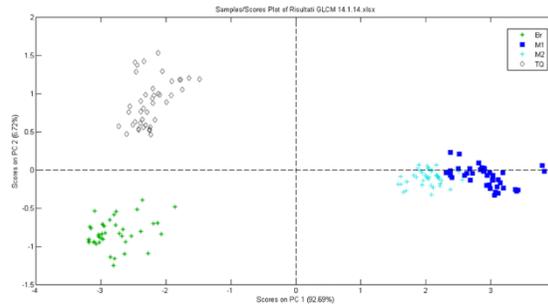


Figure 4: PCA Score plot of the steel samples, considering all the seven variables.

To better understand the importance of the single features in distinguishing the samples, the average trend of these variables for each specimen was considered. Observing these values (table 1) it is possible to find the ability of each variable in providing the distinction between the contaminated specimens, M1 and M2, and the cleaned ones. Regarding the steel, all the variables are equally useful to distinguish the specimens: i.e. all variables show a large difference between the average value of M1 and M2 and the average value of TQ and Br. In addition, PC1 loadings have similar values, explaining the importance of each variable in constituting the first principal component and thus confirming the importance of all variables considered.

In figure 4, the specimens contaminated by biofilms (M1 and M2) are clearly divided from those non-contaminated (Br and TQ). The separation between the two pairs of samples occurs along the first component, which explains 92.69% of the variance. M1 and M2 are confused, indicating that the differences between the two trials do not significantly affect the analysis. The GLCM applied to steel specimen is able to distinguish TQ, clean surfaces, from Br, surfaces contaminated by organic residues, along PC 2 (6.72% of variance explained).

Steel - Microbiological analysis

With regard to the microbiological count, the histogram (figure 3) allows to assume the behaviour of the biofilm. Each analysis was matched with the image after Safranin colouring that visually confirm the plate count.

After 48 hours (day 2), a limited adhesion on steel is observed: the sponge has removed most of the microorganisms, which leads to the hypothesis that the biofilm has not yet formed. In fact the colouring by Safranin highlights the differences between the right half and the left half of the steel specimen. At day 5 the situation is clearly different: the number of microorganisms removed from steel is reduced. At day 6, the removal from steel is higher, compared to the previous day, and growing up to day 7.

Plastic - Image texture analysis

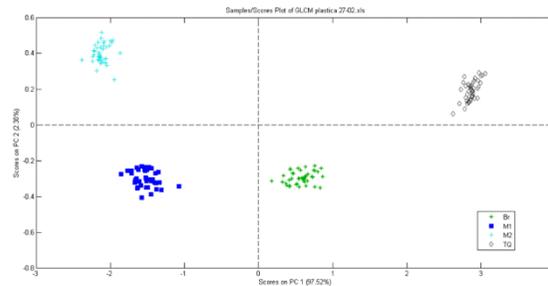


Figure 5: PCA Score plot of the plastic samples, considering the selected variables (Inverse Difference Moment, Contrast, Entropy and Homogeneity).

The PCA was carried out using all 7 features; the scores plot built allows a good separation along the abscissa axis (56% of variance explained). The homogeneous dispersion of the 4 groups along the second principal component (37%) does not appear relevant for the intended purpose (data not shown).

Considering the average trend of each variable, for both the clean and biofilm specimen, (table1) is evident that not all the variables are necessary to recognize the texture of the biofilm; some variables indeed only add variability and noise to the co-occurrence matrix. The variables useful in recognizing the biofilm on plastic samples are Inverse Difference Moment, Contrast, Entropy and Homogeneity. These four features show also higher values of the PC1 loading; this is an excellent confirmation of the variable selection.

The selection of variables is a useful practice to reduce noise and have a clearer data plotting, as shown by the scores plot of Figure 5, in which only the four selected variables are used. The samples belonging to each specimen are more grouped, along the first principal component (97% of the variance). Furthermore, it possible to distinguish the samples belonging to M1 and M2 along axis PC2; this axis explains only 2% of the variance, as desirable since M1 and M2 are replicates.

Plastic - Microbiological analysis

Focusing on the plastic histogram in figure 3, the number of cells removed at day 2 from the plastic is lower than from the other materials, evidence that the biofilm has already formed. The value of removed cells remains almost constant until day 5. The apparent detachment of cells occurs at day 6 and more completely at day 7. The Safranin colouring of the specimen at day 7 is in accordance with the microbial count (picture not shown).

Ceramic - Image texture analysis

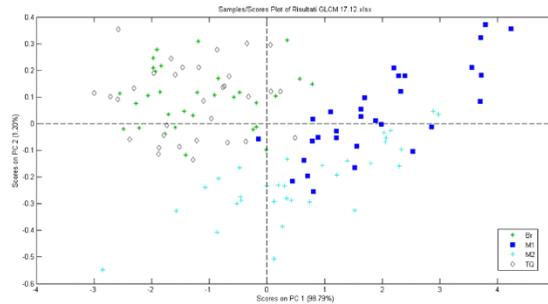


Figure 6: PCA Score plot of the ceramic samples, considering the selected variables (IDM, Contrast, and Homogeneity).

From a first evaluation of the PCA results, it is possible to highlight how the samples do not separate satisfactorily in the PCA scores plot: the specimen TQ is significantly different from the other along the second component, while Br is separate from the sample M1 but partially confused with M2.

Trying to improve the distinction of the samples, the ability of each variable distinction, was considered (table1). The features that can bring together the values of M1 and M2 and, at the same time, move them away from those of Br and TQ are IDM, Contrast and Homogeneity. In this case, the variable selection according to the features average trends is not confirmed by the PC1 loadings that show more or less all the same values. In Figure 6, only the three selected variables were used to carry out PCA. Sample distribution is extremely effective: M1 and M2 are clearly distinct from TQ and Br along PC1, which explains 99% of the differences.

Ceramic - Microbiological analysis

The number of cells removed at each sampling date decreases until day 6, as can be seen by the histogram in figure 3. Compared to the other material, the cell detachment is delayed. At day seven detachment begins to be visible but it is not complete. This is confirmed by the colorimetric test in which the intensity of the coloured part increases until day 6 and seems almost constant at day 7. Only regarding ceramic, there is no return to the initial state of cleanliness of the material.

Discussions

Principal Component Analysis shows that, depending on the material, a selection of the GLCM features might be necessary. In fact, the use of all eight variables may not allow to distinguish the different specimens, clean and covered by biofilm, adding noise to the

data and reducing the accuracy. The selection is made possible by comparing the average trend of the variables for each specimen and the PC1 loadings.

Ceramic is the material that gave less consistent results probably because of the colour of the specimens (white), which may have a negative effect on the white balance of the camera and / or on the focus accuracy.

As regards the biofilm development, the adhesion capability of the biofilm changes considerably depending on the type of material on which bacteria are developing. At day 2, it is possible to distinguish typical trends for the different materials that stay on up to the removal of biofilm, at day 7. This behaviour is probably due to lack of nutrients on the surface of the specimens with the consequent detachment from the surface. The steel is characterised by a more prolonged phase of adhesion, compared to plastic, but the biofilm that results is strong even after 6 days. The plastic instead is characterised by a rapid accession but by an equally rapid, almost complete, removal of the biofilm. Regarding the ceramic, the adhesion is slower than on the other two materials with the phase of detachment that begins at day 7.

Conclusions

It was observed that the development of the biofilm is related to the type of material of which the surface is made. Despite these differences, the technique of texture imaging has proven to be a useful and effective tool for the early detection of biofilm. The Grey Level Co-Occurrence Matrix, used in combination with Principal Component Analysis, was able to recognize a clean surface from a surface of the same material to which microbial cells had adhered, despite the differences between the surfaces were not visible by the naked eye. It is important to emphasize the non-destructiveness of this method. The method, being also rapid and simple to apply, could be very interesting for monitoring the hygienic condition of industrial surfaces, even online.

To be sure about the type of biofilm development on the surface, spectral information have to be combined with surface texture knowledge; the technique that allows to obtain these two information is hyperspectral imaging, that could be the future prospective of this case study.

References

- Allan, J.T., Yan, Z., Genzlinger, L.L., Kornacki, J.L., 2004. Temperature and biological soil effects on the survival of selected foodborne pathogens on a mortar surface. *J. Food Prot.* 67, 2661–2665.
- Bower, C.K., Daeschel, M. a., 1999. Resistance responses of microorganisms in food environments. *Int. J. Food Microbiol.* 50, 33–44. doi:10.1016/S0168-1605(99)00075-6
- Carpentier, B., Cerf, O., 1993. Biofilms and their consequences, with particular reference to hygiene in the food industry. *J. Appl. Bacteriol.* 75, 499–511. doi:DOI 10.1111/j.1365-2672.1993.tb01587.x
- Chen, J., 2007. Surface texture of foods: perception and characterization. *Crit. Rev. Food Sci. Nutr.* 47, 583–598. doi:10.1080/10408390600919031
- Chmielewski, R. a N., Frank, J.F., 2003. Biofilm Formation and Control in Food Processing Facilities. *Compr.*

- Rev. Food Sci. Food Saf. 2, 22–32. doi:10.1111/j.1541-4337.2003.tb00012.x
- Donlan, R.M., 2002. Biofilms: Microbial life on surfaces. *Emerg. Infect. Dis.* 8, 881–890. doi:10.3201/eid0809.020063
- ElMasry, G., Wang, N., ElSayed, A., Ngadi, M., 2007. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *J. Food Eng.* 81, 98–107. doi:10.1016/j.jfoodeng.2006.10.016
- Fletcher, M., 1988. Attachment of *Pseudomonas fluorescens* to glass and influence of electrolytes on bacterium-substratum separation distance. *J. Bacteriol.* 170, 2027–2030.
- Fletcher, M., Loeb, G.I., 1979. Influence of substratum characteristics on the attachment of a marine pseudomonad to solid surfaces. *Appl. Environ. Microbiol.* 37, 67–72.
- Fongaro, L., Kvaal, K., 2013. Surface texture characterization of an Italian pasta by means of univariate and multivariate feature extraction from their texture images. *Food Res. Int.* 51, 693–705. doi:10.1016/j.foodres.2013.01.044
- Fuster, N., 2006. Importancia del control higiénico de las superficies alimentarias mediante técnicas rápidas y tradicionales para evitar y/o minimizar las contaminaciones cruzadas. 160.
- Gibson, H., Taylor, J.H., Hall, K.E., Holah, J.T., 1999. Effectiveness of cleaning techniques used in the food industry in terms of the removal of bacterial biofilms. *J. Appl. Microbiol.* 87, 41–48. doi:10.1046/j.1365-2672.1999.00790.x
- Gibson, J., 1978. The Ecological Approach to the Visual Perception of Pictures. *Leonardo* 11, 227–235. doi:10.2307/1574154
- Gonzalez, R.C., 2002. Digital Image Processing, *Leonardo*. doi:10.2307/1574313
- Guidetti, R., Beghi, R., Giovenzana, V., n.d. Chapter Number Chemometrics in Food Technology 1–36.
- Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Syst. Man, Cybern.* SMC-3. doi:10.1109/TSMC.1973.4309314
- Huang, H., Liu, L., Ngadi, M.O., 2014. Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors (Basel)*. 14, 7248–76. doi:10.3390/s140407248
- IEEE, 1990. IEEE Standard Glossary of Software Engineering Terminology. Office 121990, 1. doi:10.1109/IEEESTD.1990.101064
- Jullien, C., Bénézech, T., Carpentier, B., Leuret, V., Faille, C., 2003. Identification of surface characteristics relevant to the hygienic status of stainless steel for the food industry. *J. Food Eng.* 56, 77–87. doi:10.1016/S0260-8774(02)00150-4
- Kumar, C.G., Anand, S., 1998. Significance of microbial biofilms in food industry: a review. *Int. J. Food Microbiol.* 42, 9–27. doi:10.1016/S0168-1605(98)00060-9
- Lee, S.K., Kader, A.A., 2000. Preharvest and postharvest factors influencing vitamin C content of horticultural crops. *Postharvest Biol. Technol.* 20, 207–220. doi:10.1016/S0925-5214(00)00133-2
- Pan, Y., Breidt, F., Kathariou, S., 2006. Resistance of *Listeria monocytogenes* biofilms to sanitizing agents in a simulated food processing environment. *Appl. Environ. Microbiol.* 72, 7711–7717. doi:10.1128/AEM.01065-06
- Sutherland, I.W., 2001. The biofilm matrix - An immobilized but dynamic microbial environment. *Trends Microbiol.* 9, 222–227. doi:10.1016/S0966-842X(01)02012-1
- Van Houdt, R., Michiels, C.W., 2010. Biofilm formation and the food industry, a focus on the bacterial outer surface. *J. Appl. Microbiol.* 109, 1117–1131. doi:10.1111/j.1365-2672.2010.04756.x
- Wong, a C., 1998. Biofilms in food processing environments. *J. Dairy Sci.* 81, 2765–2770. doi:10.3168/jds.S0022-0302(98)75834-5

3. HYPERSPECTRAL

Hyperspectral imaging is an alternative technique for non-destructive food analysis, enabling real-time monitoring of quality (Sun 2010). Each pixel of a hyperspectral image contains the spectrum of that specific position thus giving both spatial and spectra information of a sample. The data are three-dimensional blocks, called hypercubes, analysed with a chemometric approach that reduces dimensionality of the data while retaining the most useful spectral information (Elmasry et al. 2012).

