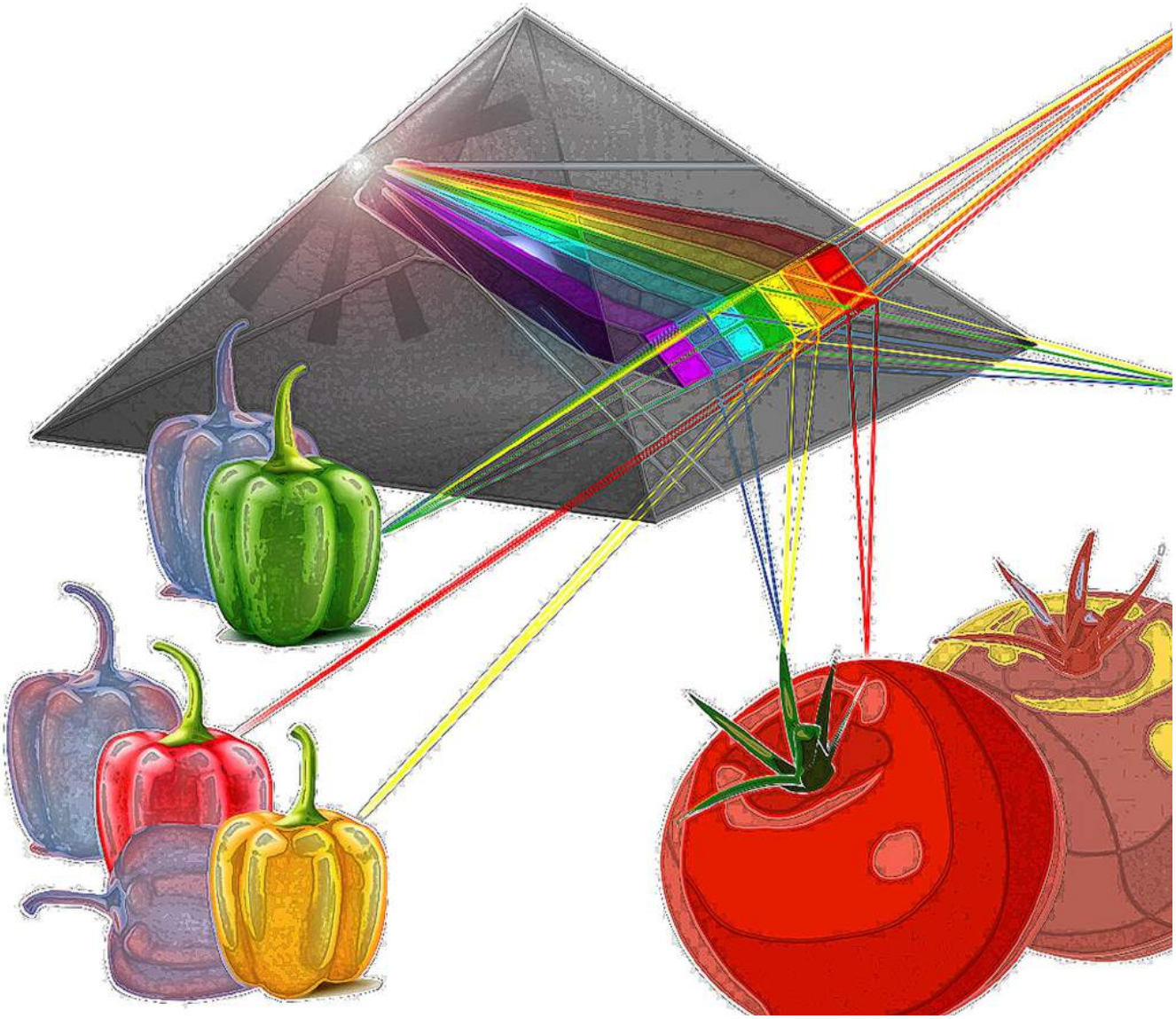
UNIVERSITY OF CÓRDOBA

Near infrared spectral sensors for the characterization, authentication and quality and safety assurance of horticultural products

PhD Dissertation

Irina Torres Rodríguez



TITULO: *Near infrared spectral sensors for the characterization, authentication and quality and safety assurance of horticultural products*

AUTOR: *Irina Torres Rodríguez*

Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

https://www.uco.es/ucopress/index.php/es/
ucopress@uco.es

UNIVERSIDAD DE CÓRDOBA

Programa de Doctorado Ingeniería Agraria, Alimentaria, Forestal y del Desarrollo Rural Sostenible por la Universidad de Córdoba y la Universidad de Sevilla

Sensores espectrales de infrarrojo cercano para la caracterización, autentificación y aseguramiento de la calidad y seguridad de productos hortofrutícolas

Near infrared spectral sensors for the characterization, authentication and quality and safety assurance of horticultural products

TESIS DOCTORAL

Irina Torres Rodríguez

Directoras:

Dra. Mª Teresa Sánchez Pineda de las Infantas
Dra. Dolores Pérez Marín

Julio 2019

Sensores espectrales de infrarrojo cercano para la caracterización, autentificación y aseguramiento de la calidad y seguridad de productos hortofrutícolas

TESIS DOCTORAL

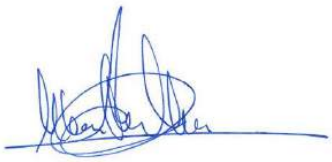
para aspirar al grado de Doctora por la Universidad de Córdoba con mención de doctorado internacional presentada por la Ingeniera Agrónoma y Máster en Proyectos y Gestión de Plantas Agroindustriales Dña. *Irina Torres Rodríguez*

La Doctoranda

Fdo.: Irina Torres Rodríguez

VºBº Las Directoras

Fdo.: Profª. Dra. Mª Teresa Sánchez Pineda de las Infantas



Fdo.: Profª. Dra. Dolores Pérez Marín

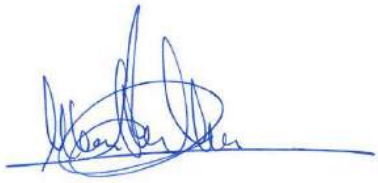
Julio 2019

Mª Teresa Sánchez Pineda de las Infantas, Catedrática de Universidad del Departamento de Bromatología y Tecnología de los Alimentos de la Universidad de Córdoba y Dolores Pérez Marín, Catedrática de Universidad del Departamento de Producción Animal de la Universidad de Córdoba

I N F O R M A N:

Que la Tesis titulada "**Sensores espectrales de infrarrojo cercano para la caracterización, autentificación y aseguramiento de la calidad y seguridad de productos hortofrutícolas**", de la que es autora Dña. Irina Torres Rodríguez, ha sido realizada bajo nuestra dirección durante los años 2015, 2016, 2017, 2018 y 2019; y cumple los requisitos académicos exigidos por la Legislación vigente para optar al título de Doctora por la Universidad de Córdoba con mención de doctorado internacional.

Y para que conste a los efectos oportunos firman el presente informe en Córdoba a 24 de junio de 2019



Fdo.: Profª. Dra. Mª Teresa Sánchez Pineda de las Infantas

Fdo.: Profª. Dra. Dolores Pérez Marín

TÍTULO DE LA TESIS:

SENSORES ESPECTRALES DE INFRARROJO CERCANO PARA LA CARACTERIZACIÓN, AUTENTIFICACIÓN Y ASEGURAMIENTO DE LA CALIDAD Y SEGURIDAD DE PRODUCTOS HORTOFRUTÍCOLAS

DOCTORANDA:

IRINA TORRES RODRÍGUEZ

INFORME RAZONADO DE LAS DIRECTORAS DE LA TESIS

La Tesis cuyo título se menciona arriba ha podido adaptarse, desde sus inicios, a la metodología y el diseño programados, derivando todo ello en la obtención de resultados de indudable relevancia científica y tecnológica.

En primer lugar, hay que destacar que del trabajo de esta Tesis se han establecido las bases científico-técnicas para el desarrollo de modelos de predicción NIRS cuantitativos y se han obtenido un amplio abanico de aplicaciones NIRS relativas a la determinación de parámetros de calidad, principalmente aquellos relacionados con parámetros morfológicos (peso, longitud, diámetro), el color, la textura, y los contenidos en materia seca, azúcares y acidez, en frutas (naranjas y mandarinas) y hortalizas (tomate, calabacín, espinaca y pimiento), para la cuantificación no destructiva de los cambios físico-químicos que tienen lugar durante la maduración de frutas y hortalizas en árbol o en la mata, y para facilitar la toma de decisiones sobre el momento óptimo de cosecha. Con este fin se han utilizado tanto estrategias de

regresión lineal (regresión en mínimos cuadros parciales modificada) como algoritmos de regresión no lineal (algoritmo LOCAL). En naranjas y mandarinas, se han realizado igualmente, modelos predictivos de calidad físico-química en árbol que pudiesen ser aplicados a frutas del Género *Citrus*, independientemente de la especie analizada.

En segundo lugar, se ha llevado a cabo la determinación cuantitativa mediante espectroscopía NIR de parámetros (contenido en nitratos) relacionados con la seguridad alimentaria con el objetivo de determinar el destino final (alimentación infantil, procesado industrial, consumo en fresco) de calabacines y espinacas. Asimismo, se han realizado modelos cualitativos destinados a evaluar la viabilidad del uso de la tecnología NIRS para determinar la posible utilización de dichas hortalizas en alimentación infantil.

Asimismo, se han realizado modelos cualitativos NIRS destinados a la autentificación del origen (cultivo al aire libre o en invernadero) en pimientos tipo California y se han desarrollo y evaluado modelos predictivos NIRS para la optimización del manejo de factores precosecha (irrigación), destinados a favorecer la toma de decisiones en campo.

Por otro lado, se han determinado, para las distintas aplicaciones desarrolladas en este Trabajo de Investigación, los instrumentos NIRS más idóneos en función, tanto del momento (durante el desarrollo y maduración de los frutos, en la cosecha, en su conservación poscosecha) como del lugar (en árbol o en la mata, en las líneas de manipulación y clasificación en la industria) para la realización de dichas determinaciones, teniendo en cuenta asimismo, los parámetros de interés elegidos y las características intrínsecas del producto. Se ha establecido una metodología de actuación para cada uno de los instrumentos NIRS evaluados.

Finalmente, se ha iniciado la puesta a punto de una metodología de análisis destinada a la determinación del rendimiento de cosecha de naranjas verdes en árbol empleando el análisis de imágenes hiperespectrales.

Lo anteriormente expuesto justifica plenamente que la forma más idónea de presentación de esta Tesis Doctoral sea el compendio de publicaciones científicas.

La doctoranda ha tenido la posibilidad de formarse, no sólo en aspectos científicos-técnicos ligados a la tecnología NIRS e imagen hiperespectral, sino también en los relacionados con la ingeniería y tecnología pre y poscosecha de frutas y hortalizas. Asimismo, la doctoranda ha complementado su formación realizando las siguientes estancias de investigación: 1) en el Agriculture Research Service (Beltsville, USDA, EE.UU.), bajo la supervisión del Dr. Moon S. Kim (Junio-Agosto 2016); 2) en el Nondestructive Bio-Sensing Laboratory, Department of Biosystems Machinery Engineering, College of Agriculture and Life Science, Chungnam National University, Daejeon, República de Corea, bajo la supervisión del profesor Byoung-Kwan Cho (Septiembre-Noviembre 2017); 3) en el Department of Food Science, University of Copenhagen, Dinamarca, bajo la supervisión del Dr. José Manuel Amigo Rubio (Julio-Septiembre 2018).

Los trabajos publicados en forma de artículos científicos relacionados con los resultados de la Tesis Doctoral son los siguientes:

1. Torres, I., Pérez-Marín, D., De la Haba, M.J., Sánchez, M.T. 2015. Fast and accurate quality assessment of Raf tomatoes using NIRS technology. Postharvest Biology and Technology 107, 9–15.

2. Sánchez, M.T., Pérez-Marín, D., Torres, I., Gil, B., De la Haba, M.J. 2017. Use of NIRS technology for on-vine measurement of nitrate content and other internal quality parameters in intact summer squash for baby food production. Postharvest Biology and Technology 125, 122–128.

3. Torres, I., Pérez-Marín, D., De la Haba, M.J., Sánchez, M.T. 2017. Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors. Biosystems Engineering 153, 140–148.

4. Pérez-Marín, D., Torres, I., Entrenas, J.A., Vega, M., Sánchez, M.T. 2019. Pre-harvest screening on-vine of spinach quality and safety using NIRS technology. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 207, 242–250.

5. Torres, I., Sánchez, M.T., Benlloch-González, M., Pérez-Marín, D. 2019. Irrigation decision support based on leaf relative water content determination in olive grove using near infrared spectroscopy. Biosystems Engineering 180, 50–58.

6. Sánchez, M.T., Torres, I., De la Haba, M.J., Chamorro, A., Garrido-Varo, A., Pérez-Marín, D. 2019. Rapid, simultaneous, and *in situ* authentication and quality assessment of intact bell peppers using near-infrared spectroscopy technology. Journal of the Science of Food and Agriculture 99, 1613–1622.

7. Torres, I., Sánchez, M.T., De la Haba, M.J., Pérez-Marín, D. 2019. LOCAL regression applied to a citrus multispecies library to assess chemical quality parameters using near infrared spectroscopy. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 217, 206–214.

8. Torres, I., Sánchez, M.T., Entrenas, J.A., Garrido-Varo, A., Pérez-Marín, D. 2019. Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors. Postharvest Biology and Technology 154, 21–30.

9. Torres, I., Sánchez, M.T., Cho, B.K., Garrido-Varo, A., Pérez-Marín, D. Setting up a methodology to distinguish between green oranges and leaves using hyperspectral imaging. Enviado a la revista Computers and Electronics in Agriculture. Julio 2019.

La doctoranda ha participado en los siguientes congresos, simposios y conferencias:

- VIII Congreso Español de Ingeniería de alimentos (CYTA/CESIA 2015), organizado por la Universidad de Extremadura. Badajoz, España. 8-10 Abril de 2015. Presentación de dos comunicaciones escritas: "Monitorización de la calidad y madurez de naranjas analizadas en árbol mediante tecnología NIRS" e "Incorporación de sensores NIRS para la categorización "on-line" de mandarinas en industria".

- 17th International Conference on Near Infrared Spectroscopy, organizada por el International Council for Near Infrared Spectroscopy (ICNIRS). Foz do Iguassu, Brazil. 18th to 23rd October 2015. Presentación de la comunicación escrita: "Near Infrared Spectroscopy for discrimination between Extra Virgin Olive Oil and Lampante Olive Oil".

- VIII International Postharvest Symposium, organizado por la Universidad Politécnica de Cartagena. Cartagena, Murcia, España. 21-24 junio 2016. Presentación de la comunicación oral: "*In situ* determination of external quality parameters in intact summer squash using near-infrared reflectance spectroscopy", y de la comunicación escrita "*In situ* prediction of internal quality parameters in intact bell peppers using NIRS technology".

- International Diffuse Reflectance Conference. Wilson College, Chambersburg, Pennsylvania, USA. July 30-August 5, 2016. Presentación de la comunicación escrita: "LOCAL algorithm for the prediction of optimum harvest dates in fruits of different genus".

- V Congreso Científico de Investigadores en Formación de la Universidad de Córdoba. "Creando Redes". Córdoba, España. 30 de Noviembre y 1 y 2 Diciembre de 2016. Presentación de la comunicación oral: "Sensores espectroscópicos NIRS para la identificación, caracterización y mejora varietal de productos hortofrutícolas".

X

- 18th International Conference on Near Infrared Spectroscopy, organizada por el International Council for Near Infrared Spectroscopy (ICNIRS). Copenhagen, Denmark. 11th to 15th June 2017. Presentación de la comunicación oral: "Application of LOCAL regression methods to NIR spectra database of citrus fruits for the assessment of internal quality", y de la comunicación escrita: "Evaluation of NIRS for the prediction of physical and sensorial quality parameters in intact Raf tomato".

- 5th Food Integrity Conference 'Assuring the Integrity of the Food Chain', organizado por Eurofins. Nantes, France. 14th to 15th November 2018. Presentación de la comunicación escrita: "*In situ* assessment of safety of vegetables by measuring the nitrate content using near infrared technology".

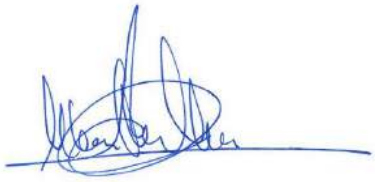
- VI International Symposium on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain - Model-IT 2019, organizado por la Università degli Studi di Foggia. Molfetta, Italy. 9-12 June 2019. Presentación de la comunicación oral: "Online postharvest assessment of quality in spinach plants using Near-Infrared Spectroscopy".

- 19th International Conference on Near Infrared Spectroscopy, organizada por el International Council for Near Infrared Spectroscopy (ICNIRS). Gold Coast, Queensland, Australia. 15th to 20th September 2019. Presentación de la comunicación escrita: "Routine safety control of vegetables at different stages of the production chain using Near Infrared Spectroscopy".

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 24 de junio de 2019

Fdo.: Profª. Dra. Mª Teresa Sánchez Pineda de las Infantas



Fdo.: Profª. Dra. Dolores Pérez Marín

"Fue el tiempo que pasaste con tu rosa…
lo que la hizo tan importante"

(El Principito)

A mi familia,
especialmente a los más pequeños

Agradecimientos

Deseo expresar mi sincera gratitud y reconocimiento a todas las personas que han hecho posible la realización de este Trabajo de Investigación:

A la Dra. María Teresa Sánchez Pineda de las Infantas, Catedrática de Universidad del Departamento de Bromatología y Tecnología de los Alimentos de la Universidad de Córdoba, y codirectora de esta Tesis.

A la Dra. Dolores Pérez Marín, Catedrática de Universidad del Departamento de Producción Animal de la Universidad de Córdoba, y codirectora de esta Tesis.

Al Grupo de Investigación AGR-128 "Ingeniería de Sistemas de Producción Agro-Ganaderos", del Departamento de Producción Animal de la Universidad de Córdoba; en particular, a los Profesores Ana Garrido Varo y José Emilio Guerrero Ginel y a la Dra. Cecilia Riccioli.

Al Grupo de Investigación AGR-193 "Tecnología de Alimentos", del Departamento de Bromatología y Tecnología de los Alimentos de la Universidad de Córdoba; en particular a la Profesora Mª José de la Haba de la Cerda.

A la Empresa Gelagri Ibérica, S.L., especialmente a D. José Luis Onetti Carranza, D. Sergio Lorente Aguado, Dª Natividad Luqui Muñoz y D. Antolín Imás Gurrea.

Al Ministerio de Educación, Cultura y Deporte, para la concesión de una ayuda para la Formación de Profesorado Universitario (FPU). Convocatoria 2014.

A mi familia, por su confianza y su apoyo incondicional durante todo el tiempo que ha durado la realización de esta Tesis Doctoral. En especial, agradezco a mi hermana, por ser mi amiga, confidente y pilar fundamental desde que tengo conciencia.

Por último, a todos aquellos que durante estos años se han interesado por el desarrollo de esta Tesis, muchas gracias.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Resumen

RESUMEN

Los productos hortofrutícolas tienen una gran importancia económica y nutricional a nivel mundial, siendo básicos en una dieta equilibrada. En este tipo de productos es clave que su recolección se realice en el momento de madurez considerado como óptimo en función de su destino final. Además, se debe llevar a cabo un control fiable de calidad y seguridad alimentaria tanto en la recepción en la industria como durante su procesado. Por tanto, los productores, la industria, los responsables de realizar las inspecciones de control de calidad y seguridad y, finalmente, los consumidores, demandan tecnologías que proporcionen información exacta y útil sobre los parámetros que influyen directamente sobre la calidad y seguridad de una fruta u hortaliza y sobre su autentificación, siendo clave el que dichas tecnologías no estén limitadas por sus costes, su carácter destructivo o sus tiempos de análisis.

La Espectroscopía de Reflectancia en el Infrarrojo Cercano (en inglés, Near Infrared Reflectance Spectroscopy, NIRS) ha demostrado su capacidad para ser utilizada con éxito en el sector agroalimentario, con ventajas específicas frente a otras técnicas analíticas como son la alta velocidad de respuesta, el ser no destructiva, respetuosa con el medioambiente, multi-producto y multi-parámetro, así como proporcionar una señal digital única de cada producto que puede ser ligada a otras tecnologías de la información y de la comunicación, proporcionando sistemas de control automatizados de última generación.

Los sensores NIRS han sido utilizados principalmente en la industria hortofrutícola en aplicaciones "at-line" o "ex-post". Recientemente, existe un gran interés, a nivel científico-técnico, por la aplicación *in situ*, tanto en campo como en la línea industrial de transformación, de este tipo de sensores, lo que permitiría la implantación, en la cadena alimentaria, de sistemas de toma de decisiones en tiempo real, aumentando la eficiencia productiva y el control de la calidad y seguridad de los productos elaborados. La evolución y mejoras en

la instrumentación, orientada hacia su miniaturización, una mayor portabilidad y estabilidad en condiciones no controladas están permitiendo que se pueda abordar la posibilidad de realizar el control de materias primas, productos y procesos *in situ* que está demandando el sector hortofrutícola. No obstante, en este ámbito de enormes potencialidades quedan numerosos aspectos en los que profundizar, principalmente los relativos a la optimización de la medida, al procesado de los datos espectrales y a su conexión con sistemas de apoyo a la decisión, que posibiliten que este tipo de aplicaciones sean una realidad.

El objetivo principal de esta Tesis Doctoral ha sido desarrollar modelos NIRS precisos y robustos para la predicción de parámetros de calidad y seguridad en frutas y hortalizas durante el seguimiento de maduración y en la recepción en la industria, así como en las líneas de clasificación. Con este propósito se han evaluado dos espectrofotómetros comercialmente disponibles, uno muy adecuado para efectuar mediciones *in situ*, directamente sobre producto en la mata (espectrofotómetro basado en la tecnología de filtros lineales variables, en inglés conocida por sus siglas LVF) y el otro idóneo para su utilización en las líneas industriales de clasificación (espectrofotómetro basado en espectroscopía NIR por Transformada de Fourier, en inglés, FT-NIR).

Asimismo, el presente Trabajo de Investigación analizó la viabilidad de utilizar un espectrofotómetro manual, portátil, basado en tecnología MEMS (en inglés, microelectrical mechanical system), para la autentificación de hortalizas en función de su origen. Igualmente, un modelo más actual del instrumento MEMS empleado en la determinación anterior, fue utilizado para el desarrollo y evaluación de modelos predictivos NIRS para la optimización del manejo de factores precosecha (irrigación), destinados a favorecer la toma de decisiones en tiempo real en campo.

También se ha llevado a cabo el desarrollo de estrategias de calibración avanzadas para la predicción de parámetros de calidad físico-química en frutas del género *Citrus*, durante su proceso de maduración en árbol, empleando para ello un instrumento NIRS manual, portátil basado en tecnología MEMS.

Finalmente, se ha iniciado la puesta a punto de una metodología de análisis destinada a la determinación del rendimiento de cosecha en naranjas verdes en árbol, empleando para ello el análisis de imágenes hiperespectrales.

Los resultados obtenidos en los distintos trabajos de investigación que forman parte de esta Tesis Doctoral han puesto de manifiesto el potencial de la tecnología NIRS para su incorporación *in situ* en el sector hortofrutícola, como sensor que proporcionará una huella espectral única de cada producto, de utilidad para la trazabilidad de los mismos, y asimismo, como un registro óptico de enorme interés para controlar que el producto cumple unos estándares de calidad y de seguridad determinados, de acuerdo a las distintas normativas que regulan su uso industrial.

Asimismo, los resultados del análisis de imágenes hiperespectrales indican que es posible utilizar un número reducido de longitudes de onda del rango del infrarrojo cercano del espectro para la estimación del rendimiento de cosecha en naranja, lo cual permitirá el desarrollo futuro de equipos de bajo coste y peso reducido para la detección de frutos verdes sanos, posibilitando su integración en vehículos aéreos no tripulados para la obtención de imágenes con una alta resolución espectral y espacial. No obstante, el trabajo desarrollado es sólo el inicio de una línea de investigación compleja, de gran actualidad y valor para las necesidades del sector agroalimentario, en general, y del citrícola, en particular.

Summary

SUMMARY

Horticultural products, which play a significant role in the human diet due to its high content of vitamins, fibre, minerals and trace elements and its extremely low caloric value, are among the world's major agricultural sectors. In this kind of products, it is essential that harvesting takes places in an optimum maturity stage according to its destination. Additionally, a feasible quality and safety control in the handling and processing industries must be implemented. Therefore, in response to growing demand from producers, consumers and the industry, recent years have seen the development of rapid, accurate, economical and above all non-destructive technologies for determining food-produce quality and safety.

Near Infrared Reflectance Spectroscopy (NIRS) is one of those flexible, fast, accurate, non-destructive and versatile technologies, which has been successfully applied in the agri-food sector. Thanks to considerable research into the spectra of a growing number of species and varieties of fruits and vegetables, together with the development of efficient algorithms for the multivariate treatment of spectral data, NIRS technology can now be used both in the field and on industrial production lines to determine a range of pre- and post-harvest indices, thus enabling optimal harvesting time and post-harvest shelf-life to be established. Likewise, NIRS technology together with other information and communication technologies is able to provide an automated control system of last generation.

NIRS sensors have usually been used in the horticultural sector at-line and in ex-post applications. However, a new generation of portable, compact and extremely light-weight NIRS instruments has recently been developed, ideally suited for use in the field and for taking *in situ* measurements. These instruments enable not only to take spectra at any time, but also to analyse the

whole surface of the product, thus obtaining more information about the quality and safety of the product to be harvested.

Lately, there has also been a growing interest from the horticultural processing industry to incorporate NIR spectroscopy as a routine technique to ensure the quality and safety of the horticultural products they process. Nevertheless, this requires the NIRS applications to be developed in advance, simulating the industrial processes of the horticultural industries. One fundamental issue in developing online NIRS applications is to make the correct choice of instrument to be used, which must be robust and stable when subjected to vibrations and thermal variations.

Therefore, the main objective of this PhD dissertation was to develop accurate and robust NIRS models for measuring major quality and safety parameters in fruits and vegetables during their ripening in the field, at harvest, and in the industry, with a view to enabling growers and producers to routinely use NIRS technology under field conditions and at the sorting lines in the processing industry. Two new generation of NIR spectrophotometers, one manual, portable instrument based on LVF technology, suitable for use in the field, to analyse the product while it is developing in the plant, and another based on FT-NIR technology which can be incorporated into the processing industry of these products, such as in the product sorting lines, and to control the quality and safety standards established by the industry, were evaluated for these purposes.

Furthermore, in this PhD dissertation the feasibility of NIR spectroscopy in providing nondestructive, *in situ* authentication for the growing system (outdoors or in a greenhouse) of bell peppers, using a handheld microelectromechanical system (MEMS)-based NIR digital transform spectrophotometer was evaluated. Likewise, a new portable NIR instrument based on MEMS technology was evaluated for helping growers to make

irrigation decisions to mitigate negative effects of stress on crop performance under future weather conditions associated to climate change.

The non-destructive on-tree measurement of the physico-chemical quality attributes of fruits belonging to the Citrus genus, using rapid spectral sensors and advanced chemometric methods was also evaluated. With this objective, the viability of using a handheld portable MEMS-NIRS instrument was also tested.

Finally, a methodology to distinguish between green oranges and leaves using hyperspectral imaging to estimate harvest yield in oranges, was set up.

The results obtained in the research papers that form part of this PhD dissertation showed that NIR spectroscopy was sensitive to physical-chemical changes of horticultural products, taking place during ripening, at harvest and during the postharvest storage. The results also confirmed the advantages of using portable NIRS sensors which allow to analyse the fruit/vegetable in the field, in order to harvest the horticultural products selectively, at the optimum time, and to obtain a product of the highest quality and safety which is intended both for fresh consumption and for the processing industry.

Likewise, the results obtained with the hyperspectral imaging (HSI) systems tested, confirm that it is possible to use a lower number of wavelengths to estimate harvest yield in oranges, which could pave the way for the future development of low-cost and low-weight equipment for the detection of green and sound fruit which could be mounted on drones.

In summary, the results showed that quality and safety in horticultural products can be measured non-destructively, with a single spectrum measurement and in a matter of seconds, during on-vine or on-tree ripening, at

harvest, and in the industrial process, paving the way for using NIRS and HSI technologies to assist growers and producers in making decisions in the horticultural sector.

Chapter 1

Chapter 1. INTRODUCTION

In horticultural products, quality is the sum of the characteristics, attributes and properties that give the product its food value (Bruhn, 2002).

The relative importance of each quality component depends on the product itself and on the use for which it is intended (i.e. fresh vs. processed) (Kader, 2002). Additionally, it must be considered that various quality components are used to evaluate horticultural products in terms of their compliance with specifications for various commercial categories, their value for genetic breeding programmes, and their response to a range of environmental factors and post-harvest treatments.

It should also be noted that consumer demand currently puts a high value on products that are local, seasonal or produced under traditional practices, being horticultural products grown outdoors favoured by these consumers. In general, consumers are interested in buying horticultural products obtained using these particular cultivation systems (for instance, those cultivated outdoors or under organic systems), attributing them higher quality standards (Meca and Cespedes, 2015). The variations in growing conditions between outdoors and greenhouse systems can make an important difference of the quality of the product, especially in terms of the organoleptic characteristics, linked to dry matter and sugar content.

For horticultural products grown outdoors and, in a greenhouse, the correct management and optimization of preharvest factors is of high importance for the quality production of fruits and vegetables. In the Mediterranean regions, characterised by hot and dry summers and mild winters, growers have to make accurate irrigation decisions depending on the water deficit of the tree, due to the fact that species could undergo frequent periods of water and heat stress, with concomitant effects on yields. The relative water

content (RWC) of a leaf is an important indicator of the plant's water status. In this sense, RWC provides a measurement of the 'water deficit' of the leaf and may indicate a degree of stress expressed under unfavourable conditions such as drought or high temperature.

Furthermore, growers are demanding methods for the on-tree monitoring of fruit quality parameters during the ripening process, with a view to identify the optimal harvest time. This is particularly critical for non-climacteric fruits such as citrus fruits since the physiological maturation process finishes at harvest (Watkins, 2008). Nowadays, in these fruits, the market price is based on physical attributes; it would be useful to introduce, in the near future, quality-based pricing systems, using both external (morphological and colour related parameters) and internal quality indices which are of particular interest to the citrus-fruit industry because they are linked to fruit yield (Obenland et al., 2009; Magwaza et al., 2012).

Likewise, over recent years, consumers have become increasingly aware of the risks involved in excessive consumption of nitrates and nitrites in water and foods. Vegetables, and specially the green leafy ones, are a major source of nitrates in the human diet (Elia et al., 1998). In response to growing public concern, the European Union passed Commission Regulation (EC) No 1258/2011 of 2 December 2011 setting maximum levels for nitrates in vegetables as a function of the final destination of the harvested product (baby foods for infants and young children, for preserved, deep-frozen or frozen vegetables and for fresh vegetables). Thus, the maximum level for nitrates in processed cereal-based foods and baby foods for infants and young children was set at 200 mg NO_3/kg while for leafy vegetables, such spinach this level was set for preserved, deep-frozen or frozen spinach at 2000 mg NO_3/kg and for fresh spinach at 3500 mg NO_3/kg (OJEU, 2011). These regulations highlight the need for nitrate content determination in horticultural products at harvest in order to establish their industrial use.

All this has prompted greater attention to fruit and vegetable safety and quality concerns. As a result, producers and industry are increasingly anxious to provide consumers with assurances regarding the safety, quality and provenance of their products. Therefore, there is a clearly need for non-destructive sensors that can be used in the field to measure nitrate content as well as other internal quality parameters (firmness, dry matter, soluble solid content, pH and titratable acidity); on the basis of the values obtained, decisions can be taken regarding optimum harvesting times and possible industrial uses.

In this sense, NIRS sensors, which combine fast spectrum acquisition, accurate measurement, versatility, simplicity of sample presentation and low cost, providing a unique digital signal of each product analysed, have shown great potential for the non-invasive monitoring of quality and safety in real-time and for ensuring traceability in horticultural products (Nicolaï et al., 2007; Sánchez and Pérez-Marín, 2011; Zhang et al., 2017).

Additionally, the incorporation of NIRS technology along the food supply chain, from farm to fork, has been favoured by the development of a new generation of portable, compact and extremely light-weight NIRS instruments, ideally suited for use in the field and for taking measurements and decisions *in situ* (Yan and Siesler, 2018). These instruments enable not only to take spectra at any time, but also to analyse the whole surface of the product, thus obtaining a more representative analysis, and therefore, more information about the quality and safety of the product to be harvested.

Lately, there has also been a growing interest from the horticultural processing industry to incorporate NIR spectroscopy to ensure the quality and safety of the fruits and vegetables they process, and this requires the NIRS applications to be developed in advance, simulating the industrial processes of the horticultural industries (Pasquini, 2018).

In the same way, it should be noted that, once the best instrument for a particular application has been chosen, the routine implementation of NIRS technology, both in the field and in the industry, requires to optimize the NIRS analysis methodology, including all the issues related to taking spectra and the selection of the optimal spectral region.

Traditional NIRS sensors are now known as 'single-point Near Infrared Spectroscopy or single channel' to distinguish them from 'multichannel spectral imaging' or 'Hyperspectral Imaging sensors'.

Hyperspectral imaging (HSI) and multispectral imaging (MSI) are two emerging techniques in the agricultural sector due to their ability to acquire both spectral and spatial information (Dale et al., 2013). HSI and MSI combine the advantages of spectroscopy and machine vision in addressing food quality, authenticity and safety problems, increasing the inspection potential, since they provide a complete spectrum at each pixel location in the product analysed. Here, signal preprocessing and data management are even more complex than with NIRS alone, since very large amounts of data are obtained, and the linkage of different optical sensors is challenging. In the specific case of HSI, although there is abundant scientific literature on spectral imaging (most of it coming from the remote sensing field) using the short wavelength infrared region up to 1100 nm, there is a lack of knowledge and experience of using the extended NIR region up to 1700 nm or even to 2500 nm to determine quality and safety parameters in food products.

In the case of citrus fruits, the estimation of green fruit yield is a key parameter for growers and the industry. The early estimation of orange yield based on the distinction between green oranges and leaves and the detection of external defects in the oranges could influence the future market price and allow producers to plan the harvest in advance, thus reducing costs. Therefore, tools are needed to identify these green fruits, to make their harvesting easier

and to optimize the process, so that fruits of the highest quality are picked in keeping with their subsequent industrial use. Drones with hyperspectral or multispectral sensors attached may be used in the near future to achieve this aim due to the rise in their popularity for agricultural applications. Nevertheless, before this final application is ready to be used is needed to develop methodologies and to deepen in the study of the spectral signal, the best processing options, the optimization of the analysis with different devices and accessories to increase the signal quality, in order to can afford the general or final goal related to the remote measurement in field.

In summary, and taking into account the above mentioned, it is necessary the use of NIR and NIR-imaging sensors for the fast and non-destructive quality and safety assessment of horticultural products along the food supply chain. This quality and safety control of raw materials, processes and products using NIR spectral sensors could pave the way for an increasing in the efficiency of the processes and their traceability.

The Departments of Animal Production and Bromatology and Food Technology of the College of Agricultural and Forestry Engineering at Cordoba University (Spain), after a long experience in the traditional analysis of horticultural products, developed a research line in the frame of several Research Projects: Excellence Program, Project P09-AGR-5129 ‘MEMS and NIRS-image sensors for the *in situ* non-destructive analysis of food and feed’; Bilateral (Spain-Republic of Korea) EUREKA R & D project ‘Development of ICT fusion smart farm technology for the intelligent production and distribution of oranges, INTELLIGENT-CITRUS’; and also with the assistance of three research projects focused on the characterization of vegetables (summer squash, spinach and pepper) grown in open-air fields in Andalusia founded by Gelagri Ibérica, S.L., in which this PhD dissertation is framed, whose objective is the establishment of a decision-making support system for the horticultural sector based on NIR spectra fingerprint and HSI system for the authentication and

quality and safety control of fruits and vegetables along the supply chain. In this content, the objectives of this PhD dissertation are as follows (Chapter 2 of this research work).

Due to the fact that the main objective of this PhD dissertation is the generation of new scientific knowledge for the horticultural sector and with the object of facilitating its dissemination, the results here obtained are shown as a compendium of research articles published in scientific journals.

In order to facilitate the reading, this PhD dissertation has been divided in different chapters:

- In Chapter 1, a general introduction to the different research papers of this PhD dissertation is written.
- In Chapter 2, the objectives of this PhD dissertation are clarified and exposed.
- In Chapter 3, instrumental comparison, NIRS analysis optimization and *in situ* prediction of quality and safety parameters in vegetables (tomato, summer squash and spinach), are carried out.
- In Chapter 4, the evaluation of the viability of NIR spectroscopy in providing *in situ* authentication for the growing system of horticultural products (bell pepper) is studied.
- In Chapter 5, the development and evaluation of NIRS predictive models for irrigation decision support in olive grove, enabling optimal and precise decision-making at field level, are carried out.
- In Chapter 6, the development and evaluation of advanced calibration strategies for the prediction of quality parameters in fruits of the *Citrus* genus analysed on-tree, are exposed.
- In Chapter 7, a methodology to distinguish between green oranges and leaves and to detect defects in oranges using visible and near-infrared hyperspectral imaging is shown.

- In Chapter 8, the conclusions of this PhD dissertation are written.
- In Chapter 9, final considerations and recommendations for future R&D&I works, are exposed.
- In Chapter 10, references used are included.

Chapter 2

Chapter 2. OBJECTIVES

2.1. General objective

The general objective of this PhD dissertation is to analyse the feasibility of using spectral sensors in the near infrared range (NIRS and NIRS-imaging) as a non-invasive, environment-friendly, flexible and versatile (applicable to multiproduct and multiparameter analysis) and accurate tool, capable of being implemented in the field and in the industrial processing lines for the characterization, authentication and quality and safety assurance of horticultural products.

2.2. Specific objectives

The specific objectives of this PhD dissertation are:

1. Development and evaluation of global NIRS calibration equations for the prediction of quality and safety of horticultural products during their ripening process in the field and at the reception in the processing industry. [*This objective was reached in the following research articles*: *'Fast and accurate quality assessment of Raf tomatoes using NIRS technology'. Postharvest Biology and Technology 107, 9-15 (2015); 'Use of NIRS technology for on-vine measurement of nitrate content and other internal quality parameters in intact summer squash for baby food production'. Postharvest Biology and Technology 125, 122-128 (2017); 'Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors'. Postharvest Biology and Technology 154, 21-30 (2019); 'Pre-harvest screening on-vine of spinach quality and safety using NIRS technology'. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 207, 242-250 (2019)*].

2. Instrumental comparison and *in situ* NIRS analysis optimization – in the field and in the handling and sorting lines – of horticultural products. [*This objective was satisfied in the research article*: *'Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors'. Postharvest Biology and Technology 154, 21-30 (2019)*].

3. Evaluation of the feasibility of using NIR spectroscopy for the *in situ* authentication of vegetables as a function of the growing system (outdoor or greenhouse). [*This objective was met in the research article: 'Rapid, simultaneous, and in situ authentication and quality assessment of intact bell peppers using NIRS technology'. Journal of the Science of Food and Agriculture 99, 1613-1622 (2019)*].

4. Development and evaluation of NIRS predictive models for irrigation decision support, enabling optimal and precise decision-making at field level. [*This objective was carried out in the article: 'Irrigation decision support based on leaf relative water content determination in olive grove using near infrared spectroscopy. Biosystems Engineering 180, 50-58 (2019)*].

5. Setting up a methodology for NIRS analysis of spinach leaves and summer squashes during the ripening process, regarding optimum harvesting times and possible industrial uses. [*This objective was reached in the research articles*: *'Pre-harvest screening on-vine of spinach quality and safety using NIRS technology'. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 207, 242-250 (2019); 'Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors'. Postharvest Biology and Technology 154, 21-30 (2019)*].

6. Development of advanced calibration strategies for the prediction of quality parameters in fruits of the *Citrus* genus analysed on-tree. [*This objective was satisfied in the articles: 'Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors'. Biosystems Engineering 153, 140-148 (2017); 'LOCAL regression applied to a citrus multispecies library to assess chemical quality parameters using near infrared spectroscopy'. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 217, 206-214 (2019)*].

7. Setting up a methodology for the on-tree estimation of green citrus fruit yield using hyperspectral imaging. [*This objective was reached in the research article: Setting up a methodology to distinguish between green oranges and leaves using hyperspectral imaging', submitted to Computers and Electronics in Agriculture*].

Chapter 3

Chapter 3. INSTRUMENTAL COMPARISON, NIRS ANALYSIS OPTIMIZATION AND *IN SITU* PREDICTION OF QUALITY AND SAFETY PARAMETERS IN VEGETABLES

Chapter 3.1.

Fast and accurate quality assessment of Raf tomatoes using NIRS technology

Irina Torres[a], Dolores Pérez-Marín[b], María-José De la Haba[a], María-Teresa Sánchez[a,*]

[a] *Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*
[b] *Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*

Postharvest Biology and Technology 107, 9-15 (2015)

Abstract

Near infrared reflectance (NIR) spectroscopy was used as a fast and accurate technology for the simultaneous measurement of color, sugar and organic acid content in intact Raf tomatoes. The potential of this method coupled with chemometric techniques based on modified partial least squares regression was assessed by comparison with the currently-used traditional method for determining color, dry matter, soluble solid content, glucose, fructose, titratable acidity, malic acid and citric acid. At the same time, the performance of two spectrophotometers, differing primarily in terms of measurement principle and wavelength range, was evaluated. A total of 165 tomatoes (cv. "Raf") were used in the construction of calibration models for all the parameters previously cited, testing various spectral signal pretreatments. The technology was well suited to sorting Raf tomatoes on the basis of color parameters (a* and a*/b*(r^2 = 0.76-0.75; SEP = 2.58-0.09, respectively), soluble solid content (r^2 = 0.75; SEP = 0.65%) and titratable acidity (r^2 = 0.69; SEP = 0.06%), and useful, though less accurate (r^2 < 0.60), for the sorting of fruits by the rest of the color parameters tested (b*, L*), as well as by sugar content (glucose and fructose), dry matter and citric and malic content, particularly when the diode array instrument was used.

Keywords: near-infrared spectroscopy, tomato, external color, internal quality, MPLS regression.

3.1.1. Introduction

The tomato is the world's most widely consumed vegetable, and thus a key product on the global agricultural market (Scibisz et al., 2011). In many countries, tomato production is largely aimed at the fresh-produce market, and therefore requires the comprehensive monitoring of external and internal quality parameters both during on-stem growth and ripening and during subsequent industrial handling (Costa and Heuvelimk, 2005; Alvés De Oliveira et al., 2014).

The tomato is composed mainly of water, soluble and insoluble solids, organic acids (principally citric acid) and micronutrients such as carotenoids and vitamins A and C (Pedro and Ferreira, 2007). Sugars and organic acids are responsible for sweetness and tartness, and also influence tomato flavor; as a result, they are the major factors affecting consumer acceptability (Baldwin et al., 2008, Kader, 2008; Causse et al., 2010). Color also has a marked influence on the initial purchasing decision by consumers, who tend to link fruit color to taste quality (Causse et al., 2010). López-Camelo and Gómez (2004) have suggested that the a*/b* ratio could be used for practical purposes as an objective ripening index, giving a realistic view of consumer perceptions.

However, since the measurement of external and internal quality parameters using traditional analytical methods is highly time-consuming, destructive, costly and contaminant, there is a clear need for fast, accurate and non-destructive analytical techniques that can be used both in the field and by the industry, and that enable individual classification of tomatoes by quality. NIRS technology meets these requirements and also offers other advantages, making it ideal for

monitoring purposes and for ensuring traceability: low per-sample cost; little or no need for sample preparation; ability to analyze a wide range of products; and a high degree of reproducibility and repeatability (Slaughter and Abbott, 2004; Garrigues and De la Guardia, 2013).

NIR technology is currently in widespread use for measuring chemical components and quality attributes in vegetable products (Saranwong and Kawano, 2007, Sánchez and Pérez-Marín, 2011). The few studies dealing with intact tomatoes, however, focus largely on measuring total soluble solid content (Slaughter et al., 1996; Flores et al., 2009), titratable acidity (Flores et al., 2009), dry matter (Khuriyati et al., 2004), color (Clement et al., 2008), firmness (Shao et al., 2007) and growing state index (Yang et al., 2011). There are no reports in the literature regarding the use of NIRS spectroscopy for measuring glucose and fructose levels or citric and malic acid content in intact tomatoes, these being key factors for assessing ripeness and postharvest life, as well as exerting a crucial influence on the consumers' decision to purchase. This is particularly true of the Raf tomato which, though outwardly ugly due to its distinctive dark-green coloring and almost-black shoulder, boasts a salinity resistance guaranteeing an exquisite flavor rarely found in other varieties. Moreover, for Raf tomatoes, fruit color is regarded as synonymous with quality and taste: darker-colored—almost bluish—fruits are likely to have the best taste qualities; the green-black shoulder, while not an essential quality indicator, shows that the fruit has received sufficient sunlight and is therefore sweeter, and is also an ideal indicator for distinguishing Raf from similar tomatoes.

No hitherto-published research has addressed the comparison of NIRS instruments differing in terms of cost, optical design, and suitability for on-site use for quality determination in tomatoes.

This study sought to assess the feasibility of using NIRS spectroscopy to measure external and internal quality attributes (color, total soluble solid content, fructose and glucose levels, titratable acidity, citric and malic acid levels and dry matter content) in intact Raf tomatoes. Data analysis included a comparison between two NIRS instruments with very different optical designs, one of which is highly suited to laboratory measurement (monochromator spectrophotometer) and the other better suited to on-line use in the packing house (diode-array spectrophotometer).

3.1.2. Material and methods

3.1.2.1. Sampling

A total of 165 tomatoes (*Lycopersicum esculentum* Mill., cv. "Raf") were harvested at commercial maturity in greenhouses in Almería (Spain) (Figure 3.1.1). On arrival at the laboratory, fruits were promptly placed in refrigerated storage at 10 ºC and 95% relative humidity. Prior to each measurement, samples were left until the near-surface fruit temperature had risen to, and stabilized at, laboratory temperature.

3.1.2.2. Reference data

Skin or external color values (L*, a*, and b*) were individually measured at the equator, turning the fruit through 120° between measurements, using a Minolta Chroma Meter CR-400 (Minolta Corporation, Ramsay, NJ) (CIE, 2004). Illuminant D65 and the 2°

standard observer were used for all measurements. The three measurements obtained per fruit for each of the color parameters tested were averaged.

After these non-destructive measurements, fruits were halved and tissue from each fruit was taken at the same positions as those for the NIRS measurements. Dry matter content was determined by desiccation at 105 ºC for 24 h (AOAC, 2000) and results were calculated as a percentage of final dry weight of the initial wet weight. Soluble solid content (SSC, in %) was measured as the refractometer reading for tomato juice, using a temperature-compensated digital Abbé-type refractometer (model B, Zeiss, Oberkochen, Würt, Germany). Titratable acidity (TA) was measured by titration with 0.1 mol L^{-1} NaOH to an end point of pH 8.1. An automatic titrator was used (Crison Micro TT 2050, Crison, Alella, Barcelona, Spain). Results were expressed as % citric acid. Sugars (glucose, fructose) and organic acids (citric and malic acids) were quantified by an enzymatic method using food-analysis kits (Boehringer Mannheim Co., Mannhein, Germany) and expressed as g kg^{-1} of fresh weight for sugars and g kg^{-1} of fresh weight for acids. These measurements were performed with a BM-704 automatic analyzer (Hitachi, Tokyo, Japan).

Each sample was analyzed in duplicate. All measurements were performed immediately after VIS/NIRS spectrum collection.

3.1.2.3. NIR analysis

NIRS analysis was performed using two instruments that differ considerably in terms of both function and optical design: a Perten DA-7000, Flexi-Mode diode array spectrophotometer (Perten Instruments

North America, Inc., Springfield, IL, USA), more suitable for "on site" measurements, and a FNS-6500 scanning monochromator (FOSS NIRSystems, Silver Spring, MD, USA), traditionally used in a laboratory setting. These instruments operate in the 400 to 1700 nm range with a 5 nm scanning interval, and in the 400 to 2500 nm range with a 2 nm scanning interval, respectively (Table 3.1.1).

Using the diode-array instrument, tomatoes were placed centrally on the fruit holder, with the stem-calyx axis vertical, calyx up, and were irradiated from above by the light source while they rotated. Three separate spectral measurements were made on each intact tomato, after a 120º sample rotation each time. The three spectra were averaged to provide a mean spectrum for each intact fruit (Figure 3.1.2).

The FNS-6500 instrument was interfaced to a remote reflectance fiber optic probe (NR-6539-A) with a 43 x 43 mm window; a dark compartment (340 x 238 x 222 mm) was used to protect the detector assembly. Each fruit was hand-placed in the probe, so that the desired fruit location was centered on, and in direct contact with, the probe. The first measurement was made at a random location on the blossom of the fruit. The next two measurements were taken on the blossom end at rotations of roughly 120º and 240º from the initial site. The three spectra were averaged to provide a mean spectrum for each tomato (Figure 3.1.2).

3.1.2.4. Spectral repeatability

Before averaging the three spectra, the spectral repeatability of intact tomatoes was evaluated using the Root Mean Squared (RMS) statistic. The RMS statistic is defined as the averaged root mean square

of differences between the different subsamples scanned at *n* wavelengths (Shenk and Westerhaus, 1995a, 1996). This statistic indicates the similarity between different spectra of a single sample, in this case between the three spectra collected per sample. One hundred and sixty-five samples were analyzed for this purpose. Three spectra were collected from each sample in the FNS-6500 and the DA-7000, in three different positions.

For each instrument and sample presentation form, the RMS for an individual subsample (*j)* of the sample (*k*), and the MEAN and STD values for a given *k* sample were calculated according to the formulae provided by Martínez et al., (1998).

$$RMS_{j,k} = \sqrt{\frac{\sum_{i=1}^{n} D_{ij}^2}{n}};\ D_{ij} = y_{ij} - \bar{y}_i$$

$$MEAN_k = \sqrt{\sum_{j=1}^{N} (RMS_j)^2 \Big/ N}$$

$$STD_k = \sqrt{\sum_{j=1}^{N} (RMS_j)^2/(N-1)}$$

where y_{ij} is log (1/R) at wavelength *i* for subsample *j*, and $\bar{y}_i$ is log (1/R) at wavelength *i* for the average spectrum of *N* subsamples of a sample *k*; *n* is the number of data points collected by the instrument (here, 1050 data points for the FNS instrument and 228 data points for the Perten instrument). The RMS value obtained in each case was multiplied by 10^6 to facilitate value management and processing.

Rosales (1993) demonstrated that the estimated value of the error variance, σ^2 for log (1/ R) $y_{i,j}$ is the corresponding to one-way ANOVA:

$$\sigma^2 = \frac{1}{n\,(N-1)} \sum_{i=1}^{n} \sum_{j=1}^{N} \left(y_{i,j} - \bar{y}_i\right)^2$$

This expression corresponds to the STD^2. The sum of squares for error (SSE) can be thus be expressed as:

$$SSE = n\,(N\text{-}1)\,STD^2$$

which approximately follows a σ^2 distribution,

$$\frac{n\,(N-1)\,(STD^2)}{\sigma_0^2} \sim \chi^2_{[n(N-1)]}$$

with σ_0^2 the parametric value of the error variance. For infinite degrees of freedom (> 100), χ^2 tends to a normal distribution. An STD limit can then be calculated for comparing the RMS values of subsamples, following the formula given by Rosales (1993).

$$STD_{limit} = 1.036 \sqrt{\sum_{k=1}^{k=m} STD_k^2/m} = 1.036\sqrt{\overline{STD^2}}$$

where *STD* is the standard deviation per sample and *m* is the number of samples. This STD limit corresponds to the calculation done for the SCAN program of the software WINISI when the spectra of subsamples are being compared. The value 1.036 corresponds to a probability level of 85%. The usefulness of the STD RMS statistic is that when known, it can be used to calculate an STD limit for subsamples of the same sample and then to use that limit on the instrument set-up

program before to scan subsamples. If the subsample spectra taken have an STD value above the limit the spectra are averaged if not, the program asks for another subsample (Martínez et al., 1998). The STD_{limit} values were thus used to establish the RMS_{cutoff} for each instrument and sample presentation. Hence, the different sources of variation which might cause irregular spectra, were controlled, since any spectra in a sample and instrument that were above this limit were eliminated, and recalculations were performed until all the values were below the RMS_{cutoff}. Then, the mean spectrum of each sample was calculated.

3.1.2.5. Population structuring and detection of spectral outliers prior to calibration

Principal Component Analysis (PCA) was performed on a set of N = 165 samples in order to decompose and compress the data matrix. After PCA, the center of the spectral population was determined in order to detect outlier samples. The Mahalanobis distance (GH) was calculated between each sample and the center; samples with a GH value greater than 3 were considered outliers (Shenk and Westerhaus, 1995a). As spectral pretreatments, the Standard Normal Variate (SNV) plus Detrending (DT) procedure (Barnes et al., 1989) was used to remove the multiplicative interferences of scatter, and one derivative mathematical treatment (Norris derivative) was performed (1,5,5,1), where the first digit is the order of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing and the fourth is the second smoothing (Shenk and Westerhaus, 1995b; ISI, 2000).

3.1.2.6. Construction and validation of prediction models by MPLS regression

Once spectral outliers had been removed (i.e. 7 of the original 165 samples), a set consisting of 158 samples was used to develop calibration models. The set was divided into two: a training set containing about 75% of the samples (N = 121) and a test set containing the remaining 25% (N = 37). These samples were selected following the method outlined by Shenk and Westerhaus (1991) using the CENTER algorithm included in the WinISI software package to calculate the Global Mahalanobis distance (GH). Samples were ordered based on the Mahalanobis distance to the center of the population, and three of every four were selected to be part of the calibration set.

Modified Partial Least Squares (MPLS) regression (Shenk and Westerhaus, 1995a) was used to obtain equations for predicting color, sugars, acids and dry matter content. Partial least squares (PLS) regression is similar to principal component regression (PCR), but uses reference data (chemical, physical, etc.) and spectral information to identify factors useful for fitting (Williams, 2001). MPLS is often more stable and accurate than the standard PLS algorithm for agriculture applications (Shenk and Westerhaus, 1995a). Six cross-validation steps were included in the process in order to avoid overfitting (Shenk and Westerhaus 1995a).

Signal noise at the beginning and end of the spectral range was eliminated for both instruments: the resulting range for the DA-7000 spectrometer was from 515 to 1650 nm, while that of the FNS-6500 monochromator was from 516 to 2200 nm.

For each analytical parameter, different mathematical treatments were evaluated. For scatter correction, the Standard Normal Variate

(SNV) and Detrending (DT) methods were tested (Barnes et al., 1989). Additionally, four derivative mathematical treatments were tested in the development of NIRS calibrations: 1,5,5,1; 2,5,5,1; 1,10,5,1; 2,10,5,1 (Shenk and Westerhaus, 1995b).

The statistics used to select the best equations were: standard error of calibration (SEC), coefficient of determination of calibration (R^2), standard error of cross-validation (SECV), coefficient of determination for cross-validation (r^2), RPD or ratio of the standard deviation of the original data (SD) to SECV, and coefficient of variation. These latter two statistics enable SECV to be standardized, facilitating the comparison of the results obtained with sets of different means (Williams, 2001).

The best models obtained for the calibration set, as selected by statistical criteria, were subjected to evaluation using samples not involved in the calibration procedure. A test set composed of 37 samples, not used previously in the model, was evaluated following the protocol outlined by Windham et al., (1989).

3.1.3. Results and discussion

3.1.3.1. Spectral repeatability

Optimization of spectrum quality and repeatability is crucial to the construction of models which are both accurate and robust. Statistical methods such as defined RMS cut-off limit can be useful for this purpose. The RMS cut-off was calculated for the two instruments as shown in section 2.4.

For the Perten DA-7000, the mean STD for the samples analyzed was 55,732 μlog (1/R), representing an RMS cut-off of 79,296 μlog (1/R). For the FNS-6500 instrument, mean STD and the RMS cut-off were 70,436 μlog (1/R) and 82,508 μlog (1/R), respectively. Any sample whose triplicated screening scans yielded an RMS above this value was eliminated and repeated until values fell below that limit, thus ensuring a high degree of spectrum repeatability.

No reference to the calculated RMS cut-off value for intact tomatoes has been found in the literature, although this statistic is essential to the generation of representative libraries.

3.1.3.2. Descriptive data for NIR calibrations and validations sets

Values obtained for range, mean, SD and CV for each of the parameters measured (calibration and validation sets) are shown in Table 3.1.2. Structured selection based wholly on spectral information, using the CENTER algorithm, proved suitable, in that the calibration and validation sets displayed similar values for range, mean and SD for all study parameters; moreover, the ranges of the validation set lay within those of the calibration set.

All parameters except color measurements L* and b* displayed marked variability, with CV values of over 18% for both the calibration and validation sets. Williams (2001) and Pérez-Marín et al., (2005) have highlighted the importance both of sample set size and of sample distribution within the calibration set, noting that sample sets for calibration should ideally ensure uniform distribution of composition across the range of the study parameter in question.

3.1.3.3. Prediction of color quality parameters using MPLS regression and NIR spectra

The best equations for measuring color-related parameters (L*, a*, b* and a*/b*) for the two instruments tested, using the combination of signal pretreatments that yielded the best results in each case, are shown in Table 3.1.3.

Models obtained using the Perten DA-7000 instrument displayed greater predictive ability, for all color parameters, than those obtained using the monochromator. Models constructed for L* and b* using the diode-array instrument enabled samples to be classified into high, medium and low values, whilst models for a* and a*/b* displayed good predictive capacity within the limits established by Shenk and Westerhaus (1996). Models obtained with the FNS-6500 monochromator only enabled samples to be classified into high and low values (Shenk and Westerhaus, 1996).

Values for a*, like those of b* and a*/b*, increase significantly during ripening due to higher carotenoid levels, and thus also provide a useful indicator of fruit ripeness (Kader et al., 1978). It should also be stressed that the diode-array instrument enables color parameters to be measured on-site, which is particularly useful for the tomato handling industry.

No references have been found in the literature to color parameter prediction in intact Raf tomatoes. However, the RPD values recorded here were lower than those of between 2.81 and 7.22 reported by Clément et al. (2008) for color prediction (L*, a*, b* and a*/b*) in Canadian tomatoes at varying degrees of ripeness using a Varian Cary

500 UV-VIS-NIR scanning spectrophotometer equipped with an integration sphere, working in the spectral region 400-1000 nm. This highlights the difficulty of measuring color parameters in Raf tomatoes, in which both form and color distribution are highly irregular (Figure 3.1.1).

Validation statistics for the prediction of these parameters in intact tomatoes are also shown in Table 3.1.3. In terms of the validation protocol recommended by Windham et al., (1989) for the routine implementation of NIRS prediction models, the only models yielding sufficiently accurate predictions were those constructed for parameters a* and a*/b* using the DA-7000 spectrophotometer.

3.1.3.4. Prediction of internal quality parameters using MPLS regression and NIR spectra

Models obtained for all internal quality parameters using the Perten DA-7000 displayed greater predictive capacity than those constructed with the FNS-6500, with the exception of dry matter content (Table 3.1.4). For predicting dry matter, the model constructed using the monochromator and D_2 log (1/R) (r^2 = 0.59; SECV = 0.26%) enabled samples to be classified into high, medium and low values, whereas the model obtained with the diode-array instrument and the same second derivative only enabled classification into high and low values (r^2 = 0.45; SECV = 0.29%).

Walsh et al., (2004) reported slightly better predictive capacity (r^2 = 0.64; SECV = 0.20%) using a Carl Zeiss MMS1 NIR-enhanced spectrometer in the spectral region from 300 nm to 1100 nm, noting that the low standard deviation value for the sample set was probably the

main cause of poor model performance. Increasing the range for this parameter could improve the predictive capacity of the models. This would be useful for the tomato packing industry, since non-destructive measurement of tomato dry matter (DM) content is essential for fruit classification purposes, ensuring that fruit batches are of similar DM levels. It may also have implications both for consumer acceptability—fruits with higher dry matter content have a better flavor—and for improving storage potential and ripe fruit quality.

Models for total soluble solid content obtained with D_2 log (1/R) using the monochromator (r^2 = 0.77; SECV = 0.64%) and the diode-array instrument (r^2 = 0.79; SECV = 0.59%) displayed good predictive capacity in terms of Shenk and Westerhaus' recommendations (1996).

Although other studies of SSC prediction in tomatoes using NIRS technology (Slaughter et al., 1996; Hong and Tsou, 1998; Walsh et al., 2004; He et al., 2005; Shao et al., 2007) report models with r^2 values ranging from 0.49 to 0.97 and SEP values of between 0.22 and 0.38 ºBrix, the models constructed here for predicting SSC displayed adequate predictive capacity, bearing in mind the irregular shape of this tomato variety, which undoubtedly influences measurements.

Fructose and glucose are components of the main sugars and carbohydrates in tomatoes. For glucose, the model obtained using the DA-7000 (r^2 = 0.61; SECV = 3.8 g kg^{-1}) displayed greater accuracy and precision than its counterpart constructed using the FNS-6500 (r^2 = 0.50; SECV = 4.1 g kg^{-1}), enabling values to be classified into high, medium and low. For fructose, results obtained using both the diode-array instrument (r^2 = 0.43; SECV = 3.7 g kg^{-1}) and the monochromator (r^2 =

0.30; SECV = 3.7 g kg^{-1}) only enabled classification into high and low values. Guthrie et al., (2005) reported that in NIR spectra of hydrated samples with large molecules such as fruit and vegetables, the effective absorption bands are relatively wide and complex, even at the fundamental (IR) frequency.

No published studies address the direct measurement of these sugars in intact tomatoes. Pedro and Ferreira (2007) reported better predictive capacity both for glucose (r^2 = 0.98; RMSEP = 0.54%) and for fructose (r^2 = 0.94; RMSEP = 0.88%) although their results are not wholly comparable, since they used a set comprising samples of tomato concentrate products with total solid content ranging from 6.9 to 35.9%, and thus worked with a more varied calibration set.

Although measurement of acidity-related parameters in intact fruit is notoriously difficult (Flores et al., 2009), the models obtained for predicting titratable acidity using both instruments displayed good predictive capacity, with values of r^2 = 0.72 and 0.70 for the diode-array and monochromator, respectively, and SECV = 0.06% in both cases. Hong and Tsou (1998) recorded an r^2 value of 0.94, i.e. higher than that obtained here, for measurements of titratable acidity, although they used chopped rather than intact tomato; the residual error reported by these authors was similar to that recorded here (0.06%).

Models constructed for citric and malic acid content using the diode-array instrument yielded r^2 values of between 0.50 and 0.49, and SECV values in the range 0.81-0.22 g kg^{-1}, whilst with the monochromator r^2 value lay between 0.30 and 0.42 while SECV values ranged from 0.96 to 0.24 g kg^{-1}. These results suggest that NIRS

technology may be only used for screening purposes, to distinguish between low and high levels of both acids, and are according with Nicolai et al., (2007) who state that the concentration of acids in most fruit and vegetables is too small to affect the NIR spectrum significantly. The water absorption bands dominate the spectrum of fruit and vegetables, and it is not likely that minor constituents such as citric and malic acids can be measured well.

There are no published reports on the measurement of malic and citric acid in intact tomatoes using NIRS technology. However, these parameters are linked to the behavior of the tomato during ripening, and may thus act as indicators of ripeness and thus of optimal harvesting time. Malic acid levels decreases significantly during the later stages of ripening, while citric acid content generally increases (Baldwin et al., 1991); non-destructive measurement of citric and malic acid content is therefore of considerable value.

Validation statistics for the prediction of internal parameters in intact Raf tomatoes using both instruments are shown in Table 3.1.4. Slight differences in accuracy were noted between models constructed using the two instruments tested, although better results were obtained with the diode-array instrument for all parameters except dry matter content.

The models constructed for predicting SSC in intact tomatoes using both instruments tested, and for predicting TA using the diode array instrument, met the validation requirements in terms of r^2 ($r^2 > 0.6$) and both the SEP(c) and the bias were within confidence limits: the equations thus ensure accurate prediction, and can be applied routinely.

For dry matter and glucose content, it should be stressed that SEP(c) and bias lay within confidence limits for both instruments, although r^2 results did not always attain recommended minimum values, indicating that the NIRS equations constructed should be regarded as a first step in the fine-tuning of NIRS technology for the on-site monitoring of internal quality parameters in this tomato.

The models predicted fructose content, citric and malic acid content in validation-set samples with low values for r^2, in neither case meeting the recommendations of Windham et al., (1989). These models are thus not suitable for routine applications.

3.1.4. Conclusions

Near infrared reflectance spectroscopy combined with multivariate analysis is a very promising tool for determining the overall composition of intact Raf tomatoes, allowing ripeness to be monitored not only in terms of visual appearance but also in terms of taste, within one minute. The results of external validation indicate that parameters such as color (a* and a*/b*), SSC and TA can be routinely predicted using the diode array instrument, thus considerably reducing analysis time and enabling incorporation of these models into on-line NIR grading systems for measuring the ripeness of individual fruits in lines of harvested tomatoes. This could in turn lead to improved taste acceptability for this product. By contrast, the models constructed were unable to accurately predict citric and malic acid levels in tomatoes. It should be stressed that the results obtained here using a diode-array sensor should be regarded as the first step in the fine-tuning of NIRS for on- site quality monitoring of the Raf tomato, a complex vegetable with

an irregular form. Over the coming years, recalibrations may be required in order to enhance the precision and accuracy of the models obtained; the variability observed in this type of tomato could be reflected by including fruits harvested in different years and from different orchards, since these factors influence the chemical composition of tomatoes.

Acknowledgements

This research was funded by the Andalusian Regional Government under the Research Excellence Program (Project No. 3713 'Safety and Traceability in the Food Chain using NIRS'). The authors thank Prof. Dr. Ana Garrido-Varo her invaluable help in clarifying the RMS statistics and to Ms. Katherine Flores-Rojas and Ms. Mª Carmen Fernández their technical assistance.

References

Alvés De Oliveira, G.A., Bureau, S., Renard, C.M.-G.C., Pereira-Netto, A.B., de Castilhos, F., 2014. Comparison of NIRS approach for prediction of internal quality traits in three fruit species. Food Chem. 143, 223-230.

AOAC, 2000. Official Methods of Analysis of AOAC International. 17th ed. AOAC, Gaithersburg, MD, USA.

Baldwin, E.A., Goodner, K.L., Plotto, A., 2008. Interaction of volatiles, sugars, and acids on perception of tomato aroma and flavor descriptors. J. Food Sci. 73, S294-S307.

Baldwin, E.A., Nisperos-Carriedo, M.O., Moshonas, M.G., 1991. Quantitative analysis of flavor and other volatiles and for certain

constituents of two tomato cultivars during ripening. J. Am. Soc. Hortic. Sci. 116, 265–269.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard Normal Variate Transformation and De-trending of near infrared diffuse reflectance spectra. App. Spectrosc. 43, 772-777.

Causse, M., Friguet, C., Coiret, C., Moneta, E., Grandillo, S. 2010. Consumer preferences for fresh tomato at the European scale: a common segmentation on taste and firmness. J. Food Sci. 75, 531-541.

CIE, 2004. Colorimetry, 3rd ed. Commission Internationale De L'eclairage, Vienna, Austria.

Clément, A., Dorais, M., Vernon, M., 2008. Nondestructive measurement of fresh tomato lycopene content and other physicochemical characteristics using Visible-NIR Spectroscopy. J. Agr. Food Chem. 56, 9813–9818.

Costa, J.M., Heuvelimk, E., 2005. Introduction: the tomato crop and industry. In: Heuvelink, E. (Ed.), Tomatoes. CABI Publishing, Oxfordshire, UK. pp. 1–19.

Flores, K., Sánchez, M.T., Pérez-Marín, D., Guerrero, J.E., Garrido-Varo, A., 2009. Feasibility in NIRS instruments for predicting internal quality in intact tomato. J. Food Eng. 91, 311-318.

Garrigues, S., De la Guardia, M., 2013. Non-invasive analysis of solid samples. TrAC-Trend Anal. Chem. 43, 161–173.

Guthrie, J.A., Walsh, K.B., Reid, D. J., Liebenberg, C.J. 2005. Assessment of internal quality attributes of mandarin fruit. 1. NIR calibration model development. Aust. J. Agric. Res. 56, 405–416.

He, Y., Zhang, Y., Pereira, A. G., Gómez, A. H., Wang, J., 2005. Nondestructive determination of tomato fruit quality

characteristics using Vis/NIR spectroscopy technique. Int. J. Inform. Technol. 11, 97–108

Hong, T. L., Tsou, S.C.S., 1998. Determination of tomato quality by near infrared spectroscopy. J. Near Infrared Spectrosc. 6, A321-A324.

ISI, 2000. The Complete Software Solution Using a Single Screen for Routine Analysis, Robust Calibrations, and Networking. Manual, FOSS NIRSystems/TECATOR. Infrasoft International, LLC, Silver Spring, MD.

Kader, A.A. 2008. Flavor quality of fruits and vegetables. J. Sci. Food Agric. 88, 1863–1868.

Kader, A.A., Morris, L.L., Stevens, M.A., Albright-Holton, M., 1978. Composition and flavor quality of fresh market tomatoes as influenced by some postharvest handling. J. Am. Soc. Hortic. Sci. 103, 6-11.

Khuriyati, N., Matsuoka, T., Kawano, S., 2004. Precise near infrared spectral acquisition of intact tomatoes in interactance mode. J. Near Infrared Spectrosc. 12, 391–395.

López-Camelo, A. F., Gómez, P. A. 2004. Comparison of color indexes for tomato ripening. Hortic. Bras. 22, 534-537.

Martínez, M.L., Garrido, A., De Pedro, E.J., Sánchez, L. 1998. Effect of sample heterogeneity on NIR meat analysis: The use of the RMS statistic. J. Near Infrared Spectrosc. 6, 313–320.

Nicolaï, 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.

Pedro, A. M. K., Ferreira, M. M. C. 2007. Simultaneously Calibrating Solids, Sugars and Acidity of Tomato Products Using PLS2 and NIR Spectroscopy. Anal. Chim. Acta 595, 221–227.

Pérez-Marín, D.C., Garrido-Varo, A., Guerrero-Ginel, J.E., Gómez-Cabrera, A. 2005. Implementation of local algorithm with NIRS for compliance assurance in compound feedingstuffs. Appl. Spectrosc. Rev. 59, 69-77.

Rosales, M. 1993. Use of the NIR Spectral Information *per se* for Quality Evaluation of Agricultural Products. Master Thesis. ETSIAM. University of Cordoba (Spain).

Sánchez, M.T., Pérez-Marín, D., 2011. Nondestructive measurement of fruit quality by NIR spectroscopy. In: Vázquez, M., Ramírez, J.A. (Eds.), Advances in Post-Harvest Treatments and Fruit Quality and Safety. Nova Science Publishers Inc., Hauppauge, NY, USA, pp. 101-163

Saranwong, S., Kawano, S., 2007. Applications to agricultural and marine products: fruits and vegetables. In: Ozaki, Y., McClure, W.F., Christy, A.A. (Eds.), Near-Infrared Spectroscopy in Food Science and Technology. John Wiley & Sons, Inc., New Jersey, Hoboken, pp. 219–242.

Scibisz, I., Reich, M., Bureau, S., Gouble, B., Causse, M., Bertrand, D., Renard, C., 2011. Mid-infrared spectroscopy as a tool for rapid determination of internal quality parameters in tomato. Food Chem. 125, 1390-1397.

Shao, Y., He, Y., Gómez, A.H., Pereira, A.G., Qiu, Z., Zhang, Y., 2007. Visible/near infrared spectrometric technique for nondestructive assessment of tomato “Heatwave” (*Lycopersicum esculentum*) quality characteristics. J. Food Eng. 81, 672-678.

Shenk, J.S., Westerhaus, M.O., 1991. Population structuring of near infrared spectra and modified partial least squares regression. Crop Sci. 31, 1548-1555.

Shenk, J.S., Westerhaus, M.O., 1995a. Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy. Monograph, NIRSystems, Inc., 12101 Tech Road, Silver Spring, MD 20904, USA.

Shenk, J.S., Westerhaus, M.O., 1995b. Routine Operation, Calibration, Development and Network System Management Manual. NIRSystems, Inc., 12101 Tech Road, Silver Spring, MD 20904, USA.

Shenk, J.S., Westerhaus, M.O., 1996. Calibration the ISI way. In: Davies, A.M.C., Williams, P.C. (Eds.), Near Infrared Spectroscopy: The Future Waves. NIR Publications, Chichester, UK, pp.198-202.

Slaughter, D.C., Abbott, J.A., 2004. Analysis of fruits and vegetables. In: Roberts, C.A., Workman, J., Reeves III, J.B. (Eds.), Near-Infrared Spectroscopy in Agriculture. ASA, CSSA and SSSA, Madison, WI, pp. 377–398.

Slaughter, D.C., Barrett, D., Boersig, M., 1996. Nondestructive determination of soluble solids in tomatoes using near infrared spectroscopy. J. Food Sci. 61, 695-697.

Walsh, K.B., Golic, M., Greensill, C.V., 2004. Sorting of fruit using near infrared spectroscopy: application to a range of fruit and vegetables for soluble solids and dry matter content. J. Near Infrared Spectrosc. 12, 141–148.

Williams, P.C., 2001. Implementation of near-infrared technology. In: Williams, P.C., Norris, K.H. (Eds.), Near-Infrared Technology in

the Agricultural and Food Industries. AACC, Inc., St. Paul, Minnesota, pp. 145-169.

Windham, W.R., Mertens, D.R., Barton II, F.E., 1989. Protocol for NIRS calibration: sample selection and equation development and validation. In: Martens, G.C., Shenk, J.S., Barton II, F.E. (Eds.), Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality. Agriculture Handbook, vol. 643. Government Printing Office, Washington, DC, pp. 96–103.

Yang, H.; Kuang, B.; Mouazen, A.M., 2011. In situ determination of growing stages and harvest time of tomato (*Lycopersicon esculentum*) fruits using fiber-optic visible-near-infrared (Vis-NIR) spectroscopy. Appl. Spectrosc. 35, 931-938.

Table 3.1.1. Basic technical characteristics of the two spectrophotometers tested.

Properties	**Scanning monochromator (FNS-6500)**	**Diode Array Instrument (Perten DA-7000 VIS+NIR)**
Detector type	Silicon, 400–1100 nm; lead sulfide, 1100–2500 nm	Silicon and InGaAs Arrays each consist of 76-elements
Wavelenght range	400–2500 nm	Dual array: 400-1700 nm
Dispersion element	Holographic grating	One stationary reflective holographic grating
Spectral resolution (average)	10 ± 1 nm in reflectance	Silicon array 7.5 nm; InGaAs Array 10 nm
Output resolution	2 nm	5 nm
Spectral data rate	1.8 scans/s	600 scans per second
Reference	Internal ceramic reference tile	Internal: continuous dual-beam correction; External: Spectralon® for base line
Chopper modulation frecuency	16 reference/sample cycles per second	30 reference/sample cycles per second
Operating environment	Temperature: 18-28 ºC; Humidity 35-70% non condensing	Temperature 10-35 ºC; Humidity 35-80% non condensing
Dispersion	Pre	Post
Light source	Full spectrum. Tungsten-halogen bulb	Full spectrum. Tungsten-halogen bulb
Analysis Mode	Reflectance	Reflectance

Table 3.1.2. Statistical analysis of calibration and validation sets: data range, mean, standard deviation (SD), and coefficient of variation (CV).

Parameter	Set	Range	Mean	SD	CV (%)
L*	Calibration	28.92-58.18	47.73	3.30	6.93
	Validation	43.28-54.32	48.03	2.65	5.53
a*	Calibration	-18.14-10.42	-8.46	4.77	56.36
	Validation	-14.96-7.36	-8.62	5.18	60.06
b*	Calibration	20.96-47.15	28.88	3.50	12.15
	Validation	23.99-35.19	28.69	2.60	9.07
a*/b*	Calibration	-0.53-0.37	-0.29	0.16	57.05
	Validation	-0.53-0.26	-0.29	0.17	59.19
Dry matter (%)	Calibration	0.60-3.03	1.55	0.42	27.58
	Validation	0.69-2.67	1.64	0.47	28.93
SSC (%)	Calibration	2.50-9.00	5.29	1.32	25.11
	Validation	2.75-8.00	5.36	1.25	23.47
Glucose (g kg^{-1})	Calibration	8.6-43.9	21.7	6.3	29.14
	Validation	10.3-38.9	21.8	6.4	29.62
Fructose (g kg^{-1})	Calibration	9.1-43.5	18.8	5.3	28.59
	Validation	9.1-30.0	18.9	5.0	26.77
Titratable acidity (%)	Calibration	0.16-0.67	0.36	0.11	30.90
	Validation	0.20-0.58	0.35	0.09	27.09
Citric acid (g kg^{-1})	Calibration	1.87-8.96	4.50	1.23	27.35
	Validation	2.51-6.48	4.47	0.97	21.74
Malic acid (g kg^{-1})	Calibration	0.52-2.82	1.35	0.34	25.48
	Validation	0.63-1.80	1.33	0.25	18.51

Table 3.1.3. MPLS regression statistics for NIR-based models for predicting external quality parameters in Raf tomatoes.

Parameter	Instrument	Spectral range (nm)	Mathematic treatment	Calibration					Validation				
				N	SECV	r^2	RPD	CV	N	r^2	SEP	SEP (c)	Bias
L*	FNS-6500	516-2200	2,10,5,1	116	2.04	0.48	1.37	5.83	36	0.31	2.06	2.09	0.12
	DA-7000	515-1650	2,10,5,1	111	1.56	0.59	1.56	5.07	35	0.50	1.85	1.84	-0.32
a*	FNS-6500	516-2200	1,10,5,1	116	3.16	0.47	1.36	48.99	37	0.37	4.15	4.21	-0.09
	DA-7000	515-1650	2,5,5,1	113	2.19	0.74	1.97	49.59	35	0.76	2.58	2.60	0.32
b*	FNS-6500	516-2200	2,5,5,1	120	2.56	0.34	1.21	10.78	37	0.16	2.46	2.50	-0.07
	DA-7000	515-1650	2,5,5,1	111	1.81	0.67	1.71	10.71	36	0.23	2.53	2.55	-0.24
a*/b*	FNS-6500	516-2200	2,10,5,1	118	0.12	0.38	1.26	52.75	37	0.44	0.13	0.13	-0.01
	DA-7000	515-1650	2,10,5,1	110	0.07	0.80	2.23	51.46	34	0.75	0.09	0.09	-0.01

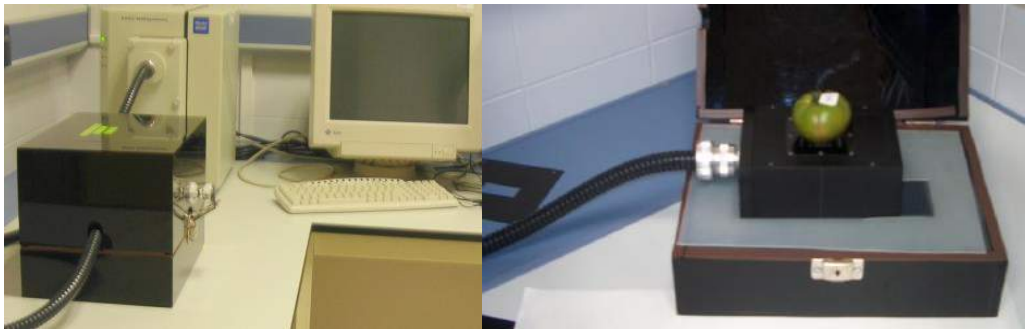
Table 3.1.4. MPLS regression statistics for NIR-based models for predicting internal quality parameters in Raf tomatoes.

Parameter	Instrument	Spectral range (nm)	Mathematic treatment	Calibration					Validation				
				N	SECV	r^2	RPD	CV	N	r^2	SEP	SEP (c)	Bias
Dry matter	FNS-6500	516-2200	2,10,5,1	116	0.26	0.59	1.54	26.39	34	0.49	0.32	0.30	0.12
(%)	DA-7000	515-1650	2,10,5,1	116	0.29	0.45	1.30	24.76	35	0.39	0.33	0.33	0.01
SSC (%)	FNS-6500	516-2200	2,5,5,1	119	0.64	0.77	2.08	25.00	36	0.60	0.83	0.84	0.01
	DA-7000	515-1650	2,5,5,1	113	0.59	0.79	2.13	24.23	36	0.75	0.65	0.65	-0.05
Glucose	FNS-6500	516-2200	1,10,5,1	113	4.10	0.50	1.40	26.71	35	0.52	4.40	4.50	0.20
(g kg^{-1})	DA-7000	515-1650	1,10,5,1	111	3.80	0.61	1.57	27.75	34	0.53	4.20	4.20	-0.70
Fructose	FNS-6500	516-2200	1,5,5,1	113	3.70	0.30	1.20	24.33	36	0.35	3.80	3.90	0.20
(g kg^{-1})	DA-7000	515-1650	2,10,5,1	116	3.70	0.43	1.29	25.57	37	0.36	4.00	4.10	-0.20
Titratable	FNS-6500	516-2200	1,5,5,1	118	0.06	0.70	1.83	30.15	37	0.56	0.07	0.07	-0.01
acidity (%)	DA-7000	515-1650	2,5,5,1	115	0.06	0.72	1.87	29.92	35	0.69	0.06	0.06	-0.01
Citric acid	FNS-6500	516-2200	2,10,5,1	119	0.96	0.30	1.18	25.57	35	0.31	0.86	0.87	0.10
(g kg^{-1})	DA-7000	515-1650	1,10,5,1	117	0.81	0.50	1.39	25.27	37	0.38	0.81	0.82	0.06
Malic acid	FNS-6500	516-2200	2,10,5,1	115	0.24	0.42	1.28	22.84	37	0.27	0.22	0.22	-0.02
(g kg^{-1})	DA-7000	515-1650	1,10,5,1	117	0.22	0.49	1.40	23.60	37	0.34	0.22	0.22	-0.03

Figure 3.1.1. Raf tomato.

Figure 3.1.2. Set-up of scanning monochromator (a) and the diode array NIR instrument (b).

a)



b)

Chapter 3.2.

Use of NIRS technology for on-vine measurement of nitrate content and other internal quality parameters in intact summer squash for baby food production

María-Teresa Sánchez[a], Dolores Pérez-Marín[b], **Irina Torres[a],** Belén Gil[a], Ana Garrido-Varo[b], María-José De la Haba[a]

[a] *Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*
[b] *Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*

Postharvest Biology and Technology 125, 122-128 (2017)

Abstract

This study sought to assess the feasibility of using NIR spectroscopy to predict the physico-chemical composition of summer squash during on-vine ripening, with a view to deciding on its possible use in baby food production depending on nitrate content at harvesting. NIR calibration models were developed using a set of 157 samples scanned *in situ* in the 1600–2400 nm region, using a portable handheld MEMS-NIR spectrophotometer working in reflectance mode. Modified partial least squares (MPLS) regression was used to interpret spectra and develop calibrations for summer squash composition. Results (R^2_{cv} = 0.83; SECV = 112.44 mg L^{-1}) showed that NIRS technology has great potential for measuring nitrate content and also other quality parameters in intact summer squashes during on-vine ripening. In addition, suitable wavelengths for nitrate content determination were identified by x-loading weights and regression coefficients. These findings suggest that NIRS may be a valuable tool for the rapid, accurate and non-destructive measurement of nitrate content, with a view to ascertaining the suitability of individual fruits for use in the production of baby foods.

Keywords: NIR spectroscopy, summer squash, on-vine, nitrate content, baby food.

.

3.2.1. Introduction

Over recent years, consumers have become increasingly aware of the risks involved in excessive consumption of nitrates and nitrites in water and foods. Vegetables are a major source of nitrates in the human diet, while nitrites are ingested mainly through canned foods. In response to growing public concern, the European Union passed Commission Regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs; the maximum level for nitrates in processed cereal-based foods and baby foods for infants and young children was set at 200 mg NO_3/kg (OJEU, 2006).

Summer squash is a common ingredient in processed vegetable-based baby foods. It is rich in polysaccharides, active proteins, essential amino acids, vitamins, carotenoids and minerals, and provides a moderate amount of dietary fiber; interest in this vegetable has increased considerably in the last few years due to its nutritional properties and health benefits (Reiss et al., 2012).

Nitrate levels at harvesting are a key issue, particularly if the summer squash is to be processed for the production of baby food. Toxicity occurs due to the conversion of nitrate to nitrite, which may lead to methemoglobin due to the oxidation of Fe^{+2} in hemoglobin. The impaired capacity of methemoglobin to deliver oxygen to tissues may lead to severe toxic effects, and may even prove fatal where methemoglobin accounts for over 70% of total hemoglobin. This occurs almost exclusively in infants and very young children, due to: lower stomach acidity (favoring the growth of bacteria able to convert nitrate to nitrite); the presence of fetal hemoglobin (which is more easily oxidized by nitrite); and lower levels of NADH-dependent methemoglobin reductase, an enzyme capable of reducing methemoglobin, which is very efficient in adults (Santamaria, 2006). In recent years, a number of studies have highlighted a possible link between nitrate exposure and childhood type 1 insulin-dependent diabetes mellitus (van Maanen et al., 2000).

All this has prompted greater attention to squash quality and safety concerns; as a result, producers are increasingly anxious to provide consumers with assurances regarding the quality and provenance of this product. Nitrate accumulation in squashes depends not only on type and genetic variety, but also on a number of other factors, including temperature, sunlight, available nitrogen and growing method (Blom-Zandstra, 1989).

There is a clearly need for non-destructive sensors that can be used in the field to measure squash nitrate content as well as other internal quality parameters (firmness, dry matter and soluble solids content, pH and titratable acidity); on the basis of the values obtained, decisions can be taken regarding optimum harvesting times and possible industrial uses.

Near-infrared spectroscopy (NIRS), in conjunction with the application of multivariate analysis strategies, is a valuable tool with great potential for the agrifood sector, ensuring rapid and reliable measurement of these parameters; over recent years, the field implementation of NIRS techniques has been helped by the development of compact, portable instruments, which may be hand-held or tractor-mounted, and can thus be readily used in the field.

There are no reports in the literature regarding the use of MEMS-NIRS instruments for the pre-harvest monitoring of summer squashes with a view to establishing the optimum time for harvesting depending on their potential destination in the industry, since research to date on the use of NIRS technology for summer squash quality control has focused only on the measurement of dry matter, hue angle h* and firmness using a NIR-AOTF spectrophotometer (Barnaba et al., 2012), and on the determination of antioxidant compound content (Blanco-Díaz et al., 2014) and mineral and carotenoid content (Martínez-Valdivieso et al., (2014a, b) using a monochromator instrument to analyze lyophilized, ground product.

Several authors have highlighted the viability of NIRS technology for the non-destructive measurement of nitrate content in various fruits and vegetables, including Japanese radishes (Ito et al., 2003), leaf stalk of Qing gin cai (Ito and Idezawa, 2006), spinach leaves (Xue and Yang, 2009), and pineapple (Srivichien et al., 2015).

This study sought to assess the feasibility of using NIR spectroscopy, with a low-cost, miniaturized, handheld, near-infrared device based on MEMS technology, for characterizing internal quality variations—particularly nitrate content—in intact summer squashes during on-vine ripening, with a view to optimizing harvesting times and enabling staggered harvesting by quality, thus allowing certain harvested squashes to be used in the production of baby foods.

3.2.2. Material and methods

3.2.2.1. Sampling

A total of 157 summer squashes (*Cucurbita pepo* subsp. *pepo* var. Mirza), grown on an open-air plantation in the district of La Montiela, Santaella (Córdoba, Spain), were harvested between May and July 2015.

3.2.2.2. Reference data

Nitrate content (mg NO_3 L^{-1}) was measured following Thompson et al., (2009), using an RQFlex reflectometer (Merck, Darmstadt, Germany). The reflectometer which measures the colour intensity of Reflectoquant ® test strips (Merck, Darmstadt, Germany) is based on a colorimetric method. For NO_3^- analysis, 50 g of sample from the equatorial part of fruit were cut into very small pieces and blended with 100, 150 and 200 ml of deionised water in a blender, depending on NO_3^- concentrations in the sample. After that, the solution was filtered using a coffee filter and let to settle for 5 minutes. Subsequently, a test strip was dipped in the supernatant for 2 s, and then the colour was allowed to develop for 1 min. The test strip was then inserted into the reflectometer and the amount of light reflected from the test strip was

measured and converted to concentration by a standard calibration previously introduced into the equipment through a bar coded plastic strip. The dilution factor was also taken into consideration.

Firmness was measured as the maximum force required to penetrate the summer squashes to a puncturing depth of 10 mm using a 3-mm cylindrical tip. Summer squashes were arranged with the stem-calyx axis horizontal; the first measurement was made at a point on the equator, and the second after turning the fruit through 180º. Texture measurements were made using a Universal Instron Texturometer (Model 3343, single-column, Instron Corporation, Norwood, MA, USA), with a head speed of 0.0008 m/s (50 mm/min) and a 1000 N load cell.

Dry matter content was determined by desiccation at 105ºC for 24 h (AOAC, 2000); final dry weight was calculated as a percentage of initial wet weight. Soluble solid content (SSC, in ºBrix) was measured as the refractometer reading for summer squash juice, using a temperature-compensated digital Abbé-type refractometer (model B, Zeiss, Oberkochen, Würt, Germany). Values for pH and titratable acidity (TA) were measured using an automatic titrator (Crison Micro TT 2050, Crison, Alella, Barcelona, Spain); TA was measured by titration with 0.1 mol L^{-1} NaOH to an end point of pH 8.1. Results were expressed as % citric acid. All the samples were taken from the equatorial zone of the fruits and analyzed in duplicate. All measurements were performed immediately after NIRS measurements.

3.2.2.3. Spectral data acquisition

NIR spectra of intact summer squashes were collected in reflectance mode (log 1/R) using a handheld micro-electromechanical system (MEMS) instrument (Phazir 2400, Polychromix, Inc., Wilmington, MA, USA).

The Phazir 2400 is an integrated near-infrared handheld analyzer that incorporates all the essential components to deliver on-vine applications. The

spectrophotometer scans at a non-constant interval of around 8 nm (pixel resolution 8 nm, optical resolution 12 nm), across the NIR wavelength range of 1600-2400 nm, with a scan time per sample of 3 s. Instrument performance was checked every 10 minutes, following the diagnostic protocols provided by the manufacturer, and white reference measurement was carried out using Spectralon as reference.

Four spectral measurements were made on each summer squash whilst on the vine, at four points located 90º from each other in the equatorial region of the fruit. The four spectra were averaged to provide a mean spectrum for each fruit.

3.2.2.4. Data analysis: definition of calibration and validation sets

Prior to carrying out NIRS calibrations, the CENTER algorithm included in the WinISI II software package ver. 1.50 (Infrasoft International LLC, Port Matilda, PA, USA) was applied to ensure a structured population selection based solely on spectral information, for the establishment of calibration and validation sets (Shenk and Westerhaus, 1991). This algorithm performs an initial principal component analysis (PCA) to calculate the center of the population and the distance of samples (spectra) from that center in an n-dimensional space, using the Mahalanobis distance (GH); samples with a statistical value greater than 3 were considered outliers or anomalous spectra.

The CENTER algorithm was applied in the spectral region 1600-2400 nm. Mathematical treatments SNV (Standard Normal Variate) and DT (De-trending) were applied for scatter correction (Barnes et al., 1989), together with the mathematical derivation treatment '1,5,5,1', where the first digit is the number of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing, and the fourth is the second smoothing (Shenk and Westerhaus, 1995b; ISI, 2000).

Once spectral outliers had been removed (i.e., 7 of the original 157 samples), a training set (C1) consisting of 150 samples was initially used to develop calibration models for all parameters tested (Table 3.2.1). After analyzing the accuracy and precision of the models obtained, new models were developed for those parameters for which the best models displayed a predictive capacity sufficient at least to distinguish high, medium and low values for that parameter; later, these were externally validated. For this purpose, and having ordered the sample set by spectral distances (from smallest to greatest distance to the center), the samples forming the validation set were selected by taking one sample out of every four in the initial set, although other alternatives for the selection of this set could have been used. After this procedure, the calibration (C2) and validation (V) sets thus comprised the samples shown in Table 3.2.1.

Data were subjected to chemometric treatment using the WinISI software package ver. 1.50 (ISI, 2000).

3.2.2.5. Data pre-processing and calibration model construction using a linear regression strategy

NIR calibration models for the prediction of quality parameters (nitrate content, firmness, dry matter, SSC, pH and TA) in intact summer squashes were initially constructed using the training set C1 (comprising all available samples) using modified partial least squares (MPLS) regression (Shenk and Westerhaus, 1995a), with subsequent cross-validation. The calibration set was partitioned into 6 groups; each group was then validated using a calibration developed on the other samples; finally, validation errors were combined to obtain a standard error of cross-validation (SECV).

A number of different pre-processing combinations were evaluated for scatter correction, including SNV and DT. Additionally, a total of four derivative mathematical treatments were tested: 1,5,5,1; 2,5,5,1; 1,10,5,1 and 2,10,5,1.

The statistics used to select the best equations were: the coefficient of determination for calibration (R^2_c), the standard error of calibration (SEC), the coefficient of determination for cross calibration (R^2_{cv}), the standard error of cross validation (SECV) and the coefficient of variation (CV), defined as the ratio between SECV and the mean value of the reference data in the calibration set. Furthermore, the Residual Predictive Deviation (RPD) was calculated as the ratio of the standard deviation (SD) of the reference data to the SECV. This statistic, together with the CV, enables SECV to be standardized, facilitating the comparison of results obtained with sets of different means (Williams, 2001).

Having analyzed the predictive capacity of the models obtained using the complete sample set (training set C1), new models were developed to predict the selected parameters using the second calibration set (C2). The best-fitting equations obtained for this new calibration set, as selected by statistical criteria, were subsequently subjected to external validation following the protocols outlined by Windham et al., (1989).

3.2.3. Results and Discussion

3.2.3.1. Spectral properties

Mean and standard deviation spectrum for summer squash at harvesting are shown in Figure 3.2.1.

In the wavelength region 1600-2400 nm, the major absorption peak at around 1920-1930 nm was mainly related to water absorption, as was the peak at around 2200 nm; this was to be expected, since summer squash is around 90% water (Osborne et al., 1993; Williams, 2001). Osborne et al. (1993) reported peaks at around 1780 nm and 2310 nm related to the first sugar-related overtone. Peaks at around 1680 nm may be linked to combination bands of proteins (Williams, 2001).

3.2.3.2. Population characterization.

Calibration (C1 and C2) and validation (V) set details, i.e. number of samples, mean, range, SD, and CV for the parameters analyzed, are shown in Table 3.2.1.

It should be stressed that structured selection using only spectral information treatment algorithms such as CENTER proved adequate and useful, since the calibration and validation sets (C2 and V, respectively) displayed similar values for mean, range and standard deviation for all study parameters, and ranges for the validation set lay within the range recorded for the calibration set C2.

As Table 3.2.1 shows, the internal quality parameter displaying the greatest variability was nitrate content, with a CV of 70.37% (training set C1) and 69.42% (training set C2); CV for the validation set was 73.73%. Nitrogen fertilization was stopped halfway through harvesting, prompting a marked drop in nitrate content thereafter. The lowest variability was recorded for pH ($CV_{calibration}$ = 2.69% and 2.85% for C1 and C2, respectively; $CV_{validation}$ = 2.54%) and titratable acidity ($CV_{calibration}$ = 11.11% and 11.25% for C1 and C2, respectively; $CV_{validation}$ = 8.89%), values for which displayed little variation over the harvesting period as a whole.

3.2.3.3. Prediction of nitrate content and other internal quality parameters in summer squash on vine

Results for the best models developed using training set C1 and various mathematical pretreatments are shown in Table 3.2.2. Statistical criteria were used to select the best model for each study parameter.

This study sought to determine whether harvested summer squashes could be used in the production of baby foods, for which nitrate content represents a major constraint. The predictive capacity of the best model for nitrate content (R^2_{cv} = 0.83; SECV = 112.44 mg NO_3 L^{-1}) may be considered

acceptable in terms of the limits recommended by Shenk and Westerhaus (1996). This result is of particular interest, in that it suggests that using a low-cost, portable NIRS instrument—suitable for use in the field—the industry can rapidly classify fruits as fit or unfit for baby food production on the basis of nitrate content.

This appears to be the first published report on the use of NIRS technology to measure nitrate content in summer squashes, although a number of authors have tested this technology in other vegetables. Ito et al. (2003) found that NIRS in conjunction with multiple linear regression enabled satisfactory determination of nitrate content in Japanese radishes, although they noted that the RMSE could be improved. Ito and Idezawa (2006) used NIRS and MLR algorithm to determine nitrate content in the leaf stalk of Qing gin cai, reporting a good match between real and Vis-NIR-calculated values for nitrate ion content for the third leaf stalk from outside and for whole nitrate ion content (0.90 and 0.76, respectively). In a 2009 study of nitrate content in spinach, Xue and Yang tested the best PLS model and PCR model in the spectral range 350-2500 nm with an independent dataset, reporting good agreement between predicted and observed values, with a correlation coefficient of 0.94 for the PLS model and 0.95 for the PCR model; the RMSE of prediction was 128.2 mg kg^{-1} for the PLS model and 120.8 mg kg^{-1} for the PCR model. In the only published study in fruit, Srivichien et al. (2015) measured nitrate content in pineapples by Vis-NIR spectroscopy, using a monochromator instrument; the results (RPD = 2.86; CV = 13.84%) were better than those obtained here, perhaps because the spectral range used was 600-1200 nm.

The predictive capacity of the models obtained for predicting dry matter content (R^2_{cv} = 0.66; SECV = 0.38% fw), total soluble solid content (R^2_{cv} = 0.68; ETVC = 0.33 ºBrix) and pH (R^2_{cv} = 0.57; ETVC = 0.11)—parameters crucial for deciding the optimum time for harvesting and determining the shelf-

life of summer squashes—enabled high, medium and low values to be distinguished (Shenk and Westerhaus, 1996).

Only one published study has addressed the prediction of dry matter content in summer squashes using NIRS technology: Barnaba et al. (2012) used an AOTF-NIRS spectrophotometer for this purpose, in conjunction with PLS regression. The predictive capacity of the calibration models developed for this parameter (R^2_{cv} = 0.81; SECV = 0.46% fw) were similar to those obtained here.

No reports have been found in the literature regarding the determination of SSC in summer squashes using NIR spectroscopy. However, Sánchez et al., (2013) measured SSC in intact mandarins using a Phazir 2400 instrument in the 1600-2400 nm spectral region, reporting results (RPD = 1.49; CV = 6.06%) very similar to those obtained in the present study.

Portable MEMS-based NIRS instruments have not hitherto been used to determine pH in summer squashes, but Sánchez et al. (2013) report poorer results for pH prediction in intact mandarins (RPD = 1.11; CV = 2.59%), highlighting the fact that pH is difficult to predict if training sets are insufficiently varied.

The predictive capacity of the models developed to predict firmness (R^2_{cv} = 0.45; SECV = 1.46 N) and titratable acidity (R^2_{cv} = 0.44; SECV = 0.01 % citric acid) enabled high and low values to be distinguished for these parameters, thus meeting the criterion recommended by Shenk and Westerhaus (1996).

The results obtained for firmness underline the difficulty in correlating a destructive measurement made at a puncturing depth of 10 mm with a non-destructive measurement; as Peirs et al. (2002) have noted, NIRS light only penetrates to a useful depth of between 1 and 5 mm, depending on the wavelength, the instrument used and the maturity of the fruit tested.

The predictive capacity of the models developed to predict titratable acidity reflects the fact that all summer squashes were harvested at commercial maturity, and thus displayed very uniform acidity values. González-Caballero et al. (2010) report that this parameter cannot be predicted using NIRS technology and MPLS regression if the training sets have low standard-deviation values, as was the case here.

Portable MEMS-based NIRS instruments have not been used to date for the prediction of firmness in summer squashes, but Pérez-Marín et al. (2010) reported a predictive capacity for firmness in intact plums (RPD = 1.18; CV = 53.10%) slightly lower than that obtained here, using the same instrument. This highlights the difficulty of correlating measurements when using an instrument working in the 1600-2400 nm spectral region.

Although there are no reports in the literature regarding the *in situ* measurement of titratable acidity in outdoor-grown summer squashes, Sánchez et al. (2013), studying quality measurements in on-tree mandarins using the Phazir 2400 instrument, obtained models whose predictive capacity (RPD = 1.68; CV = 11.93%) was better than that obtained here, perhaps because the training sets contained greater variability, since measurements were made throughout ripening.

3.2.3.4. New calibration process and external validation

Once the predictive capacity of the models using the training set C1 and cross-validation had been analyzed, only those models (nitrate content, dry matter, SSC and pH) for which $R^2_{cv} > 0.5$ were subjected to external validation.

The aim was, in the first instance, to validate the best calibration models using a sample set not included in the calibration, but similar to the calibration set. Validations of the best calibration models obtained with training set C2 were performed using a set comprising 34 samples (Table 3.2.3).

Models constructed for predicting nitrate content in intact summer squash using the MEMS instrument met the validation requirements in terms of the coefficient of determination for prediction R^2_p ($R^2_p > 0.6$), while the standard error of prediction corrected for bias or SEP(c), bias and slope were within confidence limits: the equation thus ensures accurate prediction, and can be applied routinely (Windham et al., 1989). As Table 3.2.3 shows, two samples in the initial validation set (nitrate content of 1,030 and 1,065 mg L^{-1}) were eliminated since the samples were unrepresentative of the calibration set (Figure 3.2.2), thus hindering their correct prediction.

Using the monitoring procedure, the prediction statistic values obtained for dry matter fell short of the limit recommended for routine application ($R^2_p > 0.60$). However, it should be stressed that the SEP(c) and slope values were close to confidence limits and bias was below confidence limits, suggesting that the NIRS equation for this internal parameter can be regarded as a useful preliminary trial for obtaining accurate on-vine quality predictions for intact summer squashes. Likewise, the values for bias in the models constructed for predicting SSC and pH in intact squashes using a handheld MEMS spectrophotometer lay within confidence limits, although R^2_p, SEP(c) and slope results did not always attain recommended minimum values, indicating that the NIRS equations constructed may be considered as a first step in the fine-tuning of NIRS technology for the on-vine monitoring of internal quality parameters in summer squashes.

These results highlight the importance not only of ensuring a sufficient number of samples in the calibration set, but also of guaranteeing the adequate distribution and structure of the sample set.

3.2.3.5. Analysis of sensitive wavelengths for the prediction of nitrate content

The x-loading and regression coefficient plots for the best model obtained for predicting nitrate content in intact summer squash are shown in Figure 3.2.3. These plots show the areas across the spectral range in which

variance has influenced the computing of the model to a greater or lesser degree, and the direction (positive or negative).

For the prediction of nitrate content using the Phazir 2400, representation of the six latent variables (LV) used in constructing the calibration equation and the regression coefficients shows that the areas of the spectrum exerting greatest weight on model fitting were 1712, 1776, 1850, 1920, 1984, 2008, 2100, 2152 and 2264 nm. Their influence was either positive or negative, depending on the latent variable in question (Figure 3.2.3).

3.2.4. Conclusions

These results suggest that NIRS is a very promising and useful sensor for the non-destructive quantification of changes in nitrate content and other internal quality parameters in summer squashes in the course of on-vine ripening, enabling decisions to be taken regarding the optimum harvesting time. Harvested fruit can thus be swiftly streamed, allowing batches with different nitrate contents to be processed separately.

Findings also show that the use of portable MEMS-based NIRS instruments enables nitrate content to be measured, rapidly and *in situ*, during on-vine ripening, thus providing the industry with a means of establishing the final destination of the product, since if nitrate content exceeds the maximum levels stipulated under current legislation, the fruits cannot be used in the production of baby foods.

Acknowledgements

This research was under the Research Project ‘Quality determination of summer squash grown on an open-air plantation in Santaella (Córdoba)’, funded by Gelagri Ibérica, S.L. The authors thank to Ms. Mª Carmen Fernández her technical assistance.

References

AOAC, 2000. Official Methods of Analysis of AOAC International. 17th ed. AOAC, Gaithersburg, MD, USA.

Barnaba, F.E., Bellicontro, A., Mencarelli, F., 2012. NIR-AOTF spectroscopy: Innovative technique to predict dry matter and the harvest time of zucchini. Ind. Alim. 51, 30–35.

Barnes, R.J., Dhanoa, M.S. Lister, S.J., 1989. Standard Normal Variate transformation and De-trending of near infrared diffuse reflectance spectra. App. Spectrosc.43, 772–777.