Chemometric was crucial for any steps of the PhD project and it is highlighted in a work, presented at the beginning of this paragraph that proposes a flow sheet for three-dimensional data elaboration, carrying out an example for the prediction of bread staling during storage.

Hyperspectral is an interesting approach for the evaluation of the chemical compounds distribution in heterogeneous matrices, very common in food (Amigo et al. 2013). The complexity of fruit and vegetable biochemistry is a challenge for this non-destructive technique so a fruit was chosen as sample for the experimental part of the project carried out in Brazil. The fruit selected is acerola, a typical Brazilian fruit with a huge commercial relevance in South America, which is particularly interesting for its nutraceutical activity. The distribution of ascorbic acid in acerola was the topic of the second work presented in this chapter because in literature active compounds are often quantified without evaluating their distribution in the vegetable tissues. Vitamin C is defined as the generic term for all compounds exhibiting the biological activity of L-ascorbic acid and is one of the most important nutritional quality factors in many horticultural crops (Lee and Kader, 2000). NIR spectrum embodies abundant information of hydroxyl (O–H), amino (N–H), and C–H's vibration absorption, so it is a very powerful tool for routinely studying ascorbic acid (Liu et al., 2006). For this reason, ten different fruits picked up according to two different stages of maturity were analysed. The spectra of pure vitamin C powder was used as references for computing models with two different correlation techniques: Spectral Angle Mapping (SAM) and correlation coefficient allowing the construction of a qualitative distribution map of ascorbic acid inside the fruit.

The reduction of the data dimension, already presented in the first chapter "Spectroscopy", is even more crucial on hyperspectral data (ElMasry et al. 2007). This topic is introduced in the third paragraph of this chapter where, on acerola images, a variables selection is carried out. Because the selection depends on the behaviour of spectral responses under modification of the samples, acerola post-harvest quality was evaluated along time. Hyperspectral images of 20 acerolas were acquired for five consecutive days and an investigation of trends along time was carried out to highlight

the most important three wavelengths that characterized the ripeness/degradation process of the Acerola fruit. The false-colour RGB images, derived from the composition of the three interesting wavelengths selected, enable early detection of the senescence process in a rapid and non-destructive manner.

- Amigo, J. M., Martí, I., & Gowen, A. (2013). Hyperspectral imaging and chemometrics: a perfect combination for the analysis of food structure, composition and quality. In *Data Handling in Science and Technology*. Vol 28 (pp. 343-370).
- Elmasry, G., Kamruzzaman, M., Sun, D. W., & Allen, P. (2012). Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review. *Critical Reviews in Food Science and Nutrition*, 52(11), 999-1023.
- ElMasry, G., Wang, N., ElSayed, A., & Ngadi, M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering*, 81(1), 98-107.
- Lee, S. K., & Kader, A. A. (2000). Preharvest and postharvest factors influencing vitamin C content of horticultural crops. *Postharvest biology and technology*, 20(3), 207-220.
- Liu, H., Xiang, B., & Qu, L. (2006). Structure analysis of ascorbic acid using near-infrared spectroscopy and generalized two-dimensional correlation spectroscopy. *Journal of molecular structure*, 794(1), 12-17.
- Sun D.W. *Hyperspectral imaging for food quality analysis and control*. Da-Wen. Sun (Ed.) Elsevier (2010) ISBN 978-0-12-374753-2.

3.1 HYPERSPECTRAL IMAGE ANALYSIS: A TUTORIAL

Ronan Dorrepaal^a, Cristina Malegori^b, Aoife Gowen^a

^aBiosystems Engineering, School of Agriculture, Food Science and Veterinary Medicine, University College Dublin, - Dublin - IRELAND

^b DeFENS Department of Food, Environmental and Nutritional Sciences
Università degli Studi di Milano – Milano - ITALY

Abstract

Hyperspectral imaging (HS) is an alternative technique for non-destructive food analysis, enabling real-time monitoring of quality (Sun 2010). An HS image is a large dataset in which each pixel corresponds to a spectrum, related spatially to its neighbouring point, thus providing high-quality detail of the sample surface. An HS 3D data cube, called hypercubes, combines chemical and physical information (Huang et al., 2014) and needs to be analysed with a chemometric approach that reduces dimensionality of the data while retaining the most useful spectral information (Elmasry et al. 2012). In this work, a logical step-by-step flow sheet for three-dimensional data elaboration is proposed, carrying out an example for the prediction of bread staling during storage. Starting from spectra pre-processing (image acquisition, background removal, dead pixels and spikes, masking and compression by PCA, pretreatments, binning), the typical steps encountered in hyperspectral image processing are summarised, presenting different data processing techniques concatenated PCA, classification, clustering, PLSR) that may be applied to obtain useful information. At the end of the paper, ten general rules and suggestions about hyperspectral image processing are proposed.

Introduction

Hyperspectral (HS) data are characterized by dual information, spectral and spatial, that allows one to obtain both qualitative and quantitative information from a sample (ElMasry et al., 2007). An HS image is a large dataset in which each pixel corresponds to a spectrum, related spatially to its neighbouring point, thus providing high-quality detail of the sample surface. An HS 3D data cube, combines chemical and physical information and requires a powerful statistical approach to extract the desirable knowledge (Huang et al., 2014).

Typical steps encountered in hyperspectral image processing are summarised in Figure 1. The left hand side of this schematic describes hyperspectral pre-processing, i.e. operations that are carried out on hyperspectral images prior to modelling. The rightmost side of the schematic summarises the different data processing techniques that may be applied to obtain useful information from hyperspectral images. Although this workflow is not prescriptive (the order and combination of steps can be changed) it does offer a logical step-by-step flow which should be useful for those learning how to tackle to huge amounts of data routinely encountered in HSI. With this in mind, we shall embark on a journey through most of the steps in Figure 1, using a specific dataset of bread as an example.

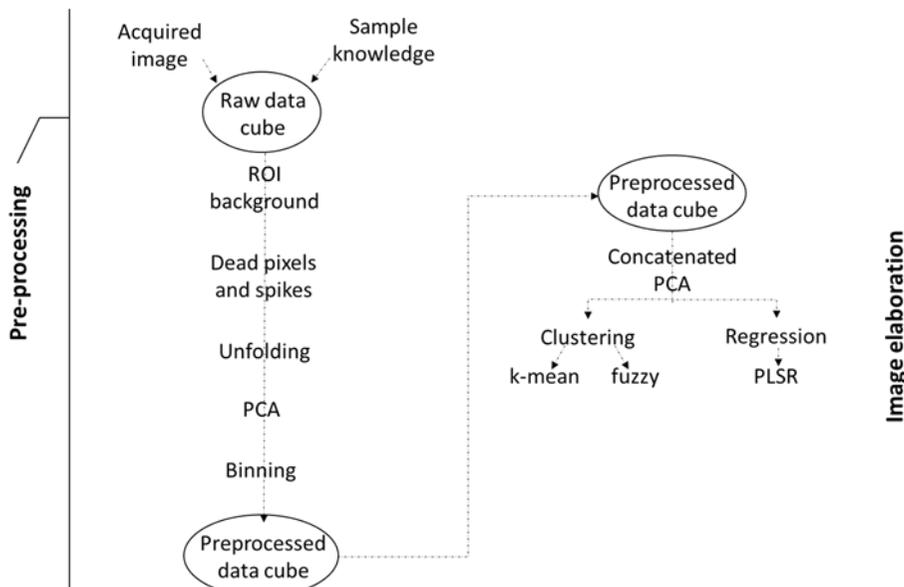


Figure 1. Typical steps involved in hyperspectral image analysis

Image acquisition

The first important step prior to data analysis is to acquire the optimal images possible. The choice of a background for your sample that is spectrally different to it (the flatter

the background's spectrum the better) and remains stable over the course of the experiment, is important. As a simple example, if imaging a white sample, such as bread, in the visible wavelength range, use a black background (black paper is fine in this case). When imaging in the NIR, a paper background is often not a great idea, since paper absorbs lots of light in the NIR (in this case black sandpaper works very well).

In addition, the direction of the light has to be kept under control to avoid saturation of the pixels causing the loss of information. A direct and unwanted reflection of the light could hide important details of the sample and mislead the data processing. The best option is to work with diffuse light but, if only direct illumination is possible, use halogen lamps angled at 45 degrees on the sample.

When setting up the acquisition parameters, it is recommended to obtain some test images before carrying out the entire experimental design and thus to avoid future acquisition errors.

Data set

In this work, a set of bread images was chosen to provide an example on which the data analysis could be carried out. Before analysing an image, it is important to have knowledge of the sample and on the aim of the processing that has to be done. The images used here are part of a dataset of six hyperspectral images recorded with a line-mapping NIR device, working in the range of 1000 to 1600 nm (142 wavelengths). The images represent bread at different days of storage. Measurement of bread hardness was made by the TPA method (Texture Profile Analysis), evaluating hardness, resilience and springiness in five different points of each slice. The dataset was provided by Dr. Jose Amigo, University of Copenhagen.

It is possible to outline different topics of the work. Firstly, a general and qualitative evaluation of the bread at different storage time is carried out, considering moisture and leavening homogeneity. Then, focusing on the staling process, the percentage of the bread getting harder during time was quantified.

Image background

Most hypercubes contain lots of redundant data in the form of image background. Since this data is usually not very interesting, one of the first steps in pre-processing is to remove it. A little foresight in sample presentation makes the subsequent jobs such as separating the image from background very straightforward.

The question of how to remove the background can be answered by taking a look at the spectrum of the background as compared to that of the sample. This will give you an indication of which wavelengths provide greatest contrast between sample and background. In the present case, there was good contrast between the sample and background at 1008.8 nm, as shown in Figure 2 (a), where it can be seen that the sample is darker than the background at this wavelength. It can also be seen that the background in this case is not homogeneous. There is an area of "shine" just above the

bread sample. This probably occurred due to poor sample presentation/imaging set-up. Such situations should be avoided. Regardless, in this case the non-homogeneous background is not a problem, since the light absorbed by the bread sample is far lower than that absorbed by any part of the background. This is further underlined in Figure 2 (b), in which the histogram of the 1008.8 nm image is shown. Three distinct groups can be seen in the histogram, corresponding respectively to bread, background and shine regions of Fig 2(a). Using this information, the creation of a mask is very straightforward. Simply setting any pixel in Fig 2 (a) with an absorbance value greater than 0.5 to zero, and the remaining pixels, representing the bread sample, to one, provides us with a good first estimate of the mask (Fig. 2(c)).