Blanco-Díaz, M.T., Del Río-Celestino, M., Martínez-Valdivieso, D., Font, R., 2014. Use of visible and near-infrared spectroscopy for predicting antioxidant compounds in summer squash (*Cucurbita pepo* ssp *pepo*). Food Chem. 164, 301–308.

Blom-Zandstra, M., 1989. Nitrate accumulation in vegetables and its relationship to quality. Ann. Appl. Biol. 115, 553–561.

González-Caballero, V., Sánchez, M.T., López, M.I., Pérez-Marín, D., 2010. First steps towards the development of a non-destructive technique for the quality control of wine grapes during on-vine ripening and on arrival at the winery. J. Food Eng. 101, 158–165.

ISI, 2000. The Complete Software Solution Using a Single Screen for Routine Analysis, Robust Calibrations, and Networking. Manual, FOSS NIRSystems/TECATOR. Infrasoft International, LLC, Sylver Spring MD, USA.

Ito, H., Idezawa, F., 2006. Non-destructive determination of nitrate ion in leaf stalk of Qing gin cai using visible (VIS)-near infrared (NIR) spectroscopy. Acta Hortic. 712, 363–370.

Ito, H., Horie, H., Ippoushi, K., Azuma, K., 2003. Potential of visible-near infrared (VIS–NIR) spectroscopy for non-destructive estimation of nitrate content in Japanese radishes. Acta Hortic. 604, 549–552.

Martínez-Valdivieso, D., Blanco-Díaz, M.T., Font, R., Del Río-Celestino, M., 2014a. Determining the mineral composition in Cucurbita pepo fruit

using near infrared reflectance spectroscopy. J. Sci. Food Agric. 94, 3171–3180.

Martínez-Valdivieso, D., Font, R., Blanco-Díaz, M.T., Moreno-Rojas, J.M., Gómez, P., Alonso-Moraga, A., Del Río-Celestino, M., 2014b. Application of near-infrared reflectance spectroscopy for predicting carotenoid content in summer squash fruit. Comp. Electr. Agric. 108, 71–79.

Osborne, B.G., Fearn, T., Hindle, P. 1993. Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. Longman Scientific and Technical, London, UK.

Official Journal of the European Union (OJEU). 2006. Commission Regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs. OJ L 364/5, 20.12.2006.

Peirs, A., Scheerlinck, N., Berna Perez, A., Jancsók, P.,. Nicolaï, B.M., 2002. Uncertainty analysis and modelling of the starch index determination during apple fruit maturation. Postharvest Biol. Technol. 26, 199–207.

Pérez-Marín, D., Paz, P., Guerrero, J.M., Garrido-Varo, A., Sánchez, M.T., 2010. Miniature handheld NIR sensor for the on-site non-destructive assessment of post-harvest quality and refrigerated storage behaviour in plums. J. Food Eng. 99, 294–302.

Reiss, R., Johnston, J., Tucker, K., DeSesso, J.M., Keen, C.L. 2012. Estimation of cancer risks and benefits associated with a potential increased consumption of fruits and vegetables. Food Chem. Toxic. 50, 4421–4427.

Sánchez, M.T., De La Haba, M.J., Pérez-Marín, D., 2013. Internal and external quality assessment of mandarins on-tree and at harvest using a portable NIR spectrophotometer. Comp. Electr. Agric. 92, 66–74.

Santamaria, P., 2006. Nitrate in vegetables: toxicity, content, intake and EC regulation. J. Sci. Food Agric. 86, 10–17.

Shenk, J.S., Westerhaus, M.O., 1991. Population structuring of near infrared spectra and modified partial least squares regression. Crop Sci. 31, 1548–1555.

Shenk, J.S., Westerhaus, M.O., 1995a. Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy. Monograph, NIRSystem, Inc., 12101 Tech Road, Silver Spring, MD, USA.

Shenk, J.S., Westerhaus, M.O., 1995b. Routine Operation, Calibration, Development and Network System Management Manual. NIRSystems, Inc., 12101 Tech Road, Silver Spring, MD 20904, USA.

Shenk, J.S., Westerhaus M.O., 1996. Calibration the ISI way. In: Davies, A.M.C., Williams, P.C. (Eds.), Near Infrared Spectroscopy: The Future Waves. NIR Publications, Chichester, West Sussex, UK, pp. 198–202.

Srivichien, S., Terdwongworakul, A., Teerachaichayut, S., 2015. Quantitative prediction of nitrate level in intact pineapple using Vis-NIRS. J. Food Eng. 150, 29–34.

Thompson, R.B., Gallardo, M., Joya, M., Segovia, C., Martínez-Gaitán, C., Granados, M.R., 2009. Evaluation of rapid analysis systems for on-farm nitrate analysis in vegetable cropping. Span. J. Agric. Res. 7, 200–211.

Van Maanen, J.M.S., Albering, H.J., de Kok, T.M.C.M., van Breda, S.G.J., Curfs, D.M.J., Vermeer, I.T.M., Ambergen, A.W., Wolffenbuttel, B.H.R., Kleinjans, J.C.S., Reeser, H.M., 2000. Does the risk of childhood diabetes mellitus require revision of the guideline values for nitrate in drinking water? Environ. Health Persp. 108, 457–461.

Williams, P.C. 2001. Implementation of near-infrared technology. In: Williams, P.C., Norris, K.H. (Eds.), Near-infrared Technology in the Agricultural and Food Industries. AACC Inc., St. Paul, MN, pp. 145–169.

Windham, W.R., Mertens, D.R., Barton II, F.E., 1989. Protocol for NIRS calibration: sample selection and equation development and validation. In: Martens, G.C., Shenk, J.S., Barton II, F.E. (Eds.), Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality. Agriculture Handbook, vol. 643. USDA-ARS, US Government Printing Office, Washington, DC, USA, pp. 96–103.

Xue, L.H., Yang, L.Z., 2009. Nondestructive determination of nitrate content in spinach leaves with visible-near infrared high spectra. GuangPuXue Yu GuangPu Fen Xi 29, 926–930.

Table 3.2.1. Number of samples (N), range, mean, standard deviation (SD), and coefficient of variation (CV) for different calibration sets (C1 and C2) and for the validation set (V).

	Parameters																	
	Nitrates (mg L^{-1})			Firmness (N)			Dry matter (% fw)			SSC (ºBrix)			pH			Titratable acidity (% citric acid)		
	C1	C2	V	C1	C2	V	C1	C2	V	C1	C2	V	C1	C2	V	C1	C2	V
N	150	116	34	150	116	34	150	116	34	150	116	34	150	116	34	150	116	34
Range	30.00-1074.00	30.00-1074.00	55.00-1068.00	0.25-9.81	0.25-9.81	1.35-9.20	3.48-6.74	3.48-6.74	3.71-5.83	3.60-6.70	3.60-6.70	3.70-6.35	5.86-6.76	5.86-6.76	5.99-6.60	0.07-0.11	0.07-0.11	0.07-0.10
Mean	410.25	403.49	433.34	5.76	5.67	6.05	4.87	4.92	4.67	4.73	4.75	4.65	6.31	6.31	6.29	0.09	0.08	0.09
SD	288.68	280.13	69.43	2.09	2.14	1.94	0.67	0.68	0.62	0.61	0.60	0.66	0.17	0.18	0.16	0.01	0.009	0.008
CV	70.37	69.42	73.73	36.28	37.74	32.07	13.76	13.82	13.28	12.90	12.63	14.19	2.69	2.85	2.54	11.11	11.25	8.89

Table 3.2.2. Calibration statistics for NIR-based models for predicting internal quality parameters in intact summer squash. Training set C1.

Parameter	Math treatment	N	Range	Mean	SD	SECV	R^2_{cv}	CV (%)	RPD
Nitrates (mg L^{-1})	2,5,5,1	139	30.00-1074.00	386.64	271.82	112.44	0.83	29.08	2.42
Firmness (N)	2,10,5,1	145	1.30-9.81	5.88	1.98	1.46	0.45	24.83	1.36
Dry matter (% fw)	2,10,5,1	145	3.48-6.74	4.84	0.65	0.38	0.66	7.85	1.71
SSC (ºBrix)	1,5,5,1	144	3.70-6.00	4.71	0.58	0.33	0.68	7.01	1.76
pH	1,5,5,1	144	5.86-6.62	6.30	0.17	0.11	0.57	1.75	1.55
Titratable acidity (% citric acid)	2,5,5,1	148	0.07-0.11	0.09	0.01	0.01	0.44	7.54	1.35

Table 3.2.3. MPLS regression statistics for NIR-based models for predicting internal quality parameters in intact summer squash.

Parameter	Math treatment	Calibration							Validation				
		N	Range	Mean	SD	SECV	R^2_{cv}	CV (%)	RPD	N	R^2_p	SEP (c)	Bias	Slope
Nitrates (mg L^{-1})	2,5,5,1	115	30.00-1074.00	398.82	276.8	145.04	0.73	36.37	1.91	32	0.67	165.55	16.25	1.10
Dry matter (% fw)	1,5,5,1	111	3.48-6.74	4.90	0.66	0.38	0.67	7.69	1.75	34	0.50	0.47	-0.09	0.87
SSC (ºBrix)	1,10,5,1	107	3.70-6.00	4.73	0.50	0.32	0.59	6.82	1.56	34	0.35	0.54	-0.02	0.86
pH	2,5,5,1	113	5.86-6.62	6.31	0.17	0.11	0.56	1.81	1.50	34	0.33	0.14	-0.01	0.76

Figure 3.2.1. Mean and standard deviation spectrum for summer squash.

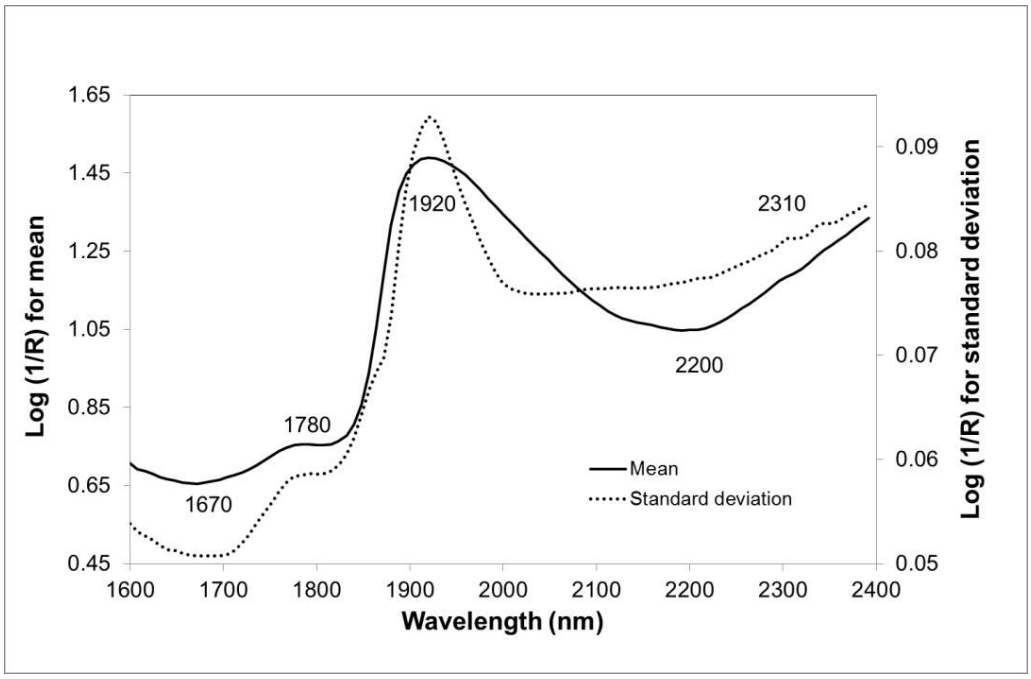
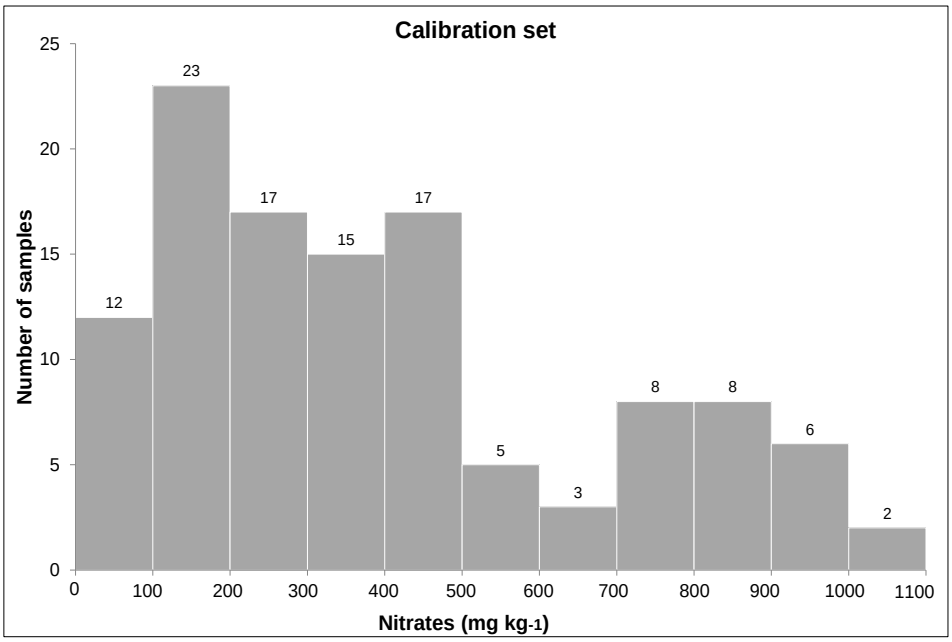


Figure 3.2.2. Distribution of nitrate content for intact summer squashes during on-vine ripening.



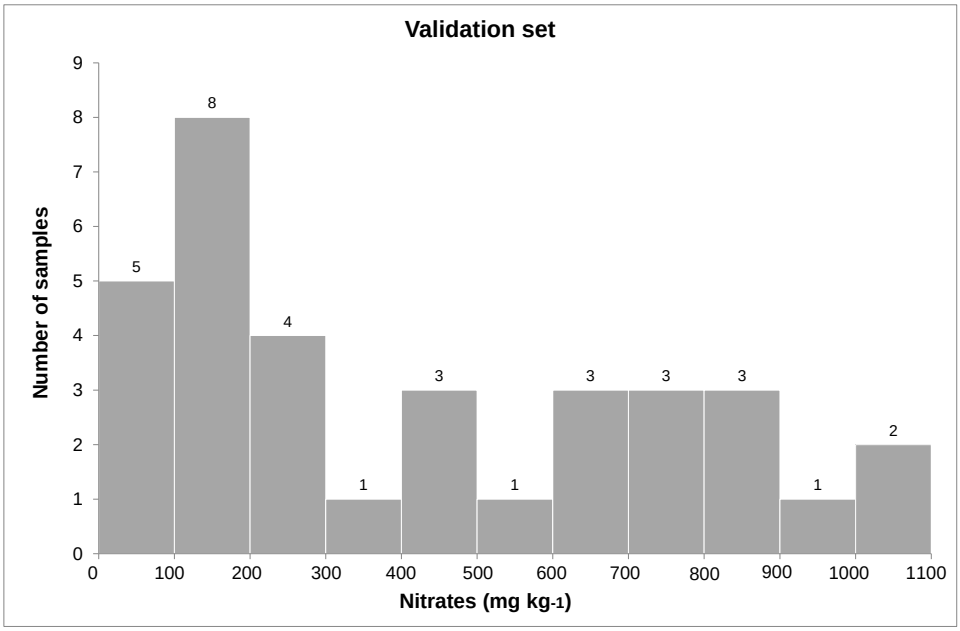
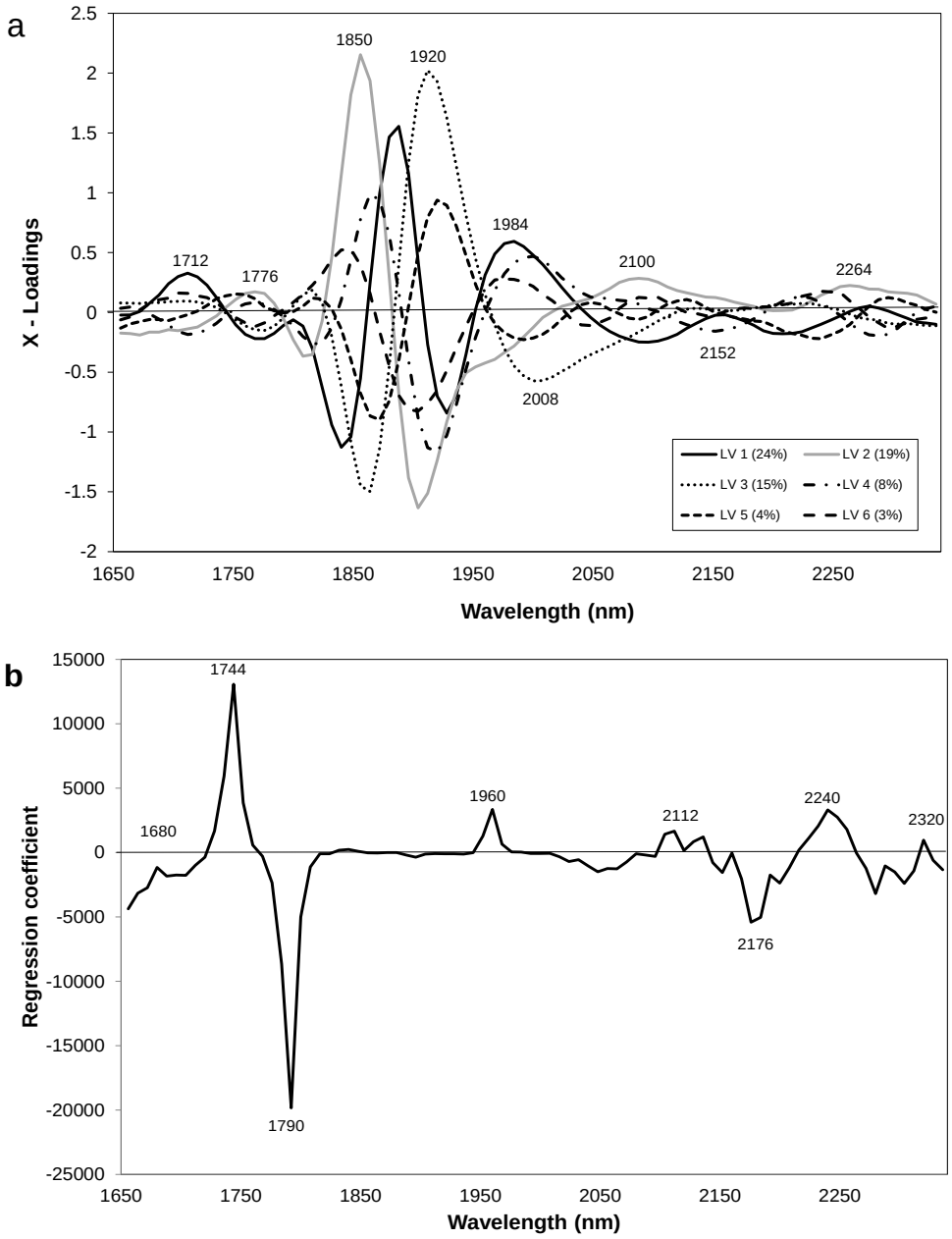


Figure 3.2.3. x-Loading weights (a) and regression coefficients (b) for summer squash nitrate content during on-vine ripening.



Chapter 3.3.

Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors

Irina Torres[a], María-Teresa Sánchez[a], José Antonio Entrenas [a], Ana Garrido-Varo[b], Dolores Pérez-Marín[b]

[a] *Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*
[b] *Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*

Postharvest Biology and Technology 154, 21-30 (2019)

Abstract

Currently, there is growing interest from the producers and the vegetable processing industry to use from farm to fork, non-destructive analysis techniques as a routine step to ensure the quality and safety of horticultural products. This interest, coinciding with the development of new instruments adapted for use both in the field and on the production line, has led to Near Infrared Spectroscopy (NIRS) becoming an increasingly practicable option for meeting the demand. The aim of this work was to develop and optimize NIRS analysis methodology using two spectrophotometers: the first is the MicroNIR™ 1700, a manual, portable instrument based on Linear Variable Filter (LVF) technology, ideally suited to analysing horticultural products in the field, and the other is the Matrix-F, based on Fourier Transform (FT) NIR technology and suitable for online analysis in the processing industries. A total of 230 summer squashes were used to predict the quality (dry matter and soluble solid content) and safety (nitrate content) parameters. For the MicroNIR™ 1700, the comparison between the equations developed confirmed that taking of point spectra (static mode) was the most suitable way of analysis to measure both the quality and safety parameters. In the case of the Matrix-F instrument, it was confirmed that a single spectrum taken online for the intact product as it moves on the conveyor belt (dynamic mode) is enough to establish the product's quality and safety during industrial processing, thus allowing it to be incorporated easily and conveniently into the production line.

Keywords: Summer squash; NIRS technology; *In situ* NIR analysis optimization; Safety and quality parameters; monitoring the food chain with sensors

3.3.1. Introduction

Summer squash, due to its high content of vitamins, fibre, minerals and trace elements and its extremely low caloric value, is an ingredient widely used in the baby food processing industry (Paris, 2008). Despite the numerous benefits accruing from the consumption of this vegetable, recently there has been growing awareness among consumers about the presence of nitrates in summer squash and about the risk involved in an excessive intake of these substances, which are reduced to nitrites in the human body (Mensinga et al., 2003).

Despite the introduction of established codes of good agricultural practices, in certain cases, some producers fail to reduce the levels of nitrates present in vegetables, mainly due to the influence of climate, and particularly, to the amount of light. Although the high light intensity typical of summer crops favours the plant's metabolism by fixing nitrogen in the form of organic nitrogenous compounds, such as aminoacids, proteins, chlorophyll, etc., and reducing its nitrate content, low light intensity (e.g. in winter crops) favours the presence of higher concentrations of nitrates (Blom-Zandstra, 1989).

The European Union is aware of these problems and has established maximum limits for nitrates content in summer squash according to their final uses (OJEU, 2011). As a result, both producers and the infant food industry must understand that only those summer squashes with a nitrate content below 200 mg kg^{-1} can be considered suitable for the production of baby food products. In addition, these regulations stress the need for producers and the agri-food industry to employ non-destructive analysis technologies in the handling and processing industries, which may be used both *in-situ* (in the field) and online, and which allow to measure the level of nitrates present in summer squash and, therefore, establish its final use, in a matter of seconds.

NIR spectroscopy is a fast, accurate, non-destructive and reliable technology which is ideal for this purpose. However, the only scientific article published to date on the non-destructive measurement of quality parameters (dry matter and soluble solid content) and safety (content of nitrates) in summer squash using NIRS technology has been that of Sánchez et al. (2017), who used a handheld MEMS (micro-electro-mechanical system)-based NIRS digital transform spectrophotometer and performed the NIRS analysis for these parameters *in-situ* on summer squash, exclusively taking of point spectral readings at the equatorial region of the fruits.

Recently, a new generation of portable, compact and extremely light-weight NIRS instruments has been developed, ideally suited for use in the field and for taking *in-situ* measurements (Yan and Siesler, 2018). These instruments enable not only to take spectra at any time, but also to analyse the whole surface of the product, thus obtaining more information about the quality and safety of the product to be harvested.

Lately, there has also been a growing interest from the vegetable processing industry to incorporate NIR spectroscopy as a routine technique to ensure the quality and safety of the vegetables they process, and this requires the NIRS applications to be developed in advance, simulating the industrial processes of the horticultural industries. One fundamental issue in developing online NIRS applications is to make the correct choice of instrument to be used, which must be robust and stable when subjected to vibrations and thermal variations (Porep et al., 2015; Garrido et al., 2018).

In the same way, it should be noted that, regardless of the type of application and instrument used, the routine incorporation of NIRS technology both in the field and in the industry requires a system to be established which includes all the issues related to taking spectra, as well as the selection of the optimal spectral region for each instrument used. For this reason, in the present research, the NIRS analysis methodology has been fine-tuned and optimized to

measure the quality and safety parameters in summer squash both *in-situ*/on the plant and online, in the industry. Two new generation of NIR spectrophotometers, one manual, portable instrument based on LVF technology (MicroNIR^{TM} 1700), suitable for use in the field, to analyse the product while it is developing in the plant, and another based on FT-NIR technology (Matrix-F) which can be incorporated into the processing industry of these vegetables, such as in the product sorting areas, and to control the quality and safety standards established by the industry.

3.3.2. Materials and methods

3.3.2.1. Sampling

A total of 230 summer squashes (135 less than 400 g in weight and 95 more than 400 g, being the minimum, maximum and mean weight of the fruits tested: 78.43 g, 1746.49 g and 454.13 g, respectively) (*Cucurbita pepo* subsp. *pepo* cv. Mirza) grown in an open-air field in the district of La Montiela, Santaella (Córdoba, Spain), were weekly harvested at commercial maturity (OJEU, 2008), from mid-May to mid-July 2017, being the number of harvests of 10. Samples were stored under refrigerated conditions (5ºC and 85% RH) until the following day, when the analysis was performed. Prior to each measurement, the samples were left to reach room temperature.

3.3.2.2. Reference data

Dry matter, soluble solid content (SSC) and nitrate content were measured following the procedures used by Sánchez et al. (2017). To analyse these parameters in summer squashes weighing more than 400 g, the fruit was cut into three sections: the stem region (upper third of the squash starting at the peduncle), the equatorial region (middle third in the equator of the fruit) and the stylar region (lower third of the fruit, starting at the pistil scar). All the analytical measurements were performed in duplicate immediately after NIR spectrum collection and the standard error of the laboratory (SEL) was estimated from these duplicates (Table 3.3.1).

3.3.2.3. Obtaining the NIR spectra

The NIR spectra of summer squashes were collected using two instruments adapted to *in-situ* and online applications, respectively.

A handheld LVF instrument (MicroNIR™ 1700 spectrophotometer, JDSU Uniphase Corporation, Milpitas, CA, USA), designed for *in-situ* analysis, in the field or in the industry, was used in reflectance mode (log 1/R). The MicroNIR™ 1700 is a miniaturised near-infrared analyser that relies on a linear variable filter as the dispersion element. This portable miniature spectrometer is extremely light (64 g). The spectrophotometer scans at a constant interval of 6.2 nm, across the NIR wavelength range of 908 to 1676 nm, and its optical window is around 227 mm^2. The instrument's performance was checked every 10 min. A white reference measurement was obtained using a NIR reflectance standard (Spectralon™) with a 99% diffuse reflectance, while a dark reference was obtained from a fixed point in the room.

The online instrument used for the spectrum acquisition was the FT-NIR spectrophotometer Matrix-F (Bruker Optik GmbH, Ettlingen, Germany). This equipment consists of a sensor head with two NIR light sources which illuminate a sampling spot of 10 mm in diameter, attached to the instrument via a 5 m-long fibre optic probe. The spectra were collected in reflectance mode in the spectral range from 4000 to 12000 cm^{-1} (834–2502.40 nm), with a resolution of 16 cm^{-1}. Furthermore, the system was equipped with a conveyor belt to move the sample, whose speed was set at 10 kHz. An internal white reference was also collected every ten minutes.

The spectra taken with the portable LVF instrument were evaluated in two different modes: static – taking of point spectra readings in the centre of the surface of the summer squashes analysed, without the instrument moving during the measurement –, and dynamic, with the sensor being moved during the spectral measurements all along the length of the summer squash being

analysed. For both analysis modes, an integration time of 11000 µs and 200 scans per spectrum were set.

To take the spectra using the MicroNIR™ 1700 instrument, the summer squash were cut into the three regions mentioned above (stem, equatorial and stylar regions), regardless of the weight of the fruit analysed. In the static analysis, 4 spectra were taken at the centre of each of the selected regions, at 90º intervals, so that 12 spectra were obtained per summer squash, while in the dynamic mode analysis, 4 spectra were also taken covering each of the regions analysed and rotating the fruit 90º between measurements, again obtaining 12 spectra per fruit analysed. In the case of summer squashes weighing less than 400 g, 12 spectra were taken per fruit. These 12 spectra were averaged to obtain the mean spectra for each fruit, obtaining a number of spectra equal to the number of analysed fruits, in this case 135 spectra.

For summer squashes weighing more than 400 g, two strategies were followed:

(I) An average was taken of the four spectra corresponding to each of the regions (stem, equatorial and stylar regions) analysed, thus obtaining one single spectrum per region.

(II) The 12 spectra taken for each summer squash were averaged, producing one spectrum per fruit.

Therefore, for Strategy I, an initial sample group consisting of 420 spectra was obtained, i.e. 95 fruits x 3 regions/fruit x 1 spectrum/region (summer squashes more than 400 g in weight) + 135 fruits x 1 spectrum/fruit (summer squashes less than 400 g in weight), while for strategy II, an initial group of spectra was obtained equal to that of the number of fruits analysed (230 spectra = 95 spectra of summer squashes more than 400 g in weight + 135 spectra of summer squashes less than 400 g in weight).

With the Matrix-F instrument, the analysis was carried out online as the product was moving on the conveyor belt. The number of scans selected per fruit was 12, covering the entire length of the fruit as it moved on the belt, and, with this instrument, no analysis of different fruit regions was carried out. A total of 4 spectra per fruit were taken, rotating the fruit 90º after each measurement, and different strategies were used to perform their chemometric analysis:

1. Selecting a single spectrum per summer squash analysed.
2. Using the average spectrum obtained after taking 2 spectra per summer squash.
3. Using the average spectrum of the 4 spectra taken for each summer squash.

In the case of strategies 1 and 2, the spectra were randomly selected by Matlab v. 2015a (The Mathworks, Inc., Natick, Massachusetts, USA).

However, regardless of the strategy followed, the total number of spectra used for the development of the predictive models was 230, which was equal to the number of fruits used in this study.

3.3.2.4. Definition of calibration and validation sets

Prior to the development of NIR calibrations, data pre-processing and chemometric treatments were performed using the WinISI software package ver. 1.50 (Infrasoft International LLC, Port Matilda, PA, USA).

Firstly, the most suitable spectral range for the instruments tested to carry out the quality and safety control of summer squashes was selected. To achieve this, the 1,1,1,1 derivation treatment was applied (the first digit being the number of the derivative, the second the gap over which the derivative is calculated, the third the number of data points in a running average or smoothing, and the fourth the second smoothing) without scatter correction,

which allows to highlight the areas of the spectrum where the signal/noise ratio is degraded (Hruschka, 2001).

Additionally, and in the case of the instrument used to simulate the analysis of summer squash in vegetable sorting lines, the Matrix-F spectrophotometer, a distance of 12 cm between the instrument head and the conveyor belt (head-fruit distance was 8 cm and 10 cm for summer squashes weighing both more and less than 400 g, respectively) was established, which remained constant throughout the process of taking spectra, regardless of the type of summer squash analysed. This attempted to simulate the existing conditions in the food processing industry, which uses fruits with different equatorial diameter values in their classification processes, and where the automatic readjustment of the distance between the instrument and the belt, depending on the different equatorial diameters presented by the summer squash analysed, is not possible. Next, principal component analysis was used to study the influence of the interaction of the light with the product when this distance was kept constant, regardless of the equatorial diameter of the fruit analysed.

Secondly, the CENTER algorithm was applied to ensure a structured population based on spectral information for the selection of calibration and validation sets. This algorithm performs principal component analysis (PCA) and determines the centre of the population and the distance between each sample and the centre [Mahalanobis distance (GH)]. Samples with a statistical value greater than 4 were considered outliers or anomalous spectra (Shenk and Westerhaus, 1995a). As spectral pre-treatments, Standard Normal Variate (SNV) and Detrending (DT) were used to remove scatter interferences (Barnes et al., 1989), together with the first derivative treatment ‘1,5,5,1’ (Shenk and Westerhaus, 1995b).

Once spectral outliers had been removed for each instrument, three of every four samples were selected to form part of the calibration set and the

remainder constituted the validation set (Table 3.3.1). Thus, for the MicroNIR™ 1700, a calibration group, C1 (305 samples) and a validation group, V1 (107 samples), for strategy I were established. For the Matrix-F, the calibration group (C2) and validation group (V2) were made up of 169 samples and 53 samples, respectively, for each of the fruit spectra-taking strategies tested (strategies 1, 2 and 3). Finally, to analyse the results obtained with the two instruments tested, the same calibration groups (C2 set) and validation (V2 set) were selected for strategy II for the MicroNIR™ 1700 as for the Matrix-F.

3.3.2.5. Construction and validation of prediction models using a linear regression strategy

Modified partial least squares (MPLS) regression (Shenk and Westerhaus, 1995a) was used to obtain NIR calibration models for the prediction of quality and safety parameters in summer squashes using the MicroNIR™ 1700 (calibration sets C1 and C2) and the Matrix-F (calibration set C2). All regression equations were obtained using SNV + DT for scatter correction (Barnes et al., 1989) and different derivative mathematical treatments were tested: 1,5,5,1; 2,5,5,1; 1,10,5,1 and 2,10,5,1 (Shenk and Westerhaus, 1995b). Calibration models were constructed eliminating physical-chemical outlier samples, if necessary.

The statistical parameters employed to select the best equations using MPLS were the coefficient of determination for calibration (R^2_c), the standard error of calibration (SEC), the coefficient of determination for cross validation (R^2_{cv}) and the standard error of cross validation (SECV). Furthermore, the Residual Predictive Deviation (RPD_{cv}) for cross-validation was calculated as the ratio of the standard deviation (SD) of the reference data to the SECV. This statistical parameter enables SECV to be standardized, facilitating the comparison of results obtained with sets of different means (Williams, 2001).

For strategy I, once the best predictive model for each parameter analysed (dry matter, SSC and nitrate content) using the instrument

MicroNIR™ 1700 in static and dynamic modes were selected by statistical criteria - and prior to external validation -, tests were run for significant differences between the predictive capacity of the quality and safety models developed for each parameter. This approach was used to identify the most suitable analysis mode for measuring *in-situ* quality and safety in summer squash. The SECV values for the best equations obtained for both analysis modes were compared using Fisher's F test (Massart et al., 1988; Naes et al., 2002). Values for F were calculated as:

$$F = \frac{(SECV_2)^2}{(SECV_1)^2}$$

where $SECV_1$ and $SECV_2$ are the standard error of cross validation of two different models and $SECV_1 < SECV_2$. F is compared to $F_{critical\ (1\text{-}P,\ n\text{-}1,\ n\text{-}1)}$ read from the table with $P = 0.05$ and n-1 degrees of freedom. If F is higher than $F_{critical}$, the two SECV values are significantly different.

For the online instrument (Matrix-F), different predictive models, without the elimination of physical-chemical outliers, were developed to test the optimum number of spectra (1, 2 or 4 spectra) per fruit that must be taken to obtain robust models so as to establish an NIR analysis methodology in summer squash which ultimately could be transferred to the horticultural processing industry. The SECV values for the best equations obtained without the elimination of physical-chemical outliers during the development of the models for each parameter with different numbers of spectra per fruit were also compared using Fisher's F test (Massart et al., 1988; Naes et al., 2002). Because in this study several SECV values were compared, a $SECV_{confidence\ limit}$ was calculated using the following formula:

$$SECV_{confidence\ limit} = SECV_{min} * \sqrt{F_{critical}}$$

where $SECV_{min}$ is the smallest SECV. Consequently, none of the models which had a SECV between $SECV_{min}$ and $SECV_{confidence\ limit}$ were significantly different ($P < 0.05$).

Once the best analysis strategy (optimal number of spectra per fruit) was chosen to measure online the quality and safety parameters of summer squash using the Matrix-F equipment, optimization of the NIR models was carried out for the parameters studied.

Finally, once the best equations for each of the instruments used were selected according to the statistical criteria and once both the best strategy of taking spectra was chosen (static or dynamic mode) when using the MicroNIR™ 1700, and the optimal number of spectra per fruit was established when using the Matrix-F, these models were subjected to an external validation process using the corresponding validation group for each instrument (V1 for the MicroNIR™ 1700 and V2 for the Matrix-F), following the protocols outlined by Windham et al. (1989).

Lastly, after analysing the results obtained and in order to carry out the analysis of the influence of the number of samples contained in the calibration group on the robustness of the developed models, new predictive models were developed for the three parameters analysed. That is, the same calibration group C2 (N = 169 samples) (Table 3.3.1) was used for the two instruments tested. Next, the two instruments were compared using the RPD_{cv} values obtained for the three parameters analysed using Fisher's F test, as mentioned above. It should be noted that, since the spectrophotometers used were originally designed for different applications, *in-situ* and online analysis, the aim of this comparison was not the choice of instrument, but the availability of the results which can be used as a reference to estimate the robustness of the models obtained.

3.3.3. Results and discussion

3.3.3.1. Optimization of the spectral analysis

Prior to performing the Principal Component Analysis (PCA) and developing the prediction models, the region of the spectra affected by noise was determined for both instruments. This aspect is especially relevant for the Matrix-F, since with this equipment, the spectral signal is transmitted by fibre optics, which commonly produce noise on extreme wavelengths (Garrido-Varo et al., 2018).

Figures 3.3.1a and 3.3.T1b show the first derivative spectra of summer squashes for the MicroNIR™ 1700 and Matrix-F instruments, respectively. In the case of the MicroNIR™ 1700, no areas of the spectrum needed to be removed due to the presence of noise, and the final range was between 908–1676 nm (Figure 3.3.1a). On the other hand, in the case of the Matrix-F, the regions between 834–1075 nm and between 2360–2502 nm were removed due to the presence of noise in the spectrum (Figure 3.3.1b).

The representation of the scores of PC1 *versus* PC2 (Figure 3.3.2) of the spectra taken with the Matrix-F instrument following strategy 3 (with a mean spectrum of 4 spectra per fruit), allowed to draw a clear distinction between the two groups of summer squashes analysed (more and less than 400 g). Because it is not possible to adjust automatically the height of the equipment based on the equatorial diameter of the fruit analysed while the industrial process is in motion, this study demonstrates the importance of optimizing the analysis distance between the head and conveyor belt, so that the difference in the amount of light which the fruits of different values of equatorial diameter are exposed to is minimal. Besides, the application of spectral pre-treatments could reduce these differences (Workman, 2008).

In the case of solid products which are analysed intact, as is the case of summer squash, part of the incident radiation can be lost due to the different

ways in which the light interacts with the sample. These include phenomena such as specular reflectance or dispersed radiation, which can lead to the responses not containing information or not reaching the detectors. These phenomena can be influenced by different factors such as the colour, brightness and shape of the sample, among others (Dahm and Dahm, 2001; Jie et al., 2014).

3.3.3.2. Characterisation of calibration and validation sets

To sort the samples according to the Mahalanobis distance (GH), the CENTER algorithm was applied to the spectral databases. Samples with GH values greater than 4 were considered outliers.

In the case of the MicroNIR™ 1700 for Strategy I (initial sample set = 420 spectra), a total of 8 outliers (4 corresponding to summer squash weighing more than 400 g y 4 corresponding to fruit weighing less than 400 g) were removed.

For the Matrix-F instrument, regardless of the number of spectra taken per sample, the number of outliers was 8 (2 corresponding to summer squash weighing more than 400 g and 6 to fruit weighing less than 400 g), and these were also removed. For the MicroNIR™ 1700 by using stratagem II, in order to obtain the same calibration and validation sets as when using the Matrix-F, the same samples considered as outliers for the Matrix-F were removed for the portable equipment.

After removing the outliers, the set for the MicroNIR™ 1700 (strategy I) was split into calibration (C1 = 305 samples) and validation (V1 = 107 samples), while for the Matrix-F and the MicroNIR™ 1700 (strategy II) the sample set was divided into calibration (C2 = 169 samples) and validation (V2 = 53 samples) sets, respectively. The statistical data for the parameters analysed (i.e. number of samples, range, mean, SD and CV) for each instrument are shown in Table 3.3.1.

For each of the parameters measured in this study, the ranges of the calibration groups included those for validation.

For both instruments, the parameter with the greatest variability was the nitrate content, with a CV between 82.71% and 83.42% and between 76.58% and 91.95% for the calibration and validation groups, respectively. This variability was due to the wide range and standard deviation obtained for this parameter because the samples were taken during the summer squash harvest period, with the nitrate content of the fruits decreasing as the harvest period progressed, due to the cessation of fertilization. However, for dry matter (CV_c = 15.78-16.43%, CV_p = 14.50-14.82%) and SSC (CV_c = 11.35-11.42%, CV_p = 11.01-12.02%) the groups show less variability, because the fruits were collected at the point of commercial maturity.

3.3.3.3. Development of predictive models and choice of the best analysis mode to predict quality and safety parameters using the MicroNIR™ 1700 instrument

The calibration statistical parameters for the best models obtained for each parameter analysed in static and dynamic modes using the MicroNIR™ 1700 instrument and the C1 set, including the F values obtained from the comparison between the SECV values for each mode of analysis are shown in Table 3.3.2.

Non-significant differences ($P > 0.05$) were obtained between the dynamic and static modes to predict all the quality (dry matter and SSC) and safety (nitrate content) parameters. The lowest SECV values were obtained in the static mode analysis for all the parameters except for SSC, where a slightly lower SECV value was obtained in the dynamic mode (Table 3.3.2).

Although it could be argued that the dynamic mode analysis resulted in a better fit, as it covered the whole area of the fruits analysed and collected more information about it, because the fact that the surface of the squash is not flat, it can lead to greater difficulty and associated error when taking the

spectra. For this reason, the results showed non-significant differences between the static analysis of the fruits and the dynamic mode. It is also important to note that, given the shape of the summer squash fruits, taking of point spectra readings of the fruit before harvest time, while it is developing on the plant, would be an easier way to use the instrument in the field. In addition, it would make it easier for the farmers to take spectra simpler, quicker and more comfortable.

After choosing the static analysis mode as the most suitable for the *in-situ* analysis of summer squash using the MicroNIR™ 1700 instrument, the statistics of the calibration models obtained with this analysis mode were analysed in more detail (Table 3.3.2).

According to Shenk and Westerhaus (1996), the model developed for the dry matter parameter presented a predictive capacity which enabled to discriminate between high, medium and low values of the parameter, while for SSC it was only possible to differentiate between high and low values.

Only one work has been found in the literature in which NIRS technology was used to predict parameters such as dry matter and SSC. In Sánchez et al. (2017), a portable equipment based on MEMS technology in the spectral range 1600-2400 nm was used to measure dry matter and SSC, and the predictive capacity of models (RPD_{cv} = 1.71 and 1.76, respectively) was very similar to the one obtained in the present study (Table 3.3.2).

As regards to the nitrate content parameter, the predictive capacity of the model obtained allowed to distinguish between high, medium and low values for this parameter (Shenk and Westerhaus, 1996). The results obtained by Sánchez et al. (2017) for this parameter using the portable instrument Phazir 2400 and with a calibration group within a range of 30.00–1074.00 mg L^{-1} (SD = 288.68), were superior (RPD_{cv} = 2.42) to those obtained in this work However, it should be noted that the Phazir 2400 has a different spectral range

compared to the instrument used in the present work (1600-2400 nm) and this instrument is no longer being used. The recent development of applications using the new manual, portable, light-weight equipment with a wider optical analysis window, such as the MicroNIR^{TM} 1700, has therefore been of great use to producers to determine the final destination of the harvested product.

3.3.3.4. Selection of the best spectrum capture strategy for online NIR analysis with the Matrix-F instrument

Table 3.3.3 shows the SECV values of the best calibration models obtained using the Matrix-F instrument (C2 set), as well as different strategies for the number of spectra to be taken (1, 2 and 4 spectra per fruit analysed), for each of the parameters studied. In order to compare the three strategies used to take spectra, the calibration models for the different parameters in the study were carried out without eliminating the physical/chemical outliers during their development. That is, the values for mean, range and standard deviation for each parameter tested were the same, regardless of the number of spectra taken per fruit.

No significant differences were found, for any of the parameters analysed, between the SECV values of the predictive models developed for the different strategies tested. Therefore, taking into account the results obtained, and in order to facilitate the use of the NIR spectroscopy in the industries where summer squashes are handled, the simplest, quickest way to measure the quality and safety parameters online, during the process of selection and classification of fruits would be to take only one spectrum per fruit, which makes this strategy highly suitable for use in sorting lines.

The results obtained agree with those reported by McCarthy and Kemeny (2008), who showed that when using FT-NIR instruments, due to the improved signal/noise ratio in these instruments, a smaller number of spectra per analysed sample was needed for the measurement to yield relevant information.

3.3.3.5. Development of new models for online prediction of quality and safety parameters in summer squash using the Matrix-F

After selecting the optimum number of spectra to be taken per fruit using the Matrix-F instrument (1 spectrum/fruit), new predictive models were developed for the parameters analysed, and these were optimized by removing any physicochemical outliers. Table 3.3.4 shows the calibration statistical data for the best models developed to predict dry matter, SSC and nitrate content in intact summer squash analysed online with this instrument.

For dry matter, the model developed enabled to differentiate between high, medium and low values, while for SSC, the model only distinguished between high and low values (Shenk and Westerhaus, 1996). These results are highly relevant because the spectra were taken on the moving fruit and that the models were developed using a single spectrum per sample, which shows the usefulness of the Matrix-F for measuring quality parameters online.

No data has been published on the best approach to use when online measurements of dry matter and SSC are taken in intact summer squashes. However, Jie et al., (2014) predicted SSC in watermelon with NIR equipment (Ocean Optics Inc., USA) adapted for online use. In transmittance mode in the spectral range 200–1100 nm: RPD_p value of 1.32 was obtained, thus illustrating the greater difficulty of taking quality parameter measurements while the fruit is moving.

The results obtained for nitrate content were limited for routine use (RPD_{cv} = 1.36). The reason for this may be that, for parameters such as nitrates, which are found in very low concentrations in summer squash (i.e., in parts per million), the fact that no contact is made with the fruit when taking the measurement may affect the results for low concentration parameters. Thus, it is essential that a large group of samples must be used in order to produce a robust calibration (Sánchez et al., 2011), and for future work, a much larger number of fruits must be provided.

It is also important to note that in this research work, spectra were also taken in summer squash (1 spectrum per fruit) with the Matrix-F instrument while the conveyor belt had stopped (data not shown). For this mode of analysis, the predictive capacity of the models obtained for each parameter (RPD_{cv} = 2.00, RPD_{cv} = 1.63 and RPD_{cv} = 1.27, for dry matter, SSC and nitrate content, respectively) was similar to that obtained in the dynamic mode (Table 3.3.4). These results are of particular interest for the industry, since they reinforce the potential of the Matrix-F equipment to be used as a tool to measure quality parameters in moving production lines.

3.3.3.6. External validation

Validations of the best calibration models obtained for the MicroNIR™ 1700 (calibration set C1) and Matrix-F (calibration set C2) were performed using the sets V1 and V2, respectively, for the two instruments tested.

For the MicroNIR™ 1700 instrument and the V1 set, according to the validation protocol established by Windham et al. (1989), the models constructed for predicting all the parameters analysed met the validation requirements in terms of $SEP_{(c)}$, bias and slope (Figure 3.3.3a). Although the R^2_p values did not attain the recommended minimum value of 0.60 (R^2_p = 0.57, 0.51, 0.54 for dry matter, SSC and nitrate content, respectively), they were closed. The equations developed can therefore be taken as an initial approximation to the *in-situ* measurement of quality and safety measurements in intact summer squash using this handheld instrument.

For the Matrix-F instrument (prediction set = V2), in the case of the model constructed for predicting dry matter, the SEP(c) does not comply with the protocol while the other statistics lie within the confidence limits (Figure 3.3.3b). SSC fails to meet the validation requirements in terms of R^2_p, while the SEP(c) and bias are below the limits. Additionally, the slope for SSC fails to comply with the values recommended (0.90-1.10). As for the nitrate content, no

external validation of the model was carried out due to its limited predictive capacity.

In general, the SEP is considered a valuable statistical parameter to evaluate the predictive capacity of an equation, and it is widely accepted that an SEP value of less than 2*SEL shows that the model has an excellent predictive capacity (Westerhaus, 1989; Williams, 2001). In this work, the SEL values for dry matter, SSC and nitrate content were 0.30 % fw, 0.07 ºBrix and 19.57 mg kg^{-1}, respectively (Table 3.3.1). For both instruments, and for the parameter dry matter, the SEP was between 1 and 2, showing the excellent capacity of the NIR model. However, the low SEL values for SSC and nitrate content in comparison with the SEP values obtained for the prediction must be correctly interpreted. It must be considered that the distribution of these components is not homogeneous in the fruit and, whereas the reference values were obtained from the summer squash juice, the spectra were taken from a specific region of the fruit. For this reason, it could be said that a sampling error occurred which was not included in the SEL value. Consequently, the NIRS model developed using both the portable and online instruments for SSC and the nitrate content were marked by their questionable performance, since the SEP value obtained exceeded 5*SEL. Nevertheless, it is important to stress that all the limits and values recommended in the scientific literature and mentioned above refer to other NIRS analysis conditions, i.e. using at-line instruments and using pre-dried and ground samples. In this study, our applications were developed with portable or online devices, using intact samples with a high level of moisture. In this case, the comparison with the limits indicated may be too restrictive.

These results suggest the importance of interpreting the SEP values correctly. It is also worth noting how important it is that the NIR spectrophotometers allow both the quality and safety of the product to be guaranteed rapidly and accurately throughout the production chain, from the field to the table, and also permit the number of samples analysed for each batch produced to be increased.

3.3.3.7. Comparison between the best models developed with the MicroNIR™ 1700 and Matrix- F instruments using the C2 calibration set

The statistical parameters for the best predictive models developed with the calibration group C2 using the MicroNIR™ 1700 instrument (spectra collection in static mode and a mean spectrum per fruit) and the Matrix-F spectrophotometer (spectra collection in dynamic mode and a mean spectrum per fruit) are shown in Table 3.3.5.

After the equations were developed for the two instruments, the RPD_{cv} values obtained for each of the parameters tested were compared using Fisher's F test. The predictive capacity of the models developed for the quality parameters with the Matrix-F equipment (RPD_{cv} = 1.98 and RPD_{cv} = 1.63 for dry matter and SSC, respectively) was higher than that obtained with the MicroNIR™ 1700 instrument (RPD_{cv} = 1.72 for dry matter and RPD_c = 1.58 for SSC), and this superiority was significant ($P < 0.05$) in the case of dry matter (Table 3.3.5). This greater predictive capacity of the Matrix-F may exist because the equipment is more robust and covers a wider spectral range (834–2502.40 nm for Matrix-FT versus 908–1676 nm for MicroNIR™ 1700) with a higher spectral resolution (1.61 nm for Matrix-FT and 6.20 nm for the MicroNIR™ 1700).

In the case of the nitrate content, there were no significant differences ($P > 0.05$) in terms of the RPD_{cv} values obtained with the Matrix-F (RPD_{cv} = 1.36) and with the MicroNIR™ 1700 (RPD_{cv} = 1.35). These results are particularly relevant, since in the case of the MicroNIR™ 1700, it can be seen that in Strategy II (C2 set = 169 fruits) the model's predictive capacity is considerably reduced (RPD_{cv} = 1.35) compared with Strategy I (C1 set = 305 fruits; RPD_{cv} = 1.85). Since the variability of the calibration group in both cases is similar (Table 3.3.1), this lower predictive capacity can be attributed to the smaller number of samples available for developing the models, making them less representative of the different values available over the whole range.

These results seem to confirm that, when it comes to measuring the nitrate content with the Matrix-F, the limited predictive capacity obtained could be due to the reduced number of samples available. Thus, if the number of calibration group samples used to develop the model were increased, the model would have been more robust. This is of particular interest to the industry, as this technique could be carried out online as a routine method of analysis, in order to measure not only quality parameters, but also safety parameters such as nitrate content.

3.3.4. Conclusions

The results obtained confirm that NIRS technology can be used as a routine analysis tool to measure quality (dry matter and SSC) in intact summer squashes, both *in-situ* in the field and online in the handling and processing industry, thus enabling to guarantee the quality of the product throughout the whole supply chain. Furthermore, these findings confirm that the two instruments tested would enable to establish adequate uses for summer squashes according to their nitrate content, provided that a sufficiently large and highly representative sample group was available.

Additionally, when measuring both quality and safety parameters using the MicroNIR™ 1700, there were no significant differences between the predictive capacity of the models obtained, when the spectra were taken in the fruits *in-situ* in static or dynamic modes. The static mode was therefore selected because its simplicity could enable growers to take the spectra in a faster and simpler way. In the case of the Matrix-F instrument, the results showed that a single spectrum taken while the product passed on the sorting belts would be enough to guarantee the product quality. This would make it extremely easy to use NIRS analysis routinely in the summer squash handling and processing industry.

Finally, it must be highlighted that the portable NIR instrument here tested could be a promising tool for its use by the growers during the development of the fruits in the field and at harvest. In addition, NIR technology by means of the incorporation of instruments such as the Matrix-F, could be used to measure the quality and safety of the fruits tested at industry level, in the sorting lines, although it is highly important the optimization of the instrument prior to its final incorporation in the industry.

Acknowledgements

This research was carried out under the research project 'Quality determination of summer squash grown in Santaella (Córdoba)', funded by Gelagri Ibérica, S.L. The authors would like to thank Mrs. María-Carmen Fernández of the Animal Production Department for her technical assistance. The authors would also like to express their gratitude to the Spanish Ministry of Education, Culture and Sports for the support offered to Irina Torres through the Training Programme for Academic Staff (FPU).

References

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra. App. Spectrosc. 43, 772–777. https://doi.org/10.1366/0003702894202201.

Blom-Zandstra, M., 1989. Nitrate accumulation in vegetables and its relationship to quality. Ann. Appl. Biol. 115, 553–561. https://doi.org/10.1111/j.1744-7348.1989.tb06577.x.

Dahm, D.J., Dahm, K.D., 2001. The physics of near-infrared scattering. In: Williams, P., Norris, K. (Eds.), Near Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, Inc. St. Paul, MN, USA, pp. 1–18.

Garrido-Varo, A., Sánchez-Bonilla, A., Maroto-Molina, F., Riccioli, C., Pérez-Marín, D., 2018. Long-length fiber optic near-infrared (NIR) spectroscopy probes for on-line quality control of processed land animal proteins. Appl. Spectrosc. 72, 1170–1182. https://doi.org/10.1177/0003702817752111.

Hruschka, W.R., 2001. Data analysis: Wavelength selection methods. In: Williams, P.C., Norris, K.H. (Eds.), Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, Inc., St. Paul, MN, USA, pp. 39–58.

Jie, D., Xie, L., Rao, X., Ying, Y., 2014. Using visible and near infrared diffuse transmittance technique to predict soluble solids content of watermelon in an on-line detection system. Postharvest Biol. Technol. 90, 1–6. http://dx.doi.org/10.1016/j.postharvbio.2013.11.009.

McCarthy, W.J., Kemeny, G.J., 2008. Fourier Transform spectrophotometers in the Near-Infrared. In: Burns, D.A., Ciurczak, E.W. (Eds.), Handbook of Near-Infrared Analysis. CRC Press, Boca Raton, Florida, USA, pp. 79–93.

Massart, D.L., Vandeginste, B.G.M., Deming, S.M., Michotte, Y., Kaufman, L., 1988. Chemometrics: A Textbook (Data Handling in Science and Technology 2). Elsevier Science, Amsterdam, The Netherlands.

Mensinga, T.T., Speijers, G.J., Meulenbelt, J., 2003. Health implications of exposure to environmental nitrogenous compounds. Toxicol. Rev. 22, 41–51. http://dx.doi.org/10.2165/00139709-200322010-00005.

Naes, T., Isaksson, T., Fearn, T., Davies, A., 2002. A User-Friendly Guide to Multivariate Calibration and Classification. NIR Publications, Chichester, UK.

Official Journal of the European Union (OJEU), 2008. Commission Regulation (EC) No 1221/2008 of 5 December 2008 amending Regulation (EC) No 1580/2007 laying down implementing rules of Council Regulations (EC) No 2200/96, (EC) No 2201/96 and (EC) No 1182/2007 in the fruit and vegetable sector as regards marketing standards. OJ L 336/1-80, 13.12.2008.

Official Journal of the European Union (OJEU), 2011. Commission Regulation (EC) No 1258/2011 of 2 December 2011 amending Regulation (EC) No 1881/2006 as regards maximum levels for nitrates in foodstuffs. OJ L 320/15-17, 3.12.2011.

Paris, H.S., 2008. Summer squash. In: Prohens, J., Nuez, F. (Eds.), Handbook of Plant Breeding: Vegetables I. Springer, New York, USA, pp. 351–379.

Porep, J.U., Kammerer, D.R., Carle, R., 2015. On-line application of near infrared (NIR) spectroscopy in food production. Trends Food Sci. Tech. 46, 211–230. https://doi.org/10.1016/j.tifs.2015.10.002.

Sánchez, M.T., De la Haba, M.J., Guerrero, J.E., Garrido-Varo, A., Pérez-Marín, D., 2011. Testing of a local approach for the prediction of quality parameters in intact nectarines using a portable NIRS instrument. Postharvest Biol. Technol. 60, 130–135. https://doi.org/10.1016/j.postharvbio.2010.12.006.

Sánchez, M.T., Pérez-Marín, D., Torres, I., Gil, B., Garrido-Varo, A., De la Haba, M.J., 2017. Use of NIRS technology for on-vine measurement of nitrate content and other internal quality parameters in intact summer squash for baby food production. Postharvest Biol. Technol. 125, 122–128. https://doi.org/10.1016/j.postharvbio.2016.11.011.

Shenk, J.S., Westerhaus, M.O., 1995a. Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy. Monograph, NIRSystem, Inc., 12101 Tech Road, Silver Spring, MD 20904, USA.

Shenk, J.S., Westerhaus, M.O., 1995b. Routine Operation, Calibration, Development and Network System Management Manual. NIRSystems, Inc., 12101 Tech Road, Silver Spring, MD 20904, USA.

Shenk, J.S., Westerhaus, M.O., 1996. Calibration the ISI way. In: Davies, A.M.C., Williams, P.C. (Eds.), Near Infrared Spectroscopy: The Future Waves. NIR Publications, Chichester, West Sussex, UK, pp. 198–202.

Westerhaus, M.O., 1989. Interpretation of regression statistics. In: Marten, G.C., Shenk, J.S., Barton II, F.E. (Eds.), Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality. Agriculture Handbook nº643.

USDA-ARS, US Government Printing Office, Washington, DC, USA, pp. 39.

Williams, P.C., 2001. Implementation of near-infrared technology. In: Williams, P.C., Norris, K.H. (Eds.), Near-Infrared Technology in the Agricultural and Food Industries. AACC Inc., St. Paul, MN, USA, pp. 145–169.

Windham, W.R., Mertens, D.R., Barton II, F.E., 1989. Protocol for NIRS calibration: sample selection and equation development and validation. In: Martens, G.C., Shenk, J.S., Barton II, F.E. (Eds.), Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality. Agriculture Handbook, vol. 643. USDA-ARS, US Government Printing Office, Washington, DC, USA, pp. 96–103.

Workman, J.J. 2008. NIR spectroscopy calibration basics. In: Burns, D.A., Ciurczak, E.W. (Eds.), Handbook of Near-Infrared Analysis. CRC Press, Boca Raton, Florida, USA, pp. 79–93.

Yan, H., Siesler, H.W., 2018. Hand-held near-infrared spectrometers: State-of-the-art instrumentation and practical applications. NIR news, 29, 8–12. https://doi.org/10.1177/0960336018796391.

Table 3.3.1. Statistical data for the calibration and validation sets selected for each instrument and standard error of laboratory.

Parameter	Statistics	Set 1 MicroNIR™ 1700		Set 2 MicroNIR™ 1700 and Matrix-F	
		Training set (C1)	Validation set (V1)	Training set (C2)	Validation set (V2)
Dry matter (% fw)	N[a]	305	107	169	53
	Range	3.22–7.56	3.53–6.94	1.31–7.34	3.67–6.22
	Mean	4.69	4.76	4.93	4.79
	SD[b]	0.74	0.69	0.81	0.71
	CV[c] (%)	15.78	14.50	16.43	14.82
	SEL[d]			0.30	
SSC (ºBrix)	N	305	107	169	53
	Range	2.80–6.50	2.80–5.70	2.80–5.63	3.37–5.30
	Mean	4.14	4.16	4.29	4.27
	SD	0.47	0.50	0.49	0.47
	CV (%)	11.35	12.02	11.42	11.01
	SEL			0.07	
Nitrate content (mg kg^{-1})	N	305	107	169	53
	Range	18.50–1979.96	20.50–1203.38	23.33–1455.27	39.50–1219.73
	Mean	345.03	375.67	371.02	314.44
	SD	285.37	287.69	309.50	289.13
	CV (%)	82.71	76.58	83.42	91.95
	SEL			19.56	

[a] *Number of samples.*
[b] *Standard deviation.*
[c] *Coefficient of variation.*
[d] *Standard error of laboratory.*

Table 3.3.2. Calibration statistics for predicting quality and safety parameters in dynamic and static modes using the MicroNIRTM 1700 instrument (calibration set C1).

Parameter	Mode	N[a]	SECV[b]	R^2_{cv}[c]	RPD_{cv}[d]	F	$F_{critical}$
Dry matter (% fw)	Dynamic	296	0.42	0.60	1.76	1.102	1.211
	Static	296	0.40	0.62	1.85		
Soluble solid content (ºBrix)	Dynamic	293	0.29	0.55	1.62	1.070	1.212
	Static	294	0.30	0.57	1.57		
Nitrate content (mg kg^{-1})	Dynamic	293	161.45	0.60	1.77	1.099	1.211
	Static	299	154.01	0.63	1.85		

[a] *Number of samples.*
[b] *Standard deviation.*
[c] *Coefficient of variation.*
[d] *Standard error of laboratory.*

Table 3.3.3. Comparison between SECV values for the best calibration models obtained using the Matrix-F and collecting different number of spectra per sample in dynamic mode. Calibration set C2.

Parameter	SECV[a] 1 spectrum	SECV 2 spectra	SECV 4 spectra	$SECV_{min}$	$SECV_{min} * \sqrt{F_{critical}}$
Dry matter (% fw)	0.63	0.63	0.58	0.58	0.66
Soluble solid content (ºBrix)	0.37	0.35	0.34	0.34	0.39
Nitrate content (mg kg^{-1})	297.31	271.03	271.16	271.03	307.84

[a] Standard error of cross validation

Table 3.3.4. Calibration statistics for predicting quality and safety parameters using the instruments Matrix-F (dynamic mode and 1 spectrum per fruit). Calibration set C2.

Parameter	N[a]	Range	Mean[b]	SD[c]	SEC[d]	R^2_c[e]	SECV[f]	R^2_{cv}[g]	RPD_{cv}[h]
Dry matter (% fw)	160	3.16–6.85	4.90	0.66	0.32	0.62	0.41	0.62	1.98
Soluble solid content (ºBrix)	162	2.88–5.63	4.30	0.45	0.27	0.57	0.30	0.57	1.63
Nitrate content (mg kg^{-1})	161	23.30–1077.71	330.31	259.69	216.74	0.25	226.76	0.25	1.36

[a] *Number of samples.*
[b] *Mean of the calibration set.*
[c] *Standard deviation of the calibration set.*
[d] *Standard error of calibration.*
[e] *Coefficient of determination of calibration.*
[f] *Standard error of cross validation.*
[g] *Coefficient of determination of cross validation.*
[h] *Residual predictive deviation for cross validation.*

Table 3.3.5. Calibration statistics for predicting quality and safety parameters using the instruments Matrix-F (dynamic mode and 1 spectrum per fruit) and MicroNIR™ 1700 (static mode and 1 spectrum per fruit). Calibration set C2.

Parameter	Instrument	N[a]	SEC[b]	R^2_c[c]	SECV[d]	R^2_{cv}[e]	RPD_{cv}[f]	F	$F_{critical}$
Dry matter (% fw)	Matrix F	160	0.32	0.62	0.41	0.62	1.98	1.31	1.299*
	MicroNIR™ 1700	161	0.44	0.58	0.47	0.58	1.72		
Soluble solid content (ºBrix)	Matrix-F	162	0.27	0.57	0.30	0.57	1.63	1.07	1.296
	MicroNIR™ 1700	164	0.28	0.57	0.31	0.57	1.58		
Nitrate content (mg kg^{-1})	Matrix-F	161	216.74	0.25	226.76	0.25	1.36	1.02	1.299
	MicroNIR™ 1700	159	221.42	0.23	228.64	0.23	1.35		

** Significant differences (P < 0.05).*

[a] *Number of samples.*
[b] *Standard error of calibration.*
[c] *Coefficient of determination of calibration.*
[d] *Standard error of cross validation.*
[e] *Coefficient of determination of cross validation.*
[f] *Residual predictive deviation for cross validation.*

Figure 3.3.1. First derivative spectra of summer squashes prior to removing the noise using the MicroNIR™ 1700 (a) and Matrix-F (b).

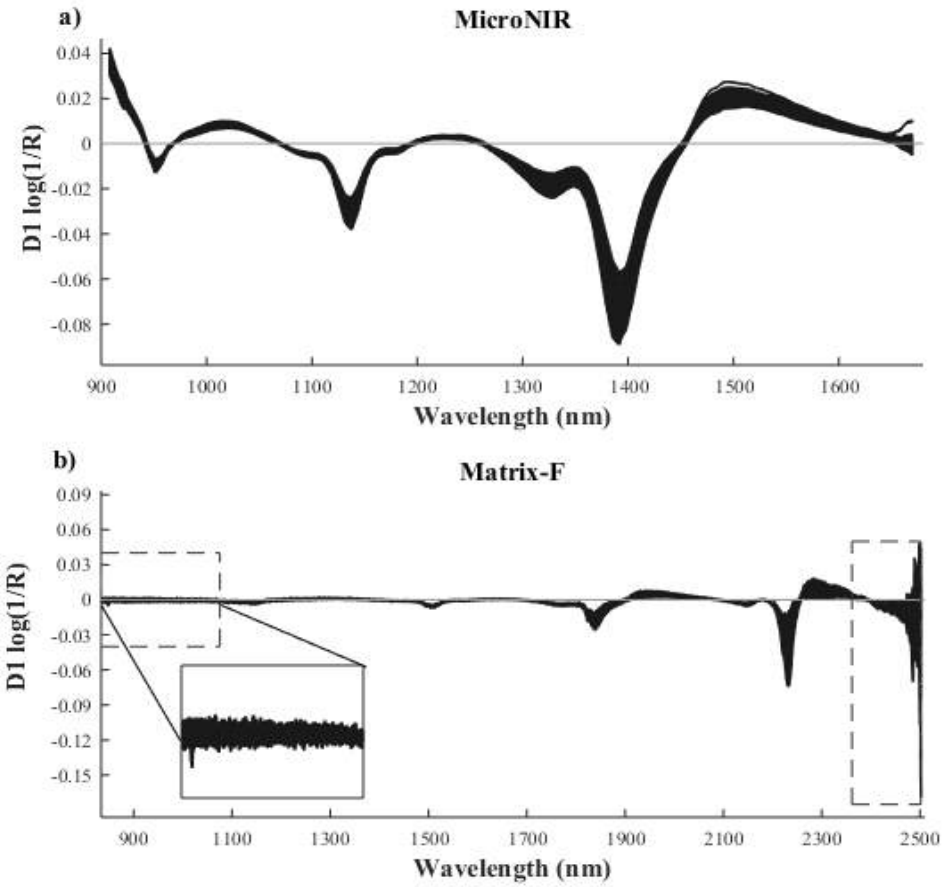


Figure 3.3.2. Scores plot for the first (PC1) and second (PC2) principal components for summer squashes using the Matrix-F instrument.

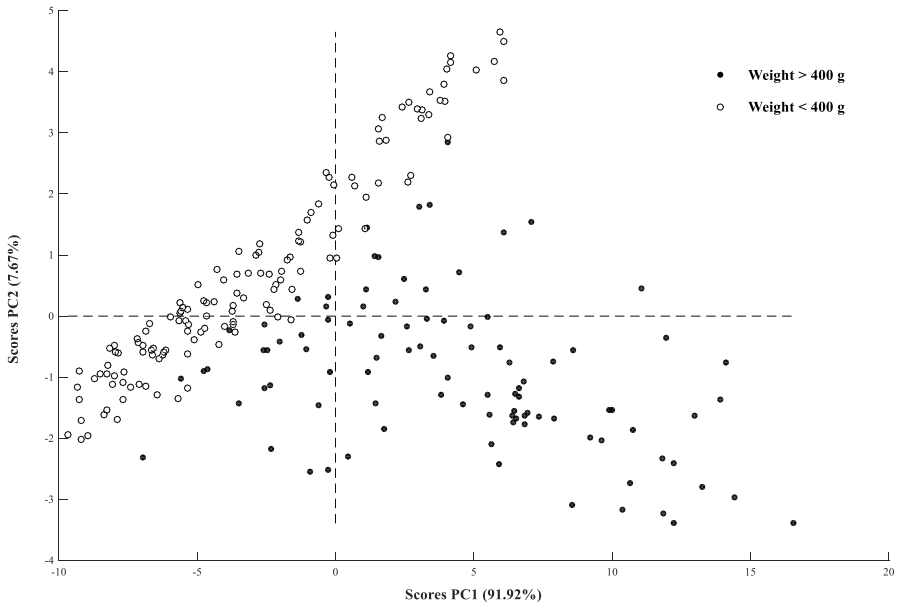
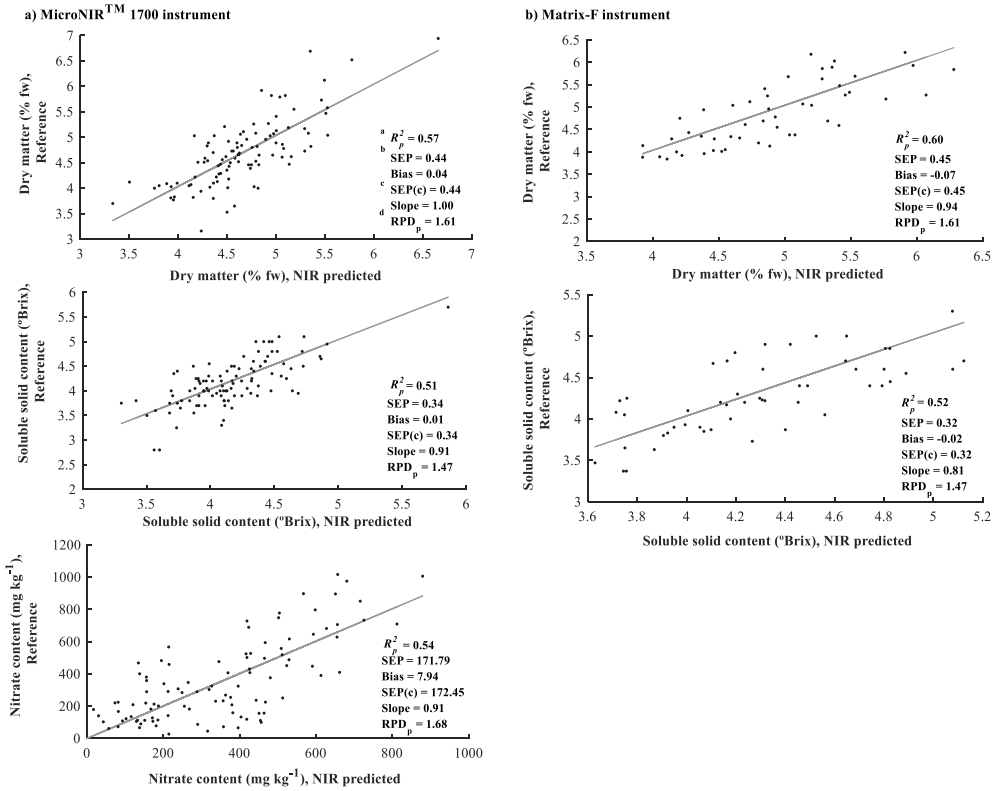


Figure 3.3.3. Reference and NIR predicted values for quality and safety parameters using the MicroNIR™ 1700 (a) and Matrix-F (b) instruments.



[a] *Coefficient of determination of prediction.*
[b] *Standard error of prediction.*
[c] *Standard error of prediction bias-corrected.*
[d] *Residual predictive deviation for prediction.*

Chapter 3.4.

Pre-harvest screening on-vine of spinach quality and safety using NIRS technology

Dolores Pérez-Marín[a], **Irina Torres[b],** José-Antonio Entrenas[b], Miguel Vega[b], María-Teresa Sánchez[b]

[a] *Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*
[b] *Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 207, 242-250 (2019)

Abstract

The study sought to perform a non-destructive and *in-situ* quality evaluation of spinach plants using near infrared (NIR) spectroscopy in order to establish its suitability for different uses once harvested. Modified partial least square (MPLS) regression models using NIR spectra of intact spinach leaves were developed for nitrate, ascorbic acid and soluble solid contents. The residual predictive deviation (RPD) values were 1.29, 1.21 and 2.54 for nitrate, ascorbic acid and soluble solid contents, respectively. Later, this predictive capacity increased for nitrate content (RPDcv = 1.63) when new models were developed, taking into account the influence on the robustness of the model exercised by the simultaneity between the NIR and laboratory analyses. Subsequently, using partial least squares discriminant analysis (PLS-DA), the ability of NIRS technology to classify spinach as a function of nitrate content was tested. PLS-DA yielded percentages of correctly classified samples ranging from 73.08-76.92% for the class 'spinach able to be used fresh' to 85.71-73.08% for the class 'preserved, deep-frozen or frozen spinach, both for unbalanced and balanced models respectively, based on N-H signal associated with proteins. Overall, the data supports the capability of NIR spectroscopy to establish the final destination of the production of spinach analysed on the plant, as a screening tool for important safety and quality parameters.

Keywords: Spinach; Safety; Quality; Portable NIRS; In-situ analysis

3.4.1. Introduction

Spinach (*Spinacea oleracea* L.) is a green, leafy vegetable, with a shiny, uniform appearance. When sold, either to industry (for preserving, freezing or deep-freezing) or as a fresh product on the market, it is essential for the leaves to look freshly-picked and tender.

Spinach has an extremely high water content (around 92%) and is very low in carbohydrates and fats. It is also a rich source of vegetable protein, with, like other vegetables, high fibre content [1].

Spinach is also noted for its relatively high content of bioactive substances, including vitamin C (ascorbic acid), which is a powerful antioxidant that participates in the scavenging of the reactive oxygen species, regenerating tocopherols from their radical forms [2, 3]. However, one drawback is that they accumulate substances which are harmful for the human organism, such as nitrates [4].

Over recent years, consumers have become increasingly aware of the presence of nitrates in foods, among which are vegetables, since nitrates are a serious threat to human health, due to the conversion of nitrate to nitrite, which may produce methemoglobin due to the oxidation of Fe+2 in haemoglobin [5]. The impaired capacity of methemoglobin to deliver oxygen to tissues may lead to severe toxic effects and may even prove fatal where methemoglobin accounts for over 70% of total haemoglobin, something which affects infants and very young children almost exclusively [6]. Similarly, a number of studies have highlighted a possible link between nitrate exposure and childhood type 1 insulin-dependent diabetes mellitus [7]. Furthermore, nitrite may react with secondary amines (HNR_2), which occur in many foods, to form nitrosamines. These substances are highly carcinogenic [8].

However, consumers are also well aware that eating vegetables with a high content of antioxidants, such as ascorbic acid, is beneficial for their health. The World Health Organization [9] showed that iron deficiency anemia is one of the most common nutritional disorders which has a major effect on both health and the economy. Main cause of anemia is not only low iron intake but also poor iron absorption [10]. This global health problem can be addressed by improving the dietary iron bioavailability, which can be altered by various components present in the food which can either enhance or inhibit iron absorption. Thus, when iron is present along with ascorbic acid, the absorption of iron has been shown to increase even in the presence of inhibitors [11].

In the case of spinach and in response to this growing public concern about nitrates, the European Union passed Commission Regulation (EC) No 1258/2011 of 2 December 2011 setting maximum levels for nitrates in this leafy vegetable [12]. Thus, the maximum level for nitrates was set for preserved, deep-frozen or frozen spinach at 2,000 mg NO_3/kg and for fresh spinach at 3,500 mg NO_3/kg.

All this has prompted greater attention to spinach safety and quality concerns. Nitrate accumulation, ascorbic and soluble solid contents in spinach depend not only on genotypic characteristics, but also on a number of other factors, including cultural practices, harvesting date and postharvest handling practices [13, 14]. As a result, producers are increasingly anxious to provide consumers with assurances regarding the safety, quality and provenance of this product.

The nitrate content in spinach when harvested is in fact key to determining the final destination of the harvested product. The ascorbic acid content is equally important because of the close relationship between this acid and the bioavailability of iron, and also the soluble solid content, which in turn is linked to the vegetable's quality and shelf-life.

NIR spectroscopy in conjunction with the application of multivariate analysis strategies is an appropriate non-destructive technology for the study of chemical constituents of vegetables at field level. This technology represents a marked change from the conventional analytical methods, because a single spectrum allows the simultaneous characterization of different chemical properties, in a matter of seconds and without sample preparation, thus allowing real-time decision making. In the field, it has become easier to use this technology by the development in recent years of compact, portable hand-held instruments which make it possible to measure the quality and safety parameters of vegetables directly on the plant, thus allowing the product to be instantly analysed.

Several authors have shown the feasibility of using NIRS technology for the non-destructive measurement of nitrate content in various fruits and vegetables, including Japanese radishes [15], the leaf stalk of Qing gin cai [16], pineapple [17] and summer squash [18]. In spinach, Xue and Yang [19] and Itoh et al. [20], are the only ones to have carried out this analysis, although these authors used NIRS instruments with optical performance and wavelength ranges different to those used here.

As regards ascorbic acid content, no references have been found where this parameter has been measured in spinach using NIRS. However, some authors [21-25] have shown how NIRS technology can be used to measure ascorbic acid in apples, zucchini, oranges, potatoes and peppers.

In the case of soluble solid content (SSC), no references have been found for measuring this parameter with NIRS in spinach, although several review works indicate that NIRS technology is a viable means of measuring this parameter in fruit and vegetables [26-28].

Taking into account the possibilities of using handheld NIRS sensors to optimize harvesting times and enable the staggered harvesting of spinach for

greater quality and safety, thus allowing certain harvested spinach to be used either in the production of baby foods, industrially processed as preserved, deep-frozen or frozen spinach or for fresh consumption, this study sought to assess the feasibility of using NIR spectroscopy~~,~~ to characterize the variations in internal safety and quality—particularly nitrate, ascorbic and soluble solid contents—in intact spinach during on-vine ripening using a low-cost, miniaturized, handheld, near-infrared device based on the MEMS system.

3.4.2. Material and methods

3.4.2.1. Sampling

A total of 128 samples of spinach plants (*Spinacia oleracea* L, cv. 'Solomon' (62 samples), 'Novico' (13 samples), 'Meerkat' (10 samples), and 'Gorilla' (43 samples), grown in an open-air plantation in the province of Córdoba (Spain), were harvested between January and March 2017.

3.4.2.2. Reference data

Nitrate content (mg NO_3/kg) was measured using an RQFlex reflectometer (Merck, Darmstadt, Germany) [29]. The reflectometer which measures the colour intensity of Reflectoquant ® test strips (Merck, Darmstadt, Germany) is based on a colorimetric method. For NO_3^- analysis, the spinach leaves were cut into very small pieces and liquified. Next, 5 ml of spinach juice was mixed in a blender with different quantities of deionised water, depending on NO_3^- concentrations. After that, the solution was filtered using a coffee filter and left to settle for 5 min. Subsequently, a test strip was dipped in the supernatant for 2 s, and then the colour was allowed to develop for 1 min. The test strip was then inserted into the reflectometer and the amount of light reflected from the test strip was measured and converted to a concentration by a standard calibration previously introduced into the equipment using a bar-coded plastic strip. The dilution factor was also taken into consideration.

Ascorbic acid content (mg/100g) was also measured using a reflection photometer (RQflex 10, Merck, Merck, Darmstadt, Germany) [30]. For the ascorbic acid analysis, the analysis procedures were the same as those for nitrate content with the exception that samples containing more than 450 mg/L of ascorbic acid were diluted with oxalic acid solution 1%. The dilution factor was also taken into consideration.

SSC (ºBrix) was measured as the refractometer reading for the spinach juice, using a temperature-compensated digital Abbé-type refractometer (model B, Zeiss, Oberkochen, Würt, Germany).

3.4.2.3. Spectral data acquisition

The NIR spectra of the spinach leaves were collected in reflectance mode (log 1/R) using a handheld MEMS instrument (Phazir 2400, Polychromix, Inc., Wilmington, MA, USA). The Phazir 2400 is a compact, low-cost near-infrared analyser that incorporates all the essential components to deliver on-vine applications. The spectrophotometer scans at a non-constant interval of approximately 8 nm, across the NIR wavelength range of 1600 to 2400 nm, with a window area of only around 4 mm^2. The sensor integration time was 600 ms. The MEMS device is equipped with quartz protection to prevent the accumulation of dirt. Instrument performance was checked every 10 min, following the diagnostic protocols provided by the manufacturer, and white reference measurement was carried out using Spectralon as reference.

Four spectral measurements were made on each spinach leaf (distal and proximal, on both sides (right and left) of the leaf blade relative to the main vein, on the adaxial side). Because between 4-10 leaves were used for the chemical analyses of the parameters of SSC, ascorbic acid and nitrate content when analysing each plant, to obtain a mean spectrum for each spinach sample, a mean spectrum was first obtained from the four spectra for each leaf, and then a mean spectrum was obtained from the four to ten mean spectra for each sample.

3.4.2.4. Spectral repeatability

The spectral repeatability of intact spinach leaves was evaluated using the Root Mean Squared (RMS) statistic. The RMS statistic is defined as the averaged root mean square of differences between the different subsamples scanned at *n* wavelengths [31, 32]. This statistic indicates the similarity between different spectra of a single sample.

For each instrument and sample presentation form, the RMS for an individual subsample (*j)* of the sample (*k*) can be calculated using the using the following expression:

$$RMS_{j,k} = \sqrt{\frac{\sum_{i=1}^{n} D_{ij}^2}{n}};\ D_{ij} = y_{ij} - \bar{y}_i$$

where y_{ij} is log (1/R) at wavelength *i* for subsample *j*, and $\bar{y}_i$ is log (1/R) at wavelength *i* for the average spectrum of a sample *k*; *n* is the number of data points collected by the instrument (here, 100 data points for the MEMS instrument). The RMS value obtained in each case was multiplied by 10^6 to facilitate value management and processing.

To determine a cut-off value (RMS_{cutoff}) of each sample presentation form, the mean RMS was calculated along with standard deviation (STD) per sample according to the formulae provided by Martínez et al. [33].

$$STD_k = \sqrt{\sum_{j=1}^{N} (RMS_j)^2/(N-1)}$$

where N are the number of sub-samples.

A STD limit can be calculated for comparing the RMS values of subsamples, following the formula provided by Rosales [34], who demonstrated that the estimated value of the error variance, σ^2 for log (1/ R) $y_{i,j}$ is the corresponding to one-way ANOVA:

$$\sigma^2 = \frac{1}{n\,(N-1)} \sum_{i=1}^{n} \sum_{j=1}^{N} \left(y_{i,j} - \bar{y}_i\right)^2$$

This expression corresponds to the STD^2. The sum of squares for error (SSE) can be thus be expressed as:

$$SSE = n\,(N\text{-}1)\,STD^2$$

which approximately follows a χ^2 distribution,

$$\frac{n\,(N-1)\,(STD^2)}{\sigma_0^2} \sim \chi^2_{[n(N-1)]}$$

with σ_0^2 the parametric value of the error variance. For infinite degrees of freedom (> 100), χ^2 tends to a normal distribution. An STD limit can then be calculated for comparing the RMS values of subsamples, following the formula given by Rosales (1993).

$$STD_{limit} = 1.036 \sqrt{\sum_{k=1}^{k=m} STD_k^2/m} = 1.036\sqrt{\overline{STD^2}}$$

where STD is the standard deviation per sample and m is the number of samples. The value 1.036 corresponds to a probability level of 85%.

The STD_{limit} value was used to establish the RMS_{cutoff} for each product and analysis mode. Hence, the different sources of variation which might cause irregular spectra were controlled, since any spectra in a sample that were above this limit were eliminated, and recalculations were performed until all the values were below the RMS_{cutoff}. Then, the mean spectrum of each leaf was calculated.

To evaluate spectral repeatability, two alternatives were available: first, analysing ten leaves and taking two spectra in each of them, at the same point, and second, analysing twenty leaves and taking four spectra in each, at different points of the leaf.

3.4.2.5. Data analysis: definition of calibration and validation sets

Principal Component Analysis (PCA) was performed on a set of N = 128 samples in order to decompose and compress the data matrix. After PCA, the center of the spectral population was determined in order to detect outlier samples. The Mahalanobis distance (GH) was calculated between each sample and the center; samples with a GH value greater than 3 were considered outliers [31]. As spectral pre-treatments, the Standard Normal Variate (SNV) plus Detrending (DT) procedure [35] was used to remove the multiplicative interferences of scatter, and first derivative mathematical treatment (Norris derivative) was performed (1,5,5,1), where the first digit is the order of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing and the fourth is the second smoothing [36, 37].

Once spectral outliers had been removed, a set consisting of 124 samples was used to develop the calibration models. The set was divided into two: a training set containing about 75% of the samples (N = 93 samples) and a test set containing the remaining 25% (N = 31 samples). These samples were selected following the method outlined by Shenk and Westerhaus [38] using the CENTER algorithm included in the WinISI software package to calculate the Global Mahalanobis distance (GH). The samples were ordered on the basis of the Mahalanobis distance to the center of the population, and three of every four were selected to be part of the calibration set.

3.4.2.6. Chemometric tools

The data were subjected to chemometric treatment using the WinISI software package version 1.50 [37].

NIR calibration models for the prediction of quality parameters (nitrate content, ascorbic content, and SSC) in intact spinach plants were constructed using modified partial least squares (MPLS) regression [31], with subsequent cross-validation. The calibration set was partitioned into 6 groups; each group

was then validated using a calibration derived from the other samples; finally, validation errors were combined to obtain a standard error of cross-validation (SECV).

For each analytical parameter, different mathematical treatments were evaluated. For scatter correction, the standard normal variate (SNV) and detrending (DT) methods were tested [35]. Additionally, a total of four derivative mathematical treatments were tested: 1,5,5,1; 2,5,5,1; 1,10,5,1 and 2,10,5,1.

The statistics used to select the best equations were: the coefficient of determination for calibration (r^2_c), the standard error of calibration (SEC), the coefficient of determination for cross calibration (r^2_{cv}), and the standard error of cross validation (SECV). Furthermore, the Residual Predictive Deviation (RPD) for cross-validation was calculated as the ratio of the standard deviation (SD) of the reference data to the SECV. This statistic enables SECV to be standardized, facilitating the comparison of results obtained with sets of different means [39].

The best-fitting equations obtained for the calibration set, as selected by statistical criteria, were subsequently subjected to external validation following the protocol outlined by Windham et al. [40].

After analysing the results obtained, and in order to test the influence of the simultaneity in time between the NIRS spectrum and the wet-chemistry analysis on the robustness of the model obtained for the prediction of nitrate content in intact spinach, new predictive models were designed for this parameter, dividing the initial total of 128 samples into 2 groups which represent two different analysis strategies.

- Strategy I. Group of samples 1 to 47. On the day the field samples arrived, the corresponding NIR spectra were taken. However, the reference

analyses were carried out 24 hours later, and the samples were stored in refrigeration conditions during that time (4ºC, RH: 85%).

- Strategy II. Group of samples 48 to 128. The NIRS spectra were taken and the wet analysis performed 24 hours after the product arrived in the laboratory, and the samples were stored in refrigeration conditions until analysis (4ºC, RH: 85%).

The same signal pre-treatments and spectral region described earlier were used here for the development of the new quantitative models.

3.4.2.7. NIRS classification models

The design of models to classify spinach by its nitrate content, in order to evaluate the viability of using NIRS technology to determine the final destination of the harvested spinach (fresh consumption or preserved, deep-frozen or frozen product), for those samples where the NIR and reference analyses were performed at the same time (Strategy II), comprised two classification groups: 1) spinach which contained 0-2,000 mg/kg NO_3^- (N = 45 samples) for preserved, deep-frozen or frozen spinach; 2) spinach with a nitrate content between 2,000-3,500 mg/kg NO_3^- (N = 36 samples) for fresh spinach.

Next, the structure and spectral variability of the sample population was studied in order to select the samples that would constitute the learning group. To do this, we used the CENTER algorithm, which is included in the WinISI version 1.50 software. This algorithm was applied separately to each of the two training groups. The mathematical treatments SNV (Standard Normal Variate) and DT (Detrend) were applied to correct any scattered radiation phenomena, together with the 1,5,5,1 derivation treatment [31].

After having ordered the sample set by spectral distances (from smallest to greatest distance from the center), a structured selection of the external validation set (10 samples for each classification group), solely on the basis of spectral data [38].

Discriminant models were constructed to classify spinach by its NO_3^- content, using PLS-DA for supervised classification. Specifically, the PLS2 algorithm was applied, using the 'Discriminant Equations' option in the WINISI version 1.50 software package [37].

All the models were constructed using six cross-validation groups. The same signal pre-treatments and spectral regions described earlier for quantitative analysis were used for qualitative model development.

The precision of the models obtained was evaluated using the percentage of correctly-classified samples, both for the global model and for each class.

The difficulty involved in obtaining balanced learning groups in terms of the number of samples per class or classification category led us to assess the influence of this factor on the development of discriminant models. In this way, the results obtained were contrasted with balanced and unbalanced classification models in terms of the number of samples per class.

The samples for the balanced groups were selected using the algorithm SELECT included in the WinISI II software package version 1.50 (Infrasoft International, Port Matilda, PA), which calculates spectral distances (Mahalanobis H), in order to detect samples whose spectrum is very similar to that of others in the population [38]. This algorithm enables spectral selection of a number of samples representative of the population as a whole, by calculating the 'NH' distance (Mahalanobis neighbour distance) between two spectra. An 'NH' of less than 0.6 implies that two spectra are too similar to each other ('neighbour'). After application of this algorithm, 26 samples of the category 'nitrate content between 0-2,000 mg/kg' were selected, thus making the number of samples of the two classes equal and allowing the classification models to be redesigned.

Next, the best classification models for each of the established types (unbalanced and balanced models) were selected and externally validated. In this case, an external validation procedure was carried out to measure the predictive capacity of the model using a sample group different to that used in the training of the model. In both models (unbalanced and balanced) 20 samples (10 per category) were selected in a structured way [38].

3.4.3. Results and discussion

3.4.3.1. Spectral repeatability

The collection of high quality spectra is crucial for the characterization of spinach plants by quality and safety characteristics and to assess its possible industrial use, as well as to construct discriminant classification models for the product depending on its possible use in the processing industry of fresh or processed vegetables (in this case, preserved, deep-frozen or frozen spinach). The RMS cut-off was calculated for the instrument MEMS used as shown in section 3.4.2.4.

For the first alternative tested (analysing ten leaves and taking two spectra in each of them, at the same point) the mean STD was 48,292 µlog (1/R), representing an RMS cut-off of 64,661 µlog (1/R). For the second alternative (analysing twenty leaves and taking four spectra in each, at different points of the leaf) the mean STD and the RMS cut-off were 118,693 µlog (1/R) and 128,437 µlog (1/R), respectively. As can be seen, the result obtained for $STD_{límit}$ in the test in which 10 samples (leaves) were used was lower than that obtained in the test which used 20 samples. This was to be expected, since the former corresponds to the analysis of 10 samples whose two subsamples were taken at the same point, while, in the latter case, 20 samples were analysed by taking the four subsamples at different points in the leaf. The repeatability results obtained therefore indicate that, in the second mode, the analysis reflects the heterogeneity of the leaf, although the reduced window of analysis presented by the spectrophotometer used must also be taken into account.

It was therefore decided to choose the second mode of analysis as the most suitable, that is, to collect 4 spectra per leaf. In this way, a more representative measurement of the analysed product was obtained. Once the RMS value did not exceed the value of the $STD_{límit}$, these spectra were then averaged in order to carry out both the quantitative and qualitative predictive models using the average spectrum of each sample.

The calculation of the RMS statistic is extremely important because it aims to ensure high spectral repeatability, which is essential to obtain quality spectral data, and therefore, constitutes an essential step in obtaining robust equations.

No values for this statistic have been found in the scientific literature for spinach analysed on the vine, although the RMS statistic is extremely useful to obtain representative spectral libraries of this vegetable, when analysed on the plant. In fact, this is the first research work to deal with measuring spectral repeatability in leafy vegetables.

3.4.3.2. Spectral features

Mean and standard deviation log (1/R) spectra for intact spinach leaves, captured by the instrument Phazir 2400, together with the most relevant absorption bands, are shown in Figure 3.4.1.

In the NIR region between 1600-2400 nm, the highest absorption peak is at the 1920 nm wavelength, which corresponds to water. This was to be expected, since spinach is made up of 90% water [41]. Osborne et al. [42] showed that the peak in the wavelength 1780 nm was directly related to the first overtone of sugars.

3.4.3.3. Descriptive data for NIR calibration and validation

Values for number of samples, range, mean, standard deviation (SD) and coefficient of variation (CV) for each of the parameters analysed using the

calibration and validation sets after application of the CENTER algorithm and removal of any spectral outliers are shown in Table 3.4.1.

The nitrate content, ascorbic acid, and SSC reference values were all fairly normally distributed around the mean values (1765.24 mg/kg, 298.97 mg/100 g and 8.84 ºBrix, respectively) with standard deviations (SD) of 1082.34 mg/kg, 66.05 mg/100g and 1.90 ºBrix for each parameter.

The ranges of nitrate content, ascorbic acid and SSC of the calibration set are 109.50-5177.00 mg/kg, 156.92-479.23 mg/100g and 5.60-14.25 ºBrix, respectively; that of the validation set are 144.00-3520.00 mg/kg, 191.83-453.85 mg/100g and 6.25-13.95 ºBrix. Since the calibration and validation sets displayed similar values for mean, range and standard deviation for all the parameters studied, a structured selection using only spectral information treatment algorithms such as CENTER proved adequate. Furthermore, the ranges of the validation set lay within those of the calibration set.

All the parameters studied in the calibration set covered a wide range of values. This was truest of nitrate content with a CV of 63.31%, while the CV of ascorbic acid and SSC were practically the same (22.09% and 21.49%, respectively). Pérez-Marín et al. [43] have highlighted the importance of the sample set and of sample distribution within the calibration set, noting that sample sets for calibration should ideally ensure an uniform distribution across the range of the study parameter in question.

For the validation group, the coefficient of variation values were set at 60.62% for nitrates, 21.10% for ascorbic acid and 22.66% for SSC.

3.4.3.4. Prediction of safety and quality parameters in intact spinach using MPLS regression

The best calibration models obtained for predicting one safety parameter (nitrate content) and quality parameters (ascorbic acid and SSC) in

spinach using different various mathematical pre-treatments are shown in Table 3.4.2. Statistical criteria were used to select the best model for each study parameter.

The models developed for nitrate content have a predictive capacity (r^2_{cv} = 0.41; RDP_{cv} = 1.29) which allows the samples to be classified under high and low values of this parameter [32, 39].

It should be noted that Xue and Yang [19] studied nitrate content in spinach (n = 58 samples), using an ASD Fieldspec FR spectroradiometer and obtained results (r^2_p = 0.88) which were higher than those obtained here, although the instrument's optical characteristics and range are significantly different from those of the Phazir 2400.

Itoh et al. [20] also measured the nitrate content in spinach leaves (n = 48 samples), using the FANTEC NIR Gun working on transmittance in a spectral range of 600-1100 nm. The authors reported values of RPD_p = 2.14 with the PCR regression and RPD_p = 2.17, using PLS regression, which were higher than those obtained in this research study. However, the size and characteristics of the sample group, the means of measurement, the window size (1 cm) and the spectral range of the instrument used, all differed from those used in this study.

As regards ascorbic content, the models designed allow the samples to be classified under high and low values for this parameter (r^2_{cv} = 0.33; RDP_{cv} = 1.21) [32, 39], although the results are limited for routine use. It must be noted that it is especially difficult to measure this parameter in vegetable products - not only with portable instruments, but also with high performance NIRS laboratory instruments.

Kramchote et al. [44] measured ascorbic acid content in cabbage and obtained predictive capacity models (RPD_P = 1.26) similar to that obtained here

using a spectrophotometer Handy Lamda II (Spectra Co., Ltd., Tokyo, Japan), in reflectance mode in the 310 and 1100 nm spectral range.

When predicting antioxidant content (beta-carotene and ascorbate) in freeze-dried leaves of *Populus* spp., Fernández-Martínez et al. [45] obtained models with an average predictive capacity (RPD_{cv} = 2.10) for beta-carotene, while for ascorbate, the models were capable of discriminating between high and low values (RPD_{cv} = 1.56) [26]. In both studies, reflectance was carried out using high performance Foss NIRS Systems 6500 laboratory equipment, in the 400-2500 nm range.

For SSC – which is a crucial parameter when choosing the optimum time for harvesting and for measuring the shelf-life of spinach, the model developed has a predictive capacity (r^2_{cv} = 0.85; RPD_{cv} = 2.54) which can be considered as good [32, 39].

No reports have been found in the literature regarding the measurement of SSC in spinach using NIR spectroscopy. However, Kramchote et al. [44], in the case of cabbage, obtained models of lower predictive capacity (RPD_p = 2.05 and 1.95) using a spectrophotometer (Handy Lamda II, Spectra Co., Ltd., Japan), in interactance and reflectance mode, respectively, in the spectral range 310-1100 nm.

The regression coefficients for the best predictive models for nitrate, ascorbic acid and soluble solids contents are illustrated in Figure 3.4.2. The figure shows that the area of the spectra around 1650–1850 nm, which is correspond to the first overtone of C–H stretching bonds [42]. The best models to predict the three parameters tested reflected variations in the wavelengths range 2000–2190 to N–H and O–H stretching modes besides C=O vibration bands [42]. Absorbance region such as 2200-2280 nm and at 2320 nm could be attributed to C–H stretch and CH2 deformation [41].

Validations of the best calibration models obtained were performed using a set comprising 31 samples (Figure 3.4.3). Models constructed for predicting SSC in intact spinach using the MEMS instrument met the validation requirements in terms of the coefficient of determination for prediction r^2_p (r^2_p > 0.6), while the standard error of prediction corrected for bias or SEP(c), bias and slope were within the confidence limits: the equation thus guarantees an accurate prediction, and can be applied routinely [40].

For nitrate content prediction, the r^2_p value does not comply with the protocol while the values shown by the other statistics lie within the confidence limits, thus complying with this validation protocol [40].

In the case of the ascorbic acid content, the r^2_p values and the slope do not comply with this protocol, while SEP (c) and the bias were below confidence limits.

3.4.3.5. Contrasting the suitability of performing reference analyses immediately after NIR measurements

Given the rough predictive capacity of the model designed to measure nitrate content in spinach, and since the non-destructive prediction of this parameter is key in determining the final destination of the harvested product, it was decided to evaluate whether, for spinach leaves, carrying out chemical analyses at times other than when the NIRS spectra are taken significantly affects the predictive capacity of the models. Several authors have pointed out the importance of performing the reference analyses immediately after the NIR analysis, especially in the case of extremely perishable vegetable products [46].

Table 3.4.3 shows the best models obtained for predicting nitrates using two analysis strategies: Strategy I: NIRS analysis, and chemical measurement of nitrate content 24 hours later; Strategy II: NIRS analysis and wet reference analysis immediately after.

Tables 3.4.2 and 3.4.3 clearly show the importance of carrying out the NIR and wet process analyses consecutively, and of using the same methodology or systematic analysis for all the samples in the study.

In this way, the joint predictive model (Table 3.4.2) shows values of r^2_{cv} = 0.41 and SECV = 836.26 mg/kg while the models developed from the two strategies considered (Table 3.4.3) had values of r^2_{cv} = 0.53 and 0.63, and SECV = 711.63 and 670.81 mg/kg, for strategies I and II, respectively.

It is important to point out that for both strategies, the SECV value was lower. In Strategy II, the SECV value fell by 19.78%, while for Strategy I, it decreased by 14.90%. Likewise, the predictive capacity of the models increased, and the models allowed to distinguish between high, medium and low values of nitrate content for both strategies [32, 39].

The results obtained indicate that growers and the agrifood industry could use NIRS technology as a screening technique, permitting a large number of plants to be tested in the field or when they are delivered to the industrial plant, providing results for this parameter and thus enabling growers to take real-time decisions as to the final destination of the harvested product.

These results also confirm the importance of performing these analyses together, in products as perishable as spinach [47] and the suitability of using the same analysis methodology throughout the trial.

3.4.3.6. Discriminant analysis

Results for the best classification models obtained, using PLS2-DA, for predicting the industrial use of spinach depending on its content in nitrates, are shown in Table 3.4.4.

The best discriminant models were obtained with D_2 log(1/R) together with SNV + DT for scatter correction (balanced and unbalanced sets).

The unbalanced model correctly classified 80.33% of the samples (85.71% in the 0-2,000 mg/kg category and 73.08% in the 2,000-3,500 mg/kg category), while the balanced model correctly classified 75.00% (73.08% in the 0-2,000 mg/kg category and 76.92% in the 2,000-3,500 mg/kg category). For validation, the percentage of correctly classified samples obtained using the unbalanced model was 70% (80% in the 0-2,000 mg/kg category and 60% in the 2,000-3,500 mg/kg category) and for the balanced model, the percentage was 70.00% (90% in the 0-2,000 mg/kg category and 50% in the 2,000-3,500 mg/kg category).

As can be seen in Table 3.4.4, the percentage of samples correctly classified by the unbalanced model was slightly higher than by the balanced model, which reflects the low sensitivity of the PLS2 algorithm to the difference between the number of samples of the classes to be discriminated [48].

A detailed study of the unbalanced model reveals that, of the 12 poorly classified samples, 5 had a nitrate content (mg/kg) of between 2,000 ± 2 * SEL (SEL = 131.6 mg/kg), a range that can be considered difficult to discriminate as it would fall within the error of the reference method, so that the error of classification may be put down to a typical laboratory error value. The remaining 7 samples did not show nitrate content within this range, and therefore their incorrect classification can be attributed to errors in the model or to their poor representation in the training set, given the limited number of samples available, which only allowed to perform a feasibility study of the potential of technology in this area. It is also important to mention that this constituent is found in ppm in spinach leaves, which makes it difficult to measure by NIRS in whole plants. However, the results obtained are promising and allow us to continue consolidating the application of NIRS technology as a screening technique in the spinach handling and processing industry, permitting, in a non-destructive way and in a matter of seconds, to assess the possible industrial destination of the spinach leaves.

Of the 13 poorly classified samples in the balanced model, 3 had a nitrate content (mg/kg) within the range of 2,000 ± 2 * SEL (SEL = 131.6 mg/kg), and any classification errors can again be attributed to the SEL value obtained.

Peaks at 1780 nm, 1860 nm and 2040 nm appeared to have more weight in the classification of intact spinach on-vine by nitrate content (Figure 3.4.4). This indicates that the discrimination of spinach by nitrate content in the NIR region of the spectrum is related to lipids, proteins and N-H combinations [41].

As regards the external validation of the classification models, a percentage of correctly classified samples of 70% was obtained for both the unbalanced and balanced models. For the first, 10 samples from the 0-2,000 mg/kg category were used, of which 8 were correctly classified, and 10 from the 2,000-3,500 mg/kg category, of which 6 were correctly classified, while of the 4 poorly classified samples, 1 had a nitrate content (mg/kg) within the range 2,000 - 2 * SEL. Similarly, in the balanced model, 9 out of 10 samples in the 0-2,000 mg/kg category were correctly classified, and 5 from the 2,000-3,500 mg/kg category.

3.4.4. Conclusions

Near infrared spectroscopy is clearly an advantageous technique for the rapid screening of quality and safety according to the SSC and nitrate levels, although further research is needed to make it robust for predicting these parameters. It has also been demonstrated that the NIRS and the laboratory analysis should be performed together.

The results obtained from the classification models of spinach leaves according to their nitrate content, which determines their possible industrial destination, also confirm the feasibility of using NIRS technology both in the

field and in the stages of selection and classification of spinach carried out during industrial processing for classification according to the quality and safety characteristics of this vegetable.

Acknowledgements

This research was under the Research Project 'Quality determination of spinach grown in Santaella (Córdoba)', funded by Gelagri Ibérica, S.L. The authors are grateful to Mrs. María Carmen Fernández of the Animal Production Department for her technical assistance. Furthermore, the authors wish to express their gratitude to the Spanish Ministry of Education, Culture and Sports for the support offered to Irina Torres in the form of the Training Programme for Academic Staff (FPU).

References

[1] G.H. Beis, A.S. Siomos, C.C. Dogras, Spinach composition as affected by leaf age and plant part, Acta Hortic. 57 (2002) 653–658. https://doi.org/10.17660/ActaHortic.2002.579.115.