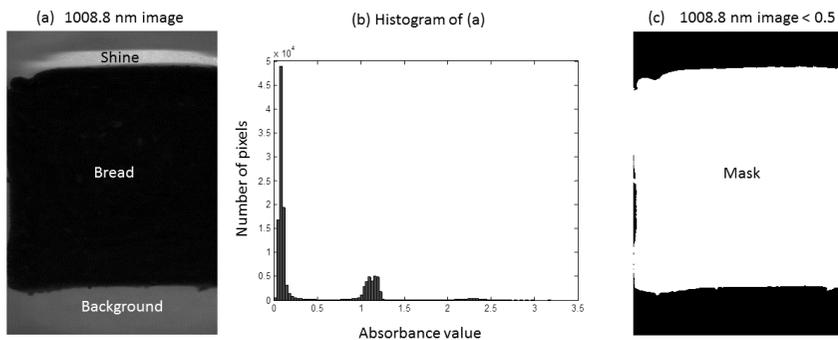


Figure 2. Masking the background from a hyperspectral image: (a) shows a single wavelength image at 1008.8 nm, (b) is a histogram showing the distribution of pixel values in (a) and (c) is a thresholded image, obtained by setting to zero all pixels in (a) that are greater than 0.5. The mask in (c) represents the sample of interest.

Dead pixels and spikes in the spectra

Dead pixels and spikes are a common occurrence in NIR hyperspectral imaging (Firtha et al., 2008; Zhang et al., 2007), where typically up to 1% of pixels will behave strangely. Image sensors, in fact, are composed of millions of photodiodes and if faulty elements occur in the sensor array, generate pixels that do not record correct information. The single isolated defective pixels located in random spatial positions are affected by impulse noise and generate black pixels, called dead pixels, and intensity peaks of the signal, known as spikes.

Many NIR cameras come with pre-programmed functions that have identified the dead pixels and compensate for them by interpolation. However, dead pixels are not well behaved. They tend to increase in number over time, so it is important to check for them prior to further data processing. A typical pixel spectrum from our bread sample is shown in Figure 3. For demonstration purposes, we have altered this spectrum to show what a dead pixel or spike might look like. Detection of such pixels is not always trivial, and many papers have been written on the topic (refs). However, if we consider the

difference spectrum, obtained by subtracting the absorbance at each wavelength from that at the preceding wavelength (lower section of Fig. 3), we see that for normal, well-behaved spectra, the difference spectrum is quite flat (thanks to the smoothly varying nature of NIR spectra in general), while the difference spectrum of spectra with the spike and dead pixel has a sharp discontinuity. This information is captured neatly in the standard deviation (SD) of each difference spectrum, with the SD of the badly behaving pixels being much larger than that of the normal one.

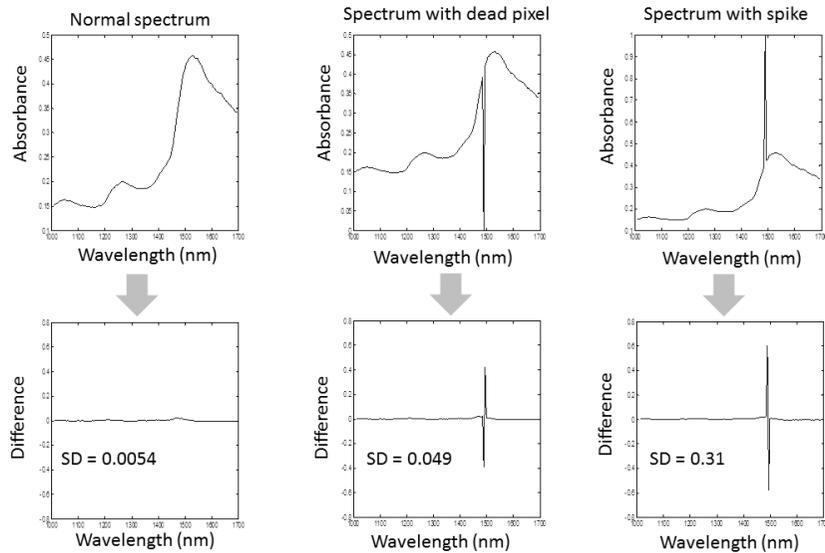


Figure 3. Normal, dead pixel and spike absorbance spectra (top row), with difference spectra and their standard deviations (SD) shown (bottom row).

These observations can be used to make a simple routine for identifying strange pixels in a hypercube. Applying this simple method to a hypercube of a slice of bread results in Figure 4 (a), where a line of strange pixels is evident on the left hand side of the image. Thresholding the SD image results in Fig.4 (b), which shows the location of the dead pixels. When we examine the spectra corresponding to these pixels (as shown in Fig. 4(c)), it is clear indeed that these pixels have been misbehaving. It is possible to replace pixels like these by interpolating from surrounding pixels. However, in this case we choose simply to leave them out of the analysis, since they are few in number.

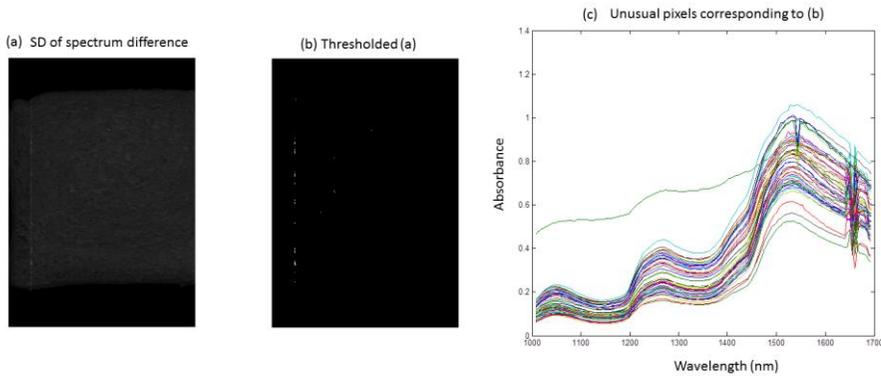


Figure 4. Identification of dead pixels. Standard deviation (SD) of difference hypercube shown in (a); thresholding (a) locates the dead pixels in (b); spectra corresponding to dead pixels shown in (c).

Unfolding

Hyperspectral images are usually called “hypercubes” due to the natural 3-D structure of their data, where each wavelength image is represented as a slice of a cube. The adaptation of the bilinear algorithms to hypercubes comes with a previous step of unfolding as shown schematically in Figure 5. In order to unfold the hypercube, each of the pixel spectra is stacked, one on top of the next (along the λ direction) to make a 2D matrix of size $(r*c)$, λ . Unfolding the hypercube into a more manageable structure is a common task in hyperspectral image analysis, so PCA can be applied. With PCs as number of principal components, the decomposition of the matrix into a scores matrix $(r*c)$, PCs , a loadings matrix PCs , λ and the residual $(r*c)$, λ was carried out. After application of PCA, the hypercube was re-folding to build the scores images, allowing the spatial visualization of the distribution of the components.

Masking and compression by PCA

Principal component analysis (PCA) is a versatile tool that can be used to summarise and explore the information contained in a hypercube. We will encounter PCA a number of times in this journey, but at this first stage we will use it to explore further the data contained in one single hypercube. After applying the masks to remove background and dead pixels, the remaining pixel spectra are organised in a 2 dimensional matrix, on which PCA is applied. The PCA loadings obtained are then applied to the original hypercube, to obtain PCA score images. The steps involved in this process are shown in the Figure 5. The upper portion of the figure explains how PCA is applied to the non-background pixels and lower section explains how the obtained PCA loadings are applied to the original hypercube, resulting in a cube of PCA score images, where each slice of the cube corresponds to a different PC.

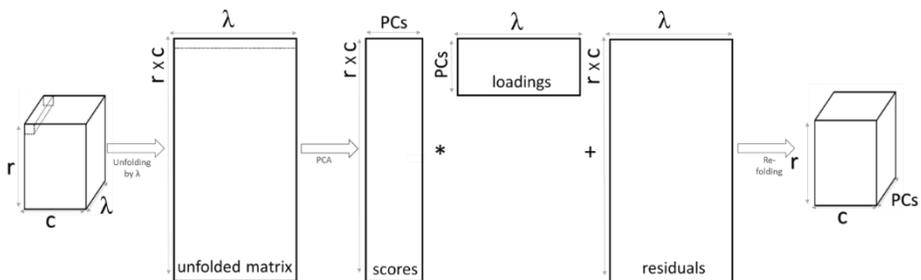


Figure 5. Application of PCA to a single hypercube extracting sample spectra from hypercube to make 2 dimensional data matrix, followed by matrix decomposition by PCA application of PCA loadings to original hypercube by unfolding, followed by re-folding to make PC scores cube.

Figure 6 shows the first four PC images, obtained by applying PCA to the non-background pixel spectra of our bread sample, which contain 99.8% of the variance of the data matrix. If we look at the PC1 image, bright pixels are apparent, both at the edge of the bread and in the upper central region. Edge pixels are something to take care of, since some edge pixels contain mixed information from both the sample and the background. The bright pixels in the central region are due to holes in the bread and contain mixed information from the bread and the background. It is important to remove such pixels prior to subsequent analysis as they can have negatively influence calibration model performance and interpretation of results (e.g. on first view, and without prior knowledge of the bread samples, the first author thought these holes were pieces of grain in the bread!). In order to do this, we examine the PC 1 image in more detail.

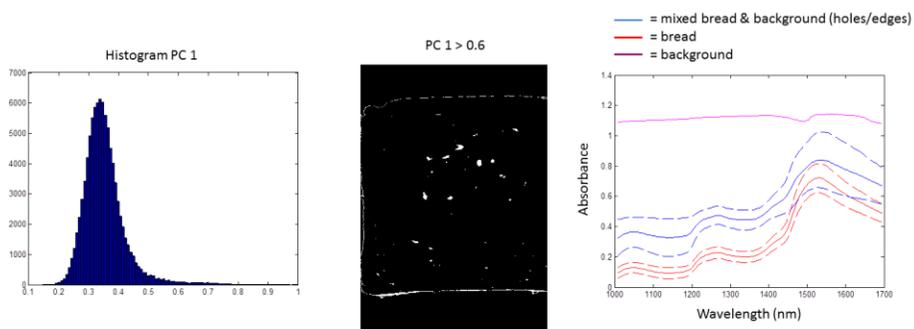


Figure 6. Principal component (PC) score images for a bread sample. Histogram of PC 1 is used to identify mixed edge pixels and holes in the bread. Spectra of bread, background and mixed pixels also shown.

The histogram of the PC 1 score image is shown in Figure 6. Bright pixels, corresponding to holes and edge regions, appear on the right hand side of the distribution, with pixel values > 0.6 . Therefore, thresholding the PC 1 image, so that all pixels > 0.6 become one,

allows us to identify the holes and edge pixels. The spectra of the holes/edges are also shown in Figure 6. It is clear that these pixel's spectra are distinct from the bread spectra, lying between the spectrum of the background and the bread. Using this information, we can now create a new mask that removes background, dead/spike pixels and holes.

We've just shown how PCA can be used to identify unwanted information (e.g. edges, holes) in a hypercube. It's useful to note also that PCA can be used as a compression method. For instance, if we apply the new mask (background, dead pixels and edges holes) to the hypercube, and then apply PCA to the unfolded hypercube, we find that the first 10 PCs contain 99.996% of the variation in the hypercube. These first 10 PC scores and loadings can be used to re-construct the hypercube, but at a lower data cost, approximately 8 % of the size of the original data cube ($10 \times 142 + 10 \times 130634 = 1307760$ data points, as opposed to $142 \times 130634 = 18550028$ data points). Since the higher PCs contain mostly noise, this also constitutes a noise reduction method, which is an added bonus.