[2] N. Smirnoff, Ascorbate biosynthesis and function in photo-protection, Phil. Trans. 355 (2000) 1455–1464. https://doi.org/10.1098/rstb.2000.0706.

[3] E. Koh, S. Charoenprasert, A.E. Mitchell, Effect of organic and conventional cropping systems on ascorbic acid, vitamin c, flavonoids, nitrate, and oxalate in 27 varieties of spinach (*Spinacia oleracea* L.), J. Agr. Food Chem. 60 (2012) 3144−3150. https://doi.org/10.1021/jf300051f.

[4] G. Jaworska, Content of nitrates, nitrites, and oxalates in New Zealand spinach, Food Chem. 89 (2005) 235−242. https://doi.org/10.1016/j.foodchem.2004.02.030.

[5] A. Elia, P. Santamaria, F. Serio, Nitrogen nutrition, yield and quality of spinach, J. Sci. Food Agric. 76 (1998) 341–346.

https://doi.org/10.1002/(SICI)1097-0010(199803)76:3<341::AID-JSFA938>3.0.CO;2-4.

[6] P. Santamaria, Nitrate in vegetables: toxicity, content, intake and EC regulation, J. Sci. Food Agric. 86 (2006) 10–17. https://doi.org/10.1002/jsfa.2351.

[7] J.M. Van Maanen, H.J. Albering, T.M. de Kok, S.G. van Breda, D.M. Curfs, I.T. Vermeer, A.W. Ambergen, B.H. Wolffenbuttel, J.C. Kleinjans, H.M. Reeser, Does the risk of childhood diabetes mellitus require revision of the guideline values for nitrate in drinking water? Environ. Health Persp. 108 (2000) 457–461 https://doi.org/10.1289/ehp.00108457.

[8] J.L. Brown, N-Nitrosamines, J. Occup. Med. 14 (1999) 839–848.

[9] World Health Organization, Iron Deficiency Anaemia: Assessment, Prevention and Control. A Guide for Programme Managers, WHO/NHD/01.3, World Health Organization, Geneva, http://apps.who.int/iris/bitstream/10665/66914/1/WHO_NHD_01.3.pdf?ua=1/, 2001 (accessed 3 June 2018).

[10] S. Bhuvaneswari, M. Joshi, A. D'Souza, Quantitative analysis of iron and ascorbic acid contents in locally consumed fruits and vegetables, Int. Res. J. Biol. Sci. 4 (2015) 42–47.

[11] L. Hallberg, M. Brune, L. Rossander, The role of vitamin C in iron absorption, Int. J. Vitam. Nutr. Res. 30 (1989) 103–108.

[12] Official Journal of the European Union (OJEU), Commission Regulation (EU) No 1258/2011 of 2 December 2011 Amending Regulation (EC) No 1881/2006 as regards Maximum Levels for Nitrates in Foodstuffs, OJ L 320/15-17, 3.12.2011, 2011.

[13] S.K. Lee, A.A. Kader, Preharvest and postharvest factors influencing vitamin C content of horticultural crops, Postharvest Biol. Technol. 20 (2000) 207–220. https://doi.org/10.1016/S0925-5214(00)00133-2.

[14] A. Ito, H. Shimizu, H. Nakashima, J. Miyasaka, K. Ohdoi, Effect of different durations of root area chilling on the nutritional quality of spinach, IFAC Proc. Vol. 47 (2014) 4406–4410. https://doi.org/10.3182/20140824-6-ZA-1003.00744.

[15] H. Ito, H. Horie, K. Ippoushi, K. Azuma, Potential of visible-near infrared (VIS–NIR) spectroscopy for non-destructive estimation of nitrate content in Japanese radishes, Acta Hortic. 604 (2003) 549–552. https://doi.org/10.17660/ActaHortic.2003.604.64.

[16] H. Ito, F. Idezawa, Non-destructive determination of nitrate ion in leaf stalk of Qing gin cai using visible (VIS)-near infrared (NIR) spectroscopy, Acta Hortic. 712 (2006) 363–370. https://doi.org/10.17660/Acta Hortic.2006.712.41.

[17] S. Srivichien, A. Terdwongworakul, S. Teerachaichayut, Quantitative prediction of nitrate level in intact pineapple using Vis-NIRS, J. Food Eng. 150 (2015) 29–34. https://doi.org/10.1016/j.jfoodeng.2014.11.004.

[18] M.T. Sánchez, D. Pérez-Marín, I. Torres, B. Gil, A. Garrido-Varo, M.J. De la Haba, Use of NIRS technology for on-vine measurement of nitrate content and other internal quality parameters in intact summer squash for baby food production, Postharvest Biol. Technol. 125 (2017) 122–128. https://doi.org/10.1016/j.postharvbio.2016.11.011.

[19] L.H. Xue, L.Z. Yang, Nondestructive determination of nitrate content in spinach leaves with visible-near infrared high spectra, Spectrosc. Spect. Anal. 29 (2009) 926–930.

[20] H. Itoh, H. Tomita, Y. Uno, N. Shiraishi, Development of method for non-destructive measurement of nitrate concentration in vegetable leaves by near-infrared spectroscopy, IFAC Proc. Vol. 44 (2011) 1773–1778. https://doi.org/10.3182/20110828-6-IT-1002.00738.

[21] A. Pissard, J.A. Fernández-Pierna, V. Baeten, G. Sinnaeve, G. Lognay, A. Mouteau, P. Dupont, A, Rondia, M. Lateur, Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy, J. Sci. Food Agric. 93 (2013) 238–244. https://doi.org/10.1002/jsfa.5779.

[22] M.T. Blanco-Díaz, M. Del Río-Celestino, D. Martínez-Valdivieso, R. Font, Use of visible and near-infrared spectroscopy for predicting antioxidant compounds in summer squash (*Cucurbita pepo* ssp *pepo*),

Food Chem. 164 (2014) 301–308. https://doi.org/10.1016/j.foodchem.2014.05.019

[23] C. Liu, S.X. Yang, L. Deng, Determination of internal qualities of Newhall navel oranges based on NIR spectroscopy using machine learning, J. Food Eng. 161 (2015) 16–23. https://doi.org/10.1016/j.jfoodeng.2015.03.022

[24] R. Tierno, A. López, P. Riga, S. Arazuri, C. Jarén, L. Benedicto, J.I. Ruiz de Galarreta, Phytochemicals determination and classification in purple and red fleshed potato tubers by analytical methods and near infrared spectroscopy, J. Sci. Food Agric. 96 (2016) 1888–1899. https://doi.org/10.1002/jsfa.7294.

[25] E.M. Toledo-Martín, M. García-García, R. Font, J.M. Moreno-Rojas, P. Gómez, M. Salinas-Navarro, M. Del Río-Celestino, Application of visible/near-infrared reflectance spectroscopy for predicting internal and external quality in pepper, J. Sci. Food Agric. 96 (2016) 3114–3125. https://doi.org/10.1002/jsfa.7488.

[26] B.M. Nicolaï, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron, J. Lammertyn, Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review, Postharvest Biol. Technol. 46 (2007) 99–118. https://doi.org/10.1016/j.postharvbio.2007.06.024.

[27] M.T. Sánchez, D. Pérez-Marín, Nondestructive measurement of fruit quality by NIR spectroscopy, in: M. Vázquez, J.A. Ramírez (Eds.), Advances in Post-Harvest Treatments and Fruit Quality And Safety, Nova Science Publishers, Inc., Hauppauge, NY, 2011, pp. 101–163.

[28] L.S. Magwaza, U.L. Opara, H. Nieuwoudt, P.J.R. Cronje, W. Saeys, B. Nicolaï, NIR spectroscopy applications for internal and external quality analysis of citrus fruit–a review, Food Bioprocess Tech. 5 (2012) 425–444. https://doi.org/10.1007/s11947-011-0697-1.

[29] R.B. Thompson, M. Gallardo, M. Joya, C. Segovia, C. Martínez-Gaitán, M.R. Granados, Evaluation of rapid analysis systems for on-farm nitrate analysis in vegetable cropping, Span. J. Agric. Res. 7 (2009) 200–211. https://doi.org/10.5424/sjar/2009071-412.

[30] K. Tsukazawa, Study of simplicity measurement methods of deoxidize type vitamin C and B-carotene content in vegetables, Bull. Saitama Pref. Agr. For. Res. Ctr. 2 (2002) 43–46.

[31] J.S. Shenk, M.O.Westerhaus, Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy, Monograph, NIRSystems, Inc., 12101 Tech Road, Silver Spring, MD 20904, USA, 1995.

[32] J.S. Shenk, M.O.Westerhaus, Calibration the ISI way, in: A.M.C. Davies, P.C. Williams (Eds.), Near Infrared Spectroscopy: The Future Waves, NIR Publications, Chichester, 1996, pp. 198–202.

[33] M.L. Martínez, A. Garrido, E.J. De Pedro, L. Sánchez, Effect of sample heterogeneity on NIR meat analysis: the use of the RMS statistic, J. Near Infrared Spectrosc. 6 (1998) 313–320. https://doi.org/10.1255/jnirs.214.

[34] M. Rosales, Use of the NIR Spectral Information per se for Quality Evaluation of Agricultural Products, Master Thesis, University of Cordoba, Cordoba, 1993.

[35] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777. https://doi.org/10.1366/0003702894202201.

[36] J.S. Shenk, M.O. Westerhaus, Routine Operation, Calibration, Development and Network System Management Manual, NIRSystem, Inc., 12101 Tech Road, Silver Spring, MD 20904, USA, 1995.

[37] ISI, The Complete Software Solution Using a Single Screen for Routine Analysis, Robust Calibrations, and Networking, Manual, FOSS NIRSystems/TECATOR, Infrasoft International, LLC, Sylver Spring MD, USA, 2000.

[38] J.S. Shenk, M.O. Westerhaus, Population structuring of near infrared spectra and modified partial least squares regression, Crop Sci. 31 (1991) 1548–1555.

[39] P.C. Williams, Implementation of near-infrared technology, in: P.C. Williams, K.H. Norris (Eds.), Near-Infrared Technology in the Agricultural and Food Industries, AACC Inc., St. Paul, 2001, pp. 145–169.

[40] W.R. Windham, D.R. Mertens, F.E. Barton II, Protocol for NIRS calibration: sample selection and equation development and validation, in: G.C. Martens, J.S. Shenk, F.E. Barton II (Eds.), Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality, Agriculture Handbook, vol. 643, USDA-ARS, US Government Printing Office, Washington, DC, 1989, pp. 96–103.

[41] J.S. Shenk, J.J.Jr. Workman, M. Westerhaus, Application of NIR spectroscopy to agricultural products, in: D.A. Burns, E. Ciurczak (Eds.), Handbook of Near-Infrared Analysis, Marcel Dekker, Basel, 2008, pp. 347–386.

[42] B.G. Osborne, T. Fearn, P.H. Hindle, Practical NlR spectroscopy with applications, in: B.G. Osborne, T. Fearn, P.H. Hindle (Eds.), Food and Beverage Analysis, Longman, Essex, 1993, pp. 11–35.

[43] D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero, Implementation of LOCAL algorithm with near-infrared spectroscopy for compliance assurance in compound feedingstuffs, Appl. Spectrosc. 59 (2005) 69–77. https://doi.org/10.1366/0003702052940585.

[44] S. Kramchote, K. Nakano, S. Kanlayanarat, S. Ohashi, K.Takizawa, G. Bai, Rapid determination of cabbage quality using visible and near-infrared spectroscopy, LWT Food Sci. Technol. 59 (2014) 695–700. https://doi.org/10.1016/j.lwt.2014.07.009.

[45] J. Fernández-Martínez, R. Joffre, M. Zacchini, B. Fernández-Marín, J.I. García-Plazaola, I. Fleck, Near-infrared reflectance spectroscopy allows rapid and simultaneous evaluation of chloroplast pigments and antioxidants, carbon isotope discrimination and nitrogen content in *Populus* spp. leaves, Forest Ecol. Manag. 399 (2017) 227–234. https://doi.org/10.1016/j.foreco.2017.05.041.

[46] D.C. Pérez-Marín, A. Garrido, J.E. Guerrero, A. Gómez-Cabrera, Near-infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs: chemical composition and open-declaration, Anim. Feed Sci. Technol. 116 (2004) 333–349. https://doi.org/10.1016/j.anifeedsci.2004.05.002.

[47] M. Cantwell, T. Suslow, Spinach: Recommendations for Maintaining Postharvest Quality, http://postharvest.ucdavis.edu/Commodity_Resources/Fact_Sheets/Datastores/Vegetables_English/?uid=32&ds=799, 2002 (accessed 5 June 2018).

[48] D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero, Optimization of discriminant partial least squares regression models for the detection of animal by-product meals in compound feeding stuffs by near-infrared spectroscopy, Appl. Spectrosc. 60 (2006) 1432-1437. https://doi.org/10.1366/000370206779321427.

Table 3.4.1. Statistical analysis of the calibration and prediction sample sets, i.e. data ranges, means and standard deviations (SD) and coefficients of variation (CV).

Parameter	Item	Calibration set	Validation set
Nitrate content (mg/kg)	Number	93	31
	Range	109.50-5177.00	144.00-3520.00
	Mean	1765.24	1795.48
	SD	1082.34	1088.46
	CV (%)	61.31	60.62
Ascorbic acid content (mg/100 g)	Number	93	31
	Range	156.92-479.23	191.83-453.85
	Mean	298.97	300.96
	SD	66.05	63.52
	CV (%)	22.09	21.10
SSC (º Brix)	Number	93	31
	Range	5.60-14.25	6.25-13.95
	Mean	8.84	8.34
	SD	1.90	1.89
	CV (%)	21.49	22.66

Table 3.4.2. Calibration statistics for the best equations obtained for the prediction of quality and safety parameters in intact spinach.

Parameter	Number of samples	Mathematical treatment	Range	Mean	SD	r^2_{cv}	SECV	RPD	SEL
Nitrate (mg/kg)	92	2,5,5,1	109.50-5177.00	1776.16	1083.10	0.41	836.26	1.29	131.61
Ascorbic acid (mg/100 g)	91	2,10,5,1	168.75-467.69	298.55	62.24	0.33	51.46	1.21	10.69
SSC (ºBrix)	92	1,10,5,1	5.60-14.25	8.85	1.91	0.85	0.75	2.54	0.07

Table 3.4.3. Calibration statistics for the best equations obtained for the prediction of nitrate content in intact spinach. Strategies I and II.

Strategy	Number of samples	Mathematical treatment	Range	Mean	SD	r^2_{cv}	SECV	RPD_{cv}
Strategy I	44	2,5,5,1	109.50-5177.00	1648.82	1037.97	0.53	711.63	1.46
Strategy II	71	1,5,5,1	122.50-3627.00	1821.91	1095.37	0.63	670.81	1.63

Table 3.4.4. Percentage of spinach plants classified correctly by nitrogen content. PLS2-DA.

Qualitative groups	**Unbalanced model**		**Balanced model**	
	Percentage of correctly-classified samples: 80.33% (49/61) Model SECV: 0.46 Number of synthetic variables: 4 Math treatment: 2,5,5,1		Percentage of correctly-classified samples 75.00% (39/52) Model SECV: 0.49 Number of synthetic variables: 5 Math treatment: 2,10,5,1	
Industrial destination of spinach according to its content in nitrates	**Training set**	**Validation set**	**Training set**	**Validation set**
Preserved, deep-frozen or frozen spinach, NO_3^-: 0-2,000 mg /kg	85.71% (30/35)	80.00% (8/10)	73.08% (19/26)	90.00% (9/10)
Fresh spinach NO_3^-: 2,000-3,500 mg /kg	73.08% (19/26)	60.00% (6/10)	76.92% (20/26)	50.00% (5/10)

Figure 3.4.1. Mean and standard deviation spectrum for spinach.

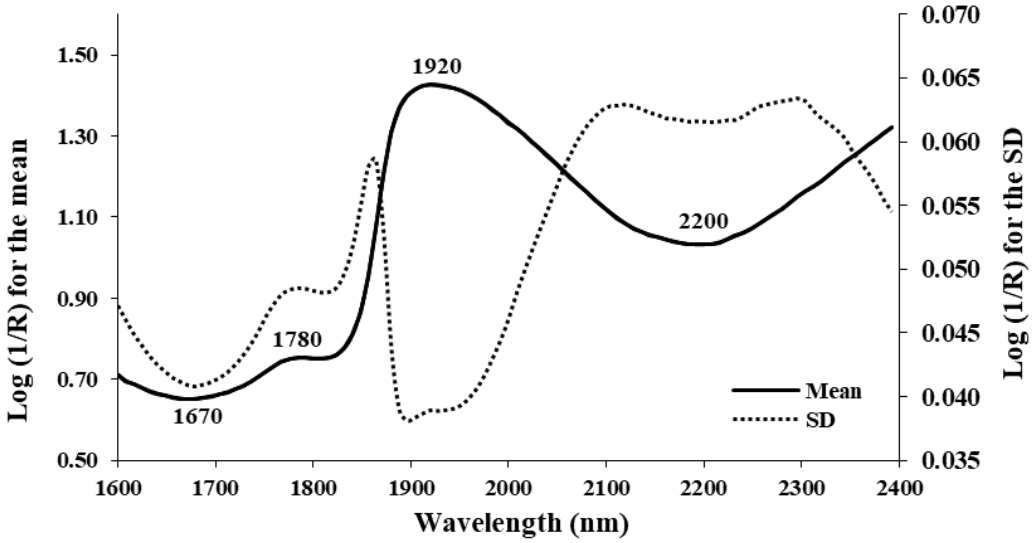
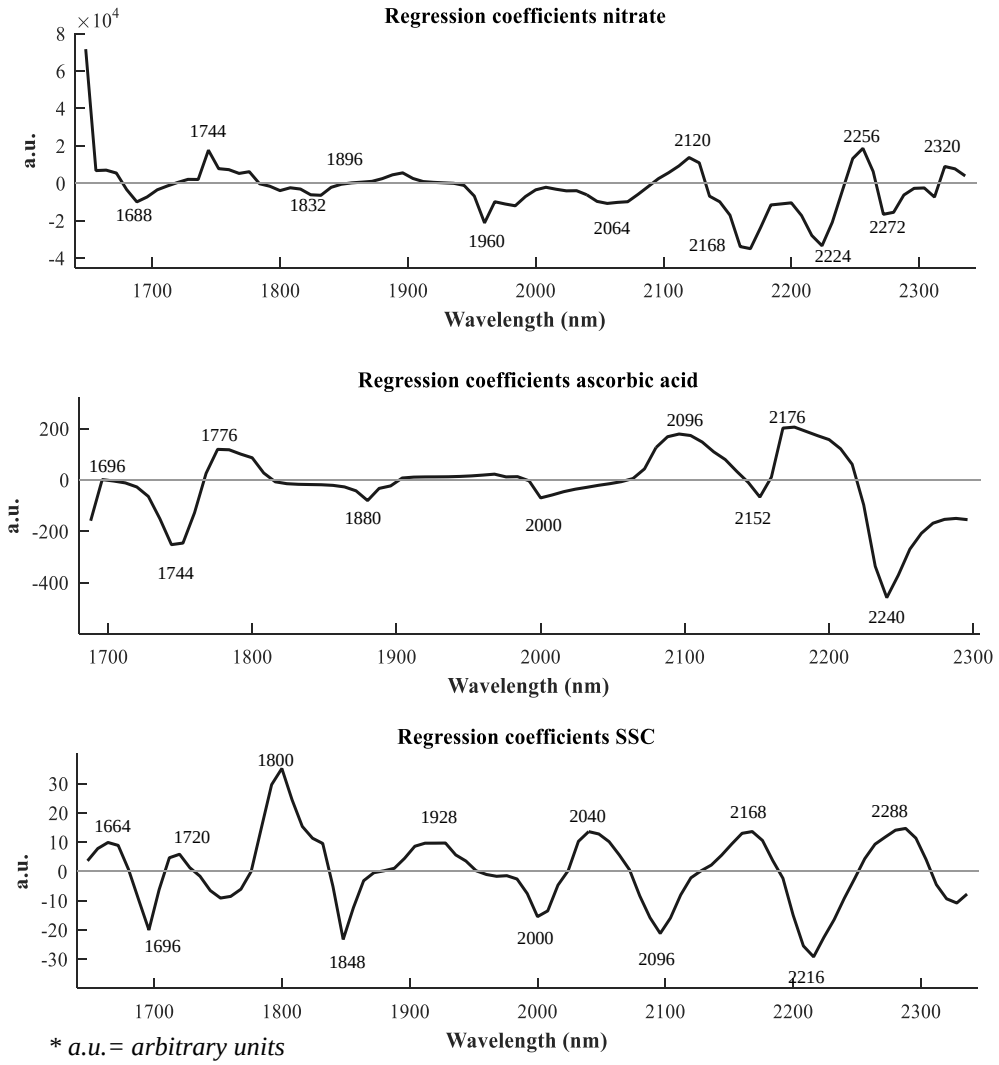
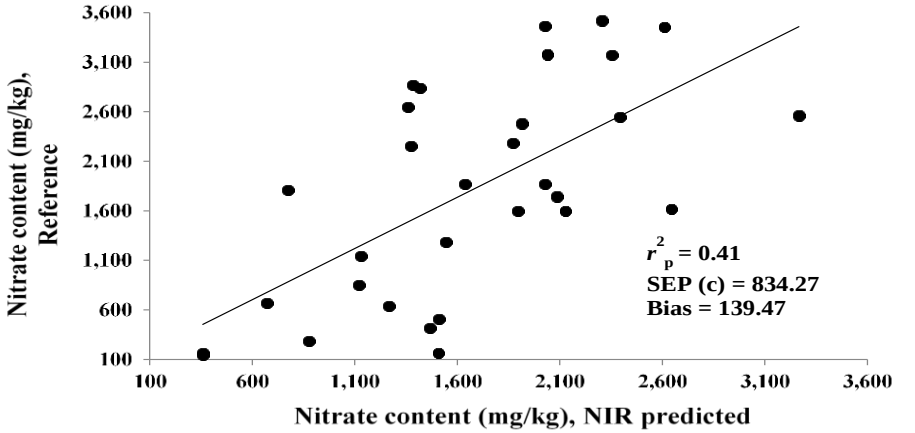


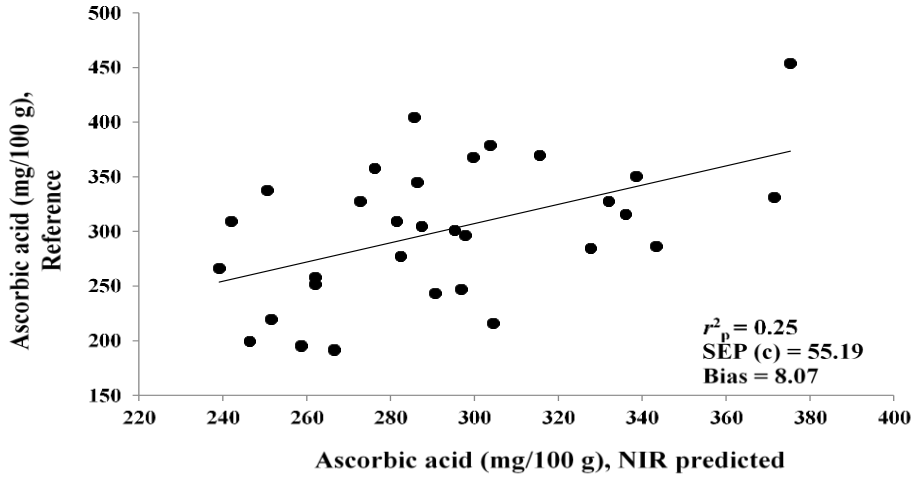
Figure 3.4.2. Regression coefficients for spinach nitrate content, ascorbic acid and SSC during on-vine ripening



* *a.u.= arbitrary units*

Figure 3.4.3. Reference and NIR predicted values for quality and safety parameters in intact spinach.





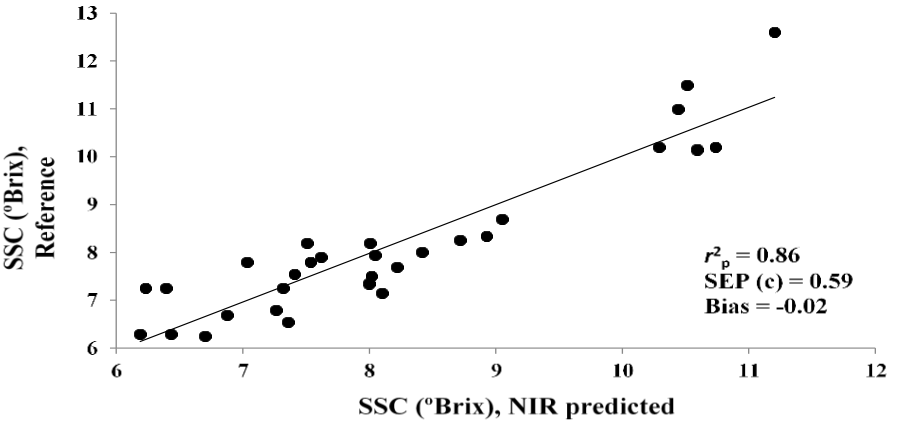
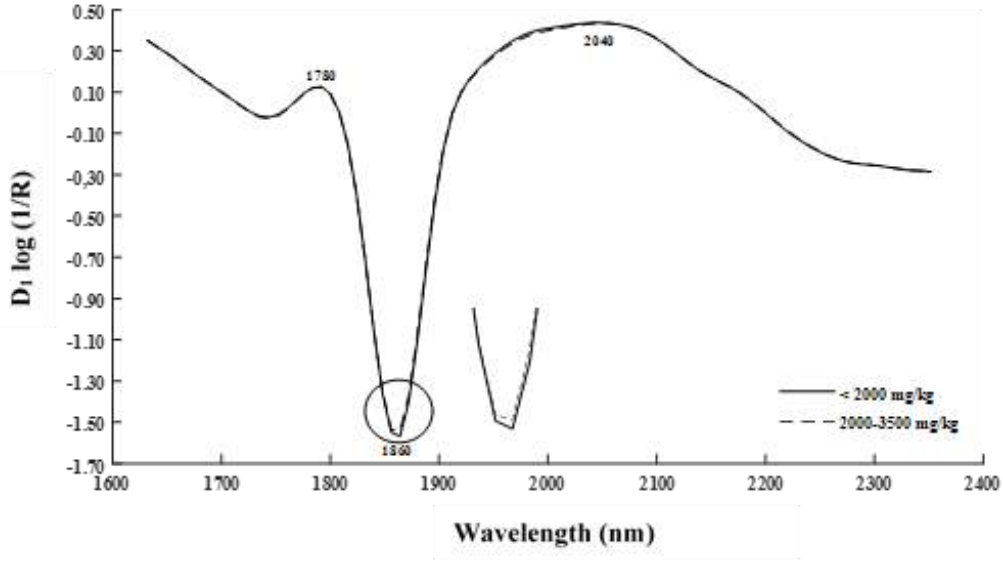


Figure 3.4.4. D_1 log(1/R) of the mean spectra for intact spinach plants with different nitrate content.



Chapter 4

Chapter 4. EVALUATION OF THE FEASIBILITY OF NIR SPECTROSCOPY IN PROVIDING *IN SITU* AUTHENTICATION FOR THE GROWING SYSTEM OF HORTICULTURAL PRODUCTS

Chapter 4.1.

Rapid, simultaneous, and in situ authentication and quality assessment of intact bell peppers using near-infrared spectroscopy technology

María-Teresa Sánchez[a], **Irina Torres[a],** María-José de la Haba[a], Ana Chamorro[a], Ana Garrido-Varo[b], Dolores Pérez-Marín[b]

[a] *Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*
[b] *Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*

Journal of the Science of Food and Agriculture 99, 1613-1622 (2018)

Abstract

BACKGROUND: The ability of near infrared (NIR) spectroscopy to authenticate individual bell peppers as a function of the growing system (outdoor or greenhouse) was tested using partial least squares discriminant analysis (PLS-DA). 394 bell peppers grown outdoors (130 samples) or in a greenhouse (264 samples) during the 2015 and 2016 seasons, were selected for this purpose and analysed using a portable, handheld MicroPhazir MEMS instrument (spectral range 1600-2400 nm), working in reflectance. Subsequently, the potential of NIRS as a non-destructive sensor for *in-situ* quality (dry matter and soluble solid content) measurements, was investigated.

RESULTS: The models correctly classified 89.73% and 88.00% of the samples by growing system, when trained with unbalanced and balanced sets, respectively, mainly due to the differences in physical-chemical attributes between bell peppers cultivated in both growing systems. Separate classification models for bell peppers grouped by ripeness (judged by the colour), allowed to classify 88.28%-91.37% of the samples correctly. The standard error of cross-validation (SECV) values for the quantitative models were 0.66% fw and 0.75 ºBrix for dry matter and soluble solid content, respectively.

CONCLUSIONS: The results showed that NIRS can be used successfully for predicting the growing systems used in bell pepper production, which is of particular value to guarantee the authentication of outdoor-grown peppers. Additionally, the results showed that NIRS can be used simultaneously as a rapid preliminary screening technique to measure quality.

Keywords: NIR spectroscopy; Bell pepper; *In situ* authentication; Quality; portable NIR device.

4.1.1. Introduction

In Spain, peppers are grown almost exclusively indoors - in greenhouses - although in some regions, as in the case of Andalusia, cultivation may take place outdoors.[1]

Bell peppers can be grown in greenhouses more compactly than outdoors. They are also pruned and trained differently: the former are cut back more severely in order to aerate the plants more inside the greenhouse, while in the latter system, the plants are allowed to grow more foliage to protect the peppers from the sun and chilly night-time temperatures.[2]

Likewise, peppers grown indoors have to be trained securely to support the fruits and prevent them from touching the ground or the branches from splitting, since the plants can reach up to 2 meters in height and the stems are far more tender than those grown outdoors. In contrast, peppers grown outdoors reach only one metre in height and do not need to be supported, as the stems are sturdier and do not grow high enough to bend or break.[3]

The variations in growing conditions between peppers grown outdoors and in a greenhouse can make an important difference to the quality of the product, especially in terms of the organoleptic characteristics linked to dry matter and sugar content. It should also be noted that consumer demand currently puts a high value on products which are local, seasonal and traditional, and peppers grown outdoors are favoured by these consumers.[4]

In general, consumers are interested in buying horticultural products obtained using this particular cultivation system and attribute higher quality standards to bell peppers grown in this specific way. It is therefore desirable for the horticultural production sector to have access to non-destructive technology which can carry out fast, highly accurate, *in-situ* analyses to guarantee the authenticity of the growing system. In this way, the consumer will receive

accurate information about the differences in quality between vegetables produced using different agronomic techniques and about their origin. In this field, NIR spectroscopy has proved to be an ideal way of providing authentication/certification of raw horticultural materials produced using different types of agricultural methods, as well as for the authentication of varieties.[5-8]

However, no papers in the scientific literature are dealing with the authentication of peppers using NIRS technology based on their origin (outdoor or greenhouse cultivation) and the possible differences in quality between fruits from the different growing systems. In fact, there are very few articles which incorporate the use of this technology to measure quality parameters in pepper. Thus, Sánchez *et al.*[9] assessed the viability of NIRS to measure pesticide residues in intact, crushed, and dry extract system for infrared analysis (DESIR) peppers, while other authors[10-13] carried out the analysis of a number of quality parameters in different types of peppers.

The aim of this study was to evaluate the viability of NIR spectroscopy in providing non-destructive, *in situ* authentication for the growing system - outdoor or in a greenhouse - of bell peppers. In addition, quantitative models were developed to predict two of the main quality parameters (dry matter content and soluble solid content) in intact bell peppers, which could help to classify peppers more easily by their origin. Particular attention was paid to the robustness of the models.

4.1.2. Materials and methods

4.1.2.1. Sampling

394 bell peppers (*Capsucicum annum* L.) of different colours (green, red and yellow, depending on the degree of ripeness), grown outdoors (N = 130 bell peppers: green = 50, yellow = 41, red = 39) and in a greenhouse (N = 264 peppers: 88 of each colour), picked in the 2015 and 2016 seasons, were

analysed. The greenhouse samples were grown in the Region of Murcia (Spain), while the outdoor peppers were harvested in Santaella (Córdoba, Spain).

4.1.2.2. Spectral set

A handheld MEMS (micro-electro-mechanical system)-based NIR digital transform spectrometer (MEMS-NIRS) (MicroPhazir, Polychromix Inc., Wilmington, MA, USA), working in reflectance mode in the spectral range 1600-2400 nm with a non-constant interval of around 8 nm was used to collect the NIR spectra of all the samples in reflectance mode. Sensor integration time was 600 ms. The device was equipped with quartz protection to prevent dirt accumulation. Each spectrum was the mean of 5 scans with a lamp warm-up time of 45 seconds. To obtain the NIR spectra, four measurements were taken at the equatorial region of the fruits, which were then rotated 90° after each measurement. The four spectra were averaged to provide a mean spectrum for each fruit.

4.1.2.3. Measurement of physical-chemical quality parameters

Dry matter content was measured by desiccation at 105ºC for 24 h[14]; the final dry weight was calculated as a percentage of initial wet weight. Soluble solid content (SSC) in ºBrix was measured as the refractometer reading for the pepper juice, using a temperature-compensated digital Abbé-type refractometer (model B, Zeiss, Oberkochen, Würt, Germany).

All the samples were analysed in duplicate and the standard error of laboratory (SEL) was estimated from these duplicates (Table 4.1.7). All the measurements were performed immediately after taking the NIRS measurements.

4.1.2.4. Data Processing

Before the spectral data was processed and using the WinISI II software package version 1.50 (Infrasoft International, Port Matilda, PA, USA),

a study was conducted to select the most suitable spectral range for the instrument to carry out the authentication and quality control of peppers. To achieve this, the 1,1,1,1 derivation treatment was applied (the first digit being the number of the derivative, the second the gap over which the derivative is calculated, the third the number of data points in a running average or smoothing, and the fourth the second smoothing) without scatter correction, which allows to highlight the areas of the spectrum where the signal/noise ratio is degraded.[15]

4.1.2.4.1. Spectral repeatability

Spectrum quality was evaluated using the Root Mean Square (RMS).[16-17] This statistic indicates the similarity between different spectra of a single sample: in this case, between the four spectra collected per sample. To establish a threshold for this statistic, 36 bell peppers were selected, from which four spectra were taken in the equatorial region, rotating the fruit 90º after each measurement. An admissible limit for spectrum quality and repeatability was set following the procedure described by Martínez *et al.*[18] to calculate the standard deviation (STD) limit from the RMS statistic and obtain an RMS cut-off value.

4.1.2.4.2. Principal component analysis

In order to study the relationship between the quality (dry matter and SSC) of the bell pepper and the growing system used (outdoor or greenhouse), Principal Component Analysis (PCA) was carried out.

PCA is a mathematical procedure that reduces the dimensionality of the data to uncorrelated variables, including in each component the maximum residual variance of the data, and each component therefore contains a representation of the data variation.[19] The PCA scores represent the weighted sums of the original variables without significant loss of useful information, and loadings (weighting coefficients) were used to identify the major variables

responsible for specific features appearing in the scores. Matlab software (version 2015a, The Mathworks, Inc., Natick, Massachusetts, USA) was used to conduct PCA, using the mean centre, by which the mean spectrum of the group is subtracted from each spectrum, as a pre-treatment.[20]

4.1.2.4.3. Authenticating bell peppers by growing method using NIR spectroscopy

To carry out the non-destructive authentication of bell peppers according to their growing system, discriminant models were designed to classify the peppers into two groups: bell peppers grown outdoors and bell peppers grown in a greenhouse.

Firstly, the spectral structure and variability of the sample population was studied to select the samples which would make up the training group. The CENTER algorithm was used for this, which is included in the WinISI II version 1.50 software. This algorithm was applied separately to each of the two training groups (130 outdoor-grown bell peppers and 264 greenhouse bell peppers). The mathematical treatments SNV (Standard Normal Variate) and DT (Detrend) for scatter correction were applied,[21] and the 1,5,5,1 derivate mathematical treatment.[22-23] After PCA, the center of the spectral population was determined in order to detect outlier samples. The Mahalanobis distance (GH) was calculated between each sample and the center; samples with a GH value greater than 4.5 were considered outliers or anomalous spectra.[16] After discarding outliers, the sampling groups consisted of 128 samples of bell pepper grown outdoors and 259 samples of greenhouse-grown bell pepper.

After ordering the sample set by spectral distances (from smallest to greatest distance from the center), a structured selection of an external validation set (28 samples for each classification group) was performed following Shenk and Westerhaus.[24]

The difficulty involved in obtaining balanced learning groups in terms of the number of samples per class or classification category meant that its influence on the predictive capacity of the models had to be assessed. The results obtained were therefore contrasted with balanced and unbalanced classification models as regards the number of samples per class.

The samples of the balanced groups were selected using the SELECT algorithm included in the WinISI II software package version 1.50.[24] This algorithm enables spectral selection of a number of samples which are representative of the population as a whole, by calculating the 'NH' distance (Mahalanobis neighbour distance) between two spectra. An 'NH' of less than 0.6 implies that two spectra are too similar to each other ('neighbours'). After this algorithm was applied, 100 samples of the 'greenhouse-grown peppers' group were selected, thus leaving the number of samples of the training group for the two classes equal, and the classification models were then developed.

Discriminant models were constructed to authenticate bell peppers according to their growing system, using PLS-DA for supervised classification.[23] Briefly, PLS-DA uses a training set to develop a qualitative prediction or calibration model that may subsequently be applied for the classification of new unknown samples. This model seeks to correlate spectral variations (X) with defined classes (Y), attempting to maximise the covariance between the two types of variable. In this type of approach, the Y variables used are not continuous, as they are in quantitative analysis, but rather categorical 'dummy' variables created by assigning different values to the different classes to be distinguished. Specifically, the PLS2 algorithm was applied, using the ''Discriminant Equations'' option in the WINISI II version 1.50 software.

All the models were designed using four cross validation groups (i.e. the calibration set was divided into four groups, and each group was then

predicted using a calibration obtained from the other samples), a spectral range from 1600 to 2168 nm, signal noise eliminated at the end of the spectral range, and combined SNV+DT treatment for scatter correction. First- and second-derivative treatments were tested by applying 1,5,5,1 and 2,5,5,1.[22]

The precision of the models obtained was evaluated using the percentage of correctly-classified samples, both for the model and for each class. In addition, the standard error of cross validation (SECV) was evaluated. Most of the papers use the value of 1.5 as discrimination limit, so that, if one sample obtain a variable value over the limit for a given class, it will be classified as belong to this class. However, in this paper it was also used the minimum difference (MD) value, calculated as the product of the value of the model's uncertainty factor (1.5) by the SECV, for the detection of uncertain samples when interpreting the results obtained. Samples with a MD higher than the MD value calculated should be considered as uncertain.[25] Regression coefficients were also used to discuss the contributions of individual wavelengths to the qualitative PLS models. [26]

Next, after selecting the best classification model for each of the established types (unbalanced and balanced models), these were validated. In this case, an external validation procedure was also carried out to determine the predictive capacity of the model using a sample group different to that used in the training of the model. In both models (unbalanced and balanced), 56 samples were selected in a structured way (28 samples for each of the culture systems: outdoor or greenhouse).[24]

Then, after analysing the results of the statistical tests which evaluated the influence of the cultivation systems and the state of ripeness (reflected by colouration) on the quality of the bell peppers harvested, new classification models were developed for the peppers according to the growing system, but also taking into account the colouration (green, yellow and red).

For each of the colours analysed, the structure and spectral variability of the sample population was studied for each of the growing systems, using the same methodology described above. A structured selection of the external validation set (5 samples for each classification group due to the low number of samples in the 'outdoor-grown' category, once the bell peppers were separated by colour), solely on the basis of spectral data, was performed following Shenk and Westerhaus.[24]

4.1.2.4.4. Quantitative models: sets, calibration development and validation procedure

Quantitative models were built to predict the parameters of dry matter and SSC, using all the available bell pepper samples, independently of the cultivation system used. The samples for the calibration and validation groups were selected by applying the CENTER algorithm in the 1600-2168 nm spectral range. The pre-treatments SNV and DT were used for scatter correction,[21] and one derivative mathematical treatment (Norris derivative) was performed (1,5,5,1).[22-23] Thus, having ordered the sample set by spectral distances (from smallest to greatest distance to the centre) and once outlier spectrum samples were eliminated (N = 2), the 130 samples forming the validation set were selected by taking one of every 3 samples in the initial 392-sample set; the calibration set thus comprised the remaining 262 samples.

Modified partial least squares (MPLS) regression[16] was used to obtain equations for predicting dry matter and SSC. All the models were constructed using four cross-validation groups. The same signal pre-treatments and spectral region described earlier for authentication analysis were used for designing the quantitative models. The statistics used to select the best equations were: standard error of calibration (SEC), coefficient of determination of calibration (r^2_c), standard error of cross-validation (SECV), coefficient of determination for cross-validation (r^2_{cv}), RPD_{cv} or ratio of the standard deviation of the original data (SD) to SECV and the coefficient of variation (CV), defined as the percentage ratio of the SECV to the mean value of the reference data for the

calibration set. These latter two statistics enable SECV to be standardized, facilitating the comparison of the results obtained with sets of different means.[27] Regression coefficients were also used to discuss the contributions of individual wavelengths to the quantitative models.[26]

The best models obtained for the calibration set, as selected by statistical criteria, were subjected to evaluation using samples not involved in the calibration procedure. A test set composed of 130 samples, not used previously in the model, was evaluated following the protocol outlined by Windham *et al*.[28]

4.1.2.5. Statistical analysis

All the quantitative analyses were expressed as mean values ± standard deviation. The data for each attribute (dry matter and SSC) for outdoor and greenhouse bell peppers were analysed statistically by analysis of variance (ANOVA) using Statgraphics Centurion XV (StatPoint Inc., Warrenton, North Virgina, USA), and initially considering the origin of the pepper (outdoor or greenhouse cultivation) as a factor. Next, in order to study the influence of both the growing technique and the pepper colouring in the dry matter and soluble solids contents, a two-factor ANOVA variance analysis was carried out.

In both cases, the difference between the means were compared with the Fisher's Least Significant Difference (LSD) test, and differences at $P < 0.05$ were considered to be significant.

4.1.3. Results and discussion

4.1.3.1. Optimal spectral region and spectral repeatability

Prior to the model development, it was necessary to optimise the NIRS analysis by means of the spectrum quality and repeatability measurement.

The existence of noise in the spectrum was evaluated (spectral range 1600-2400 nm). To this end, the derivative treatment 1,1,1,1 was applied in order to determine the area of the spectral range affected by noise, given that it degrades the signal/noise relationship.[15] After this process, the spectral range between 2169–2400 nm was eliminated, and all the models were designed using the spectral range 1600–2168 nm.

Spectral repeatability is crucial to the construction of models that are both accurate and robust. Statistical methods such as a defined RMS cut-off limit can be useful for this purpose.

The mean STD for the samples analysed was 108,733 µlog (1/R), representing an RMS cut-off of 122,144 µlog (1/R).[18] Any sample whose quadruplicated screening scans yielded an RMS above this value was eliminated and the scan was repeated until values fell below that limit, thus ensuring a high degree of spectrum repeatability.

No reference to the calculated RMS cut-off value for intact peppers has been found in the literature, although this statistic is essential for generating the representative libraries.

The mean spectrum of the four replicates of each sample was used for further analysis.

4.1.3.2. Principal Component Analysis

PCA was performed on the set comprising the spectra recorded for each growing system (outdoor or greenhouse) of intact bell peppers.

Figure 4.1.1a displays scores of the second and third components of the PCA model. These two components were chosen because, although the first two principal components (PC1 and PC2) represented a high proportion of the

explained variance 94.18% and 5.46%, respectively, they did not facilitate the grouping of the samples according to the growing system used; this grouping does however seem to become more evident when the latent variables PC2 and PC3 are used. Figure 4.1.1b shows the PCA loadings for intact bell peppers in the spectral range 1600-2168 nm.

The graphic representation of the loadings for PC2 and PC3 (Figure 4.1.1b) shows that the main absorption peaks for differentiating between the two growing systems of the bell peppers are those related to carbohydrates and water, respectively. The PC3 weighting coefficient exhibits a band of water around 1930 nm.[29] The peak points down so more water (less dry matter) means a more negative score on PC3, which is exactly what the greenhouse-grown peppers show (Table 4.1.1). PC2 exhibits a band that is characteristic of carbohydrates (~1680 nm)[29].

In the light of the PCA scores (Fig 1a) and bearing in mind the results of the ANOVA and LSD tests (Table 4.1.1) about the similarities or not in physical–chemical composition between bell peppers cultivated outdoors or in a greenhouse, it may be said that dry matter is indeed related to PC3 and significant differences ($P < 0.05$) were found for dry matter between both types of bell peppers. The positive PC3 scores are associated with fruits of higher dry matter content, while the negative PC3 scores are linked to fruits with lower dry matter values. As has already been mentioned, PC2 may be linked to carbohydrate content, and considering that no significant difference ($P > 0.05$) was found for SSC between outdoor and greenhouse bell peppers (Table 4.1.1), no grouping of samples by SSC was apparent using this component.

4.1.3.3. Authentication of bell peppers by NIRS

Values obtained for number of samples (N), range, mean, standard deviation (SD), and coefficient of variation (CV) for each of the quality parameters measured for the training and validation sets used in the

discriminant models for the authentication of bell peppers by growing system are shown in Table 4.1.2.

Table 4.1.3 shows the results for the best classification models obtained, using PLS-DA, to authenticate the origin of the intact bell peppers analysed (grown outdoors or in a greenhouse).

The most accurate models were achieved using D_1 log (1/R), for both unbalanced and balanced sets. The total percentages of correctly classified samples were 89.73% and 88.00% for the unbalanced and balanced model, respectively. These results, regardless of the population size, confirm those obtained by Pérez-Marín *et al.*[30], who showed that PLS2 is less sensitive to the fact that the populations are unbalanced.

For the unbalanced model, 74 samples of the 100 forming the training group of outdoor-grown peppers were correctly classified, while for the greenhouse-grown peppers, 223 samples out of 231 were correctly classified. It is also important to note that of the 26 samples poorly classified in the outdoor-grown bell pepper category, 17 were within the 1.5 ± MD limit, while for peppers grown in the greenhouse, 7 out of the 8 poorly classified samples are also within this limit.

For the balanced model, 88 samples of the 100 contained in each of the two established training groups (outdoor and greenhouse) were correctly classified. In this case, 11 out of the 12 samples poorly classified in the 'outdoor-grown bell pepper' category were within the limits established by the uncertainty factor ± MD, while for the greenhouse pepper category, the 12 misclassified samples were also within this limit.

The models were then validated, using samples not included in their design. In the models created from the unbalanced populations, the percentage of correctly classified samples was 78.57% and 100.00%, for the outdoor and

greenhouse cultivation systems, respectively (Figure 4.1.2 and Table 4.1.3). Out of the 6 badly-classified samples in the 'outdoor-grown' bell pepper category, 4 were in the interval between the uncertainty factor ± MD.

In the case of the balanced populations, 85.71% of the peppers grown outdoors and 96.43% of greenhouse-grown bell peppers were correctly classified (Figure 4.1.2 and Table 4.1.3). It is important to note how the 4 poorly classified samples in the 'outdoor-grown bell pepper' category were within the range of 1.5 ± MD, which was the same case as the single badly-classified sample from the 'greenhouse-grown bell pepper' category.

For the balanced model, the point clouds hardly change, but the threshold moves towards the outdoor-grown samples in that case (Figure 4.1.2). The consequence of this is that for the smaller group the total accuracy is 78.57% when using the unbalanced set and increases to 85.71% when using the balanced set. Despite a low reduction in the accuracy of the larger set (100% *versus* 96.43%) when using the balanced set, the results for the smaller group set improve.

To examine more deeply the results of the classification models obtained, the results of the ANOVA (dry matter and SSC) tests and the LSD (dry matter) test (Table 4.1.1) were also considered, along with the results of the PCA (Figure 4.1.1). Significant differences ($P < 0.05$) were detected in terms of the dry matter content between both types of bell peppers (the dry matter content was significantly higher in peppers grown outdoors), and the SSC content was higher - although not significantly ($P > 0.05$) - in the outdoor group. Likewise, as stated above, it is the PC3 related to water content and, therefore, to dry matter content, which facilitates the classification of bell peppers according to the cultivation system in which they are grown.

An ANOVA analysis was later carried out to study both the influence of the cultivation system used and the colouring of the pepper, which indicates

its state of ripeness, on the dry matter and SSC in the bell peppers analysed. The results of the ANOVA test for the parameter dry matter content pointed to the existence of significant differences ($P < 0.05$) between the cultivation systems and colouration, as well as in the interaction between the cultivation system and the colouration. For SSC, no significant differences ($P > 0.05$) were detected between the cultivation systems and in the interaction between the cultivation system and the colouration. However, significant differences were detected ($P < 0.05$) between peppers of different colourations. The results of the Fisher's tests are shown in Table 4.1.4.

After analysing the results of the ANOVA and Fisher's tests, new models were designed to classify the peppers according to the cultivation system used and taking the colour into account.

Values obtained for number of samples (N), range, mean, SD and CV for each of the quality parameters measured for the training and validation sets used in the discriminant models for the authentication of green, yellow and red bell peppers by growing system are shown in Table 4.1.5.

The results obtained for the best classification models for bell peppers according to the cultivation system used and taking the colour into account are shown in Table 4.1.6.

For peppers with green colouration, 113 samples of the 128 available were correctly classified; of these, the model correctly classified 35 of the 45 samples in the outdoor-grown category and 78 of the 83 samples in the greenhouse category. When these models were externally validated, all the samples were correctly classified in the right category.

For yellow bell peppers, the percentage of samples correctly classified in the training group was 91.37% (106 out of 116), with percentages of 87.87% (29 of 33) and 92.77% (77 of 83) for peppers grown outdoors and in the greenhouse, respectively. When the models were validated, the 5 selected

samples of outdoor-grown peppers were correctly classified, while 80% of the greenhouse peppers were correctly classified.

In the case of red bell peppers, 107 out of the 118 samples were correctly classified (90.67%). Category by category, 28 out of 35 samples were in the outdoor category and 79 out of 83 in the greenhouse category. When the models were validated, 80% and 100%, respectively, of the peppers from the outdoor and greenhouse categories were correctly classified.

The results of the classification models obtained (Tables 4.1.3 and 4.1.6) show that using NIR technology to predict the cultivation system of the intact bell peppers is a feasible option and it can be used to authenticate the origin of these vegetables.

Figure 4.1.3 shows characteristic peaks and valleys that indicate which wavelength ranges are important for the balanced classification model of bell peppers by growing system. The figure indicates that the most relevant regression coefficients are located in the region 1660-1880 nm which is associated to the absorption band of a C-H stretching first overtone corresponding to sugars. [31-32] Other relevant coefficients appear in the regions 1930-1990 nm, related to water absorption [31] and 2064-2144 nm, also related with different types of sugars. [32]

In the scientific literature, no predictive models have been found based on NIRS to authenticate the origin of bell peppers depending on the culture system used. Only Sánchez *et al.*[9] assessed the feasibility of using NIR spectroscopy to classify peppers according to the presence of pesticide residues, confirming that NIRS technology may be used to provide swift, non-destructive preliminary screening for pesticide residues.

4.1.3.4. Predicting quality parameters in bell peppers using MPLS regression

Table 4.1.7 shows the characteristics of the calibration and validation sets used to develop the predictive models for dry matter and SSC.

Structured selection based wholly on spectral information, using the CENTER algorithm, proved suitable, in that the calibration and validation sets displayed similar values for range, mean and SD for all the study parameters; moreover, the established ranges of the validation lay within those of the calibration set.

Table 4.1.8 shows the best calibration equations for the two quality parameters selected in bell peppers. For predicting dry matter and SSC in bell peppers, the models constructed allow to discriminate between high, medium and low values of these parameters.[17, 27]

As regards the dry matter parameter, Ignat *et al.*[12] reported predictive capacity (RPD_{cv} = 3.8) higher than those obtained here using a diode array instrument (spectral range: 477-950 nm), although these authors used a wider calibration set since they chose fruits picked during the growing season, from the 34th day after anthesis (DAA) until full ripening (88th DAA), and when fully grown.

For the SSC parameter, Penchaiya *et al.*[10] used a diode array spectrophotometer (spectral range 780-1690 nm) to obtain predictive capacity (RPD_{cv} = 2.08) superior to that of this research work, although the window for the spectrophotometer used (Corona Fiber VIS / NIR, Carl Zeiss Jena GmbH, Germany) was much wider than that of the instrument used here, and its measurement range was also different. In addition, these authors used a wide range of sample attribute in the calibration set, obtained by random harvesting at various stages of ripeness.

Also, for SSC, Ignat *et al.*[12] used the same instrument and spectral range and obtain predictive capacity (RPD_{cv} = 3.9); it is important to stress the greater variability of the fruits used, which also affected the 'dry matter' parameter, as aforementioned.

Toledo-Martín *et al.*[13], using an instrument based on MEMS technology with a 1000-1800 nm spectral range, obtained models for SSC with a predictive capacity (RPD_{cv} = 1.7) very similar to that obtained in this work.

When these results are compared with those of other authors, the importance of the spectrophotometer's measurement window can be seen for the robustness of the developed models. While MEMS instruments perform isolated readings on the product being studied with measurement windows of an area of only around 4 mm^2, the diode array instruments tested by the authors quoted above perform a scan of the whole sample, which is of vital importance in hollow, irregularly-shaped vegetables such as bell peppers.

Validations of the best calibration models obtained were performed using a set comprising 130 samples (Figure 4.1.4).

For dry matter and SSC, it should be stressed that bias lay within confidence limits for both parameters, although SEP(c) and r^2_p results did not attain the recommended values for their routine use in equations,[28] indicating that the NIRS equations constructed should be regarded as a first step in the finetuning of NIRS technology for the *in-situ* monitoring of internal quality parameters in this type of pepper.

The SEL values for dry matter and SSC were 0.21% fw and 0.06 ºBrix, respectively (Table 4.1.7). Such a small SEL for SSC must be correctly interpreted when it is compared with the SEP value obtained for the prediction model. Firstly, it must be considered that sugar distribution is heterogenous in the fruit. It is for this reason that in the NIR analysis four spectra were taken in

the equatorial region of the fruit. However, the reference value was obtained as the refractometer reading for the pepper juice. It means that the sampling error is included in the SEP value but not in the SEL value. Consequently, NIRS model developed for SSC was characterised by questionable performance, since SEP value obtained exceeded 5*SEL.[15] For dry matter, SEP fell between 3 and 4 SEL, indicating acceptable performance of NIRS model developed.

These findings must be considered for the correct interpretation of the statistic SEP in intact fruits and vegetables. Likewise, the use of handheld NIRS spectrometers is justified given the fact that they ensure in a short period of time, a more precise and accurate guarantee of internal quality of the horticultural product analysed, allowing increased sampling either on the surface of the product tested or in the batch produced.

Finally, the regression coefficients for the best predictive models for dry matter and SSC are illustrated in Figure 4.1.5. These regression coefficients show significant importance for the region around 1650–1850 nm which correspond to the first overtone of the C-H stretching bonds and at around 1920-1960 nm due to O-H group contribution. The absorbance region at 2040–2100 nm could be attributed to NH and OH stretching modes besides C=O vibration bands. [29]

It is also important to point out that the most relevant peaks and valleys coincide in Figure 4.1.3 (qualitative model) and in Figure 4.1.5 (quantitative models). These results reinforce the idea that the discrimination between outdoor-grown and greenhouse-grown bell peppers has a scientific explanation based on the differences in dry matter and SSC between both type of bell peppers. Nicolaï *et al*. [33] indicate that the water absorption bands dominate the spectrum of fruit and vegetables, and it is not likely that minor constituents can be measured well. The authors also state that evidently, when the concentration of such a minor constituent is correlated to, e.g., sugar content, the calibration results may seem reasonable. From the observation of Figure 4.1.5, it can be

detected that while peaks and valleys at 1680-1690 nm and 1800 nm are relevant for both, dry matter and SSC, the valley at 1776 nm only dominates in the SSC spectrum. The scientific literature concerning absorptions bands in fruits and vegetables are dominated by papers which use a more limited range of wavelengths that the considered in this study, due to many of them use transmittance instruments. Therefore, further studies are needed to confirm the bands indicated as the most relevant ones in the spectral region analysed.

4.1.4. Conclusions

The results confirm that NIR spectroscopy using a portable manual instrument based on MEMS technology can be used at any time in the food chain (from the field to the dinner table) to authenticate intact bell peppers depending on the type of cultivation (outdoor *versus* greenhouse) used for growing the crop. Also, NIRS technology could be used as a fast and *in-situ* preliminary screening technique for the classification of bell peppers by dry matter and SSC. However, further research is needed to make the quantification of these parameters more robust.

Acknowledgements

This research was carried out under the Research Project ‘Quality determination of pepper grown in Santaella (Córdoba)’, funded by Gelagri Ibérica, S.L. The authors are grateful to Mrs. María Carmen Fernández of the Animal Production Department for her technical assistance. Furthermore, the authors wish to express their gratitude to the Spanish Ministry of Education, Culture and Sports for the support offered to Irina Torres in the form of the Training Programme for Academic Staff (FPU).

References

1 López-Marín J, Porras I, Ros C and Brotons-Martínez JN, Study of the profitability of pepper cultivation (*Capsicum annum*) in greenhouse with the use of shading. *ITEA-Inf Tec Econ Ag* **112**:57 – 71 (2016) [in Spanish].

2. Maroto JV, *Special Herbaceous Horticulture*. Mundi-Prensa, Madrid (2002) [in Spanish].

3. Serrano Z, *Handbook of Pepper Cultivation*. Consejería de Agricultura y Pesca, Junta de Andalucía, Seville (2011) [in Spanish].

4. Meca D and Céspedes A, Varietal trends in protected horticulture. *Vida Rural* **406**:54 – 58 (2015) [in Spanish].

5. Pérez D, Sánchez MT, Cano G and Garrido A, Authentication of green asparagus varieties by Near-Infrared Reflectance Spectroscopy. *J Food Sci* **6**:323 – 237 (2001).

6. Sánchez MT, De La Haba MJ, Benítez-López M, Fernández-Novales J, Garrido-Varo A and Pérez-Marín D, Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. *J Food Eng* **110**:102 – 108 (2012).

7. Sánchez MT, Garrido-Varo A, Guerrero JE and Pérez-Marín D, NIRS technology for fast authentication of green asparagus grown under organic and conventional production systems. *Postharvest Biol Technol* **85**:116 – 123 (2013).

8. Torres I, Pérez-Marín D, De la Haba MJ and Sánchez MT, Fast and accurate quality assessment of Raf tomatoes using NIRS technology. *Postharvest Biol Technol* **107**:9 – 15 (2015).

9. Sánchez MT, Flores-Rojas K, Guerrero JE, Garrido-Varo A and Pérez-Marín D, Measurement of pesticide residues in peppers by near infrared reflectance spectroscopy. *Pest Manag Sci* **66**:580 – 586 (2010).

10. Penchaiya P, Bobelyn E, Verlinden BE, Nicolaï BM and Saeys W, Non-destructive measurement of firmness and soluble solids content in bell pepper using NIR spectroscopy. *J Food Eng* **94**:267 – 273 (2009).

11. Ignat T, Schmilovitch Z, Feföldi J, Bernstein N, Steiner B, Egozi H *et al.*, Nonlinear methods for estimation of maturity stage, total chlorophyll, and carotenoid content in intact bell peppers. *Biosyst Eng* **114**:414 – 425 (2013).

12. Ignat T, Alchanatis V and Schmilovitch Z, Maturity prediction of intact bell peppers by sensor fusion. *Comput Electron Agr* **104**:9 – 17 (2014).

13. Toledo-Martín EM, García-García MC, Gómez P, Moreno-Rojas JM, González A, Moya M *et al.*, Application of visible/near-infrared reflectance spectroscopy for predicting internal and external quality in pepper. *J Sci Food Agric* **96**:3114 – 3125 (2016).

14. AOAC, *Official Methods of Analysis of AOAC International*, 17th ed. AOAC, Gaithersburg, MD (2000).

15. Hruschka WR, Data analysis: Wavelength selection methods, in *Near-Infrared Technology in the Agricultural and Food Industries*, ed. by Williams PC and Norris KH. American Association of Cereal Chemists, Inc., St. Paul, MN, pp. 39 – 58 (2001).

16. Shenk JS and Westerhaus MO, *Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy*. NIRSystems, Silver Spring, MD (1995).

17. Shenk JS and Westerhaus MO, Calibration: the ISI way, in *Near Infrared Spectroscopy: The Future Waves*, ed. by Davies AMC and Williams PC. NIR Publications, Chichester, pp. 198 – 202 (1996).

18. Martínez ML, Garrido A, De Pedro EJ and Sánchez L, Effect of sample heterogeneity on NIR meat analysis: the use of the RMS statistic. *J Near Infrared Spectrosc* **6**:313 – 320 (1998).

19. Mark H, Data analysis: Multilinear regression and Principal Component Analysis, in *Handbook of Near-Infrared Analysis* ed. by Burns DA and Ciurczak EW. CRC Press, Florida, Vol. 8, pp. 151 – 188 (2001).

20. Wise BM, Gallagher NB, Bro R, Shaver JM, Windig W and Koch RS, *PLS_ToolBox 4.0. Manual for Use with MATLAB (TM) [Computer software]*. Eigenvector Research, Inc., Wenatchee, WA (2006).

21. Barnes RJ, Dhanoa MS and Lister SJ, Standard Normal Variate transformation and De-trending of near infrared diffuse reflectance spectra. *Appl Spectrosc* **43**:772 – 777 (1989).

22. Shenk JS and Westerhaus MO, *Routine Operation, Calibration, Development and Network System Management Manual*. NIRSystems, Silver Spring, MD (1995).

23. ISI, *The Complete Software Solution Using a Single Screen for Routine Analysis, Robust Calibration, and Networking. Manual, FOSS NIRSystems/TECATOR*. Infrasoft International, LLC, Silver Spring, MD (2000).

24. Shenk JS and Westerhaus MO, Population structuring of near infrared spectra and modified partial least squares regression. *Crop Sci* **31**:1548 – 1555 (1991).

25. Murray I, Aucott LS and Pike HI, Use of discriminant analysis on visible and near infrared reflectance spectra to detect adulteration of fishmeal with meat and bone meal. *J Near Infrared Spectrosc* **9**:297 – 311 (2001).

26. Martens H and Naes T, *Multivariate Calibration.* John Wiley & Sons, Chichester (1989).

27. Williams PC, Implementation of near-infrared technology, in *Near Infrared Technology in the Agricultural and Food Industries*, ed. by Williams PC and Norris KH. American Association of Cereal Chemists, Inc., St Paul, MN, pp. 145 – 171 (2001).

28. Windham WR, Mertens DR and Barton II FE, Protocol for NIRS calibration: sample selection and equation development and validation, in *Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality (USDA-ARS Agriculture Handbook No. 643)*, ed. by Martens GC, Shenk JS and Barton II FE. US Government Printing Office, Washington, DC, pp. 96 – 103 (1989).

29. Shenk J, Workman JJJr and Westerhaus M, Application of NIR spectroscopy to agricultural products, in *Handbook of Near-Infrared Analysis*, ed. by Burns DA and Ciurczak E. Marcel Dekker, Basel, pp. 419 – 474 (2008).

30. Pérez-Marín D, Garrido-Varo A and Guerrero JE, Optimization of discriminant partial least squares regression models for the detection of animal by-product meals in compound feeding stuffs by near-infrared spectroscopy. *Appl Spectrosc* **60**:1432 – 1437 (2006).

31. Osborne BG, Fearn T and Hindle P, *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Longman Scientific and Technical, London (1993).

32. Downey G, Fouratier V and Kelly JD, Detection of honey adulteration by addition of fructose and glucose using near infrared transflectance spectroscopy. *J Near Infrared Spectrosc* **11**:447 – 456 (2003).

33. Nicolaï BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI and Lammertyn J, Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol Technol* **46**:99 – 118 (2007).

Table 4.1.1. Dry matter and SSC in outdoor-grown and greenhouse-grown bell peppers.

Growing system	Parameter	
	Dry matter (% fw)	**SSC (ºBrix)**
Outdoor	7.02 (1.30) [a]	6.38 (1.38) [a]
Greenhouse	6.63 (1.10) [b]	6.37 (1.30) [a]

[1] *Standard deviations in brackets*

[2] *Different letters in the same column indicate statistical significance (P < 0.05)*

Table 4.1.2. Number of samples (N), range, mean, standard deviation (SD), and coefficient of variation (CV) of the quality parameters for the different training and validation sets used in the discriminant models for the authentication of bell peppers by growing system.

	Training set								Validation set			
	Unbalanced model				Balanced model							
	Outdoor		Greenhouse		Outdoor		Greenhouse		Outdoor		Greenhouse	
	DM[1]	SSC[2]	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC
N	100	100	231	231	100	100	100	100	28	28	28	28
Range	4.48-11.37	3.85-9.50	4.52-9.94	3.85-10.05	4.48-11.73	3.85-9.50	4.74-9.24	3.90-10.05	4.74-9.49	4.20-9.15	5.05-8.67	4.50-8.50
Mean	7.03	6.44	6.65	6.41	7.03	6.44	6.51	6.23	7.07	6.19	6.48	6.11
SD	1.27	1.38	1.13	1.34	1.27	1.38	1.07	1.35	1.42	1.41	0.93	1.07
CV (%)	18.07	21.43	16.99	20.90	18.07	21.43	16.44	21.67	20.08	22.78	14.35	17.51

[1] *DM: Dry matter (% fw)*

[2] *SSC: Soluble solid content (ºBrix)*

Table 4.1.3. Discriminant models for the authentication of bell peppers by growing system. PLS-DA.

Qualitative Group	**Unbalanced model**		**Balanced model**	
	Percentage of correctly-classified samples: 89.73% (297/331) Model SECV: 0.32 Number of synthetic variables: 11 Mathematical treatment: 1,5,5,1-SNV+DT		Percentage of correctly-classified samples: 88.00% (176/200) Model SECV: 0.35 Number of synthetic variables: 9 Mathematical treatment: 1,5,5,1-SNV+DT	
Growing system	**Training set**	**Validation set**	**Training set**	**Validation set**
Outdoor	74.00% (74/100)	78.57% (22/28)	88.00% (88/100)	85.71% (24/28)
Greenhouse	96.54% (223/231)	100.00% (28/28)	88.00% (88/100)	96.43% (27/28)

Table 4.1.4. Dry matter and SSC in outdoor-grown and greenhouse-grown bell peppers.

Growing system	Parameter					
	Dry matter (% fw)			SSC (ºBrix)		
	Green	Yellow	Red	Green	Yellow	Red
Outdoor	5.91 (0.70) [b]	7.46 (0.96) [d]	7.96 (1.16) [e]	5.06 (0.63) [a]	6.77 (0.81) [b]	7.61 (1.00) [c]
Greenhouse	5.62 (0.59) [a]	6.58 (0.67) [c]	7.69 (0.83) [d.e]	4.92 (0.49) [a]	6.62 (0.65) [b]	7.59 (0.88) [c]

[1] *Standard deviations in brackets*

[2] *Means with different superscripts differ significantly (P < 0.05)*

Table 4.1.5. Number of samples (N), range, mean, standard deviation (SD), and coefficient of variation (CV) of the quality parameters for the different training and validation sets used in the discriminant models for the authentication of bell peppers of different colours by growing system.

	Green bell peppers								Yellow bell peppers								Red bell peppers							
	Training set				Validation set				Training set				Validation set				Training set				Validation set			
	Outdoor		Greenhouse		Outdoor		Greenhouse		Outdoor		Greenhouse		Outdoor		Greenhouse		Outdoor		Greenhouse		Outdoor		Greenhouse	
	DM[1]	SSC[2]	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC	DM	SSC
N	45	45	83	83	5	5	5	5	33	33	83	83	5	5	5	5	35	35	83	83	5	5	5	5
Range	4.48-7.68	3.85-7.90	4.52-7.23	3.85-6.10	5.26-7.33	4.60-6.00	5.33-6.21	4.65-5.60	4.97-9.64	4.80-8.35	5.14-8.33	5.30-8.40	7.04-8.63	6.40-7.40	5.53-7.69	5.90-8.25	5.11-11.73	5.05-9.20	5.66-9.50	5.60-10.05	8.09-8.82	7.30-9.50	6.80-9.94	6.70-8.80
Mean	5.90	5.06	5.61	4.92	6.07	5.10	5.76	4.95	7.43	6.80	6.56	6.60	8.05	6.93	6.80	6.91	7.87	7.47	7.67	7.57	8.28	8.38	7.98	7.80
SD	0.69	0.75	0.60	0.50	0.83	0.58	0.35	0.38	0.93	0.82	0.66	0.64	0.64	0.42	0.85	0.87	1.22	0.98	0.81	0.88	0.30	0.88	1.23	0.87
CV (%)	11.68	14.93	10.74	10.21	13.75	11.43	6.15	7.69	12.56	12.01	10.12	9.71	8.00	6.03	12.56	12.60	15.50	13.10	10.52	11.67	3.66	10.49	15.37	11.18

[1] *DM: Dry matter (% fw)*

[2] *SSC: Soluble solid content (ºBrix)*

Table 4.1.6. Discriminant models for the authentication of bell peppers of different colours by growing system. PLS-DA.

Qualitative Group	**Green bell peppers**		**Yellow bell peppers**		**Red bell peppers**	
	Percentage of correctly-classified samples: 88.28% (113/128)		Percentage of correctly-classified samples: 91.37% (106/116)		Percentage of correctly-classified samples: 90.67% (107/118)	
	Model SECV: 0.34		Model SECV: 0.35		Model SECV: 0.33	
	Number of synthetic variables: 3		Number of synthetic variables: 10		Number of synthetic variables: 3	
	Mathematical treatment: 2,5,5,1-SNV+DT		Mathematical treatment: 1,5,5,1-SNV+DT		Mathematical treatment: 1,5,5,1-SNV+DT	
Growing system	**Training set**	**Validation set**	**Training set**	**Validation set**	**Training set**	**Validation set**
Outdoor	77.77% (35/45)	100.00% (5/5)	87.87% (29/33)	100.00% (5/5)	77.77% (28/35)	80.00% (4/5)
Greenhouse	93.97% (78/83)	100.00% (5/5)	92.77% (77/83)	80.00% (4/5)	93.97% (79/83)	100.00% (5/5)

Table 4.1.7. Number of samples (N), range, mean, standard deviation (SD), and coefficient of variation (CV) for the calibration and validation sets and standard error of laboratory (SEL).

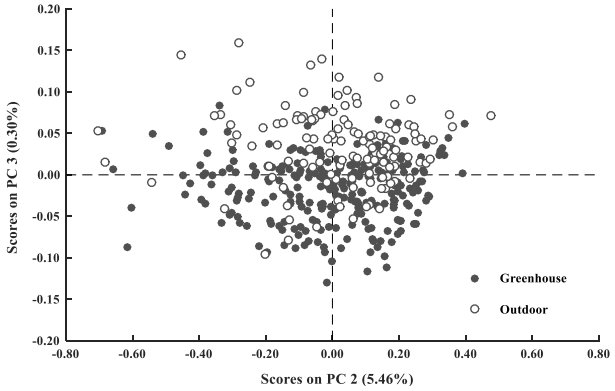
Parameter	Calibration set					Validation set					SEL
	N	Range	Mean	SD	CV	N	Range	Mean	SD	CV	
Dry matter (% fw)	262	4.48-11.73	6.78	1.16	17.10	130	4.52-9.64	6.68	1.20	17.96	0.21
SSC (ºBrix)	262	3.85-10.05	6.39	1.29	20.18	130	3.85-10.05	6.33	1.39	21.95	0.06

Table 4.1.8. Calibration statistics for NIR-based models for predicting quality parameters in intact bell peppers.

Parameter	Math treatment	Mean	SD	SECV	r^2_{cv}	RPD_{cv}	CV (%)
Dry matter (% fw)	1,5,5,1-SNV+DT	6.72	1.08	0.66	0.62	1.64	9.82
SSC (ºBrix)	1,5,5,1-SNV+DT	6.31	1.24	0.75	0.63	1.65	11.88

Figure 4.1.1. Scores plot (a) and loadings weight (b) for the second (PC2) and third (PC3) principal components for intact bell peppers.

a)



b

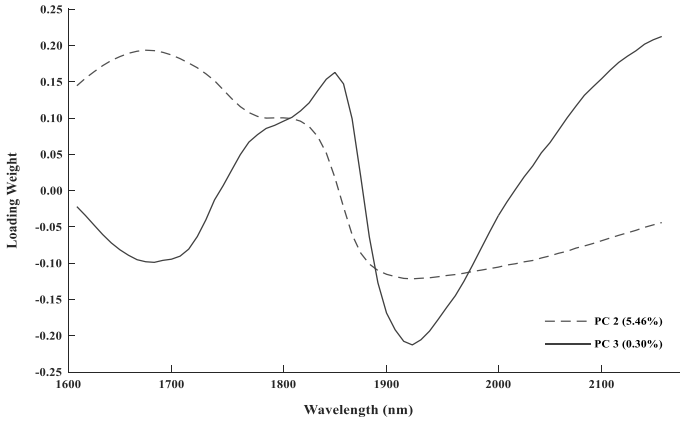


Figure 4.1.2. Values of the discriminatory variable obtained for the different validation groups. Unbalanced and balanced models.

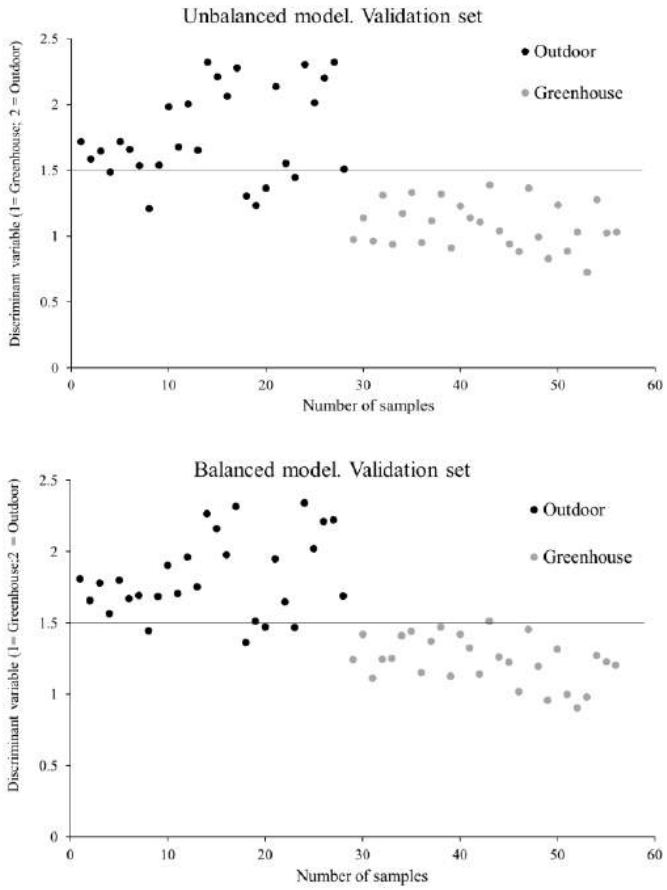


Figure 4.1.3. Regression coefficients for the bell pepper discriminant analysis. Balanced model.

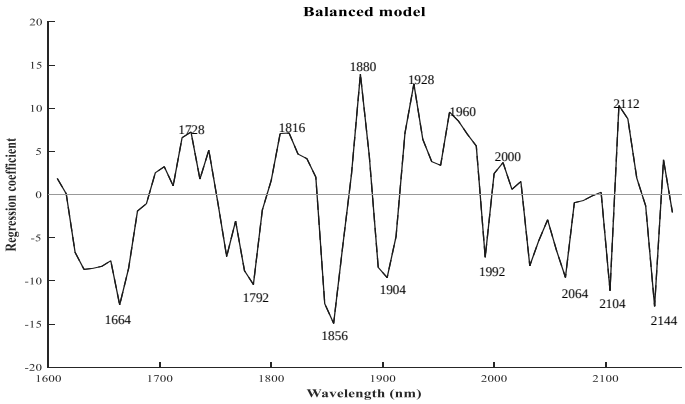


Figure 4.1.4. Reference *versus* NIR predicted concentration of dry matter (a) and SSC (b) in bell pepper.

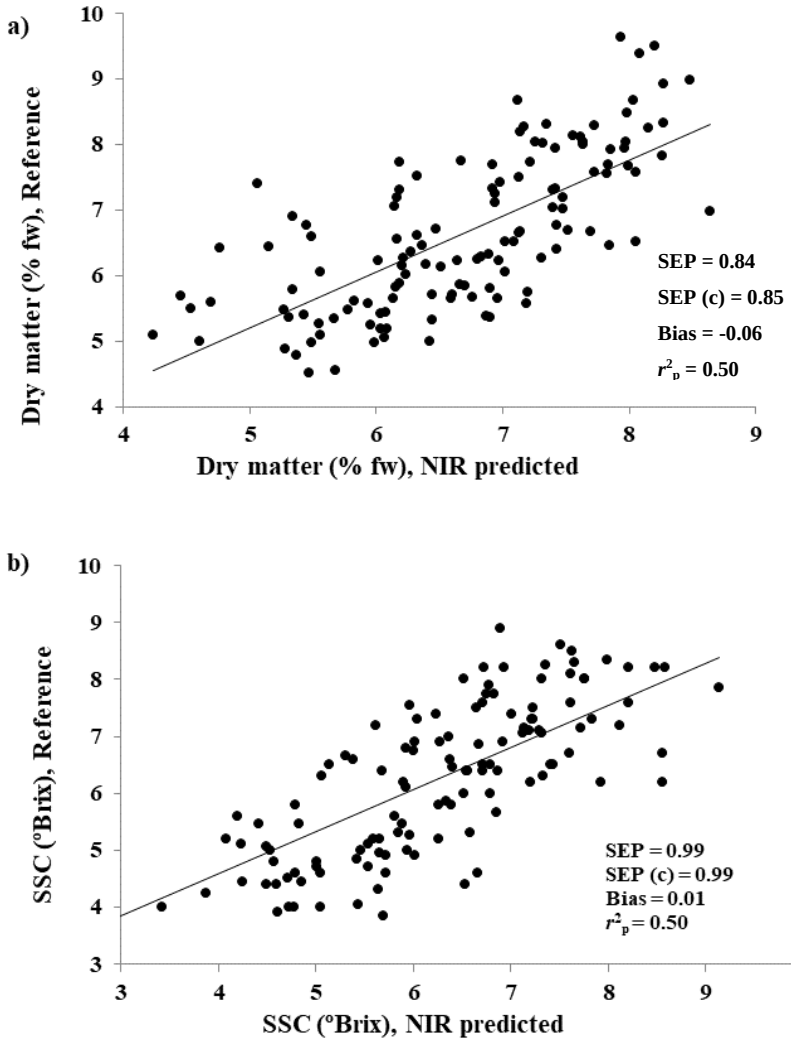
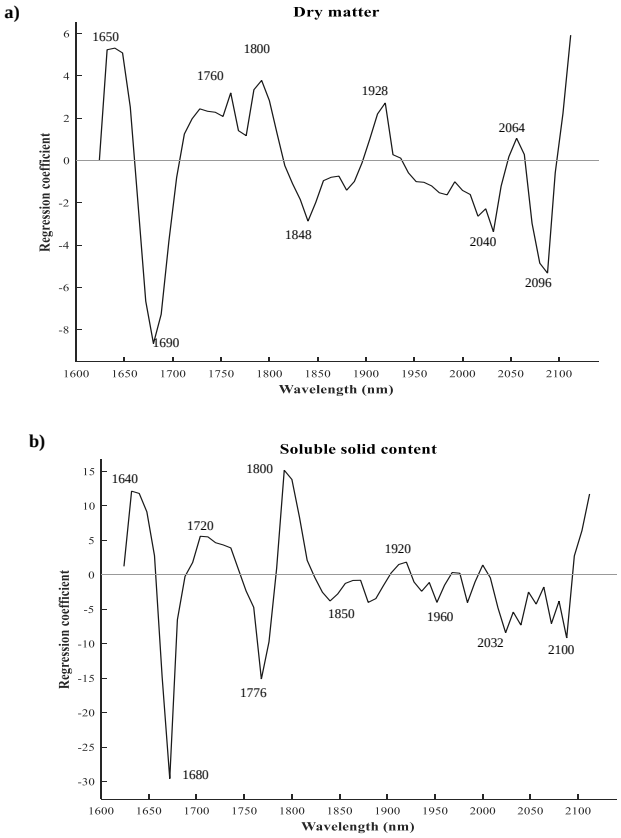


Figure 4.1.5. Regression coefficients for bell pepper dry matter and soluble solid content.



Chapter 5

Chapter 5. DEVELOPMENT AND EVALUATION OF NIRS PREDICTIVE MODELS FOR IRRIGATION DECISION SUPPORT, ENABLING OPTIMAL AND PRECISE DECISION-MAKING AT FIELD LEVEL

Chapter 5.1.

Irrigation decision support based on leaf relative water content determination in olive grove using near infrared spectroscopy

Irina Torres[a], María-Teresa Sánchez[a], María Benlloch-González[b], Dolores Pérez-Marín[c]

[a] *Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*
[b] *Department of Agronomy, Faculty of Agricultural and Forestry Engineering, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*
[c] *Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*

Biosystems Engineering 180, 50-58 (2019)

Abstract

The relative water content (RWC) provides a measurement of the water deficit of the leaf and may indicate a degree of stress endured under conditions of drought and high temperatures, its measurement therefore, being essential for the appropriate management of irrigation. This study sought to ascertain the viability of near infrared spectroscopy (NIRS), using a handheld portable NIR instrument for the non-destructive and *in situ* determination of RWC in olive tree leaves cultivated under higher temperatures than ambient. Different combinations of pre-treatments and first and second derivative were assayed to obtain information of spectral data and to develop calibration models. A calibration equation with enough prediction performance for supporting irrigation decision-making (standard error of cross-validation, SECV = 1.52%; r^2_{cv} = 0.61; residual predictive deviation for cross-validation, RPD_{cv} = 2.01) was obtained. The findings obtained from the external validation of the model (standard error of prediction, SEP = 1.63%; r^2_p = 0.64; residual predictive deviation for prediction, RPD_p = 2.17) suggest the viability of the on-tree use of NIRS technology for the instant measurement of RWC in olive groves, ensuring a major saving in time and avoiding the disadvantage of transporting samples to the lab, thereby favouring real-time decision-making in the field regarding the optimal amounts of irrigation to be applied; this is of enormous significance for the future, given that the availability of irrigation water for such vital crops to the Mediterranean region as the olive could be limited in years to come by a gradual increase in planetary temperatures.

Keywords: Olive grove; *In situ* RWC measurement; NIRS technology; Irrigation management; Climate change

5.1.1. Introduction

Olive (*Olea europaea* L.) is the most prevalent crop in the Mediterranean basin and has enormous ecological and economic importance to the region. It is well suited to the Mediterranean climate, which is characterised by hot and dry summers, mild winters, and relative lack of rainfall. However, the climatic conditions of this region are expected to change in the near future due to global warming. Climate experts have predicted an increase in average air temperature in the range of 2–5ºC (Giorgi, 2006; Gualdi et al., 2013; IPCC, 2014) together with more frequent occurrence of extreme events such as droughts and heat-waves (Giorgi & Lionello, 2008; Tanasijevic, Todorovic, Pereira, Pizzigally, & Lionello, 2014). In the climatic conditions being predicted for the region therefore – lower precipitation and higher temperatures – it is likely that this species will undergo frequent periods of water and heat stress, with concomitant effects on yields.

The leaf is the organ of the olive tree that is most responsive to environmental conditions (Nevo et al., 2000). The RWC of a leaf is an important indicator of a plant's water status. In this sense, RWC provides a measurement of the 'water deficit' of the leaf and may indicate a degree of stress expressed under unfavourable conditions such as drought or high temperature (Barrs & Weatherley, 1962; Barrs, 1968). This parameter has long been used as a reliable indicator of plant wellbeing and could be highly useful in ascertaining whether olive trees subjected to the climate conditions of the future are suffering from stress at any of the phenological stages of their reproductive cycle (Mullan & Pietragalla, 2012; Rallo & Cuevas, 2017). It can be useful for indicating plant water needs (Jones, 2004, 2007) aimed at reducing potential stressful situations for olive trees, especially in those phenological stages where the species is more vulnerable to extreme conditions.

The traditional method used to determine RWC is by measuring the differences in weight between the fresh, dry and turgid leaf (Stocker, 1929). This method is time-consuming– it requires more than 24 hours – and labour intensive in the laboratory. Moreover, although the procedure is straightforward, the taking of samples is prone to errors, because it can be accompanied by a modification in the water content prior to the start of the analysis. There is therefore a need for a fast and efficient method for the determination of the RWC in a way that is non-destructive and *in situ* (on-tree), allowing growers to make accurate irrigation decisions depending on the water deficit of the tree. In this context, near-infrared spectroscopy has significant potential as an appropriate method, since it is a non-invasive, rapid, economical and accurate alternative to traditional methods. The technology is simple, so fewer errors are introduced than in conventional analytical techniques (Osborne, Fearn, & Hindle, 1993). At the same time, NIR spectroscopy is a powerful tool for general process monitoring in real time (De la Roza et al., 2017; Zhang et al., 2017); this is of particular interest for many agricultural practices such as irrigation.

NIRS technology has been successfully used to determine various parameters in the leaves of a range of species, using both laboratory (Menesatti et al., 2010; Fernández-Martínez et al., 2017) and portable equipment (Itoh, Tomita, Uno, & Naomasa, 2011; Steidle-Neto, Lopes, Pinto, & Zolnier, 2017). In the case of olive leaves, the research that has been published makes reference to measuring nutrient content (Fernández-Cabanás, Garrido-Varo, Delgado-Pertiñez, & Gómez-Cabrera, 2008; Rotbart et al., 2013) and differentiation between juvenile and adult leaves (León & Downey, 2006), both carried out in laboratory conditions. However, there is no trace in the scientific literature of any research into the measurement of RWC in olive leaves using NIRS technology. Several authors have demonstrated the feasibility of NIRS technology in the non-destructive measurement of RWC in the fresh leaves of *Epipremnum aureum* and *Miscanthus (M. sinensis, M. sacchariflorus, M. lutarioriparia, M. floridulus and M. giganteus)* (Zhang, Li, & Zhang, 2012; Jin,

Shi, Yu, Yamada, & Sacks, 2017) using lab-based monochromator instruments, and in seedling eucalyptus leaves using a portable MEMS-NIRS instrument (Warburton, Brawner, & Meder, 2014).

The aim of this study was to evaluate the feasibility of using NIRS technology for determining RWC in olive leaves growing *in situ*. The goal is to help growers to make irrigation decisions to mitigate negative effects of stress on crop performance under future weather conditions associated to climate change.

5.1.2. Material and methods

5.1.2.1. Plant material

Olive (*Olea europaea* L.) leaves from cultivars 'Picual' (N = 178 samples) and 'Arbequina' (N = 72 samples) were analysed. Each sample consisted of four fully-expanded leaves, which were located at the middle position of the canopy and exposed to sunlight. Samples were sequentially collected from March 2016 to July 2017 on 17 different days, covering the range of the distinct phenological phases of the olive tree (Table 5.1.1).

These olive trees were located in an experimental field at the Rabanales Campus of Córdoba University (Spain) and exposed to different temperature treatments (ambient temperature *versus* 4 ºC above ambient temperature), by the use of open top chambers equipped with heating and ventilation devices. These systems are able to maintain permanently a day/night temperature gradient between the tree and the surrounding environment of 4 °C throughout the complete reproductive cycle of this species (Benlloch-Gónzalez, Sánchez-Lucas, Benlloch, & Fernández-Escobar, 2018).

5.1.2.2.NIRS analysis

A handheld Micro-Electro-Mechanical System (MEMS) spectrometer (MicroPHAZIR™, Thermo Fisher Scientific, Wilmington, MA, USA) was

used to collect the spectra of olive leaves *in-situ*. This instrument operates in reflectance mode (log 1/R) across the spectral range of 1600–2400 nm every 8 nm. Internal white reference was automatically collected every ten minutes.

Olive tree leaves are small and thick, so in order to avoid the loss of light during spectra collection and to ensure that the field analysis was carried out correctly, without detaching the leaf from the tree, a circular (15 cm of diameter) black metal plate was used to hold the leaf.

At first, with the aim of establishing which side of the leaf was most appropriate for recording spectra, NIRS readings were carried out both on the adaxial and abaxial side of the leaf. Three spectral measurements were made per leaf (at the upper, middle and bottom parts) and per side (adaxial and abaxial). Since, four leaves were analysed per each olive tree, and a total of 12 spectra were obtained for each sample and for each leaf side. These 12 spectra per side were averaged to provide a mean spectrum for each olive tree, a mean for each sample and, initially, for each side.

5.1.2.3. Reference method

RWC was determined in accordance with the procedure set out by Stocker (1929). Briefly, leaves were collected at solar noon and quickly put inside a 10 ml-test tube, which was hermetically sealed with a lid and placed in a container filled with ice to avoid loss of leaf moisture. Once in the laboratory, the olive leaves were weighed (FW) and then rehydrated by adding 1 ml of deionised water to the test tube. After incubation at 4 ºC for 24 h, the leaves were re-weighed to determine the turgid weight (TW) and thereafter put into an oven at 70 ºC for 48 h to determine the dry weight (DW). The leaf RWC (%) was calculated as follows:

$$RWC\ (\%) = ((FW-DW)/(TW-DW))\ x\ 100$$

For the purposes of this research the Standard Error of Laboratory (SEL) was estimated by analysing 10 duplicated samples. In order to calculate the error, both the sampling error (selection of two consecutives leaves to analyse) and the error arising from the process of analysis in the laboratory (analysis was done by duplicated) were determined. Once these two errors had been calculated, the SEL value was obtained in accordance with Fearn (1986).

5.1.2.4. Spectral repeatability

The spectral repeatability was evaluated using the root mean squared (RMS) statistic, is defined as the averaged root mean square of differences between the different subsamples scanned at n wavelengths (Shenk & Westerhaus, 1995a, 1996). It indicates the similarity between different spectra of a single sample, in this case between the three spectra collected per sample. For this purpose, 10 leaves were selected from which three spectra were taken in the upper, middle and lower parts using the MEMS-NIR instrument.

An admissible limit for spectrum quality and repeatability was set following the procedure described by Martínez, Garrido, De Pedro and Sánchez, (1998) to calculate the standard deviation (STD) limit from the RMS statistic and obtain an RMS cut-off value.

5.1.2.5.Data processing

5.1.2.5.1. Principal component analysis

With the goal of studying the relationship between the RWC and the distinct phenological states in the olive tree's cycle, as well as conducting the possible identification of anomalous samples, Principal Component Analysis (PCA) was carried out. In this work, PCA was performed using the mean spectrum derived from each of the days being analysed.

Matlab software (version 2015a, The Mathworks, Inc., Natick, Massachusetts, US) was used to conduct PCA, using mean centre, which

subtracts the mean spectrum of the group from each spectrum, as a pre-treatment (Wise et al., 2006).

5.1.2.5.2. Selection of the calibration and validation sets

Data pre-processing and chemometric treatments were performed using the WinISI software package ver. 1.50 (Infrasoft International LLC, Port Matilda, PA, USA). For the development of the model, the total set was divided into a calibration and a validation set. The selection of these sets was based on spectral information, using the CENTER algorithm (Shenk & Westerhaus, 1995a).

As spectral pre-treatments, Standard Normal Variate (SNV) and Detrending (DT) were used to remove scatter interferences (Barnes, Dhanoa, & Lister, 1989) together with the first derivative treatment '1,5,5,1', where the first digit is the number of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing, and the fourth is the second smoothing (Shenk & Westerhaus, 1995b).

Having ordered the population by spectral distances, samples that displayed GH values > 3 were removed. The validation set was selected by taking one sample out of every four in the initial set; the remainder constituted the calibration set.

5.1.2.5.3. Calibration development and validation procedure

Calibration models for the prediction of the RWC of the olive leaf were developed using Modified Partial Least Squares (MPLS) regression (Shenk & Westerhaus, 1995a) with six cross-validation groups to avoid overfitting. SNV and DT and Multiplicative Scatter Correction (MSC) were used as pre-processing for scatter correction (Barnes et al., 1989; Dhanoa, Lister, Sanderson, & Barnes, 1994). Additionally, four derivative mathematical treatments were tested: 1,5,5,1; 1,10,5,1; 2,5,5,1; and 2,10,5,1.

Best equations were selected according to the following statistics: coefficient of determination for calibration (r^2_c), standard error of calibration (SEC), coefficient of determination for cross-validation (r^2_{cv}) and standard error of cross-validation (SECV). However, in order to standardise the SECV value, another statistic, the residual predictive deviation (RPD), calculated as the ratio between the standard deviation (SD) of the calibration set to the SECV, was also calculated.

The best model obtained for the calibration set, as selected by statistical criteria, was subjected to external validation and evaluated in accordance with the protocol outlined by Windham, Mertens, and Barton (1989).

5.1.3. Results and discussions

5.1.3.1. Optimisation of in-situ olive tree analysis

After the spectra taken from both sides of the leaf at the beginning of the study, it was decided to take spectra only from the adaxial side, because the leaf of the olive has a highly-pronounced central vein on the abaxial side, causing greater dispersion of light during analysis. The procedure of taking spectra only from the adaxial side of the leaf is consistent with the practice of such authors as Zhang et al. (2012) in *Epipremnum aureum*, and Warburton et al. (2014) and Yang et al. (2017). in *Eucalyptus* leaves. Specifically, the study carried out by Warburton et al. (2014) on *Eucalyptus* seedlings, aimed at determining which side of the leaf was most appropriate for NIRS analysis, concluded that there were no significant differences enabling a particular part of the leaf to be established for recording spectra, although it is important to note that *Eucalyptus* leaves do not exhibit the very prominent central vein that is a feature of olive leaves.

After that and prior to the model development, it was necessary to optimise the NIRS analysis by means of the spectrum quality and repeatability measurement.

Firstly, the existence of noise in the spectrum was evaluated (spectral range 1600–2400 nm). To this end, the derivative treatment 1,1,1,1 was applied in order to determine the area of the spectral range affected by noise, given that it degrades the signal/noise relationship (Hruschka, 2001). After this process, the spectral range between 2312–2400 nm was eliminated (Figure 5.1.1).

Secondly, spectral repeatability which is crucial to the construction of models that are both accurate and robust was evaluated. Statistical methods such as defined RMS cut-off limit can be useful for this purpose. The RMS cut-off was calculated as described in Section 2.4.

The STD_{limit} for the samples analysed using the handheld instrument was 42,663 µlog (1/R). Despite the importance of this parameter for fine-tuning new analytical methodologies and ensuring more robust models, no references have been found in the scientific literature that calculate STD_{limit} for the *in situ* analysis of olive leaves. In the present research, any sample whose triplicated screening scans yielded an RMS above this value was eliminated and repeated until values fell below that limit, thus ensuring a high degree of spectrum repeatability. It was found for example that the samples taken on 16 March 2017 exhibited values far higher than the established STD_{limit}, despite the analysis of the leaves being repeated on numerous occasions. A detailed study was carried out of the various factors that could have affected the analysis on that day, arriving at the conclusion that the variation arose from the fact that a few days prior to the analysis a copper-based treatment was been applied, with the consequence that the particles deposited on the leaves caused the analysis to be distorted. The samples taken on that particular day were therefore eliminated, leaving a set consisting of the 235 remaining samples.

5.1.3.2. Principal Component Analysis (PCA)

PCA was performed on the set comprising the spectra recorded per day (N = 16), after eliminating those mentioned in section 3.1. Figure 5.1.2(a) shows the PCA loadings for intact olive leaves in the spectral range 1600–2312

nm, while Figure 5.1.2(b) displays scores of the second and third components of the PCA model. These two components were chosen because although the first two principal components (PC1 and PC2) represented a high proportion of the explained variance (82.23% and 16.57%, respectively), they did not facilitate the grouping of the samples in accordance with the phenological state; this grouping does however seem to become evident when the latent variables PC2 and PC3 are used.

The graphic representation of the loadings for PC2 and PC3 shows that the main absorption peaks for differentiating between the various phenological states of the olive tree are those related to water and carbohydrates respectively. Whereas the PC2 weighting coefficient exhibits a peak of water around 1900 nm, PC3 exhibits a band that is characteristic of carbohydrates (~1780 nm) (Shenk et al., 2008). The accumulation of carbohydrates in the plant differs in accordance with the phenological state that the plant is in at that time; thus, during the period of fruit formation and ripening; nutrients and carbohydrates will migrate from the leaf towards the fruit, accumulating in the latter (Fernández-Escobar, Moreno, & García-Creus, 1999). It therefore follows that the carbohydrate content in the leaf, represented by the third principal component, aids discrimination between the states the plant happens to be in.

Score plotting revealed apparent grouping by phenological stages (Figure 5.1.2(b)), as shown in Table 5.1.1. Six groups emerge, which range from the period of winter dormancy to the maturation of the fruit, encompassing the intermediate phases of flowering, setting and growth of the fruit (Rallo & Cuevas, 2017).

In light of the PCA scores and bearing in mind the data set out in Table 5.1.1, it may be said that the phases of winter dormancy and flowering, which fundamentally occurs during the spring, when evapotranspiration is low (a rainy season), are related to PC2. The negative PC2 scores are associated with times of restricted water, which place the plants in situations of more acute

hydrological stress. As it has already been mentioned, PC3 may be linked to carbohydrate content. This becomes particularly evident when analysing the group pertaining to the swelling of the fruit, which exhibits a positive PC3 score (Figure 5.1.2(b)), setting it apart from the other samples and highlighting that in this phase there is a movement of carbohydrates from the plant's various organs towards the fruit, where it is subsequently assimilated (Fernández-Escobar et al., 1999).

León and Downey (2006) used PCA to differentiate between young and adult leaves in olive trees. They proposed that water content and various chemical compounds, particularly pigments, were responsible for this separation between the various ages of the leaf. In accordance with these authors, the distinction between the various phenological states could be due to the water and carbohydrate content of the leaf, although a depth study of the spectral characteristics of each state of the plant should be considered in future research.

5.1.3.3. Population characterisation

After applying the CENTER algorithm to the overall set (N = 235), two samples were identified as anomalous spectra. Once spectral outliers were removed, a set consisting of 233 samples was used to develop calibration models. As described in section 2.5, the set was divided into a training set (N = 174) and a test set (N = 59).

The distribution and statistics of the calibration and validation sets (mean, SD and CV) for the RWC are shown in Figure 5.1.3. The structured selection based only on the spectral information treatments, such as CENTER algorithm, proved to be useful because the statistics for both sets were similar and the range in the calibration set encompassed the validation set.

Although *a priori* it may seem that the RWC parameter exhibits a wide range, both for the calibration (77.23–96.24 %) and for the validation set

(78.22–95.61 %), this parameter actually exhibits severely restricted variability, as is evident from the low coefficients of variation obtained (Figure 5.1.3). For the calibration set, 93% of the samples recorded an RWC of between 85% and 95%, while in the validation set 88% of the samples fell within this range, with very few samples (9 out of 174 and 5 out of 59 for the calibration and validation sets, respectively) recording RWC scores below 85%.

The low variability (CV_c = 3.37% and CV_v = 3.95%) is due to the RWC in olive-tree leaves not subjected to controlled water stress being around 90-95%, so this variation only derives from periods in which olives are suffering from water stress. Olive trees are drought tolerant, and leaves can reach extremely low relative water contents (75-80%) before losing turgor (Lo Gullo & Salleo, 1988). Therefore, values below 80% may correspond to extreme temperature events, which generally occur during the long dry season of the Mediterranean areas, where symptoms of dehydration are frequently observed and are generally associated with a low-potassium nutritional status (Fernández-Escobar, García, & Benlloch, 1994), something that was not applicable in the case of the current trial.

5.1.3.4. Calibration and validation for the prediction of the relative water content

Statistics for the best models obtained using the various pre-treatments to determine RWC in olive leaves measured on-tree are shown in Table 5.1.2.

According to Shenk and Westerhaus (1996) and Williams (2001), all models obtained enable classification of the RWC parameter between high, medium and low values ($0.50 < r^2_{cv} < 0.69$), being the best of them the one obtained using MSC and the first derivative of the spectrum (SECV = 1.52%; r^2_{cv}= 0.61; RPD_{cv} = 2.01).

In the present study, the estimated SEL was 0.87%. According to Fearn (1986), the SECV is determined not only by the SEL but also reflects the error

of the NIRS method and the chemometric method. If the value of SECV is less than two times the SEL of the reference method, the NIRS equation is fit for use (Windham et al., 1989), meaning that this would be considered as appropriate for use in the field.

In order to compare the results obtained here to those obtained by other authors in leaves, the RPD_{cv} statistic was used to standardise the SECV value.

No other results have been found for determining RWC in olive leaves. However, various authors have used the technique to determine this parameter in a range of crops, initially using monochromator instruments in the laboratory. Zhang et al. (2012) reported good predictive capability (RPD_{cv} = 2.73) in determining the RWC in *Epipremnum Aureum* subjected to various water stress treatments, using a monochromator instrument with a spectral range of 200–1100 nm and a resolution of 1 nm. Jin et al. (2017) reported superior results to those obtained here (RPD_{cv} = 2.75) for *Miscanthus* leaves, using a monochromator instrument for the NIRS analysis with a wide spectral range (400–2500 nm, every 2 nm). These authors also had a calibration set for the parameter being studied that exhibited greater variability (CV = 6.53%), compared to the present case (CV = 3.37%), something that enables more robust models to be obtained (Shenk, Westerhaus, & Berzaghi, 1997). It is important to point out that both studies mentioned above carried out their RWC determinations with NIRS in the laboratory, whereas in the present study the analysis was conducted directly on the tree, with the MEMS-NIR instrument previously described. Moreover, the difference in predictive capacity between the first two spectrophotometers and the handheld instrument may reflect differences in spectral ranges, spectral resolution and in measuring area; the MEMS device measures an area of only around 4 mm^2, whereas both monochromators scan the whole sample.

While there are no reports of the use of portable instruments to measure RWC in olive leaves, various authors have used this type of instrument to

measure RWC in the leaves of *Eucalyptus* seedlings. Thus, Warburton et al. (2014) measured RWC using a MEMS-NIR (MicroPhazir^{TM} NIR spectrometer) instrument in the 1600–2400 nm spectral range; the results were better (r^2_{cv} = 0.88 and RER = 10.45) than those obtained here (r^2_{cv} = 0.61 RER = 12.51), possibly owing to the fact that they had a calibration set with a greater range (15.40–99.30%) than the one in the present study (77.23–96.24%). According to Fearn (2014), although r^2_{cv} can be useful for studying the predictive capability of the model, this is closely linked to the range of reference values, and this may provide a reason why the aforementioned authors reported a higher determination coefficient than that obtained here. In a similar trial and using a NIRS instrument that worked in the same spectral range (1600–2400 nm), Yang et al. (2017) obtained, for the *in situ* measurement of RWC in *Eucalyptus* seedlings, a predictive capability model (RPD_c = 2.59) that was slightly higher to the one obtained here (RPD_c = 2.09). This may be due to the fact that the authors in question had a calibration set with greater variability (SD = 6.33% and CV = 7.9%) than that in the present study (SD = 3.05% and CV = 3.37%), as well as the difficulties implicit in olive leaves in terms of thickness, sheen, enervation, etc., compared to *Eucalyptus* leaves, something that may have effect on NIRS analysis.

It should be noted that all these authors have conducted their experiments under controlled environmental conditions (temperature, humidity, irrigation, etc.), with situations involving induced water stress, thereby ensuring a set with a good and even coverage of the range. As Pérez-Marín, Garrido-Varo, and Guerrero (2005) point out, the distribution of samples within the calibration set is of great importance, because a uniform distribution throughout the range of the parameter being studied helps to obtain robust models.

Finally, Figure 5.1.4 shows the regression coefficients for the best predictive model for the RWC parameter. The figure illustrates that the areas of the spectrum with greater weight in the model are located around 1720 nm, related to the C-H stretch first overtone and around 1936 nm, which

corresponds to O-H bend second overtone (Osborne et al., 1993). This makes sense, because the RWC in olive leaves is very high, at around 90-95%. Furthermore, the area at around 2200 nm could be attributed to the C=O second overtone (Shenk, Workman & Westerhaus, 2008).

5.1.3.5. External validation procedure

After the development and analysis of the calibration models, the best model was subjected to external validation. For this purpose, a sample set not included in the calibration was used. Validation was performed using a set initially comprising 59 samples. Prior to the validation procedure, four samples were excluded from the validation set because they displayed values of RWC (78.22, 79.08, 95.60 and 95.61%) beyond the range obtained after the development of the equation (83.34–95.42%) for the parameter analysed. A graphic representation of the reference values *versus* the NIR predicted values for RWC in olive leaves is shown in Figure 5.1.5.