Pretreatments

Spectral pretreatments are commonly employed in NIR chemical imaging and spectroscopy to overcome variations in spectra caused by physical effects (e.g. light scattering, variable focal length due to non-flat samples, ref). The most commonly used spectral pretreatments are standard normal variate SNV (Taghizadeh et al. 2011), first Savitzky–Golay derivative (SG1) and second Savitzky–Golay derivative (SG2) (Rinnan et al. 2009), although many more are available. It is not always straightforward to measure the effectiveness of a pretreatment (although efforts have been made in this direction, see *Carlos paper*), and for this reason, a common practise is to try a variety of pretreatments and compare them in terms of model performance, the pretreatment resulting in the lowest error being optimal. Application of the SNV pretreatment results in a clear decrease of the spectral variability, while SG1 and SG2 emphasize the regions of absorbance in the spectrum. It is not clear which among the three pretreatments is optimal at this stage, so we will reserve judgement until we develop some objective criteria such as model performance, as discussed in the next section.

Binning, if necessary

Hypercubes are data heavy – for example, after masking the background, spikes, edges and holes, we are left with more than 90,000 spectra to play with. When we wish to analyse multiple hypercubes simultaneously (as we will do in the next section), the data load can quickly make computations become painfully slow. In order to reduce the time required to do such computations, some further data reduction techniques be applied (Schmalzl, 2003; Sethi et al., 2011). One of the simplest is known as binning: this consists of constructing a “bin” of a certain size (e.g. $a \times b$ pixels) and averaging the components of the bin to make a new pixel (Srinivas et al. 2004). Moving the bin around the original image results in a new “binned” image, smaller than the original by a factor of $1/(a \times b)$.

However, caution is required when doing this. Binning is dangerous near edges and holes, as it can result in further spectral mixing, between the background and sample. For example, consider the mask we have created (Fig 8(a)). This image was binned using a [5x5] window (Fig. 8(b)). The original mask contained only 0's and 1's, but the binned mask also exhibits some intermediate values (grey pixels in the image), corresponding to bins containing both background and sample pixels. Including such pixels in further analysis would be akin to self-imposition of mixed pixels and should be avoided. A simple way to do this is by setting all pixels in the binned mask that are less than one to zero, resulting in a new binned mask (Fig. 8(c)). This new mask should be applied to the original hypercube after binning.

Concatenated PCA

We return now to PCA – this time as applied to a set of hypercubes. After removing the background, spikes, dead pixels, edge and holes and binning we are left with a set of cleaned hypercubes and ready to start some data processing. A common first step is to apply PCA to the entire set of hypercubes, in order to study the variation in the dataset. This can be carried out in at least 2 different ways. The first variation of PCA takes all of the sample spectra from each hypercube (omitting the background, dead pixels etc.) and stacks them on top of each other to make a 2D matrix (X), on which PCA is applied. The second way is to calculate the mean spectrum of the sample spectra from each hypercube and compile the mean spectra into a matrix on which PCA is applied. The second way is obviously faster, since fewer spectra are analysed. Whichever method is used, after obtaining the PCA loadings, PCA score images can be obtained in the same way as described in Figure 5(ii). After calculating the PCA scores it is possible to calculate an average value for each sample, and this can be compared to any measured variables that are available. This is useful for investigating whether the PC scores are correlated to measurements of interest.

Table 1 shows the correlation between mean PC score values and measured hardness for each pretreatment and the 2 variants of PCA applied (as described above and in Fig. 10). Straight away it is clear that the two variants of PCA produce different results, i.e. they are not equivalent. It is also apparent that there is a strong correlation between the measured hardness and the average PC 4 score after SG1 pretreatment, using the first variant of PCA. This is interesting, since this PC only describes a small amount of the total variation in the dataset (1.2 %). The higher PCs (representing 78, 12 and 5% of the variance) are related to variations in surface morphology and lighting.

Pretreatment	PC	R2 - PCA 1	R2 - PCA 2
None	1	0.409	0.411
	2	0.054	0.026
	3	0.36	0.276
	4	0.087	0.26
SNV	1	0.051	0.624
	2	0.281	0.269
	3	0.469	0.083
	4	0.182	0.001
SG1	1	0.497	0.524
	2	0.439	0.317
	3	0.147	0.07
	4	0.922	0.054
SG2	1	0.479	0.559
	2	0.359	0.104
	3	0.741	0.277
	4	0.332	0.014

Table 1. Correlation between mean PC score values and measured hardness, using 2 variants of PCA (PCA 1 and 2 as described in Fig. 10 (i) and (ii) respectively).

Classification

There are two categories of classification techniques: supervised and unsupervised classification (Amigo et al., 2013).

In supervised classification, the analyst has sufficient known pixels to generate representative parameters for each class of interest. This set is called training and it is used as reference for the classification of all other pixels in the image. The user also sets the bounds for how similar other pixels must be to group them together. Once trained, the classifier is then used to attach labels to all the image pixels according to the trained parameters.

Unsupervised classification is where the grouping of pixels is carried out using clustering algorithm to determine which pixels are related and groups them into classes. The user can specify which algorithm the software will use and the desired number of output classes but otherwise does not aid in the classification process. However, the user must have knowledge of the area being classified.

Clustering: k-means

Unsupervised clustering algorithms such as K-means can be applied to segment the hyperspectral data into regions of spectral similarity (ref). It is usually necessary to unfold the hypercube (as shown in Fig. 5(ii)) and concatenate the unfolded matrixes for each sample, prior to applying any segmentation. In the present case, k-means was applied to the spectra of the entire dataset. In k-means clustering, deciding on the number of clusters to include is crucial. In the example dataset we present here, bread is undergoing the process of staling over time, which could be interpreted as changing

from one state (not stale) to another (stale), or 2 clusters. We include the background in this analysis, giving a total of 3 clusters. We also compared the results with using 4 clusters. In order to check if the cluster size was related to the measured variable (hardness), the percentage of each cluster on the bread sample was calculated and linear regression was applied between the calculated percentage and sample hardness. The clusters were labelled according to the region they represented (e.g. background = 0, edge = 1, inner region = 2, central region = 3) as shown in Figure 12.

Table 2 shows the correlation between percentage of each cluster and measured hardness, with 3 and 4 clusters. Generally, imposing 4 clusters resulted in better correlation with sample hardness, with the percentage of cluster 3 on the SG 2 pre-treated data giving the best correlation ($R^2 = 0.86$). It can be seen that, for the 3 cluster models, the correlation between cluster 2 or 3 and hardness is the same. This is because they are inversely proportional to each other. The cluster map for the SG2 pretreated data with 4 clusters is shown in Figure 12. It is clear that the distribution of clusters shown there is very different to the distributions predicted by the PC 4 image. Which one is the real distribution of hardness? As we can see in Tables 1 and 2, correlation between the proportion of each cluster or mean PC score and hardness is relatively low. This highlights the limitation of carrying out unsupervised analysis such as k-means. Luckily we have at our disposal some reference data which can be used in supervised analysis of the data, as described in the next section.

Table 2. Correlation between percentage of cluster in image and measured hardness, using 2 k-means clustering with 3 and 4 clusters.

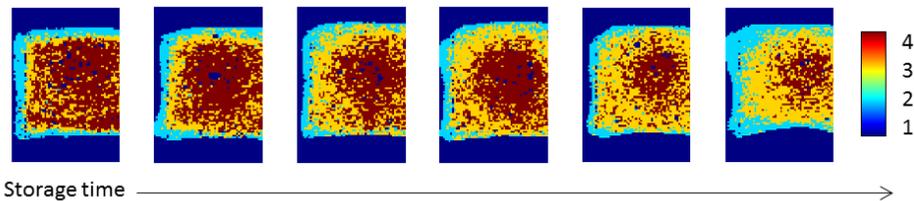


Figure 12. K-means with 4 clusters applied to SG2 pre-treated data.

Regression by PLSR

Calibration methods can be applied to hyperspectral data in a straightforward manner (ElMasry et al., 2007), by obtaining the mean spectrum of each sample, as shown schematically in Figure 14. An important point to note is that all calibration models should be validated with independent data sets. In the present case, the 1st 2 slices at each time point were used for model calibration and the 3rd was used for model validation. This is not a truly independent validation set, but we are somewhat justified, in that this dataset is used purely for demonstrating typical steps in hyperspectral image analysis. The number of latent variables to include was decided by 2-fold cross validation on the calibration set. Model performance metrics (shown in Table 3) can be used to compare the effectiveness of the pretreatments. We can see that the SNV pretreatment

clearly results in the lowest error and highest correlation ($R^2 = 0.99$). The SG1 pretreatment also resulted in a low error and high correlation ($R^2 = 0.97$) with hardness, however the SNV model required fewer latent variables (3 v's 4).

When we look at the prediction maps resulting from the PLSR model (shown in Fig. 13) we can see that the predictions for SNV and SG1 are similar in terms of the distribution of hardness in the sample (harder pixels are in red, softer are blue). This distribution is also quite similar to that of the PC4 image, but different to that for the k-means analysis (Fig. 12). The k means analysis suggests that the staling process is radial, while the PCA and PLS analysis suggest a skewed distribution, with greater staling occurring at the top of the bread. Deciding which distribution represents best reality requires some prior knowledge on the process (e.g. it could be explained if the upper portion of the bread was exposed to air during storage). However, the higher correlation resulting from the PCA and PLSR analysis make the distributions more convincing than those resulting from the k-means analysis.

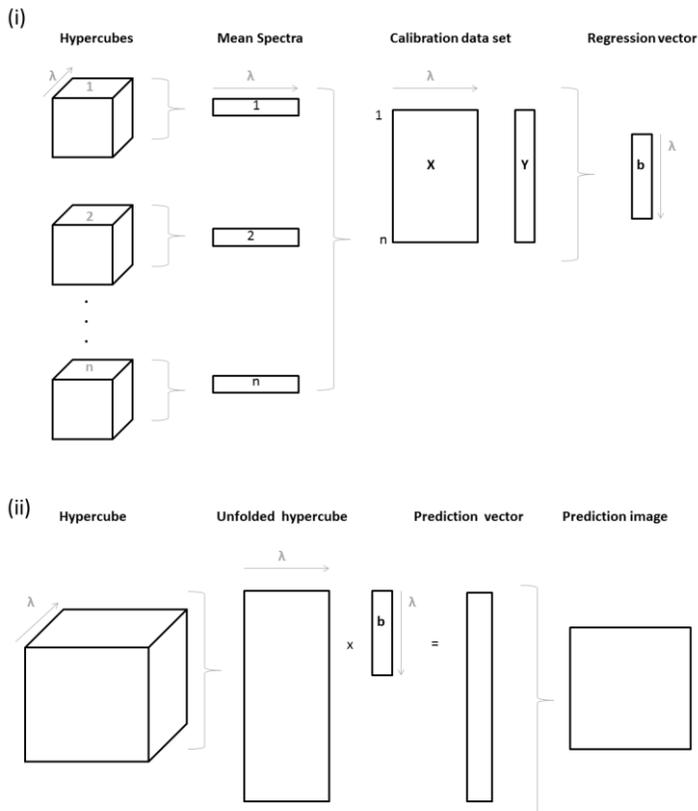


Figure 13. Application of regression modelling to hyperspectral images ('hypercubes'). (i): calculation of mean spectra from sample images and construction of regression vector (b), (ii): application of regression model to hypercube, resulting in prediction image.

Table 3. Prediction metrics for application of PLSR to mean spectra of bread samples

Pretreatment	R2	RMSEP	RPD	#LV
None	0.7	125.9	1.9	4
SNV	0.99	36.5	6.7	3
SG1	0.97	42.9	5.7	4
SG2	0.96	43.2	5.7	4

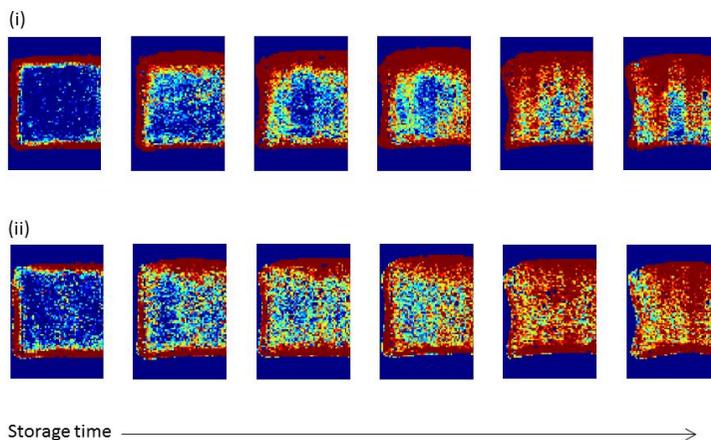


Figure 14. Prediction maps for 3LV SNV (i) and 4 LV SG1 (ii) model. Red = hard, blue = less hard

General rules and suggestions about hyperspectral image processing

1. Have an adequate knowledge of the sample.
2. Be careful not to be deceived by problems of light reflection.
3. Before start the processing, have a look of the raw data and reduce the image size (if necessary).
4. Pay attention at the spikes and dead pixels resulting from the CCD of the camera.
5. Try different spectral pre-processing and combination, choosing the one that underlines in the images the object of your study.
6. Find outliers (whether there are).
7. Chose the best technique that help you to rich the aim of the work.
8. Know very well all the outputs of the algorithm used, in order to draw correct conclusions.
9. Remember that no one solution is wrong because an unexpected result could masks some interesting trend of the data.
10. Enjoy your image processing!

References

- Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, *TrAC - Trends in Analytical Chemistry* 28 (10) (2009) 1201–1222.
- Amigo, J.M., Babamoradi, H., Elcoroaristizabal, S. (2015). References Hyperspectral image analysis. A tutorial. *Analytica Chimica Acta*. Vol. 896, Pages 34–51. doi:10.1016/j.aca.2015.09.030.
- Amigo, J.M., Martí, I., Gowen, A., 2013. Hyperspectral Imaging and Chemometrics. A Perfect Combination for the Analysis of Food Structure, Composition and Quality. *Data Handl. Sci. Technol.* 28, 343–370. doi:10.1016/B978-0-444-59528-7.00009-0
- Elmasry, G., Kamruzzaman, M., Sun, D. W., & Allen, P. (2012). Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review. *Critical Reviews in Food Science and Nutrition*, 52(11), 999-1023.
- ElMasry, G., Wang, N., ElSayed, A., Ngadi, M., 2007. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *J. Food Eng.* 81, 98–107. doi:10.1016/j.jfoodeng.2006.10.016
- F. Firtha, A. Fekete, T. Kaszab, B. Gillay, M. Nogula-Nagy, Z. Kovacs, D.B. Kantor, Methods for improving image quality and reducing data load of NIR hyperspectral images, *Sensors* 8 (2008) 3287–3298.
- Huang, H., Liu, L., Ngadi, M.O., 2014. Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors (Basel)*. 14, 7248–76. doi:10.3390/s140407248
- J.M. Amigo, Practical issues of hyperspectral imaging analysis of solid dosage forms, *Analytical and Bioanalytical Chemistry* 398 (2010) 93–109.
- L. Zhang, M.J. Henson, A practical algorithm to remove cosmic spikes in Raman imaging data for pharmaceutical applications, *Applied Spectroscopy* 61 (2007) 1015–1020.
- Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., García-Navarrete, O. L., & Blasco, J. (2012). Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology*, 5(4), 1121-1142.
- M. Taghizadeh, A.A. Gowen, C.P. O'Donnell, The potencial of visible-near infrared hyperspectral imaging to discriminate between casing soil, enzymatic browning and undamaged tissue on mushrooms (*Agaricus bisporus*) surfaces, *Computers and Electronics in Agriculture* 77 (2011) 74–80.
- N. Sethi, R. Krishna, R.P. Arora, Image compression using Haar wavelet transform, *Computer Engineering and Intelligent Systems* 2 (2011) 1–5.
- Schmalzl, Using standard image compression algorithms to store data from computational fluid dynamics, *Computer & Geosciences* 29 (2003) 1021–1031.
- Sun D.W. Hyperspectral imaging for food quality analysis and control. Da-Wen. Sun (Ed.) Elsevier (2010) ISBN 978-0-12-374753-2.
- Y. Srinivas, D.L. Wilson, Quantitative image quality evaluation of pixel-binning in a flat-panel detector for x-ray fluoroscopy, *Medical Physics* 31 (2004) 131–141.

3.2 VITAMIN C DISTRIBUTION IN ACEROLA FRUIT BY HSI

Cristina Malegori^a, Silvia Grassi^a, Emanuel Marquez^b, Maria Fernanda Pimentel^b, Sergio Tonetto de Freitas^c,
Celio Pasquini^d, Ernestina Casiraghi^a

^a DeFENS Department of Food, Environmental and Nutritional Sciences,
Università degli Studi di Milano – Milano - ITALY

^b Departamento de Engenharia Química, Universidade Federal de Pernambuco – Recife (PE) – BRAZIL

^c Brazilian Agricultural Research Corporation, Embrapa Semiárido, Petrolina (PE) – BRAZIL

^d Instituto de Química, Universidade Estadual de Campinas, Campinas (SP) - BRAZIL

Abstract

Acerola (*Malpighia Emarginata*) is a fruit widely known and consumed throughout Brazil. Its main nutritional value is related to its high content of ascorbic acid. It is known that the content of vitamin C decrease during ripening on the tree, from acerolas picked up green to the purple ones. Ten different acerola fruits picked up according to two different stages of maturity, based on the colour of the peel (5 green and 5 red acerola), were analysed. The hyperspectral images of sliced acerola were caught with a SisuChema Spectral Camera and a NIR device was used to acquire the spectra of vitamin C powder, pure and added with acerola juice (5% w/v). These spectra were used to select the most informative range along the spectrum to discriminate vitamin C contribution. Moreover, spectrum of vitamin C powder was used as reference for computing models with two different correlation techniques: Spectral Angle Mapping (SAM) and correlation coefficient (CC). This work allowed to develop a methodology for a qualitative distribution map of vitamin C in the fruit. Regarding the acerola hyperspectral images, the results demonstrated that HSI on small range of the spectrum strictly linked to vitamin C absorption bands, combined with correlation techniques is a promising approach for the evaluation of vitamin C distribution in acerola fruit.

Introduction

Acerola (*Malpighia emarginata*) is a fruit widely known and consumed throughout Brazil. Its main nutritional value is related to its high content of ascorbic acid, so that the pure acerola juice is consumed “in natura” or used to enrich other foods. Acerola is round shaped, with diameter varying from 3 to 6 cm; a very thin protection peel that quickly ripens encases its fleshy and succulent pulp. At the initial stages of ripening, the fruit is a full green colour, changing to yellow-reddish and finally to red or purple when completely ripened. The colour of the fruits is not only a sign of pigment transformation on the outer surface but it is also related to complex biochemical changes during ripening on the tree. Acerola fruit is rich in many nutrients such as protein, carotenes, thiamine, riboflavin, niacin, proteins, calcium and phosphorus but its main appealing feature is the high Vitamin C content (De Assis et al., 2001).

Vitamin C is defined as the generic term for all compounds exhibiting the biological activity of L-ascorbic acid (Lee and Kader, 2000b). Including ascorbic acid and dehydroascorbic acid, is one of the most important nutritional quality factors in many horticultural crops. It is required for the prevention of scurvy and maintenance of healthy skin, gums and blood vessels. Vitamin C is also known to have many biological functions in collagen formation, absorption of inorganic iron, reduction of plasma cholesterol level, inhibition of nitrosamine formation, enhancement of the immune system, and reaction with singlet oxygen and other free radicals. Vitamin C, as an antioxidant, reportedly reduces the risk of arteriosclerosis, cardiovascular diseases and some forms of cancer (Loewus, 1999). Summarizing, the vitamin C is important in the development and maintenance of the human body. A intake of 100–200 mg/day has been suggested, since stress in modern life is known to increase the requirement for vitamin C (Lee and Kader, 2000b).

The content of vitamin C in fruits and vegetables can be influenced by various factors such as genotypic differences, pre-harvest climatic conditions and cultural practices, maturity and harvesting methods, and postharvest handling procedures. The higher the intensity of light during the growing season, the greater is vitamin C content in plant tissues. Temperature management after harvest is the most important factor to maintain vitamin C of fruits and vegetables; losses are accelerated at higher temperatures and with longer storage durations, low relative humidity, physical damage and chilling injury. Ascorbic acid is easily oxidized, especially in aqueous solutions, and in presence of oxygen, heavy metal ions, especially Cu^{2+} , Ag^{+} , and Fe^{3+} , and by alkaline pH and high temperature. Processing methods and cooking procedures can result in significant losses of vitamin C (Lee and Kader, 2000b).

Vitamin C is a six carbon keto-lactone, a strong reducing agent, serves as an antioxidant and as a cofactor in hydroxylation reactions. The hydrogen donation from ascorbic acid

is thought to be primarily responsible for the anti-oxidant properties attributed to ascorbic acid. NIR spectrum embodies abundant information of hydroxyl (O–H), amino (N–H), and C–H's vibration absorption, so it is a very powerful tool for routinely studying ascorbic acid (Liu et al., 2006). NIR technology is a selective, non-destructive and accurate method based on the application of spectrophotometry together with chemometric reducing cost and time of the analysis and without generate chemical waste. To avoid the interference from other chemical compounds in the sample, a selection of regions of the spectra that could be attributed to vitamin C is advised.

Considering the economic and nutritional importance of the acerola, there is a big interest for both producers and industry to develop analytical physical-chemical methodologies for determining quite rapidly the vitamin C content. Therefore, the aim of this work is to evaluate, by HIS, vitamin C distribution in acerola fruit as discriminant quality parameter so to properly deliver the product, increasing the commercialization potential of the fresh acerola (Nicoli et al., 1999).

Material and methods

Data acquisition

Ten different acerola fruits from Junco cultivar, picked up according to two different stages of maturity based on the colour of the peel (5 green and 5 red acerola), were analysed. The hyperspectral images of acerola were caught with a Sisuchema NIR Spectral Camera (spectral range 900 - 2500 nm; spectral sampling 4 nm; spectral resolution 6 nm) equipped with a 50mm lens with a minimum resolution of 150 μm (Figure 1). To better understand the distribution of vitamin C inside each fruit acerolas were cut in a standard way avoiding the kernel damage. A NIR device Perkin Elmer Frontier FT-Spectrometer equipped with a Reflectance accessory (spectral range 4000-12000 cm^{-1} ; 64 scans; spectral resolution 8 cm^{-1}) was used to acquire the spectra of vitamin C powder, pure and added in increasing quantity, up to 5% w/v, to acerola juice (Marques et al., 2007).

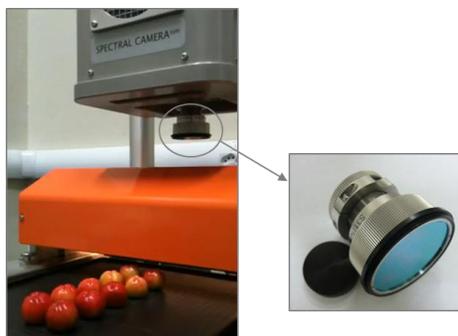


Figure 2: SisuChema NIR Spectral Camera during the acquisition of acerola samples and a detail of the 50 mm lens.