The model developed for the prediction of the RWC complies with the limit established in terms of r^2_p for its implementation in routine ($r^2_p > 0.60$), as well as the confidence control limits for bias and SEP(c). The SEP value obtained shows a minor difference (0,09 %) compared to the SECV, and around 0.12% compared to the mean of the parameter, thereby confirming that the SECV provides a good estimate of the SEP (Shenk et al., 2008). In addition, the slope (slope = 1.09) also falls within the established slope values (0.90–1.1) (Windham et al., 1989).

These findings suggest that the NIRS equation obtained may be considered as a first step for the *in situ* measurement of RWC in olive leaves. This could eventually enable growers to ascertain the plant's degree of water stress in real time, and to take appropriate and informed decisions about the irrigation of the crop.

Under future scenarios, growers could use leaf RWC measures by NIRS technology to quickly determine *in situ* whether olive is suffering from water shortage, trying to prevent stressful conditions and supporting irrigation scheduling.

In a practical sense, the best strategy to follow is to make a protocol in which the value of RWC corresponding to each phenological stage and specie is established. Values of leaf RWC rapidly measured using NIRS technology which were below those indicate that irrigation treatments would be necessary. This would be an excellent complement to the different routine scanning usually made, such as soil water content and tree evotranspiration demand.

5.1.4. Conclusions

The results of this study, which used a handheld NIR spectrophotometer, confirmed the viability of NIRS technology for the measurement of RWC in olive leaves on the tree. Non-destructive and rapid determination of this parameter provides a quantitative measure of the hydration status of the olive tree in the field, enabling optimal and precise management of irrigation, something that will prove of great importance to olive cultivation in Mediterranean countries. Climate change forecasts are predicting major periods of drought and an increase in temperatures in the region, where water will become an increasingly scarce resource; this will make it imperative to be able to determine the RWC of olive trees with a view to maintaining the efficiency of photosynthesis and crop productivity.

Over the coming years, further studies will be needed in order to improve the calibration specificity, accuracy and robustness of this procedure.

Acknowledgement

The authors would like to thank Ms. Mª Carmen Fernández for her technical support. Furthermore, the authors wish to express their gratitude to the Spanish Ministry of Education, Culture and Sports for the support offered to Irina Torres Rodríguez in the form of the Training programme for Academic Staff (FPU).

References

Barnes, R.J., Dhanoa, M.S., & Lister, S.J. (1989). Standard Normal Variate Transformation and De-trending of near infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43, 772–777.

Barrs, H.D. (1968). Determination of water deficits in plant tissues. In T. T. Kozlowski, (Ed.), *Water deficits and plant growth* (Vol. 1, pp. 235–368). New York: Academic Press.

Barrs, H.D., & Weatherley, P.E. (1962). A re-examination of the relative turgidity technique for estimating water deficits in leaves. *Australian Journal of Biological Sciences*, 15, 413–428.

Benlloch-González, M., Sánchez-Lucas, R., Benlloch, M., & Fernández-Escobar, R. (2018). An approach to global warming effects on flowering and fruit set of olive trees growing under field conditions. *Scientia Horticulturae*, 240, 405–410.

De la Roza-Delgado, B., Garrido-Varo, A., Soldado, A., Arrojo, A.G., Valdés, M.C., Maroto, F., & Pérez-Marín, D. (2017). Matching portable NIRS instruments for *in situ* monitoring indicators of milk composition. *Food Control*, 76, 74–81.

Dhanoa, M.S., Lister, S.J., Sanderson, R., & Barnes, R.J. (1994). The link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) transformations of NIR spectra. *Journal of Near Infrared Spectroscopy*, 2, 43–47.

Fearn, T. (1986). Some statistical comments on the errors in NIR calibrations. *Analytical Proceedings Articles*, 23, 123–125.

Fearn, T. (2014). The overuse of R^2. *NIR News*, 25–32.

Fernández-Cabanás, V.M., Garrido-Varo, A., Delgado-Pertiñez, M., & Gómez-Cabrera, A. (2008). Nutritive evaluation of olive tree leaves by near-infrared spectroscopy: effect of soil contamination and correction with spectral pretreatments. *Applied Spectroscopy*, 62, 51–58.

Fernández-Escobar, R., García, T., & Benlloch, M. (1994). Estado nutritivo de las plantaciones de olivar en la provincia de Granada. *ITEA*, 90, 39–49.

Fernández-Escobar, R., Moreno, R., & García-Creus, M. (1999). Seasonal changes of mineral nutrients in olive leaves during the alternate-bearing cycle. *Scientia Horticulturae*, 82, 25–45.

Fernández-Martínez, J., Joffre, R., Zacchini, M., Fernández-Marín, B., García-Plazaola, J.I., & Fleck, I. (2017). Near-infrared reflectance spectroscopy allows rapid and simultaneous evaluation of chloroplast pigments and antioxidants, carbon isotope discrimination and nitrogen content in *Populus* spp. leaves. *Forest Ecology and Management*, 399, 227–234.

Giorgi, F. (2006). Climate change hot-spots. *Geophysical Research Letter*, 33, L08707, 1–4.

Giorgi, F., & Lionello, P. (2008). Climate change projections for the Mediterranean region. *Global and Planetary Change*, 63, 90–104.

Gualdi, S., Somot, S., Li, L., Artale, V., Adani, M., Bellucci, A., Braun, A., Calmanti, S., Carillo, A., Dell'Aquila, A., Déqué, M., Ruti, C., Sanna, A., Sannino, G., Scoccimarro, E., Sevault, F., & Navarra, A. (2013). The CIRCE simulations: regional climate change projections with realistic representation of the Mediterranean Sea. *Bulletin of the American Meteorological Society*, 94, 65–81.

Hruschka, W.R. (2001). Data analysis: Wavelength selection methods. In P.C. Williams, & K.H. Norris (Eds.), *Near-infrared technology in the agricultural and food industries* (pp. 35-55). St. Paul, MN: AACC, Inc.

Intergovernmental Panel on Climate Change (IPCC). (2014). Climate change 2014: Synthesis Report. In R.K. Pachauri & L.A. Meyer (Eds.), *Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change* (pp. 151). Geneva, Switzerland: IPCC.

Itoh, H., Tomita, H., Uno Y., & Naomasa, S. (2011). Development of method for nondestructive measurement of nitrate concentration in vegetable leaves by near infrared spectroscopy. *IFAC Proceedings Volumes*, 44, 1773–1778.

Jin, X., Shi, C., Yu, C.Y., Yamada, T., & Sacks, E.J. (2017). Determination of leaf water content by visible and near-infrared spectrometry and multivariate calibration in *Miscanthus*. *Frontiers in Plant Science*, 8(271), 1–8.

Jones, H.G. (2004). Irrigation scheduling: advantages and pitfalls of plant-based methods. *Journal of Experimental Botany*, 55, 2427–2436.

Jones, H.G. (2007). Monitoring plant and soil water status: established and novel methods revisited and their relevance to studies of drought tolerance. *Journal of Exerimental Botany*, 58, 119–130.

León, L., & Downey, G. (2006). Preliminary studies by visible and near-infrared reflectance spectroscopy of juvenile and adult olive (*Olea europea* L.) leaves. *Journal of the Science and Agriculture*, 86, 999–1004.

Lo Gullo, M.A., & Salleo, S. (1988). Different strategies of drought resistance in three Mediterranean sclerophyllous trees growing in the same environmental conditions. *New Phytologist*, 108, 267–276.

Mark, H. (2001). Data analysis: Multilinear regression and Principal Component Analysis. In D.A, Burns, & E.W. Ciurczak (Eds.), *Handbook of near-infrared analysis*, No. 8 (pp. 151–188). Florida, USA: CRC Press.

Martínez, M.L., Garrido, A., De Pedro, E.J., & Sánchez, L. (1998). Effect of sample heterogeneity on NIR meat analysis: the use of the RMS statistic. *Journal of Near Infrared Spectroscopy*, 6, 313–320.

Menesatti, P., Antonucci, F., Pallottino, F., Roccuzzo, G., Allegra, M., Stagno, F., & Intrigliolo, F. (2010). Estimation of plant nutritional status by Vis–NIR spectrophotometric analysis on orange leaves [*Citrus sinensis* (L) Osbeck cv Tarocco]. *Biosystems Engineering*, 105, 448–454.

Mullan, D., & Pietragalla, J. (2012). Leaf relative water content. In A. Pask, J. Pietragalla, D. Mullan, & M. Reynolds (Eds.), *Physiological breeding II: A field guide to wheat phenotyping* (pp. 25-27). Mexico: CIMMYT.

Nevo, E., Bolshakova, M.A., Martyn, G.I., Musatenko, L.I., Sytnik, K., Pavlíèek, T., & Beharav, A. (2000). Drought and light anatomical adaptive leaf strategies in three woody species caused by microclimatic selection at 'Evolution Canyon', Israel. *Israel Journal of Plant Sciences*, 8, 33–46.

Osborne, B.G., Fearn, T., & Hindle, P. (1993). *Practical NIR spectroscopy with applications in food and beverage analysis*. Harlow, UK: Addison-Wesley Longman Ltd.

Pérez-Marín, D., Garrido-Varo, A., & Guerrero, J.E. (2005). Implementation of LOCAL algorithm with near-infrared spectroscopy for compliance assurance in compound feedingstuffs. *Applied Spectroscopy*, 59, 69–77.

Rallo, L., & Cuevas, J. (2017). Fructificación y producción. In D. Barranco, R. Fernández-Escobar, L. Rallo (Eds.), *El cultivo del olivo* (pp. 147-189). Madrid: Mundi-Prensa.

Rotbart, N., Schmilovitch, Z., Cohen, Y., Alchanatis, V., Erel, R., Ignat, T., Shenderey, C., Dag, A., & Yermiyahu, U. (2013). Estimating olive leaf nitrogen concentration using visible and near-infrared spectral reflectance. *Biosystems Engineering*, 114, 426–434.

Shenk, J.S., & Westerhaus, M.O. (1995a). *Analysis of agriculture and food products by near infrared reflectance spectroscopy*. Silver Spring: Monograph, NIRSystems, Inc.

Shenk, J.S., & Westerhaus, M.O. (1995b). *Routine operation, calibration, development and network system management manual*. Silver Spring: NIRSystems, Inc.

Shenk, J.S., & Westerhaus, M.O. (1996). Calibration the ISI way. In A.M.C. Davies, & P. Williams (Eds.), *Near infrared spectroscopy: The future waves* (pp.198-202). Chichester: NIR Publications.

Shenk, J.S., Westerhaus, M.O., & Berzaghi, P. (1997). Investigation of a LOCAL calibration procedure for near infrared instruments. *Journal of Near Infrared Spectroscopy*, 5, 223–232.

Shenk, J.S., Workman, J., & Westerhaus, M. (2008). Application of NIR spectroscopy to agricultural products. In D.A. Burns, & E.W. Ciurczac (Eds.), *Handbook of near infrared analysis* (pp. 347-386). Boca Raton, FL: CRC Press, Taylor & Francis Group.

Steidle-Neto, A.J., Lopes, D.C., Pinto, F.A.C., & Zolnier, S. (2017). Vis/NIR spectroscopy and chemometrics for non-destructive estimation of water and chlorophyll status in sunflower leaves. *Biosystems Engineering*, 155, 124–133.

Stocker, O. (1929). Das wasserdefizit von gefässpflanzen in verschiedenen llimazonen. *Planta*, 7, 382–387.

Tanasijevic, L., Todorovic, M., Pereira, L.S., Pizzigalli, C., & Lionello, P. (2014). Impacts of climate change on olive crop evapotranspiration and irrigation requirements in the Mediterranean region. *Agricultural Water Management*, 144, 54–68.

Warburton, P., Brawner, J., & Meder, R. (2014). Technical Note: Handheld near infrared spectroscopy for the prediction of leaf physiological status in tree seedlings. *Journal of Near Infrared Spectroscopy*, 22, 433–438.

Williams, P.C. (2001). Implementation of near-infrared technology. In P.C. Williams, & K.H. Norris (Eds.), *Near-infrared technology in the agricultural and food industries* (pp. 145-169). St. Paul, MN: AACC, Inc.

Windham, W.R., Mertens, D.R., & Barton II, F.E. (1989). Protocol for NIRS calibration: sample selection and equation development and validation. In G.C. Martens, J.S. Shenk, & F.E. Barton II (Eds.), *Near infrared*

spectroscopy (NIRS): Analysis of Forage Quality, No. 643 (pp. 96-103). Washington, DC: USDA-ARS, US Government Printing Office.

Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, W., & Koch, R.S. (2006). *PLS_ToolBox 4.0. Manual for use with MATLAB (TM)* [Computer software]. Wenatchee, WA: Eigenvector Research, Inc.

Yang, G.L., Lu, Y.L., Luo, J.Z., Wang, C.B., Meder, R., Warburton, P., & Arnold, R.J. (2017). Monitoring water potential and relative water content in *Eucalyptus camaldulensis* using Near Infrared Spectroscopy. *Journal of Tropical Forest Science*, 29, 121–128.

Zhang, Q., Li., Q., & Zhang, G. (2012). Rapid determination of leaf water content using VIS/NIR spectroscopy analysis with wavelength selection. *International Journal of Spectroscopy*, 27, 93–105.

Zhang, Y., Luo, L., Li, J., Li, S., Qu, W., Ma, H., & Ye, X. (2017). In-situ and real-time monitoring of enzymatic process of wheat gluten by miniature fiber NIR spectrometer. *Food Research International*, 99, 147–154.

Table 5.1.1 - Olive phenological stages on date analysis.

Measurement date	Number of samples	Phenological stage	Mean temperature (ºC)	Mean relative humidity (%)	Leaf RWC (%)		
					Min	Max	Mean
1. 03/16/2016	10	Bud dormancy	14.40	74.10	89.81	93.31	92.40
2. 03/30/2016	15	Flower development	16.60	61.90	89.40	93.10	92.10
3. 04/06/2016	15	Flower development	17.60	54.80	90.40	95.10	93.20
4. 04/21/2016	5	Flower development	16.30	68.60	91.50	96.20	93.50
5. 04/28/2016	15	Flower development	19.20	64.80	91.70	95.80	93.50
6. 05/04/2016	8	Flower development	20.50	41.00	86.80	93.00	90.50
7. 05/23/2016	15	Fruit set	23.00	44.40	89.80	94.80	91.30
8. 06/09/2016	16	Fruit set	29.50	42.10	85.50	89.40	91.80
9. 06/15/2016	15	Fruit set	23.30	47.80	86.90	95.60	90.70
10. 06/20/2016	15	Fruit growth	27.70	37.30	89.00	95.40	91.70
11. 06/30/2016	15	Fruit growth	28.20	50.10	86.60	95.60	92.10
12. 07/26/2016	22	End of stone hardening	32.60	31.80	83.60	90.00	87.40

Table 5.1.1 - Olive phenological stages on date analysis (continuation).

Measurement date	Number of samples	Phenological stage	Mean temperature (ºC)	Mean relative humidity (%)	Leaf RWC (%)		
					Min	Max	Mean
13. 10/06/2016	22	Fruit ripening	24.30	52.90	77.20	88.70	84.10
14. 11/02/2017	22	Fruit ripening	18.80	67.10	84.60	92.40	89.00
15. 03/02/2017	15	Flower development	13.70	72.30	90.40	92.60	91.60
16. 03/16/2017	15	Flower development	12.70	65.60	90.90	95.50	92.40
17. 04/05/2017	10	Flower development	17.10	49.40	90.40	92.90	91.70

Table 5.1.2. MPLS regression statistics for NIR-based models for predicting RWC in olive leaves.

Scatter correction	Math treatment	N	Mean	SD	SEC	r^2_c	SECV	r^2_{cv}	RPD_{cv}
SNV + DT	1,5,5,1	167	90.82	2.55	1.68	0.57	1.78	0.52	1.71
	1,10,5,1	163	90.94	2.45	1.70	0.51	1.73	0.50	1.76
	2,5,5,1	163	90.71	2.50	1.51	0.64	1.58	0.61	1.93
	2,10,5,1	164	90.84	2.45	1.63	0.56	1.68	0.53	1.82
MSC	**1,5,5,1**	**161**	**90.79**	**2.42**	**1.46**	**0.64**	**1.52**	**0.61**	**2.01***
	1,10,5,1	163	90.81	2.45	1.58	0.59	1.63	0.57	1.87
	2,5,5,1	162	90.75	2.46	1.49	0.64	1.54	0.61	1.98
	2,10,5,1	166	90.85	2.54	1.82	0.49	1.86	0.47	1.64

** Best model for RWC prediction.*

Figure 5.1.1. First derivative spectra of olive leaves prior to removing the noise.

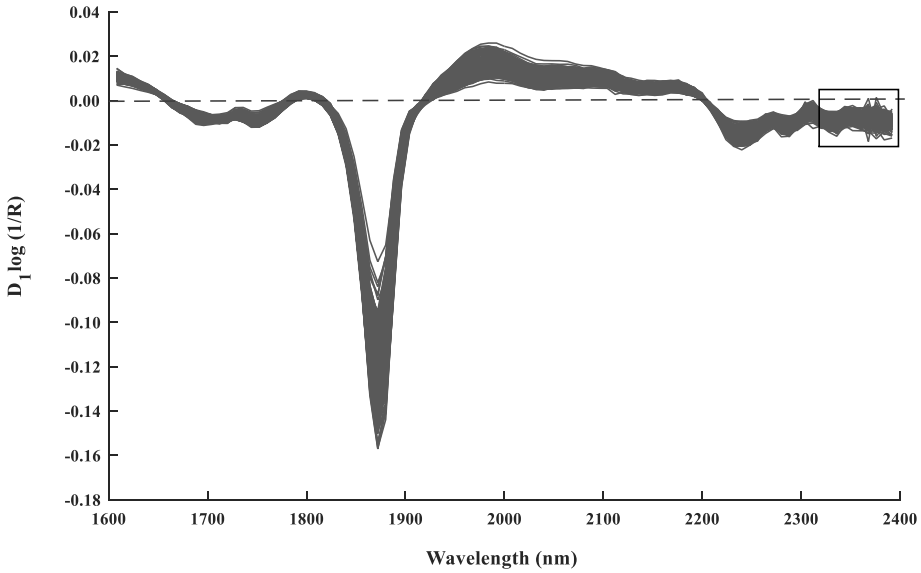
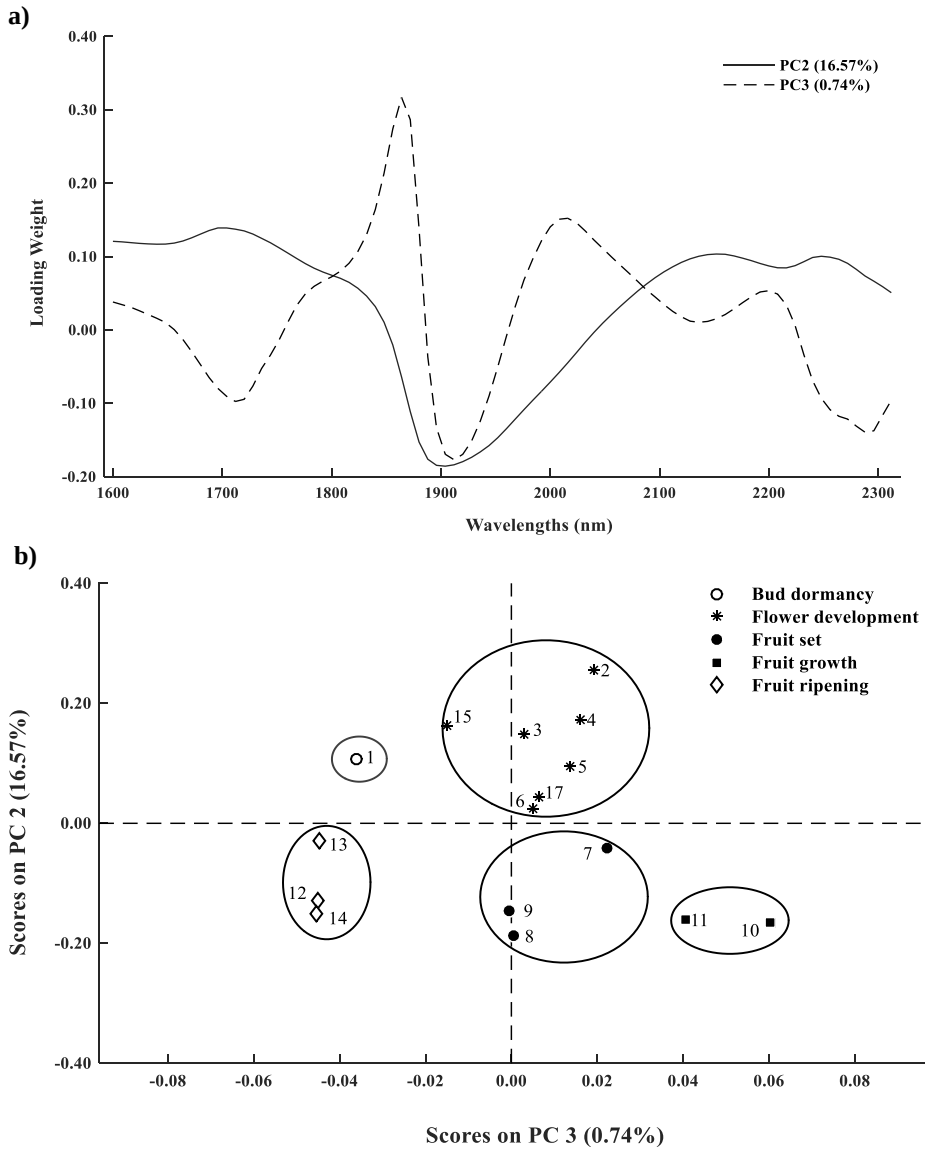


Figure 5.1.2. Loadings weight (a) and score plot (b) for the second (PC2) and third (PC3) principal components for olive leaf spectra.



**More information is displayed in Table 5.1.1.*

Figure 5.1.3. Calibration and validation sets structure for the RWC.

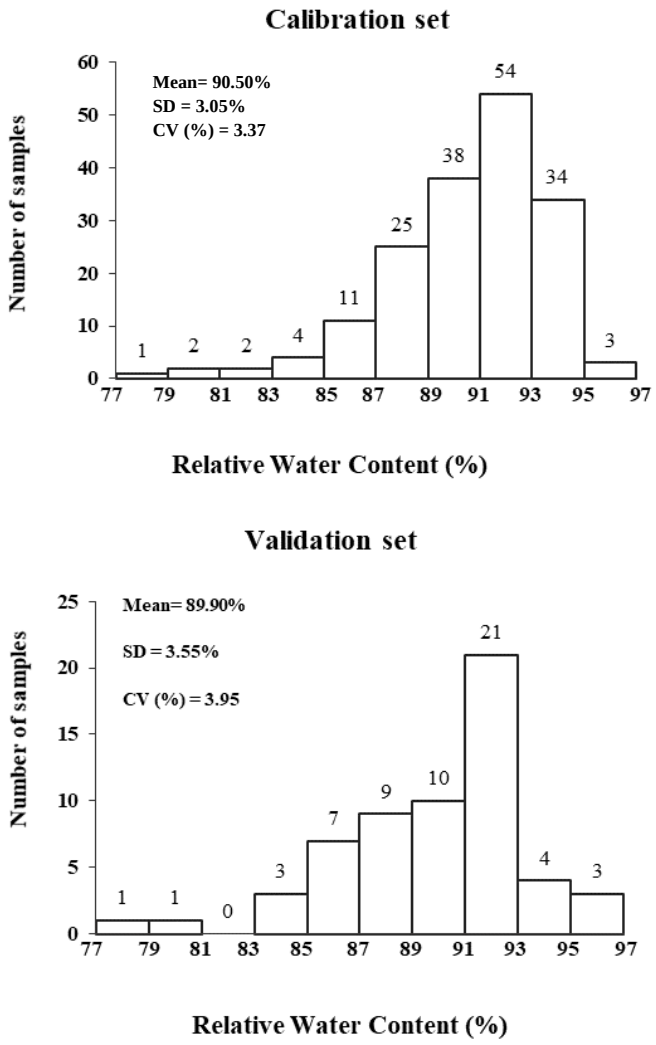


Figure 5.1.4. Regression coefficients for the RWC predictive model.

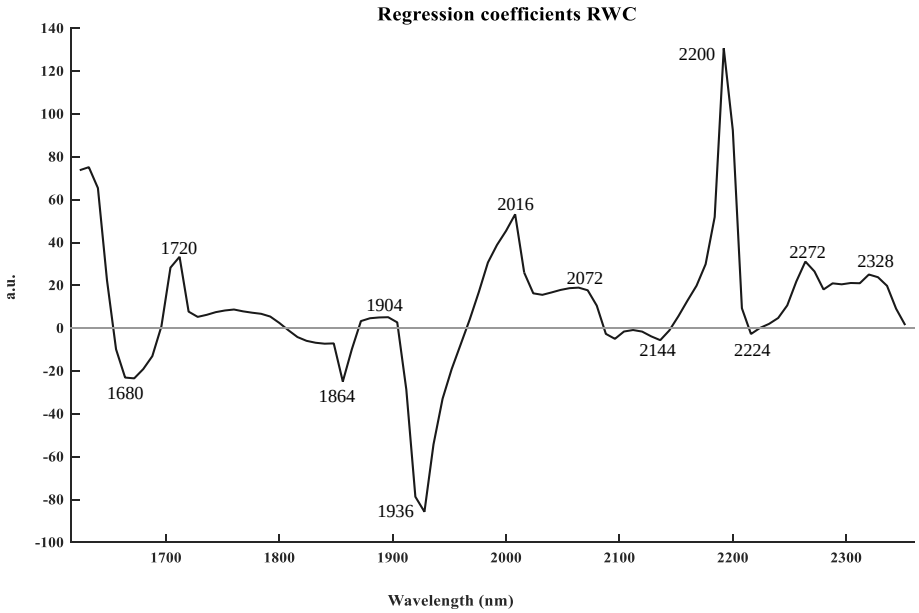
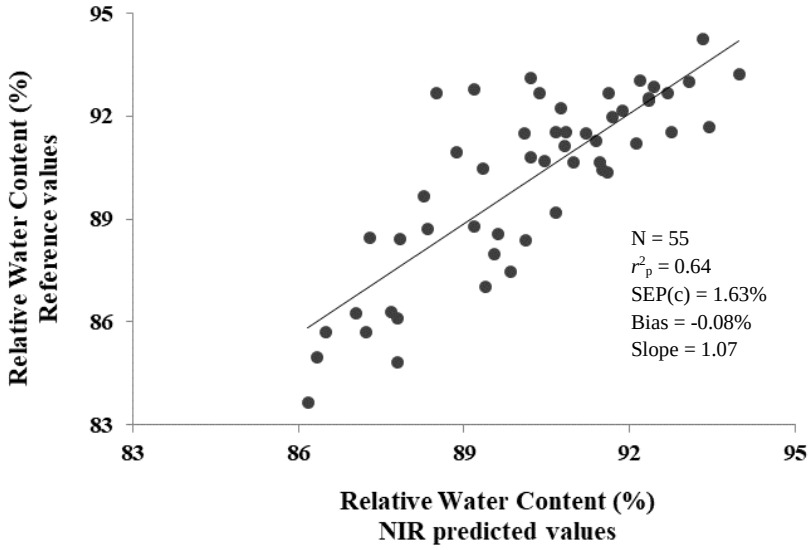


Figure 5.1.5. Reference vs. NIR predicted data for the validation set.



Chapter 6

Chapter 6. DEVELOPMENT AND EVALUATION OF ADVANCED CALIBRATION STRATEGIES FOR THE PREDICTION OF QUALITY PARAMETERS IN FRUITS OF THE CITRUS GENUS ANALYSED ON-TREE

Chapter 6.1.

Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors

Irina Torres[a], Dolores Pérez-Marín[b], María-José De la Haba[a], María-Teresa Sánchez[a]

[a] *Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*
[b] *Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*

Biosystems Engineering 153, 140-148 (2017)

Abstract

The citrus sector seeks rapid, economical, environmentally-friendly and non-destructive technologies for monitoring external and internal changes in physical quality taking place in fruit during on-tree growth, thus allowing fruit quality to be evaluated at any stage of fruit development. The use of portable near-infrared reflectance spectroscopy (NIRS) sensors based on micro-electro-mechanical system (MEMS) technology, in conjunction with chemometric data treatment models, has already been studied for quality-control purposes in two citrus species: oranges and mandarins. The critical challenge is to develop robust and accurate universal models based on hundreds of highly heterogeneous citrus samples in order to design quality prediction models applicable to all fruits belonging to the genus citrus, rather than models that can only be applied successfully to a single citrus species. This study evaluated and compared the performance of Modified Partial Least Squares (MPLS) and LOCAL regression algorithms for the prediction of major physical-quality parameters in all citrus fruits. Results showed that, while models developed using both linear (MPLS) and non-linear regression techniques (LOCAL) yielded promising results for the on-tree quality evaluation of citrus fruits, the LOCAL algorithm additionally increased the predictive capacity of models constructed for all the main parameters tested. These findings confirm that NIRS technology, used in conjunction with large databases and local regression strategies, increases the robustness of models for the on-tree prediction of citrus fruit quality; this will undoubtedly be of benefit to the citrus industry.

Keywords: NIRS; citrus; physical quality; universal models; MPLS regression; LOCAL algorithm

6.1.1. Introduction

Citrus fruits, which play a significant role in the human diet (Liu, Heying, & Tanumihardio, 2012), are among the world's major crops, and the highest value fruit crop in international trade (Magwaza et al., 2012b). The market price of harvested citrus fruit is at present based largely on external colour, size and mass (Olmo, Nadas, & García, 2000; Nicolaï et al., 2007; Magwaza et al., 2012a); it would be useful to introduce, in the near future, quality-based pricing systems, using both external (morphological and colour related parameters) and internal quality indices which are of particular interest to the citrus-fruit industry because they are linked to fruit yield.

In response to growing demand from producers, consumers and the industry, recent years have seen the development of rapid, accurate, economical and above all non-destructive technologies for determining food-produce quality. NIRS is one flexible and versatile technology, which has been successfully applied for the prediction of quality parameters in various citrus fruit species, and especially in oranges and mandarins. Numerous authors, including Fraser, Jordan, Künnemeyer, & McGlone (2003), Guthrie, Walsh, Reid, & Lienberg (2005a), Guthrie, Reid, & Walsh (2005b), Hernández-Gómez, He, & Pereira (2006), Sun, Zhang, & Liu (2009), Liu, Sun, Zhang, & Aiguo (2010b), Antonucci et al., (2011), Magwaza et al., (2012b, 2013b, 2014), Magwaza, Opara, Cronje, Landahl, & Terry (2013a) and Sánchez, De la Haba, & Pérez-Marín (2013a) have confirmed the potential of NIRS for predicting quality in mandarins; similar findings have been reported for oranges by Cayuela (2008), Cayuela & Weiland (2010), Liu, Sun, & Ouyang (2010a), Zheng et al., (2010), Magwaza et al., (2013c) and Sánchez, De la Haba, Serrano, & Pérez-Marín (2013b).

At the same time, the citrus sector is increasingly demanding methods for the on-tree monitoring of fruit quality parameters throughout the ripening

process, with a view to identifying the optimal harvesting time depending on the final destination of the product. Advances in NIRS instrumentation include the development of handheld and portable equipment, some of which has already been applied successfully for on-tree quality measurements in mandarins (Sánchez, De la Haba, & Pérez-Marín, 2013a) and, separately, in oranges (Sánchez, De la Haba, Serrano, & Pérez-Marín, 2013b). However, the predictive capacity and robustness of the models thus developed could be improved by using larger and more varied sample sets. In this sense, universal models applicable to any citrus fruit species would be particularly useful, and would favour the uptake of this technology by the citrus sector. However, when using what might be termed "multi-product sample sets", the relationship to be modelled may not always be linear; as a result, classical regression methods are not always the most suitable (Pérez-Marín, Garrido-Varo, & Guerrero, 2007). Barton II, Shenk, Westerhaus, & Funk (2000) suggested that one option in these cases could be to use local approaches based on the development of specific calibrations for each sample to be predicted, enabling existing nonlinearity to be addressed through the production of "local" linear models. LOCAL algorithm obtains, for each unknown sample to be predicted, a specific equation from a group of samples selected from the database on the basis of their spectral similarity to the unknown sample, which could afford a high degree of accuracy in the predictions (Pérez-Marín, Garrido-Varo, & Guerrero, 2007).

The aim of this study was to evaluate the LOCAL algorithm using a citrus-fruit database for the development of models to predict physical quality parameters during on-tree ripening—regardless of species, growing-season and crop practices—using a handheld MEMS-NIRS spectrophotometer.

6.1.2. Materials and methods

6.1.2.1. Fruit samples and reference data

The initial sample set comprised 611 samples belonging to the genus Citrus: 378 oranges (*Citrus sinensis* L. cv. 'Powell Summer Navel') (191 harvested in 2010 and 187 in 2011) and 233 mandarins (*Citrus reticulata* Blanco cv. 'Clemevilla') harvested in 2010, both grown on a commercial plantation near the village of La Campana (Seville, Spain). A total of 191 of the oranges were those used by Sánchez, De la Haba, Serrano, & Pérez-Marín (2013b), while all the mandarins were used by Sánchez, De la Haba, & Pérez-Marín (2013a).

Harvested oranges and mandarins were kept in refrigerated storage at 5ºC and 90% RH until the following day, when laboratory testing was performed. Prior to each test, samples were allowed to reach room temperature. All physical tests were performed at 20ºC.

For mandarins, external (mass, equatorial and axial diameters, colour (lightness (L*), red-greenness (a*), blue-yellowness (b*), Chroma (C*, (a^2 + b^2)^(1/2), hue angle (h*, $\tan^{-1}$ (b*/a*) and colour index, 1,000 a*/L*b*), and internal (firmness, pericarp thickness and juice mass) physical-quality parameters were measured following Sánchez, De la Haba, & Pérez-Marín (2013a); the same external and internal physical-quality parameters for oranges were measured following Sánchez, De la Haba, Serrano, & Pérez-Marín, (2013b).

6.1.2.2. NIR analysis

NIR spectra of mandarins and oranges were collected in reflectance mode (log 1/R) using the Phazir 2400 (Polychromix, Inc., Wilmington, MA, USA). The Phazir 2400 is an integrated near-infrared handheld analyser that incorporates all the essential components to deliver on-tree applications. It combines a digital transform spectrometer (DTS) engine, a reflectance probe, rechargeable batteries, an integrated computer, colour LCD display, and software into one unit that can be used remotely, such as in field applications. The reflected light is collected and measured by a single InGaAs photodetector,

and the instrument has no moving parts. The spectrophotometer scans at a non-constant step of about 8 nm interval (optical resolution 12 nm), across a range of near infra-red wavelengths (1600–2400 nm). Sensor integration time was 600 ms. The MEMS device measures an area of around 4 mm^2 and is equipped with quartz protection to prevent dirty accumulation. Instrument performance was checked every 10 minutes, following the diagnostic protocols provided by the manufacturer, and white reference measurement was carried out using Spectralon as reference.

Four spectral measurements were made for each fruit on the tree, taking orientation (north, south, east and west) into account. The four spectra were averaged to provide a mean spectrum for each sample.

6.1.2.3. Definition of calibration and validation sets

Prior to carrying out NIRS calibrations, the CENTER algorithm included in the WinISI II software package, version 1.50 (Infrasoft International, Port Matilda, PA, USA) was applied to ensure a structured population selection based solely on spectral information for the establishment of calibration and validation sets (Shenk & Westerhaus, 1991, 1995). This algorithm performs an initial principal component analysis (PCA) to calculate the centre of the population and the distance of samples (spectra) from that centre in an n-dimensional space, using the Mahalanobis distance (GH); samples with a statistical value greater than 4 were considered outliers or anomalous spectra.

The standard normal variate (SNV) and detrending (DT) methods were applied for scatter correction (Barnes, Dhanoa, & Lister, 1989), together with the mathematical derivation treatment 1, 5, 5, 1; where the first digit is the number of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing, and the fourth is the second smoothing (Shenk & Westerhaus, 1995; ISI, 2000).

Once each sample in the global set (N = 611) was ordered by its spectral distance (from shorter to longer) from the centre of the global population, outlier spectrum samples (GH > 4) were eliminated (N = 7) and the remaining samples (N= 604) were split into two: a training set containing about 75% of the samples (N = 457) and a test set containing the remaining 25% (N = 147).

Data were subjected to chemometric treatment using the WinISI II software package, version 1.50.

6.1.2.4. Construction of prediction models for major physical quality parameters in intact citrus fruits on-tree using the LOCAL algorithm

The LOCAL algorithm operates by searching and selecting samples in large databases that have spectra similar to the sample being analysed. The selected samples are then used to compute a specific calibration equation, based on Partial Least Squares (PLS) regression, for predicting the constituents of an unknown sample (Shenk, Westerhaus, & Berzaghi, 1997).

Different parameters have to be evaluated in order to optimise the LOCAL algorithm (Pérez-Marín, Garrido-Varo, & Guerrero, 2007). In the present study, an optimisation design for the LOCAL algorithm was set up by varying the number of calibration samples (*k*) from 80 to 120 in steps of 20, and the number of factors (*l*) from 14 to 16 in steps of 1. This gave a factorial design of 3 x 3 or 9 runs. Finally, the number of PLS factors discarded was set at the first four.

For each analytical parameter, different mathematical treatments were evaluated for scatter correction, including SNV and DT methods (Barnes, Dhanoa, & Lister, 1989). Additionally, four derivative mathematical treatments were tested: 1, 5, 5, 1; 2, 5, 5, 1; 1, 10, 5, 1; 2,10,5,1 (Shenk & Westerhaus, 1995).

Global calibration using the same pre-treatments used in LOCAL was performed (WinISI, II software package, version 1.50 (Infrasoft International, Port Matilda, PA, USA) in order to compare results obtained using the non-linear regression algorithm with those yielded by the classical prediction strategy based on MPLS regression. The same validation file for the genus citrus was then predicted using both regression algorithms. The results provided by the models constructed using non-linear regression for mandarin and orange were also compared with those obtained for mandarin alone (Sánchez, De la Haba, & Pérez-Marín, 2013a) and for orange alone (Sánchez, De la Haba, Serrano, & Pérez-Marín, 2013b), in both cases using MPLS regression. Standard errors of prediction (SEP), coefficients of determination (r^2), and the RPD (Residual Predictive Deviation) values calculated as the ratio of the standard deviation (SD) of the reference data to the SEP, using the LOCAL procedure and MPLS regression were compared.

6.1.3. Results and discussion

6.1.3.1. Descriptive data for NIR calibration and validation

After applying the CENTER algorithm to the overall set (N = 611), a total of 7 samples (2 oranges and 5 mandarins) were identified as anomalous spectra. Analysis showed that six of these displayed extreme values for the parameter a*, three being very green (2 mandarins and 1 orange at the start of harvesting), and three (mandarins) displaying a marked reddish hue at the end of harvesting. The other anomalous orange sample displayed an abnormally high value for pericarp thickness.

Values (range, mean, standard deviation and coefficient of variation, CV) obtained for each physical-quality parameter in the calibration and validations sets, after removing outliers, are shown in Table 6.1.1. Structured selection based on spectral information, using the CENTER algorithm proved suitable, in that the calibration and validation sets displayed similar values for

range, mean and SD for all study parameters. Furthermore, the ranges of the validation set lay within those of the calibration set.

All physical parameters tested, except three of the colour-related parameters (L*, b* and C* for the calibration and validation sets), displayed marked variability, with CV values of over 12% for both sets, covering a wide range of values. Other parameters also recorded CV values of over 40% in both sets, including mass, a*, colour index, firmness and juice mass.

Pérez-Marín, Garrido-Varo, & Guerrero (2005) have highlighted the importance of sample set and of sample distribution within the calibration set, noting that sample sets for calibration should ideally ensure uniform distribution of composition across the range of the study parameter in question.

6.1.3.2. Prediction of physical quality parameters in citrus fruits using the LOCAL algorithm

Results for the prediction of citrus-fruit physical quality parameters using LOCAL algorithm are shown in Table 6.1.2.

It should be noted that for predicting the external validation set, the LOCAL algorithm used only between 80 and 100 samples to predict most of the parameters tested and only 120 samples for mass and L* prediction, rather than using all 457 samples in the calibration set (as was the case for MPLS regression); only those samples whose spectra were considered representative of the calibration set were used.

The results obtained using the LOCAL algorithm were better than those achieved with MPLS regression (Table 6.1.2) for universal citrus models; robustness was increased by minimising prediction error and increasing the coefficient of determination for prediction.

The accuracy of the predictions obtained using the LOCAL algorithm was greater (i.e. SEP values were lower) than that of those obtained using the MPLS regression for all parameters tested in mandarin and orange. The greatest reduction in SEP using the LOCAL algorithm was recorded for the a* parameter (34.60%), followed by colour index (31.20%) and b* (23.91%). The smallest reductions in SEP using the LOCAL algorithm were recorded for equatorial diameter and juice mass (4.79% and 6.20%, respectively, with respect to MPLS regression).

An overall increase in the coefficient of determination was recorded for models obtained using the LOCAL algorithm with respect to those using MPLS. The most significant increases in value for r^2 were recorded for all colour-related parameters (r^2_{MPLS} = 0.30-0.44; r^2_{LOCAL} = 0.50-0.63), firmness (r^2_{MPLS} = 0.08; r^2_{LOCAL} = 0.28), fruit mass (r^2_{MPLS} = 0.65; r^2_{LOCAL} = 0.73), and axial diameter (r^2_{MPLS} = 0.74; r^2_{LOCAL} = 0.82).

However, neither of the strategies yielded results for L*, a*, C*, h* and firmness that lay within the limits recommended by Windham, Mertens, & Barton (1989) for the coefficient of determination ($r^2 > 0.60$). Even so, the LOCAL algorithm improved the coefficient of determination by 29.55% for L*, 37.83% for colour index, 52.94% for a*, 69.70% for C*, 66.67% for h* and by 250% for firmness, compared to the MPLS regression.

6.1.3.2.1. Morphological parameters

For morphological parameters (mass, equatorial and axial diameters) the citrus universal calibrations using MPLS performed worse in terms of accuracy and precision of prediction (Table 6.1.2), whilst the use of LOCAL reduced the SEP value by 13.39% for mass, by 4.79% for equatorial diameter, and by 17.80% for axial diameter. Moreover, the predictive models obtained for mass and equatorial diameter using the global strategy and MPLS regression only enabled fruit to be classified as high, medium or low, whereas the predictive capacity using the LOCAL algorithm may be considered good

according to the limits defined by Shenk & Westerhaus (1996). For axial diameter, the LOCAL strategy yielded an r^2 value of 0.82 compared to 0.74 for MPLS, i.e. an increase of 10.81%.

Comparison of the results obtained using the LOCAL algorithm for universal models (i.e. mandarin and orange) with those yielded by MPLS for mandarins alone (Sánchez, De la Haba, & Pérez-Marín, 2013a) and for oranges alone (Sánchez, De la Haba, Serrano, & Pérez-Marín, 2013b) showed that the r^2 values recorded for individual species were below the minimum recommended by Windham, Mertens, & Barton II (1989) for routine use of predictive models in the citrus sector, whereas models constructed using LOCAL regressions strategies for the three morphological parameters studied displayed r^2 values of over 0.70, and were therefore suitable for routine use. However, SEP values for the accuracy of predictive models developed using LOCAL strategies were slightly higher than those recorded using the linear regression models for the individual species tested, due to higher SD values in the universal equations.

6.1.3.2.2. Colour-related parameters

As Table 6.1.2 shows, the precision of the models constructed for colour parameters (L*, a*, b*, C*, h*, colour index) using the LOCAL algorithm may be considered acceptable for screening purposes ($0.50 \leq r^2 \leq 0.63$), enabling values for citrus fruits to be classified as high, medium and low; by contrast, the precision of the universal models developed using MPLS ($0.30 \leq r^2 \leq 0.44$) enabled only classification into high or low (Shenk & Westerhaus, 1996). The LOCAL-based model enabled routine prediction of parameter b* (blue–yellow), while values for the other parameters came close to threshold values for this purpose. The ability to measure, using a single NIRS instrument, the changes in colour from green-yellowish tones (negative a* and positive b*) to orange-reddish tones (positive a* and b*) typically occurring in the course of on-tree ripening, together with the non-destructive estimation of selected morphological parameters is undoubtedly of considerable interest in order to determine the optimal harvesting time.

For all study parameters, application of the LOCAL algorithm improved the accuracy of predictive models; reduction of the SEP for parameter a* was particularly noteworthy (SEP_{LOCAL} = 7.24; SEP_{MPLS} = 11.07). Comparison of LOCAL results for mandarin and orange with those obtained for individual species using MPLS showed that precision was greater with LOCAL for all parameters except a*, h* and colour index in mandarins. SEP values for the universal equations using LOCAL algorithm were also better, except for L*, b* and h* which were better in MPLS models for mandarin and also the colour-parameters tested in orange. RPD values for orange and mandarin equations using LOCAL algorithm were also higher, except for h* in mandarin.

6.1.3.2.3. Internal physical parameters

Results obtained for the prediction of firmness using the LOCAL algorithm indicate that the predictive capacity of the model, though very low (r^2 = 0.28, SEP = 11.02 N), was higher than that obtained with MPLS; the standard error was reduced by 12.88% and the coefficient of determination increased by 250%. Though increased by the application of non-linear regression algorithms, this low predictive capacity underlines the difficulty in correlating destructive measurements made to a puncturing depth of 10 mm with non-destructive NIR measurements, particularly for thick-peel fruits such as this orange variety (Sánchez, De la Haba, Serrano, & Pérez-Marín, 2013b). As Peirs, Scheerlinck, Touchant, & Nicolaï (2002) have noted, NIR light will only penetrate usefully down to a depth of between 1 and 5 mm, depending on the wavelength, the instrument and the fruit ripeness stage.

For pericarp thickness and juice mass, the robustness of universal models was enhanced by application of the LOCAL algorithm; SEP values were reduced by 8.91% for pericarp thickness and by 6.20% for juice mass, whilst r^2 was increased by 11.29% and 7.46%, respectively. Non-destructive prediction of both parameters is of particular interest to the citrus sector, which prizes fruit with reduced peel thickness and high juice content.

Comparison of the results obtained here with those reported by Sánchez, De la Haba, & Pérez-Marín (2013a) and by Sánchez, De la Haba, Serrano, & Pérez-Marín (2013b) confirms the view expressed by Williams (2001) and Pérez-Marín, Garrido-Varo, & Guerrero (2005), among others, regarding the importance of using a sufficiently-large and sufficiently-varied calibration set for developing global calibration equations. Here, increased sample size and greater uniformity in terms of the number of samples available across the whole range of the test parameter improved the predictive capacity of the models.

The frequency histogram for juice mass is shown in Figure 6.1.1. Juice mass is one of the parameters most affected by sample distribution over the entire range, especially when two different citrus species are tested together; here, the range for oranges (19.33-282.96 g) was much wider than for mandarins (2.62-90.69 g). The effect of combining the two species in calibration sets, in terms of increased range and improved distribution for the juice-mass parameter was evident when comparing the results obtained by Sánchez, De la Haba, & Pérez-Marín (2013a) and by Sánchez, De la Haba, Serrano, & Pérez-Marín (2013b) for individual species (mandarin: r^2 = 0.28; SEP = 14.71 g; RPD = 1.17; orange: r^2 = 0.28; SEP = 24.07 g; RPD = 1.27) with those obtained here for combined orange-mandarin sets and LOCAL algorithm (r^2 = 0.72; SEP = 28.13 g; RPD = 1.86). Shenk, Westerhaus, & Berzaghi (1997) suggest that the samples selected for calibration should include all possible sources of variation encountered during prediction, in order to increase the robustness of the calibration, although this usually decreases the accuracy of prediction. Likewise, Bellon-Maurel, Fernandez-Ahumada, Palagos, Roger & McBratney (2010) insist on the fact that the SEP value generally increases when the measurement range of this parameter – or the mean of this range – increases, decreasing the accuracy of the prediction. However, the use of the LOCAL algorithm obviates the need to choose between accuracy and robustness of a calibration.

6.1.3.3. Matching calibration samples for the external prediction of physical quality parameters in citrus fruits using the LOCAL algorithm

It was considered useful to determine the percentage of each fruit species in the training set used by the LOCAL algorithm to develop prediction models for that species in a combined validation set. Juice mass was the parameter selected for this purpose. Results are shown in Figure 6.1.2. Additionally, typical log (1/R) spectra for oranges and mandarins, obtained on the Phazir instrument are shown in Figure 6.1.3.

The LOCAL algorithm applied to the validation set (N = 147 samples; 88 oranges and 59 mandarins) used 80 samples to predict juice mass, rather than the 457 samples used in MPLS regression. In most cases, moreover, samples belonged to the species to be predicted. As Figure 6.1.2(a) shows, 72 (81.82%) of the 88 oranges in the validation set were predicted with between 80% and 100% of oranges in the training set. In two cases, oranges in the validation set were predicted with less than 40% of orange samples, and between 62% and 71% of mandarin samples from the training set. As Figure 6.1.2(b) shows, 36 of the 59 mandarins in the validation set (61.02%) were predicted using over 80% of mandarins in the training set.

No previously-published research has addressed the use of non-linear regression methods such as LOCAL to develop predictive models in other fruit species, but Sánchez, De la Haba, Serrano, & Pérez-Marín (2013b) used this algorithm to predict the same quality parameters tested here, in oranges, also reporting that LOCAL improved the predictive capacity of models for all parameters with respect to MPLS.

Sánchez, De la Haba, Guerrero, Garrido-Varo, & Pérez-Marín (2011) also found that the use of LOCAL rather than MPLS regression improved models for predicting quality parameters in nectarines using on-tree measurements.

6.1.4. Conclusions

These findings confirm that NIRS technology using the LOCAL algorithm is a promising tool for the development of universal quality-prediction models for different fruit species belonging to the same genus, thus obviating the need to develop specific models for each species. The results also confirm the viability of NIRS technology, using latest-generation portable instruments, for the development of models enabling monitoring of the physical changes taking place during on-tree ripening. The LOCAL non-linear regression algorithm proved to be considerably more effective for this purpose than MPLS regression. To our knowledge, this is the first attempt to develop universal quality models using on-tree NIR spectroscopy for the genus Citrus. Over the coming years, however, recalibration may be required, increasing the number of samples in the calibration set by adding other species of this genus such as lemons, pomegranates, etc.

Acknowledgements

Authors acknowledge Prof. Elías Fereres and Dr. Victoria González-Dugo. Ms. Mª Carmen Fernández is also acknowledged for her technical support. This work was funded by the Spanish Ministry of Science and Innovation for the projects CONSOLIDER CSD2006-0067 and AGL2012-40053-C03-01, with participation of funds from FEDER (European Union) and by the Andalusian Regional Government under the Research Excellence Program, project no. P09-AGR-5129 MEMS and NIRS-image sensors for the in situ non-destructive analysis of food and feed. The authors have no conflict of interest to declare.

References

Antonucci, F., Pallottino, F., Paglia, G., Palma, A., D'Aquino, S., & Menesatti, P. (2011). Non-destructive estimation of mandarin maturity status through portable VIS-NIRS spectrophotometer. *Food and Bioprocess Technology*, 4, 809–813.

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43, 772–777.

Barton II., F. E., Shenk, J. S., Westerhaus, M. O., & Funk, D. B. (2000). The development of near infrared wheat quality models by locally weighted regressions. *Journal of Near Infrared Spectroscopy*, 8, 201–208.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., & McBratney, A. (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. Trends in Analytical Chemistry, 29, 1073-1081.

Cayuela, J. A. (2008). Vis–NIR soluble solids prediction in intact oranges (*Citrus sinensis* L.) cv. Valencia Late by reflectance. *Postharvest Biology and Technology*, 47, 75–80.

Cayuela, J. S., & Weiland, C. (2010). Intact orange quality prediction with two portable NIR spectrometers. *Postharvest Biology and Technology*, 58, 113–120.

Fraser, D. G., Jordan, R. B., Künnemeyer, R., & McGlone, V. A. (2003). Light distribution inside mandarin fruit during internal quality assessment by NIR spectroscopy. *Postharvest Biology and Technology*, 27, 185–196.

Guthrie, J. A., Walsh, K. B., Reid, D. J., & Lienberg, C. J. (2005a). Assessment of internal quality attributes of mandarin fruit I. NIR calibration model development. *Australian Journal of Agriculture Research*, 56, 405–416.

Guthrie, J. A., Reid, D. J., & Walsh, K. B. (2005b). Assessment of internal quality attributes of mandarin fruit II. NIR calibration model robustness. *Australian Journal of Agriculture Research*, 56, 417–426.