Data analysis

The NIR spectra of acerola juice were pre-processed with SNV, smoothing (11 pts, 2 polynomial order) and first derivative (11 pts, 2 polynomial order) to enhance the spectral trend; whereas the HS images were previously masked for background removal and then equally pre-processed. A priori variables selection between 7200 and 6700 cm^{-1} was carried out, focusing on the most characteristic ascorbic acid absorbance (Liu et al., 2006). Vitamin C powder spectra were used as references for computing models with two different correlation techniques: Spectral Angle Mapping (SAM), that calculate the cosine of the angle between the target spectrum and each spectrum of the images to describe the dissimilarities, and the correlation coefficient (CC), a useful correlation technique when having a priori information.

The abovementioned data analyses were performed using Matlab environment (v. 2014b, The MathWorks).

Results

Variable selection

First, the spectrum of the pure vitamin C powder was acquired to understand the interaction between the NIR source and the molecules that have to be detected (figure 2). To take into account the influence of the analysed matrix on the absorption bands, the spectra of acerola juice enriched in vitamin C (0.5 and 5%) were acquired, for a total of ten spectra. The need of enrichment of the samples of acerola extracts with solutions of pure vitamin C was due to the low variation in the content of this vitamin in the samples (2844 ± 448 mg/100g), evaluated after harvesting. Perhaps the low range observed was due to the fact that samples belong to just one crop and to trees of the same cultivar (Marques et al., 2007). Spectra analysis allowed us to select the range 6700-7200 cm^{-1} of the spectra, that modifies according to increasing amount of vitamin C in acerola juice. Besides ascorbic acid, as a binary acid, can form four kinds of

intermolecular hydrogen bond (OH and CaO) and the OH connected with C3 and C2 have lower vibrational frequency because the conjugate action of double bond and carbonyl. Because of the formation of hydrogen bonds, the absorption of ascorbic acid near 7000 cm^{-1} is a broad peak that cannot be assigned to certain structure exactly (Liu et al., 2006).

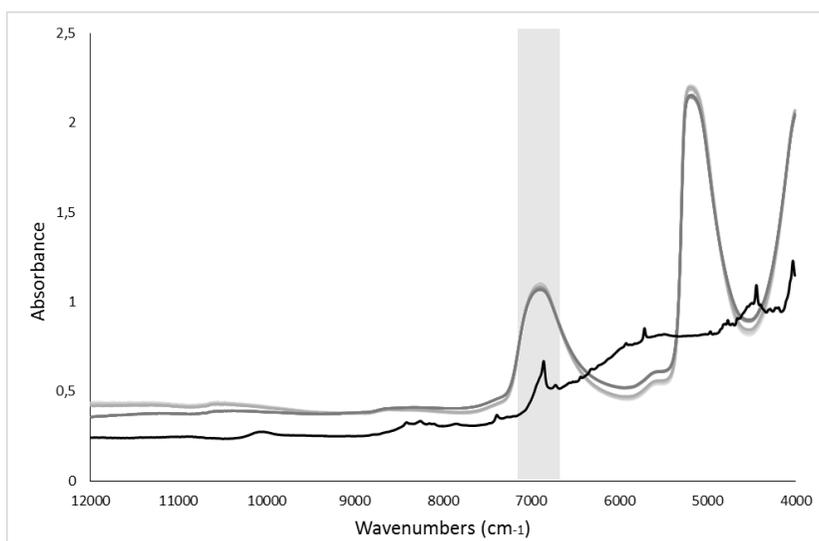


Figure 3: spectra of the acerola juice added with increasing amount of vitamin C (in grey) and spectrum of the ascorbic acid powder (in black). The selected wavelengths are highlighted in grey.

Hyperspectral Image pre-processing

Working on the images, the spatial information was reduced removing the background, thanks to the histogram of the intensity, while the spectral information was highlighted thanks to the pre-processing of the spectra. The best combination of pre-processing was:

- standard normal variate (SNV), in which every data point of the spectrum is subtracted from the mean and divided by the standard deviation,
- Savitzky–Golay smoothing (11 point, 2nd polynomial order) reducing the noise of the spectra,
- Savitzky–Golay second derivative (11 point, 2nd polynomial order) that enhance the information contained in the peak selected.

In figure 3, the spectra of acerola after pre-processing are presented (a) with a zoom between 7200- 6700-cm⁻¹ (b). The stack of spectra represents the absorption of each pixel of the acerola fruit image, after background removal. The colours presented in the picture help to visualize the differences inside the stack, between the spectra of single pixel of the image with higher or lower intensity in the region ascribable to the vitamin C absorption bound. The dotted line in figure 3b is the vitamin C powder spectrum, normalised and pre-processed in the same way of the acerola spectra.

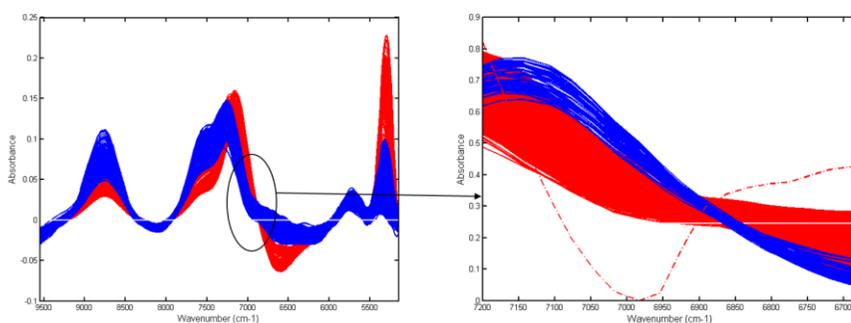


Figure 4: pre-processed spectra of acerola with a zoom in the selected range of the spectra. The dotted line in the right part of the figure is the vitamin C powder spectrum, pre-processed in the same way of the acerola data.

Correlation analysis

After the spectra pre-processing, the removal of the background and the selection of the most informative region of the NIR spectrum, the 3D data cube was ready to be analysed using two different correlation algorithms. The output of both these techniques is the correlation maps between each pixel of the image and the pure spectra of the vitamin C powder, allowing a qualitative evaluation of the vitamin C distribution inside the different tissues of the fruit in an understandable way and without needing further modelling.

The two algorithms applied are Correlation Coefficient (CC) and Spectral Angle Mapping (SAM) using as a priori information the spectrum of the ascorbic acid powder. CC is a mathematical parameter that measures the similarity between two spectra calculating the Pearson's correlation coefficient, however, because many of the spectra in an image are highly correlated with each other, the interpretability of the correlation map may be complicated (REF). Through SAM, the cosine of the angle between the target spectrum (vitamin C) and each pixel-spectrum of the images is calculated, treating the spectra as vectors in a space with dimensionality equal to the number of bands thus permitting to describe the dissimilarities (Park et al. 2004). Smaller angles represent closer matches

to the reference spectrum. For more details, see the reference (Kruse et al. 1993). The coefficients calculated with both the algorithms were scaled from 0 (blue pixels in the images of Figure 4) to 1 (yellow pixels in the images of Figure 4), corresponding to lowest and highest similarity, respectively. In this way, a good contrast for better interpretation of the images was obtained.

This approach was applied only in a qualitative manner although it is possible to set a threshold on the enhanced correlation maps to highlight the percentage of pixels that are correlated with the reference spectrum more than a selected value. Nevertheless, setting this threshold is not an easy task and to our knowledge no objective method has been reported so far (Cairós et al., 2009).

Discussion and conclusion

In this study, two techniques for classifying hyperspectral images were compared: SAM and CC; both the algorithms give similar results but SAM seems to overestimate the vitamin C content compared with the same image elaborated with CC. In literature, it is well known that SAM may provide the sensitivity needed to discriminate between minor changes (Yuhás et al., 1992) and allows to use laboratory spectra directly compared to remotely sensed apparent reflectance spectra (Kruse et al., 1993). Pearson's correlation coefficient (CC) is widely used in the entire research field, in particular for the evaluation of regression and correlation models, but this algorithm is not well-known as correlation technique.

Regarding the application of this type of algorithms on hyperspectral data, SAM have found widespread acceptance in the remote sensing community (Sohn and Rebello, 2002), including monitoring of land cover change (Petropoulos et al., 2010). To our knowledge, this kind of approach has not been applied in the food field and, in particular, no one application for the development of a qualitative distribution map of a particular compound was found in literature. Both CC and SAM are proved to be a tool that permits rapid mapping of the spectral similarity of image spectra to reference spectra. Therefore, due to the similarity of the results achieved with both the algorithms we can conclude that the decision of which algorithm has to be used is often based on experience and available software implementation (Dennison et al., 2004).

Another novelty of this work is the study of the vitamin C distribution inside the fruit. This determination, in fact, is normally obtained on the entire sample, without focusing on the distribution inside the different vegetal tissues (Lee and Kader, 2000a). Regarding the technique used for ascorbic acid determination, the traditional approach is widespread but we can find also application of the NIR spectroscopy for the non-destructive estimation of this nutraceutical compound (Yang and Irudayaraj, 2002).

In particular, it is possible to find one article on the application of NIR spectroscopy for the evaluation of vitamin C content in acerola juice (dos Santos Garcia et al., 2013). According to the lacking references about acerola fruit (Righetto, 2005; Vendramini and Trugo, 2000), we can conclude that the vitamin C content decreases according to the ripeness degree with high level of ascorbic acid in green fruit with a reduction along the evolution of acerola from yellow to red.

No one reference is already present in literature about the application of hyperspectral imaging on this matrix. Moreover the idea of develop a distribution map of a nutraceutical compound in a tropical fruit is completely new.

References

- Cairós, C., Amigo, J.M., Watt, R., Coello, J., Maspocho, S., "Implementation of enhanced correlation maps in near infrared chemical images: Application in pharmaceutical research", *Talanta* 79, 657–664 (2009). doi:10.1016/j.talanta.2009.04.042
- De Assis, S.A., Lima, D.C., De Faria Oliveira, O.M.M., "Activity of pectinmethylesterase, pectin content and vitamin C in acerola fruit at various stages of fruit development", *Food Chem.* 74, 133–137 (2001). doi:10.1016/S0308-8146(01)00104-2
- Dennison, P.E., Halligan, K.Q., Roberts, D. a., "A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper", *Remote Sens. Environ.* 93, 359–367 (2004). doi:10.1016/j.rse.2004.07.013
- Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., Goetz, A.F.H., "The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data". *Remote Sens. Environ.* 44, 145–163 (1993). doi:10.1016/0034-4257(93)90013-N
- Lee, S.K., Kader, A. a., "Preharvest and postharvest factors influencing vitamin C content of horticultural crops". *Postharvest Biol. Technol.* 20, 207–220 (2000). doi:10.1016/S0925-5214(00)00133-2
- Liu, H., Xiang, B., Qu, L., "Structure analysis of ascorbic acid using near-infrared spectroscopy and generalized two-dimensional correlation spectroscopy" *J. Mol. Struct.* 794, 12–17 (2006). doi:10.1016/j.molstruc.2006.01.028
- Loewus, F.A., "Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi". *Phytochemistry*, 52(2), 193-210 (1999). doi:10.1016/S0031-9422(99)00145-4
- Marques, L.G., Ferreira, M.C., Freire, J.T., Freeze-drying of acerola (*Malpighia glabra* L.). *Chem. Eng. Process. Process Intensif.* 46, 451–457 (2007). doi:10.1016/j.cep.2006.04.011
- Nicoli, M.C., Anese, M., Parpinel, M., "Influence of processing on the antioxidant properties of fruit and vegetables". *Trends Food Sci. Technol.* 10, 94–100 (1999).
- Petropoulos, G.P., Vadrevu, K.P., Xanthopoulos, G., Karantounias, G., Scholze, M., "A comparison of spectral angle mapper and artificial neural network classifiers combined with landsat TM imagery analysis for obtaining burnt area mapping". *Sensors* 10, 1967–1985 (2010). doi:10.3390/s100301967
- Righetto, a. M., "Chemical Composition and Antioxidant Activity of Juices from Mature and Immature Acerola (*Malpighia emarginata* DC)". *Food Sci. Technol. Int.* 11, 315–321 (2005). doi:10.1177/1082013205056785
- Sohn, Y., Rebello, N.S., "Supervised and unsupervised spectral angle classifiers". *Photogramm. Eng. Remote Sensing* 68, 1271–1280 (2002).
- Vendramini, A.L., Trugo, L.C., "Chemical composition of acerola fruit (*Malpighia punicifolia* L.) at three stages of maturity". *Food Chem.* 71, 195–198 (2000). doi:10.1016/S0308-8146(00)00152-7
- Yang, H., Irudayaraj, J., "Rapid determination of vitamin C by NIR, MIR and FT-Raman techniques". *J. Pharm. Pharmacol.* 54, 1247–1255 (2002). doi:10.1211/002235702320402099
- Yugas, R., Goetz, F.H., Boardman, J.W., "Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm", in: *Summaries of the Third Annual JPL Airborne Geoscience Workshop*. pp. 147–149 (1992).

3.3 SELECTION OF NIR WAVELENGTHS FROM HYPERSPECTRAL IMAGING DATA FOR QUALITY EVALUATION OF ACEROLA FRUIT

Cristina Malegori^{a*}, Aoife Gowen^b, Emanuel Marquez^c, Maria Fernanda Pimentel^c, Sergio Tonetto de Freitas^d, Celio Pasquini^e, Ernestina Casiraghi^a

^a DeFENS Department of Food, Environmental and Nutritional Sciences,
Università degli Studi di Milano – Milano – ITALY

^b Biosystems Engineering, School of Agriculture, Food Science and Veterinary Medicine,
University College Dublin, - Dublin - IRELAND

^c Departamento de Engenharia Química, Universidade Federal de Pernambuco – Recife (PE) – BRAZIL

^d Brazilian Agricultural Research Corporation, Embrapa Semiárido, Petrolina (PE) – BRAZIL

^e Instituto de Química, Universidade Estadual de Campinas, Campinas (SP) – BRAZIL

* Corresponding author: cristina.malegori@unimi.it

Abstract

Hyperspectral imaging (HSI) is a powerful analytical tool but the large amount of data typically generated during HSI acquisition can limit subsequent industrial applications. In order to extract the relevant information, it is possible to adapt the classical chemometric techniques for the identification of key wavelengths for a given quality control issue. This would allow the development of multispectral devices, intended for specific product control. Acerola is a typical Brazilian fruit that at the initial stages of ripening is green, changing to yellow-reddish colour and finally to purple when it is completely ripened. Due to its high moisture content, a rapid deterioration is commonly observed on ripening. This work is focused on the changes in post-harvest quality of Acerola fruits, with the aim of developing an easy way to visualize the global fruit status during ripening. An investigation of 20 samples for five consecutive days was carried out to highlight the most important wavelengths that characterized the maturity/senescence process of the Acerola fruit, investigating the evolution of acerolas mean spectra during time. The three wavelengths selected were 1883 nm, 1407 nm and 1136 nm, related to absorption and overtones of OH and CH bonds. Grey scale images at these selected wavelengths were combined to provide a false RGB image that allowed evaluation of the ripening process in a rapid and non-destructive manner. These results will facilitate the development of a low cost multispectral imaging system characterized by a simple image based output that could improve quality monitoring of acerola.

Keywords: hyperspectral. Acerola, *Malpighia emarginata*, wavelengths selection, ripeness

Introduction

Hyperspectral imaging (HSI) is a non-destructive technique to explore surfaces in more detail than single point spectroscopy. Each pixel of a hyperspectral image, usually highly correlated to its neighbours, contains the spectrum of that specific position. Thus, the hyperspectral image contains both spatial and spectral information of a sample. The data are organized in three-dimensional blocks, called hypercubes. Many devices for acquiring hyperspectral images have been manufactured and there is an increasing interest for improving the data analysis techniques applied to such complex datasets (Vidal and Amigo, 2012). Despite being one of the main advantages of hyperspectral systems, the large datasets routinely encountered in HSI, can complicate the extraction of useful information since much of the information obtained is redundant (Shaw, 2003). Multivariate data analysis, or Chemometrics, is a suitable approach to reduce the dimensionality of the data while retaining the most useful spectral information. A useful methodology that can reduce the data dimension is the variable selection. In variable selection wavelengths that would be most influential on product quality evaluation are selected, removing the wavelengths that have no or low discrimination power. In this way, the data dimension is reduced while preserving the most useful information (ElMasry et al., 2007). The variables selected depend on the behaviour of spectral responses under modification of the samples and on differences among them.

Fruit quality is defined by a series of external characteristics that make the product more or less attractive to the consumer, including suitability to be eaten as fresh or stored for reasonable period without deterioration (Kays, 1999). Fruit quality could be considered as a multivariate concept encompassing the physical, physiological, nutritional, and pathological attributes that affect shelf life (ElMasry et al., 2007). The ripe phenotype is the summation of biochemical and physiological changes that occur at the terminal stage of fruit development, rendering it edible and desirable to seed-dispersing animals. These changes, although variable among species, generally include modification of cell wall ultrastructure and texture, conversion of starch to sugars, increased susceptibility to post-harvest pathogens, alterations in pigment biosynthesis and accumulation, and heightened levels of flavour and aromatic volatiles (Giovannoni, 2001). Nowadays, fruit are managed manually or automatically on the basis of external quality features but the high risk of human error in the classification process has been underlined as one of the most important drawbacks that machine vision can help preventing (Paulus et al., 1997). Vision systems for fruit sorting were traditionally based on video cameras working in the visible wavelength range, limited to obtain information on the external aspect like colour or damages presence. More information about sample composition might be obtained by computer vision, which can acquire a set of optimised monochromatic images at few selected wavelengths, and can make possible to estimate or discover features, difficult to uncover with traditional vision systems like dry matter, total soluble solids and acidity (Lorente et al., 2012). The use of NIR spectroscopic information and hyperspectral imaging as a non-destructive measurements of quality attributes has been implemented only recently in the post-harvest field (<https://www.tomra.com/en>; <http://www.aweta.nl/it>).

Acerola (*Malpighia emarginata*, Junco cultivar) is a fruit native to Central America and

Northern South America, with some of the largest plantings occurring in Brazil, and it has recently been introduced in subtropical areas throughout the world, including Southeast Asia, India and South America. The plant provides flowers and fruits at different stages and, consequently, long fruiting periods are observed during the year. The fruit presents a short postharvest shelf-life (2 to 3 days) at room temperature (Vendramini and Trugo, 2000). The ripening process of acerola fruit involves a succession of complex biochemical reactions including starch hydrolysis, conversion of chloroplasts into chromoplasts with the transformation of chlorophyll and the production of carotenoids, anthocyanins, phenolics and volatile compounds that result in a colour change (Speirs and Brady, 1991). It has been increasingly recognized that in addition to vitamin C acerola contains other functional constituents such as carotenes, thiamine, riboflavin, niacin, proteins, and mineral salts, mainly iron, calcium and phosphorus which are beneficial to human health (Hanamura, 2008). For these reasons Acerola is considered a super-fruit, nevertheless research on this Brazilian fruit is very limited and it has not yet been studied using non-destructive techniques for quality estimation. Furthermore, its biochemical evolution is not completely known and its perishability is very rapid, preventing the export of the fresh fruit. Table 1 shows the chemical composition of unripe and ripe acerolas.

The aim of this work is to lay the foundation for the development of a multispectral camera for the quality evaluation of acerola fruit at different ripeness stages. This overarching aim was achieved by meeting two specific objectives: (1) to evaluate the differences in hyperspectral images of acerolas at different maturity stages for the selection of useful wavelengths; (2) to develop a simple imaged based approach to allow global, rapid and non-destructive evaluation of the maturity/senescence degree of acerolas.

Parameters	Acerola juice	
	Immature	Mature
pH	3.62	3.40
Total solids (g/100g juice)	5.4 ± 0.2	5.5 ± 0.02
Soluble solids (°Brix)	5.1 ± 0.0	5.7 ± 0.0
Sugars (g/100g of juice)		
Fructose	2.14 ± 0.04	3.33 ± 0.02
Glucose	0.99 ± 0.04	0.88 ± 0.02
Sucrose	Nd	0.02 ± 0.01
Acids (g/100g juice)		
Malic	0.25 ± 0.04	0.38 ± 0.03
Citric	0.012 ± 0.001	0.003 ± 0.003
Tartaric	0.01 ± 0.00	0.002 ± 0.000
Ascorbic	1.85 ± 0.02	0.80 ± 0.02
Vitamin C	1.9 ± 0.028	0.97 ± 0.031
Total phenolics (mg catechin/g juice)	3.8 ± 0.02	1.35 ± 0.01
Composition (% dw basis)		
Sugars	58.5	75.55
Acids	39.3	21.6
Vitamin C	35.18	17.54
Total phenolics (mg catechin/g juice)	7.0	2.4

Figure 5: Chemical composition of acerola juices at different stages of maturity (Righetto, 2005)

Materials and methods

Hyperspectral images of 20 acerolas were acquired for 5 consecutive days, giving a total of 100 images. The fruits were selected at different initial ripeness degrees, based on the colour of the peel, and were stored at room temperature ($25 \pm 2^\circ\text{C}$). According to literature, green was chosen as the initial stage, yellow as the middle and red as the last stage of maturity (Vendramini and Trugo, 2000). A SisuChema Hyperspectral Imaging System (900 - 2500 nm) was used, equipped with a 50 mm lens with a minimum spatial resolution of 150 μm . One of the main aspects to be taken into account is the spherical shape of acerolas. In this case, two problems arise: the presence of bright spots caused by the reflection of light and the progressive darkness of the borders, both caused by the effect of Lambert's cosine law (Lorente et al., 2012). To elaborate the huge amount of data and try to overcome the light artefacts, a chemometric approach was applied. A binary mask was first created to produce an image containing only the fruit, avoiding any interference from the background. For this task, the image at the wavelength in which the fruit appeared opaque, compared with the background, and can be segmented easily by simple thresholding, was selected (1111 nm). The spectra of the fruit (without background) were pre-processed using automatic baseline correction (SNV) and multiplicative scatter correction (MSC) and then the data were smoothed (Savitzky-Golay smoothing, 11 points) and mean centred. Other pre-processing treatments, such as first and second derivatives were tested, but the results were not improved; no outliers, dead pixels or spikes were detected in the hyperspectral images. An investigation of the change in acerolas mean spectra during time was carried out to highlight the most important three wavelengths that characterized the ripeness process of the acerola fruit. This was done by subtracting the mean spectrum at day 1 from all subsequent mean spectra, for a given fruit. Consequently, for each fruit, the images at the three wavelengths with the widest variation along time, were plotted in grey scale and combined to make a false RGB image allowing an intuitive understanding of the qualitative status of each fruit.

Results

Figure 2 shows the mean spectra of all the fruits analysed during the five days shelf life. Simple visual approaches to evaluate the changes in spectra were not satisfactory since did not highlight remarkable differences. For this reason, the mean spectrum of each fruit at day 1 was subtracted from all subsequent images of the same fruit (day 2, 3, 4, 5), to identify the main changes, as shown in Figure 2,. As it can be seen in the right part of figure 2, the spectrum of the initially less ripe acerola does not change much during time, while from the image of the initially more ripe acerola (that in five days became senescent) it is possible to highlight the most important three wavelengths that characterized the acerola spectra. The three wavelengths, that appear to exhibit consistent and large changes with time, were 1883 nm, due to the second overtone of the C=O double bond, 1407 nm, the absorption band of the CH first overtone combination and first OH overtone, and 1136 nm characterized by the second overtone of CH bond. It difficult to interpret why these bands are changing with ripening due to the biochemical complexity of the ripening process that as previously said, has not been widely studied. To describe acerola using these most informative wavelengths, a RGB

false image was created using the selected wavelengths as the red, green and blue channel.

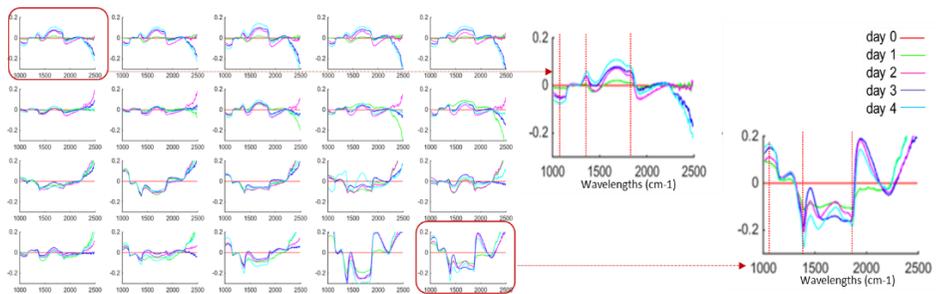


Figure 2: mean spectra of all the fruits analysed during the five days shelf life with a zoom on acerola number 1 and acerola number 20

As an example, the false colour images of the less ripe acerola number 1, and the more ripe , acerola number 20, (that in five days passed from ripe to overripe), are presented in figure 3 allowing monitoring of the ripening/senescence process. It is evident that the changes of the peel colour in acerola number 1 during ripening are not well represented in the false images while the senescence/degradation, which occurs on acerola number 20, is well-highlighted (figure 3).

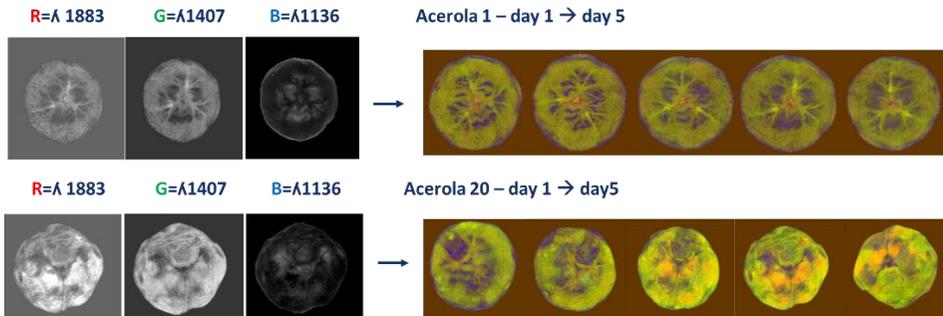


Figure 3: image of acerola 2 and acerola 20 at those selected wavelengths were combined to produce an RGB false image that was followed during 5 days

Discussion and Conclusion

The methodological approach for the wavelengths selection allowed the initial aim to be reached: the evaluation of Acerola post-harvest quality during ripening/senescence process, in a view of developing a simple multispectral device for qualitative evaluation. During the work, unexpected variability of the acerola NIR profile was highlighted during time. In fact, a colour change from green to red, which is considered as a ripeness evolution index, does not correspond to a modification of the fruit composition, commonly reflected in NIR spectra. Spectral modification, that allowed variable selection, appears more evident when acerola colour turns from red to purple-brown

during the degradation process of the fruit. As a matter of fact, the distinction between ripening and senescence has never been finely drawn (Brady 1987) so the phenomena that we highlighted with the hyperspectral approach could be defined as overripe or initial stage of senescence. Due to the poor knowledge about acerola, a step back is necessary in order to well delineate the biochemical evolution during ripening. First of all, it will be important to focus on ethylene production and cellular respiration defining if acerola could be classified as a climacteric fruit. So far, there is no evidence in literature, but if acerola could be considered a climacteric fruit, like banana, apple and mango, it could be harvested unripe and easily exported to other countries. This will be of paramount importance and will widely increase the potential market of this super-fruit.

Acknowledgments

The corresponding author thanks the ICNIR for the John Shenk Travel Grants and the SISNIR for funding the participation to the NIR 2015 Conference.

References

- ElMasry, G., Wang, N., ElSayed, A., Ngadi, M. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *J. Food Eng.* 2007 81, 98–107. doi:10.1016/j.jfoodeng.2006.10.016
- Giovannoni, J. Molecular Biology of Fruit Maturation and Ripening. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 725–749, 2001. doi:10.1146/annurev.arplant.52.1.725
- Kays, S.J. Preharvest factors affecting appearance. *Postharvest Biol. Technol.* 15, 233–247, 1999. doi:10.1016/S0925-5214(98)00088-X
- Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., García-Navarrete, O.L., Blasco, J. Recent Advances and Applications of Hyperspectral Imaging for Fruit and Vegetable Quality Assessment. *Food Bioprocess Technol.* 5, 1121–1142 2012.
- Paulus, I., De Busscher, R., Schrevers, E. Use of Image Analysis to Investigate Human Quality Classification of Apples. *J. Agric. Eng. Res.* 68, 341–353 1997. doi:10.1006/jaer.1997.0210
- Shaw, P.J.A. *Multivariate statistics for the Environmental Sciences*, New York 2003.
- Vendramini, A.L., Trugo, L.C. Chemical composition of acerola fruit (*Malpighia puniceifolia* L.) at three stages of maturity. *Food Chem.* 71, 195–198 2000. doi:10.1016/S0308-8146(00)00152-7
- Vidal, M., Amigo, J.M. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemom. Intell. Lab. Syst.* 117, 138–148 2012. doi:10.1016/j.chemolab.2012.05.009
- Speirs, J., & Brady, C. J. Modification of gene expression in ripening fruit. *Functional Plant Biology*, 18(5), 519-532 1991.
- Hanamura, T., Uchida, E., & Aoki, H. Changes of the composition in acerola (*Malpighia emarginata* DC.) fruit in relation to cultivar, growing region and maturity. *Journal of the Science of Food and Agriculture*, 88(10), 1813-1820, 2008.

4. OVERALL CONCLUSIONS

The three non-destructive techniques delineated in this work have proved to be one of the most efficient and advanced tools for safety and quality evaluation in food industry answering the need for accurate, fast and objective food inspection methods that ensure safe production throughout the entire production process. Food quality could be easily monitored using spectroscopic instruments with the final goal of the development of simplified device and the plant hygiene management could be improved using fast, green and on-line detection methods based on spectroscopy information.

Nowadays, there is a great interest about hyperspectral imaging thank to the added spatial dimension that enables the mapping of chemical components in a non-homogeneous sample. The combination of spectral and spatial information could open new possibility for characterize heterogeneous matrix. In light of the presented applications, not only spectroscopy but also hyperspectral imaging could be implemented in the future in a view of a simplified and low-cost devices based on the selection of few interesting wavelengths.

d. COMUNICATIONS AND POSTERS

NIR 2013 - 16th International Conference on Near Infrared Spectroscopy – la Grande Motte, France:

“Wavelengths selection with a view to a simplified handheld optical system to estimate grape ripeness” Malegori, C., Beghi, R., Giovenzana, V., Civelli, R., Guidetti, R., Casiraghi, E. **Poster**

10th AIIA Conference, Associazione Italiana Ingegneria Agraria – Viterbo, Italy:

“Selection of optimal wavelengths for decay detection in fresh-cut Valerianella locusta Laterr” Malegori, C., Giovenzana, V., Beghi, R., Civelli, R., Guidetti, R. **Poster**

18th Workshop on the Developments in the Italian PhD Research on Food Science Technology and Biotechnology. Università degli Studi di Padova – Italy:

“Evaluation of safety, quality and shelf life of agro-food product using non-destructive optical and instrumental-sensory devices” Malegori, C., Casiraghi, E., Guidetti, R. **Poster**

19th Workshop on the Developments in the Italian PhD Research on Food Science Technology and Biotechnology, University of Bari - Bari, Italy:

“IR spectroscopy and Hyperspectral imaging for the non-destructive evaluation of fruit quality: Acerola” Malegori, C., Casiraghi, E., Guidetti, R. **Poster**

6^a Simposio Italiano di Spettroscopia NIRItalia 2014, Modena – Italy:

“La spettroscopia infrarosso per discriminare microrganismi responsabili di contaminazioni nell'industria alimentare” Malegori, C., Grassi, S., Foschino, R., Casiraghi, E. **Poster**

2nd International Conference on Optical Characterization of Materials - OCM2015 - Karlsruhe

“Non-destructive evaluation of Acerola quality using IR spectroscopy” Malegori, C., Marquez, E., Pimentel, M.F., Tonetto de Freitas, S., Pasquini, C., Casiraghi, E. **Poster**

“Image texture potential for the early detection of biofilm on different surfaces” Malegori, C., Franzetti, L., Guidetti, R., Casiraghi, E. **Oral Presentation**

5th MoniQA International Conference 2015 "Food and Health - Risks and Benefits" - Porto, Portugal

“Optical techniques for food safety purpose” Malegori, C., Grassi, S., Casiraghi, E. **Oral presentation**

20th Workshop on the Developments in the Italian PhD Research on Food Science Technology and Biotechnology, Università di Perugia - Perugia, Italy:

“Spectroscopy, image analysis and hyperspectral imaging for food quality and plant hygiene: a chemometric approach” Malegori, C., Casiraghi, E., Guidetti, R. **Oral Presentation**

NIR2015 - 17th International Conference on Near Infrared Spectroscopy - Foz do Iguacu (Brazil)

“Comparison between FT-NIR and Micro-NIR in the evaluation of Acerola fruit quality, using PLS and SVM regression algorithms” Malegori, C., Marquez, E., Pimentel, M.F., Tonetto de Freitas, S., de França Souza, F., Pasquini, C., Casiraghi, E. **Poster**

“Selection of NIR wavelengths from hyperspectral imaging data for the quality evaluation of Acerola fruit” Malegori, C., Gowen, A., Marquez, E., Pimentel, M.F., Tonetto de Freitas, S., Pasquini, C., Casiraghi, E. **Poster**

“HIS for quality evaluation of vitamin C content in Acerola fruit” Malegori, C., Grassi, S., Marquez, E., Pimentel, M.F., Tonetto de Freitas, S., Pasquini, C., Casiraghi. **Oral presentation**

e. COURSES AND COLLABORATION

Experience abroad in Chile (from March to April 2012) at INIA, Instituto Nacional de Investigaciones Agropecuarias - Centro Regional Quilamapu (Chile) - working on the research project SOQUIC

Attendance to the course "Chemometrics" held by Professor Roberto Todeschini at the Università degli Studi di Milano - Bicocca.

Attendance to the course "Fundamentals of Matlab" held by Professor Jose Amigo at the Università degli Studi di Milano.

Attendance to the pre-course "Introduction to Multivariate Image Analysis (MIA)" held by Barry Wise at the conference NIR2013 –La Grande Motte (France).

Attendance to the workshop "Hyperspectral imaging" at ITIA, Institute of Industrial Technologies and Automation - National Research Council, Milan.

Attendance to the course "Hyperspectral and Multichannel Image analysis" held by Professor Jose Amigo at the University of Copenhagen.

Experience abroad in Brazil (from March to September 2014) at UFPE- Universidade Federal de Pernambuco, under the supervision of Professor Fernanda Pimentel, and at UNICAMP - Universidade Estadual de Campinas, under the supervision of Professor Celio Pasquini.

Attendance to the course "Hyperspectral and Multichannel Image analysis - advanced" held by Professor Jose Amigo at the UFPE Univesidade Federal de Pernambuco - Brazil.

Attendance to the "Winter School – Le tecniche spettroscopiche: strumenti innovativi applicati all'analisi dei settori ambientale e agro-alimentare – nuove sfide per il futuro" organized by SISNIR Società Italiana di Spettroscopia NIR.

Cooperation with Professor Aiofe Gowen (April 2015) for the elaboration of hyperspectral images using Matlab environment.

Attendance to the pre-course intitled "Near Infrared Hyperspectral imaging" held by José Amigo and Aoife Gowen at the conference NIR2015 – Foz do Iguaçu (Brazil)