Hernández-Gómez, A. H., He, Y., & Pereira, A. G. (2006). Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR spectroscopy techniques. *Journal of Food Engineering*, 77, 313–319.

ISI. (2000). *The complete software solution using a single screen for routine analysis, robust calibrations and networking. Manual*. Silver Spring: Foss NIRSystems/Tecator, Infrasoft International.

Liu, Y., Sun, X., & Ouyang, A. (2010a). Non-destructive measurements of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLS and PCA-BPNN. *LWT-Food Science and Technology*, 43, 602–607.

Liu, Y., Sun, X., Zhang, H., & Aiguo, O. (2010b). Nondestructive measurement of internal quality of Nanfeng mandarin fruit by charge coupled device near infrared spectroscopy. *Computers and Electronics in Agriculture*, 71(S1), S10–S14.

Liu, Y., Heying, E., & Tanumihardjo, S. A. (2012). History, global distribution, and nutritional importance of citrus fruits. *Comprehensive Reviews in Food Science and Food Safety*, 11(6), 530-545.

Magwaza, L. S., Opara, U. L., Nieuwoudt, H., Cronje P. J. R., Saeys, W., & Nicolaï, B. M. (2012a). NIR spectroscopy applications for internal and external quality analysis of citrus fruit—a review. *Food and Bioprocess Technology*, 5, 425-444.

Magwaza, L. S., Opara, U. L., Terry, L. A., Landahl, S., Cronje, P. J., Nieuwoudt, H., Mouazen, A. M., Saeys, W., & Nicolaï, B. M. (2012b). Prediction of "Nules Clementine" mandarin susceptibility to rind breakdown disorder using Vis/NIR spectroscopy. *Postharvest Biology and Technology*, 74, 1-10.

Magwaza, L. S., Opara, U. L., Cronje, P. J. R., Landahl, S., & Terry, L. A. (2013a). Canopy position affects rind biochemical profile of "Nules

Clementine" mandarin fruit during postharvest storage. *Postharvest Biology and Technology*, 86, 300-308.

Magwaza, L. S., Opara, U. L., Cronje, P. J. R., Nieuwoudt, H. H., Landahl, S., & Terry, L. A. (2013b). Quantifying the effects of fruit position in the canopy on physical and biochemical properties and predicting susceptibility to rind breakdown disorder of "Nules Clementine" mandarin (*Citrus reticulate* Blanco) using Vis/NIR spectroscopy. *Acta Horticulturae* 1007, 83-91

Magwaza, L. S., Opara, U. L., Terry, L. A., Landahl, S., Cronje, P. J. R., Nieuwoudt, H. H., Hanssens, A., & Saeys, W. (2013c). Evaluation of Fourier transform-NIR spectroscopy for integrated external and internal quality assessment of Valencia oranges. *Journal of Food Composition and Analysis*, 31, 144-154.

Magwaza, L. S., Opara, U. L., Cronje, P. J. R., Nieuwoudt, H. H., Mouazen, A. M., Nicolaï, B. M., & Terry, L. A. (2014). Assessment of rind quality of "Nules Clementine" mandarin using postharvest storage: 1. Vis/NIR PCA models and relationship with canopy position. *Scientia Horticulturae*, 165, 410-420.

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., & Theron, I. K., & Lammertyn, J. (2007). Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46, 99–118.

Olmo, M., Nadas, A., & García, J. M. (2000). Nondestructive methods to evaluate maturity level of oranges. *Journal of Food Science*, 65 (2), 365-369.

Peirs, A., Scheerlinck, N., Touchant, K., & Nicolaï, B. M. (2002). Comparison of Fourier transform and dispersive near-infrared reflectance spectroscopy for apple quality measurements. *Biosystems Engineering*, 81, 305–311.

Pérez-Marín, D., Garrido-Varo, A., & Guerrero, J. E. (2005). Implementation of local algorithm with near-infrared spectroscopy for compliance

assurance in compound feedingstuffs. *Applied Spectroscopy*, 59, 69–77.

Pérez-Marín, D., Garrido-Varo, A., & Guerrero, J. E. (2007). Non-linear methods in NIRS quantitative analysis. *Talanta*, 72, 28–42.

Sánchez, M. T., De la Haba, M. J., Guerrero, J. E., Garrido-Varo, A., & Pérez-Marín, D. (2011). Testing of a Local approach for the prediction of quality parameters in intact nectarines using a portable NIRS instrument. *Postharvest Biology and Technology*, 60, 130-135.

Sánchez, M. T., De la Haba, M. J.; & Pérez-Marín, D. (2013a). Internal and external quality assessment of mandarins on-tree and at harvest using a portable NIR spectrophotometer. *Computers and Electronics in Agriculture*, 92, 66-74.

Sánchez, M. T., De La Haba, M. J., Serrano, I., & Pérez-Marín, D. (2013b). Application of NIRS for non-destructive measurement of quality parameters in intact oranges during on-tree ripening and at harvest. *Food Analytical Methods*, *6*, 826-837.

Shenk, J. S., & Westerhaus, M. O. (1991). Population definition, sample selection and calibration procedures for near infrared spectra and modified partial least squares regression. *Crop Science*, 31, 469–474.

Shenk, J. S., & Westerhaus, M. O. (1995). *Routine operation, calibration, development and network system management manual*. Silver Spring: NIRSystem, Inc.

Shenk, J. S., & Westerhaus, M. O. (1996). Calibration the ISI way. In A. M. C. Davies & P. C. Williams (Eds.), *Near infrared spectroscopy: The future waves* (pp. 198–202). Chichester: NIR Publications.

Shenk, J. S., Westerhaus, M. O., & Berzaghi, P. (1997). Investigation of a LOCAL calibration procedure for near infrared instruments. *Journal of Near Infrared Spectroscopy*, 5, 223–232.

Sun, X., Zhang, H., & Liu, Y. (2009). Nondestructive assessment of quality of 'Nanfeng' mandarin fruit by a portable near infrared spectroscopy. *International Journal of Agricultural and Biological Engineering*, 2, 65–71.

Williams, P. C. (2001). Implementation of near-infrared technology. In P. C. Williams & K. H. Norris (Eds.), *Near-infrared technology in the agricultural and food industries* (pp. 145–171). St. Paul, MN: AACC, Inc.

Windham, W. R., Mertens, D. R., & Barton, F. E. II, (1989). Protocol for NIRS calibration: Sample selection and equation development and validation. In C. Martens, J. S. Shenk, & F. E. Barton, II (Eds.), *Near infrared spectroscopy (NIRS): Analysis of forage quality. Agriculture handbook, No. 643* (pp. 96–103). Washington, DC: USDA-ARS, US Government Printing Office.

Zheng, Y., He, S., Yi, S., Zhou, Z., Mao, S., Zhao, X., & Deng, L. (2010). Predicting oleocellosis sensitivity in citrus using vis–NIR reflectance spectroscopy. *Scientia Hoticulturae*, 125, 401–405.

Table 6.1.1. Number of samples (N), range, mean, standard deviation (SD) and coefficient of variation (CV) in calibration and validation sets.

Parameter	Set	Range	Mean	SD	CV (%)
Mass (g)	Calibration	44.20-598.30	243.93	108.35	44.42
	Validation	54.93-561.00	239.46	113.27	47.30
Equatorial diameter (mm)	Calibration	38.05-108.34	76.52	12.60	16.47
	Validation	41.37-107.18	76.10	12.93	16.99
Axial diameter (mm)	Calibration	42.10-113.92	73.71	16.24	22.03
	Validation	45.47-107.20	72.76	16.24	22.32
L*	Calibration	46.12-79.52	65.53	4.23	6.46
	Validation	48.61-70.81	65.23	3.81	5.84
a*	Calibration	-16.34-42.41	21.40	12.44	58.13
	Validation	-15.43-41.35	21.63	13.17	60.89
b*	Calibration	34.89-78.14	64.77	7.30	11.27
	Validation	36.94-76.49	64.42	6.84	10.62
C*	Calibration	37.61-81.98	69.24	8.11	11.71
	Validation	38.92-80.42	69.10	7.79	11.27
h*	Calibration	51.74-112.40	72.71	10.95	15.06
	Validation	52.77-108.36	72.43	11.42	15.77
Colour index	Calibration	-8.73-13.56	4.86	3.50	72.02
	Validation	-6.83-13.18	5.01	3.59	71.66
Firmness (N)	Calibration	2.07-79.88	19.21	14.60	76.00
	Validation	2.65-62.18	16.92	12.63	74.65
Pericarp thickness (mm)	Calibration	1.59-10.27	5.02	1.64	32.67
	Validation	2.25-9.19	5.01	1.63	32.53
Juice mass (g)	Calibration	2.62-282.96	102.79	52.21	50.79
	Validation	17.79-260.67	101.72	52.38	51.49

Table 6.1.2. Statistics for validation of citrus samples using LOCAL and MPLS regression strategies.

	Best of LOCAL				MPLS regression		
	SEP	**r^2**	**RPD**	**Settings[a]**	**SEP**	**r^2**	**RPD**
Morphological parameters							
Mass (g)	58.86	0.73	1.92	120, 14	67.96	0.65	1.66
Equatorial diameter (mm)	6.96	0.71	1.86	80, 14	7.31	0.68	1.77
Axial diameter (mm)	6.88	0.82	2.36	80, 15	8.37	0.74	1.94
Colour parameters							
L*	2.52	0.57	1.51	120, 14	2.96	0.44	1.29
a*	7.24	0.52	1.82	80, 16	11.07	0.34	1.19
b*	4.17	0.63	1.64	100, 16	5.48	0.39	1.25
C*	5.19	0.56	1.50	100, 16	6.55	0.33	1.19
h*	8.09	0.50	1.41	80, 16	9.81	0.30	1.16
Colour index	1.94	0.51	1.85	80, 15	2.89	0.37	1.24
Physical internal parameters							
Firmness (N)	11.02	0.28	1.15	80, 14	12.65	0.08	1.00
Pericarp thickness (mm)	0.92	0.69	1.77	100, 16	1.01	0.62	1.62
Juice mass (g)	28.13	0.72	1.86	80, 15	29.99	0.67	1.75

[a]*LOCAL settings: number of selected samples, number of PLS factors.*

Figure 6.1.1. Distribution of juice mass (g) for mandarins, oranges and mandarins and oranges during on-tree ripening.

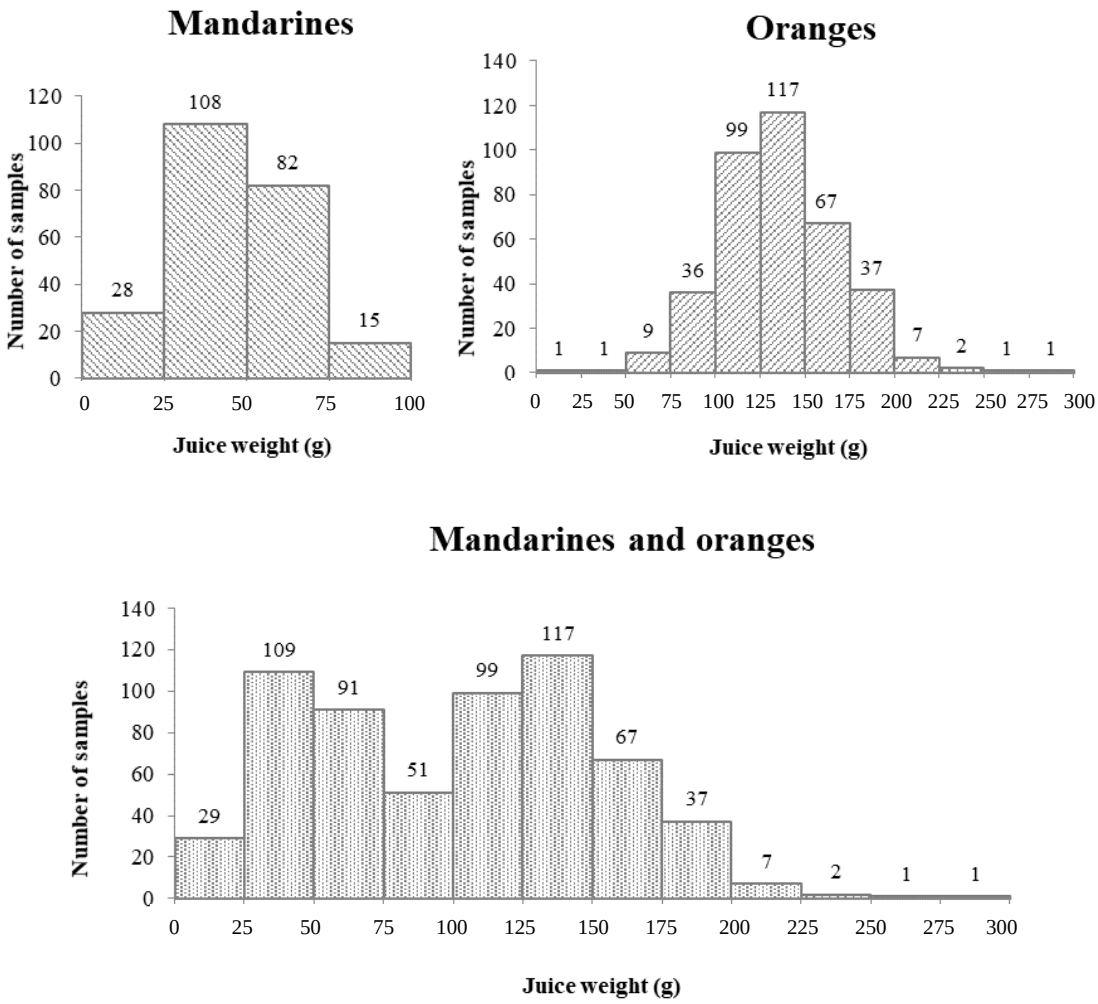
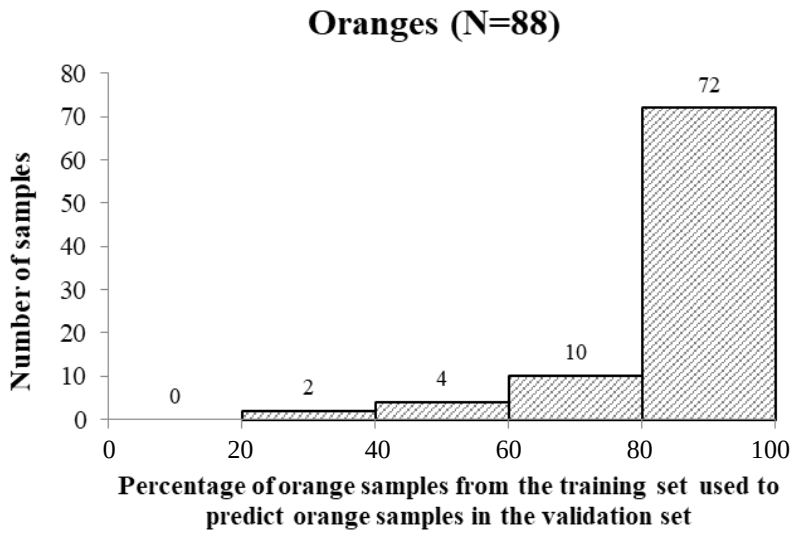


Figure 6.1.2. Prediction of the validation set for juice mass using LOCAL algorithm.

a



b

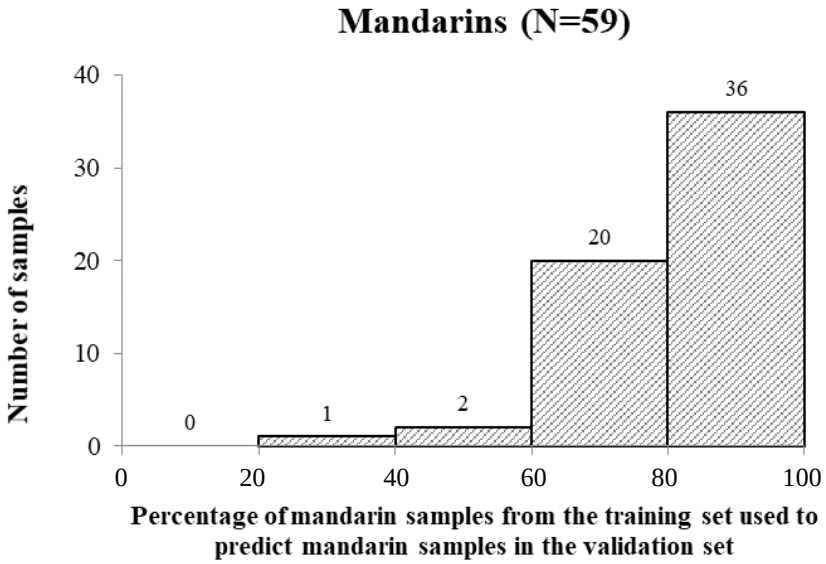
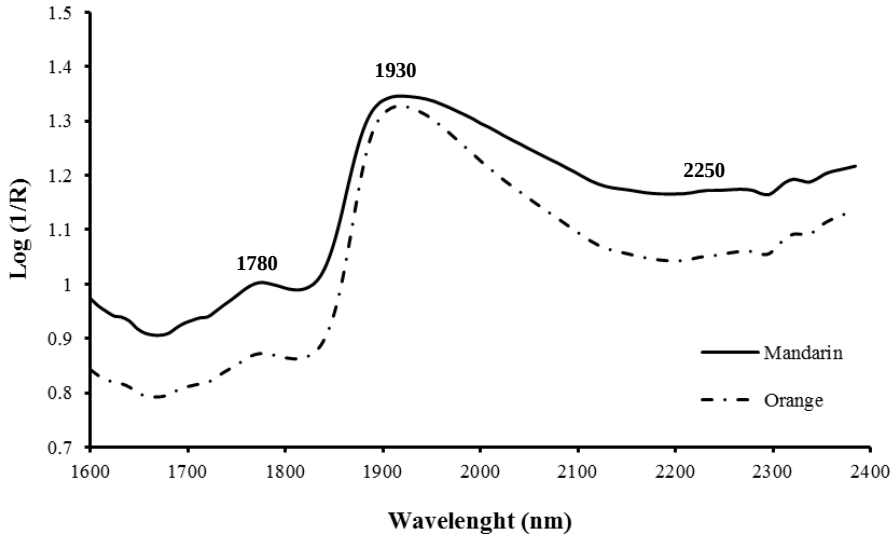


Figure 6.1.3. Typical log(1/R) spectra for mandarins and oranges.



Chapter 6.2.

LOCAL regression applied to a citrus multispecies library to assess chemical quality parameters using near infrared spectroscopy

Irina Torres[a], María-Teresa Sánchez[a], María-José De la Haba[a], Dolores Pérez-Marín[b]

[a] *Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*
[b] *Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.*

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 217, 206-214 (2019)

Abstract

The non-destructive on-tree measurement of the chemical quality attributes of fruits belonging to the Citrus genus using rapid spectral sensors is of vital interest to citrus growers, allowing them to carry out a selective harvest of any species of citrus fruit. With this objective, the viability of using of a handheld portable near infrared spectroscopy (NIRS) instrument to predict soluble solid content (SSC), pH, titratable acidity (TA), maturity index and BrimA, in order to measure the optimum harvest time in a group made up of 608 samples belonging to the Citrus genus (378 oranges and 230 mandarins) was evaluated. For each of the parameters analysed, both non-linear regression (LOCAL algorithm) and linear regression (Modified Partial Least Squares, MPLS) strategies were designed and compared. The use of the LOCAL algorithm in the sample group of oranges and mandarins for all the parameters analysed allowed to obtain more robust models than those obtained with MPLS regression, and it could also be extended more easily when routinely applied. The results confirm that NIRS technology combined with non-linear regression strategies such as the LOCAL algorithm can indeed respond to the needs of the citrus growers and help them to set the optimum harvest time, in this case of oranges and mandarins, by predicting the chemical quality parameters *in situ*.

Keywords: NIR spectroscopy; Citrus genus; *In situ* analysis; Chemical quality; LOCAL algorithm; Optimum harvest time.

6.2.1. Introduction

Since oranges and mandarins are non-climacteric fruits, the harvest should be timed for when the fruit has reached its commercial maturity [1].

For both these citrus species, the harvest indices generally used are based on maturity index values (ratio of SSC to TA), BrimA (an abbreviation for Brix minus Acids) and a minimum yellow-orange colour of the peel [2–5].

On-tree intact measurement of these harvest indices for all the fruits is particularly critical for these non-climateric fruits due to the fact that the physiological maturation process has finished at harvest, and since flavour perception of these fruits is closely linked to these quality attributes (SSC and TA) [4, 5]. Since consumer acceptance of the fruits is based on flavour and sweetness, measuring these values of the fruits on the tree would allow them to be harvested selectively and then sold according to their quality [6].

Therefore, due to the need to test and measure the chemical quality parameters of individual Citrus fruits, the Citrus sector requires the introduction of non-destructive, fast, versatile, environmentally-friendly and cost-effective technologies such as NIR Spectroscopy, which allows to measure the quality of the fruit directly on the tree during the maturation process, regardless of the species analysed.

Most applications which use NIR spectroscopy to measure quality chemical parameters (SSC, pH, TA, maturity index and vitamin C) in fruits of the Citrus genus refer to studies carried out with laboratory equipment for a single species using linear regression techniques, such as Partial Least Squares (PLS), Multiple Linear Regression (MLR) and Principal Component Regression (PCR) [7–10].

However, very few works have focused on taking readings directly on the tree to establish the optimum harvest time [11-13]. Sánchez et al. [11, 12] used portable equipment based on micro-electro-mechanical system (MEMS) technology, with a 1600-2400 nm spectral range to measure quality parameters (SSC, TA, pH, maturity index) in mandarins and oranges separately. Similarly, Cavaco et al. [13] measured the on-tree quality of oranges through internal quality parameters (SSC, pH, TA and maturity index), using portable VIS/NIR equipment with a charge-coupled device array detector and a reduced range of measurement (680-1100 nm).

In addition to *in situ* measurements, it would be highly advantageous in practical and commercial applications to be able to use universal equations for different citrus species to measure physical-chemical quality, thus permitting the staggered collection of the fruits depending on when they reach their full maturity. Despite this need for universally-applicable equations, there are few published works which refer to developing NIRS models for multi-product groups in plants [14-17], and the work published by Torres et al. [17] is the only one dealing with analysing citrus species intact on the tree as a way of measuring the morphological and physical quality of the fruits.

In the case of heterogeneous spectral libraries (multispecies libraries), the application of non-linear regression methods based on local calibrations allow a better management of the population available, since the characteristics of the samples selected by the algorithm to be used for calibration are specific in each case and for each of the samples to be predicted, thus making it easier for producers to develop models [17-20].

A number of works, in products which are not fruits, have confirmed that the use of non-linear regression techniques with multispecies libraries allows to obtain models with a higher predictive capacity and, most importantly, facilitates the routine management of prediction models and especially their recalibration, since it is simply a case of expanding the

calibration database, rather than having to recalculate the models as in classic global strategies. Thus, Godin et al. [21] compared their results obtained by applying non-linear and linear regression methods (LOCAL *versus* MPLS algorithms) to predict neutral and acid detergent fibre residues, acid detergent lignin and mineral compound content in a set composed of different fibrous plants. They concluded that the reliability of non-linear models is greater, since they fit in better with the non-homogeneity associated with a multispecies database.

Similarly, in fruits, the potential of local regression techniques for increasing the robustness of prediction models has been demonstrated by different authors, although these models have been developed for individual species [12, 22-24].

In the particular case of citrus species (oranges and mandarins), Torres et al. [17] applied the LOCAL algorithm in a previous work to measure morphological parameters (weight, equatorial and axial diameter), colour (L*, a*, b*, C* and h*) and physical parameters (firmness, pericarp thickness and juice mass). When this non-linear regression algorithm was applied instead of the MPLS regression, the predictive capacity of the models increased for all parameters and the prediction error decreased.

The aim of this study was to develop predictive models based on non-linear regression strategies (LOCAL algorithm), in order to measure the main chemical quality parameters which indicate the optimum harvest time and allow to carry out selective harvesting in fruits of the Citrus genus, regardless of the species, growing-season and crop practices, using NIRS technology together with a handheld portable MEMS-NIR spectrophotometer.

6.2.2. Material and Methods

6.2.2.1. Fruit Samples

The initial sample set comprised 608 samples belonging to the genus Citrus – 378 oranges (*Citrus sinensis* L. cv. 'Powell Summer Navel') and 230 mandarins (*Citrus reticulata* Blanco cv. 'Clemevilla') – grown in a commercial plantation in La Campana (Seville, Spain), under four different irrigation regimes.

In the case of orange, each experimental plot comprised three rows of four trees, with four repetitions for each irrigation regime; oranges were monitored on the two central trees in each plot. Thus, ripening was monitored on eight trees for each of the four irrigation regimes, giving a total of 32 trees. A total of twelve oranges were labeled on each of the 32 trees, thus giving a total of 384 oranges. However, in the course of the study, six ripe oranges dropped off the tree and were thus excluded. The final sample set thus comprised 378 oranges.

For mandarins, each experimental plot comprised three rows of four trees, with four repetitions for each irrigation regime; mandarins were monitored on the two central trees in each plot. Thus, ripening was monitored on eight trees for each of the four irrigation regimes, giving a total of 32 trees. A total of eight mandarins were labeled on each of the 32 trees, thus giving a total of 256 mandarins. However, in the course of the study, twenty-six ripe mandarins dropped off the tree and were thus excluded. The final sample set thus comprised 230 mandarins.

On arrival at the laboratory, the harvested oranges and mandarins were kept in refrigerated storage at 5ºC and 90% RH until the following day, when laboratory testing was performed. Prior to each test, the samples were allowed to reach room temperature of 20ºC, suitable for conducting the analysis.

6.2.2.2. Reference Data

The chemical parameters (SSC, TA and pH) of the oranges and mandarins were measured in the same way as Sánchez et al. [11]. The maturity index was also calculated as an SSC/TA ratio and the BrimA index was calculated using the equation described by Jordan el al. [4]:

$$BrimA = SSC - k(TA),$$

where k is a constant that reflects the tongue's higher sensitivity to TA compared to SSC. The value of the constant *k* was 4, which was suggested for oranges by Obenland et al. [5] in order to avoid the generation of negative BrimA values.

6.2.2.3. NIR Analysis

NIR analysis of both fruits were performed in reflectance mode (log 1/R) using a handheld MEMS spectrophotometer Phazir 2400 (Polychromix, INC., Wilmington, MA, USA) that incorporates all the essential components to deliver on-tree applications. This instrument scans at 8 nm non-constant intervals in the spectral range 1600-2400 nm. Four spectral measurements were made for each fruit (orange and mandarin) in the equatorial zone whilst on the tree, taking orientation (north, south, east and west) into account. The four spectra were averaged to provide a mean spectrum for each sample.

6.2.2.4. Definition of Calibration and Validation Sets

Principal component analysis (PCA) was performed on each individual data set (378 oranges and 230 mandarins) in order to structure and compress the data matrix. After PCA, the centre of the spectral population was fixed in order to detect outlier samples. The Mahalanobis distance (GH) was calculated between each sample and the centre of the population. Samples with a GH value greater than 4 were considered outliers [25]. As signal spectral pre-treatments, the standard normal variate (SNV) plus detrending (DT) procedures

[26] were used to remove the multiplicative interferences of scatter, and the Norris first derivative mathematical treatment was performed (1,5,5,1), where the first digit is the order of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing and the fourth is the second smoothing [27].

After removing the outliers (in this case, 3 oranges and 1 mandarin), each of the resulting sets, consisting of 375 oranges and 229 mandarins, was divided into two: a calibration set containing about 75% of the samples and a validation set containing the remaining 25%. These samples were selected following the method outlined by Shenk and Westerhaus [28] using the CENTER algorithm included in the WinISI II software package version 1.50 to calculate the distance to the centre of the population based on the Global Mahalanobis distance (GH), with three out of every four samples selected to be part of the calibration set [29]. Additionally, the calibration and validation sets of oranges and mandarins were merged to make new calibration and validation sets of citric fruits with the two species tested together. The differences in the number of samples available for the different parameters analysed in both the calibration and validation groups were due to the fact that, in some of them, the pH and TA measurements or the parameters derived from titratable acidity (maturity index and Brim A) could not be recorded since the fruits had a very low juice content.

6.2.2.5. Construction of Prediction Models using the LOCAL Algorithm. Comparison with Models Obtained Using Linear Regression Strategy

The LOCAL algorithm was performed for each dataset (oranges, mandarins and oranges and mandarins). LOCAL operates by searching for, and selecting, samples in large databases containing spectra similar to the sample being analysed. The selected samples are then used to compute a specific calibration equation, based on PLS regression, to predict the constituents of an unknown sample [18].

Selection of the calibration samples is controlled by the value of the coefficient of correlation between the spectrum of the unknown sample and those comprising the spectral database [18]; the samples with the highest correlation are selected. A minimum correlation cut-off is available to ensure that the selected samples are highly correlated [30].

Different parameters must be evaluated in order to optimize the LOCAL algorithm. In this work, an optimization design was set up by varying the number of calibration samples (k) from 80 to 140 in steps of 20, and the number of factors (l) from 14 to 16 in steps of 1. This gave a factorial design of 4 x 3 or 12 runs. Finally, it was established that the first four PLS factors should be removed.

Furthermore, for each parameter analysed, the different mathematical signal pre-treatments were evaluated. For scatter correction, the SNV and DT methods were tested [26]. Additionally, four derivative mathematical treatments were tested in the development of NIR calibrations: 1,5,5,1; 2,5,5,1; 1,10,5,1; 2,10,5,1 [27].

The effect of the different settings on the performance of the LOCAL algorithm was evaluated by comparing the standard error of prediction (SEP) obtained for each set, the coefficient of regression for external validation (r^2_p) and the RPD_p (ratio of the standard deviation (SD) of the reference data for validation to the SEP).

In addition, in order to compare the results obtained with the LOCAL algorithm, global models using linear regression were developed.

To achieve this, MPLS regression was used to obtain equations for each data set and for each parameter analysed [25]. During the development of the MPLS equation, the same signal pre-treatments used with LOCAL algorithm were used (SNV + DT, and the four derivative mathematical

treatments). The best predictive models obtained for the calibration sets, selected by statistical criteria (the standard error of cross validation (SECV) and the coefficient of determination for cross validation (r^2_{cv}), were subjected to evaluation using the validation sets, which consisted of samples not involved in the calibration procedure.

The SEP values of the predictive models for the parameters tested obtained using the LOCAL and MPLS regression algorithms were statistically compared using Fisher's F test [31]. The values for F were calculated as:

$$F = \frac{SEP_2^2}{SEP_1^2}$$

where SEP_1 and SEP_2 are the standard error of prediction of two different models and $SEP_1 < SEP_2$. F is compared to $F_{critical}$ (1- *P*, n_1-1, n_2-1) as read from the table, with $P = 0.05$ and n_1 the number of times the measurement is repeated with method 1; n_2 is the number of times the measurement is repeated with method 2. If F is higher than $F_{critical}$, the two SEP values are significantly different.

6.2.3. Results and Discussion

6.2.3.1. Population Distribution of Chemical Quality Parameters

Perez-Marín et al. [20] showed the importance of the population distribution used in calibration to obtain robust models. For multispecies or multiproduct groups, using local rather than global calibrations has particular advantages in those parameters where different populations are observed for each species [17].

The distribution of the chemical quality parameters tested for oranges, mandarins, and oranges and mandarins, is shown in Figure 6.2.1, together with their mean and standard deviation. Since the maturity index and BrimA parameters are obtained from the SSC and TA content, in the discussion the distributions shown for the latter, together with their pH values are focused on.

For SSC, the set composed of oranges shows a non-normal distribution, more similar to a bimodal distribution, with a valley around 12% and a maximum around 10.5%, while mandarins show a normal distribution, with a maximum around 12.5%. It could be said that mandarins (ranging from 9.95 to 15.65) were sweeter than oranges (ranging from 6.80 to 15.30). If the groups of oranges and mandarins are joined, a new group is formed (oranges and mandarins) with a distribution close to normal, with a range between 6.80 and 15.6% and a maximum around 12.5%.

In the case of pH and similar to SSC, the mandarins group shows a normal distribution with a range from 2.08 to 4. The oranges group also has a normal distribution, with a range of 3.01-4.15. Since there are more oranges than mandarins, when the two groups are joined, the average value (3.53) is closer to that of the oranges group (3.69), and its deviation (0.30) is higher than that of both groups (0.20 in both cases) and losing the normal distribution.

Taking the groups of oranges and mandarins individually, they show a normal distribution for titratable acidity, with maximum values of 0.60 and 1.10% of citric acid for oranges and mandarins, respectively. For both groups together, there is a positive asymmetric distribution, with a clear maximum value around 0.60% for citric acid and a standard deviation of 0.34% citric acid.

6.2.3.2. Descriptive Data for NIR Calibration and Validation Sets

As it was explained in the Material and Methods section, the CENTER algorithm was applied to the individual spectral databases in order to structure the populations according to GH. A total of 3 oranges and 1 mandarin presented values of GH greater than 4, and these were therefore considered outliers. A detailed analysis of the chemical characteristics of these samples could determine that these samples have different characteristics from the rest; the three oranges considered as outliers had low values of SSC (7, 7.35 and 9.11%, respectively), being cases of samples collected before complete

maturation, whereas the mandarin sample showed a high value of SSC (15.45%), being a sample collected in an over-ripe state.

Once the outliers have been removed, the remaining samples were used to create the calibration and validation sets. The statistics obtained (number of samples, range, mean, standard deviation and coefficient of variation) for each of the parameters analysed in the calibration and validation sets for oranges, mandarins and the set composed by oranges and mandarins are shown in Table 6.2.1. For each parameter, the ranges for the validation set lay within the range for the calibration set; it could be affirmed that the validation set comprised representative samples of the whole variance. Furthermore, both sets of the same group of samples displayed similar values for mean, SD and CV.

For both the calibration and validation groups, the group that has the greatest variability is the one consisting of oranges and mandarins for the TA, pH and maturity index and in the case of SSC and BrimA, the variability of the oranges set is practically identical to that of the oranges and mandarins set.

6.2.3.3. Optimization of Settings for the Development of Predictive Models using the LOCAL Algorithm

The SEP values obtained for the best mathematical treatments for the set composed of oranges and mandarins using the LOCAL algorithm, for each one of the combinations of the number of samples (k) and the number of PLS factors, are shown in Figure 6.2.2. It must be highlighted that LOCAL was tuned (i.e. the pre-treatments, numbers of factors and calibration set size) on the validation set. This could give LOCAL a slight advantage over PLS; in this case PLS was tuned by the cross-validation.

As regards the SSC parameter, it can be seen in Figure 6.2.2 that, when 16 PLS factors are used, the SEP value increases as the number of samples increases, while for 14 and 15 factors, there is a slight decrease in SEP when the number of samples reaches 120; the lowest SEP value is obtained when 80

samples and 16 PLS factors are used. This shows that when there is a group with a uniform distribution (Figure 6.2.1), the LOCAL algorithm used fewer samples (80 samples) for predicting the external validation set than the global regression techniques (456 samples), since only those samples whose spectra were considered representative of the sample of the calibration set to be predicted were used. It should also stress the importance of having a large sample group with a wide variability in order to obtain robust prediction models, since having a wide, varied spectral library available, thanks to the samples selected for development from the specific models carried out by the LOCAL algorithm, allows to obtain better prediction results [23].

As it is shown in Figure 6.2.2, the pH does not follow a fixed trend in terms of the evolution of SEP values obtained and the number of samples used to develop the models, and the lowest SEP value (0.15) is obtained when 100 samples and 16 PLS terms were used. For titratable acidity, the lowest SEP value (0.14% citric acid) is obtained when 80 samples and 14 PLS factors are used. In general, it could be said for both parameters that the more samples used, the higher the value of SEP obtained.

For maturity index and BrimA, the SEP values decrease as the number of samples used increases, and the lowest SEP values for both parameters are obtained initially when 140 samples are used (Figure 6.2.2). The need for a greater number of samples shows that these modelling parameters are more complex, since they are derived from the relationship between simpler ones, such as SSC and TA. In addition, since in this case it was not clear if the minimum SEP value had been obtained with the number of samples tested (up to 140), it was decided to extend the number of samples used to evaluate this optimization parameter of the model (number of samples, k) to 200. For the maturity index, the minimum SEP value was obtained with 160 samples and 16 PLS factors, while for the BrimA parameter, the lowest SEP value was obtained with 140 samples and 14 PLS factors. It can therefore be confidently

asserted that the lowest SEP values for maturity index and BrimA are 2.98 and 0.84, respectively (Figure 6.2.2 and Table 6.2.2).

6.2.3.4. Validation Statistics for Predicting Chemical Quality Parameters in Citrus Fruits using the LOCAL and MPLS Algorithms

The validation statistics used to predict the chemical quality parameters in oranges, mandarins, and oranges and mandarins using LOCAL and MPLS regression algorithms are shown in Table 6.2.2. This table shows SEP, r^2_p, RPD_p and the settings (LOCAL algorithm) used for the best mathematical treatment for both regression strategies.

The set including all the samples (oranges and mandarins) obtained a good predictive capacity for all the parameters tested using the LOCAL algorithm, displaying values of r^2_p between 0.72 and 0.84 [32]. In general, the values of r^2_p obtained with the non-linear regression algorithm for the set composed of both species are greater than the values obtained for the individual sets, except for the set of oranges in the case of SSC and BrimA, and the set of mandarins for pH and maturity index, whose r^2_p values are slightly higher.

Furthermore, the validation statistics used to predict the chemical quality parameters show that models obtained using the LOCAL algorithm improved the predictive capacity (higher values of r^2_p) and the accuracy (lower values of SEP) with respect to MPLS regression for all the parameters, except for titratable acidity and maturity index in the set composed of oranges, whose predictive ability (r^2_p values) using LOCAL algorithm fell by 4% and 3%, respectively. For the other models developed, the improvement obtained with the LOCAL algorithm was 7–17% for r^2_p, with the mandarins group the highest for the SSC parameter and the oranges group for pH, with 46% and 67%, respectively; in the same way, the decrease in SEP values when applying the non-linear regression algorithm ranged from 4 to 18%, except in the case of pH for the mandarins group and titratable acidity in oranges, where there was no difference in terms of the errors obtained with the algorithms tested.

On the other hand, comparisons using Fisher's F test of the SEP values in the models obtained for the different parameters analysed, using different regression strategies (LOCAL and MPLS algorithms) for the groups of oranges, mandarins, and oranges and mandarins, pointed to the existence of significant differences ($P < 0.05$) for the SSC parameters in the oranges group, and for titratable acidity and maturity index both in the mandarins and the oranges and mandarins groups. For the other remaining parameters, the differences in SEP values were not significant ($P > 0.05$) (Table 6.2.2).

As regards the SSC and BrimA parameters, although there were no significant differences between the SEP values when applying the LOCAL algorithm or MPLS in the group of oranges and mandarins, Figure 6.2.1 clearly shows that the range available for the oranges group covers that of the mandarins and makes no distinction between the populations. For this reason, there are no important benefits is applying local regressions, except for the advantages of a routine handling of the spectral databases and the possibility of updating the models more easily if LOCAL is used.

In terms of r^2_p and considering the LOCAL algorithm, the SSC models obtained a good predictive capacity for oranges ($r^2_p = 0.81$) and for the set composed of oranges and mandarins ($r^2_p = 0.78$), whereas in the case of mandarins, the model constructed could only distinguish between low, medium and high values ($r^2_p = 0.57$) [32]. However, according to Nicolaï et al., [33] the RPD_p values obtained for the models developed for oranges (RPDp = 2.23) and for the oranges and mandarins group ($RPD_p = 2.09$) indicate that coarse quantitative predictions are possible for this parameter ($RPD_p = 2–2.5$), while the model obtained for mandarins ($RPD_p = 1.51$) can discriminate low from high values ($RPD_p = 1.50–2.00$). This reduced capacity obtained for the mandarins group can be attributed to its lower variability, according to the CV value given (Table 6.2.1). As shown in Table 6.2.2, the predictive capacity obtained for the oranges and mandarins group is very similar to that of the oranges group, and there are no significant differences ($P > 0.05$) between their

SEP values, which stresses the effectiveness of the LOCAL algorithm to measure SSC in two species simultaneously, using the same equipment and prediction model.

The only study found in the bibliography which measures SSC in a multispecies group of the Citrus genus was the work by Clark [15], who analyzed a group made up of samples of grapefruit, interspecific hybrids (including kumquats, orangequats and citranges), lemon-lime, mandarins and oranges, using FT-NIR (Bruker Alpha spectrometer) equipment and applying PLS regression. This author, however, analyzed samples of the juice, which is much more homogeneous than the whole fruit.

For the prediction of pH and titratable acidity, the results obtained for the oranges and mandarins group show a good predictive capacity for both parameters (r^2_p = 0.72 and RPD_p = 1.93 for pH and r^2_p = 0.84 and RPD_p = 2.43 for TA) using the LOCAL algorithm [32], while for RPD_p, the models developed for these parameters allow to distinguish between high and low pH values and to make a coarse prediction for TA [33].

With the LOCAL algorithm, the predictive capacity improves considerably both for pH and for titratable acidity in the oranges and mandarins group compared with the oranges group. When both species are taken together, r^2_p increases by 188% and 87%, for pH and TA respectively, compared with the oranges group, which could be due to the increase in range which occurs when mandarins are added to the oranges group.

In the same way, there is also a 10% improvement in the accuracy of the model for titratable acidity compared with the mandarins group (r^2_p= 0.76), while for pH, with both groups combined, there is a significant increase in the SEP value (around 36%) compared with the mandarins group, which may be caused by the fact that, when both species are taken together, the mean value is higher than that of the latter group.

As regards the maturity index and BrimA parameters for the oranges and mandarins group, both parameters have r^2_p values of 0.70 - 0.90, thus showing a good predictive capacity [32]. In terms of SEP, when LOCAL is applied to all the samples, the error decreases relative to the oranges group, while there is a significant increase in the error ($P < 0.05$) compared with the mandarins group: 163% and 20% for maturity index and BrimA, respectively. However, these SEP values refer to mean values of uncertainty, which means that they vary depending on the mean of the calibration group used to produce each individual model, although individual uncertainty values can vary, being in some cases higher and in others lower [34]. Nevertheless, this lack of precision is to a large extent compensated for by the opportunity of having a model which includes different species, which is of great interest to the citrus fruit industry. In the same way, although maturity index and BrimA are two parameters related to the perception of sweetness or tartness in the fruit, different authors have defined the latter as more useful [4,5], and it obtained a slightly higher predictive capacity than that of the maturity index ($RPD_p = 2.15$ for BrimA *versus* $RPD_p = 2.08$ for maturity index) when LOCAL algorithm is applied.

In general, it is important to stress the usefulness of the LOCAL regression algorithm compared with the linear regression algorithm MPLS to predict chemical quality parameters in the oranges and mandarins group. In particular, as mentioned by other authors [12, 23, 35], the most important factor is the increased robustness attained when applying the LOCAL algorithm to measure quality parameters in fruits, which is notable in this work in the case of pH and titratable acidity parameters, which are both of great interest for the industry and the consumers of these products.

There are no references in the bibliography to authors applying LOCAL regression models in order to measure chemical parameters in groups made up of several species of citrus fruit. However, a number of authors have demonstrated the potential of local regression techniques to measure chemical

parameters in oranges [12], grapes [22], nectarines [23], and apples [24], all of which show increased precision and accuracy when non-linear regression techniques are used, as opposed to linear ones.

6.2.3.5. Effective Wavelengths for the parameter BrimA

Given the value of the BrimA parameter to the citrus industry [5], it was considered important to study the wavelengths that influence its measurement.

To do this, the loading plot corresponding to the best model obtained using MPLS regression to predict BrimA in a set composed of oranges and mandarins using the Phazir 2400 is shown in Figure 6.2.3. This figure shows the areas of the spectral range where covariance has influenced the computing of the MPLS model to a greater or lesser degree, and the direction (positive or negative). A representation of the latent variables (LV5 to LV8) used in constructing the calibration equation shows that the areas of the spectrum exerting higher weight on model were 1730, 1830, 1900 and 2350 nm, related to the absorption of glucides and water [36].

6.2.4. Conclusions

These results confirm that NIR spectroscopy could be an advantageous technique to predict chemical parameters in a set composed by two species belonging to the Citrus genus using the LOCAL regression algorithm in order to establish the quality and maturity indexes of the citrus fruits on-tree. Using the LOCAL algorithm not only represents an improvement in the predictive capacity of the models obtained, but also allows to use multispecies spectral libraries. This is extremely important for the citrus fruit sector, as the libraries can easily be extended to include other citrus species, thus allowing us to obtain universal models. In addition, the results confirm the advantages of using portable equipment which allows to analyse the fruit in the field, in order to harvest the fruits selectively at the optimum time and to obtain a product of

the highest quality which is intended both for fresh consumption and for the processing industry.

From a practical point of view, this could be extremely useful for citrus growers, since it permits them to measure maturity indices such as BrimA quickly and without damaging the fruit, which is essential for setting the optimum harvest time and producing fruit which is acceptable to the consumers.

Acknowledgements

The authors would like to thank Prof. Elías Fereres and Dr. Victoria González-Dugo. We are also grateful to Mrs. Mª Carmen Fernández of the Animal Production Department for her technical assistance. This work was funded by the Spanish Ministry of Science and Innovation for the projects CONSOLIDER CSD2006-0067 and AGL2012-40053-C03-01, with participation of funds from FEDER (European Union) and from the Andalusian Regional Government under the Research Excellence Program, project no. P09-AGR-5129 'MEMS and NIRS-image sensors for the in situ non-destructive analysis of food and feed'. Furthermore, the authors wish to express their gratitude to the Spanish Ministry of Education, Culture and Sports for the support offered to Irina Torres in the form of the Training Programme for Academic Staff (FPU).

References

[1] C.B. Watkins, Postharvest ripening regulation and innovation in storage technology, Acta Hortic. 796 (2008) 51–58, DOI: 10.17660/ActaHortic.2008.796.4.

[2] M.L. Arpaia, A.A. Kader, Orange: recommendations for maintaining postharvest quality.

http://postharvest.ucdavis.edu/Commodity_Resources/Fact_Sheets/Datastores/Fruit_English/?uid=41&ds=798, 1999 (accessed 18 February 2019).

[3] M.L. Arpaia, A.A. Kader, Mandarin/Tangerine: recommendations for maintaining postharvest quality. http://postharvest.ucdavis.edu/Commodity_Resources/Fact_Sheets/Datastores/Fruit_English/?uid=36&ds=798, 1999 (accessed 18 February 2019).

[4] R.B. Jordan, R.J. Seelye, V.A. McGlone, A sensory-based alternative to brix/acid ratio, Food Technol. 55 (2001) 36–44.

[5] D. Obenland, S. Collin, B. Mackey, J. Sievert, K. Fjeld, M.L. Arpaia, Determinants of flavour acceptability during the maturation of navel oranges, Postharvest Biol. Tec. 52 (2009) 156–163, DOI: 10.1016/j.postharvbio.2009.01.005.

[6] E. Arendse, O.A. Fawole, L.S. Magwaza, U.L. Opara, Non-destructive prediction of internal and external quality attributes of fruit with thick rind: A review, J. Food Eng. 217 (2018) 11–23, DOI: 10.1016/j.jfoodeng.2017.08.009.

[7] L.S. Magwaza, U.L. Opara, H. Nieuwoudt, P.J.R. Cronje, W. Saeys, B. Nicolaï, NIR spectroscopy applications for internal and external quality analysis of citrus fruit – a review, Food Bioprocess Tech. 5 (2012) 425–444, DOI: 10.1007/s11947-011-0697-1.

[8] L.S. Magwaza, U.L. Opara, L.A. Terry, S. Landahl, P.J.R. Cronje, H. Nieuwoudt, A. Hanssens, W. Wouter-Saeys, B.M. Nicolaï, Evaluation of Fourier Transform-NIR spectroscopy for integrated external and internal quality assessment of Valencia oranges, J. Food Compos. Anal. 31 (2013) 144–154, DOI: 10.1016/j.jfca.2013.05.007.

[9] C. Liu, S.X. Yang, L. Deng, A comparative study for least angle regression on NIR spectra analysis to determine internal qualities of navel

oranges, Expert Syst. Appl. 42 (2015) 8497–8503, DOI: 10.1016/j.eswa.2015.07.005.

[10] K. Ncama, U.L. Opara, S.Z. Tesfaym, O.A. Fawole, L.S. Magwaza, Application of Vis/NIR spectroscopy for predicting sweetness and flavour parameters of 'Valencia' oranges (*Citrus sinensis*) and 'Star Ruby' grapefruit (*Citrus x paradise* Macfad), J. Food Eng. 193 (2017) 86–94, DOI: 10.1016/j.jfoodeng.2016.08.015.

[11] M.T. Sánchez, M.J. De la Haba, D. Pérez-Marín, Internal and external quality assessment of mandarins on-tree and at harvest using a portable NIR spectrophotometer, Comput. Electron. Agr. 92 (2013a) 66–74, DOI: 10.1016/j.compag.2013.01.004.

[12] M.T. Sánchez, M.J. De La Haba, I. Serrano, D. Pérez-Marín, Application of NIRS for non-destructive measurement of quality parameters in intact oranges during on-tree ripening and at harvest, Food Anal. Method. 6 (2013b) 826–837, DOI 10.1007/s12161-012-9490-7.

[13] A.M. Cavaco, R. Pires, M.D. Antunes, T. Panagopoulos, A. Brázio, A.M. Afonso, L. Silva, M. Rosendo-Lucas, B. Cadeiras, S.P. Cruz, R. Guerra, Validation of short wave near infrared calibration models for the quality and ripening of "Newhall" orange on tree across years and orchards, Postharvest Biol. Tec. 141 (2018) 86–97, DOI: 10.1016/j.postharvbio.2018.03.013.

[14] M.K.D. Rambo, E.P. Amorim, M.M.C. Ferreira, Potential of visible near infrared spectroscopy combined with chemometrics for analysis of some constituents of coffee and banana residues, Anal. Chim. Acta. 775 (2013) 41–49, DOI: 10.1016/j.aca.2013.03.015.

[15] C.J. Clark, Fast determination by Fourier-Transform infrared spectroscopy of sugar-acid composition of citrus juices for determination of industry maturity standards, New Zeal. J. Crop Hort. 44 (2016) 69–82, DOI: 10.1080/01140671.2015.1131725.

[16] D.A. Santos, K.P. Lima, V. Cavalcante, A. Coqueiro, M.F. Barriquello-Consolin, N. Consolin, P.H. Março, P. Valderrama, Multiproduct, multicomponent and multivariate calibration: a case study by using Vis-NIR spectroscopy, Food Anal. Method. 11 (2018) 1915–1919, DOI: 10.1007/s12161-017-1099-4.

[17] I. Torres, D. Pérez-Marín, M.J. De la Haba, M.T. Sánchez, Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors, Biosyst. Eng. 153 (2017) 140–148, DOI: 10.1016/j.biosystemseng.2016.11.007.

[18] J.S. Shenk, M.O. Westerhaus, P. Berzaghi, Investigation of a LOCAL calibration procedure for near infrared instruments, J. Near Infrared Spec. 5 (1997) 223–232, DOI: 10.1255/jnirs.115.

[19] P. Berzaghi, J.S. Shenk, M.O. Westerhaus, LOCAL prediction with near infrared multi-product databases, J. Near Infrared Spec. 8 (2000) 1–9, DOI: 10.1255/jnirs.258.

[20] D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero, J.C. Gutiérrez, Implementation of LOCAL algorithm with Near-Infrared Spectroscopy for compliance assurance in compound feedingstuffs, Appl. Spectrosc. 60 (2005) 1062–1069, DOI: 10.1366/0003702052940585.

[21] B. Godin, R. Agneessens, J. Délcarte, P. Dardenne, Prediction of chemical characteristics of fibrous plan biomasses from their near infrared spectrum: comparing local versus partial least square models and cross-validation versus independent validations, J. Near Infrared Spec. 23 (2015) 1–14, DOI: 10.1255/jnirs.1138.

[22] R.G. Dambergs, D. Cozzolino, W.U. Cynkar, L. Janik, M. Gishen, The determination of red grape quality parameters using the LOCAL algorithm, J. Near Infrared Spec. 14 (2006) 71–79, DOI: 10.1255/jnirs.593.

[23] M.T. Sánchez, M.J. De la Haba, J.E. Guerrero, A. Garrido-Varo, D. Pérez-Marín, Testing of a local approach for the prediction of quality parameters in intact nectarines using a portable NIRS instrument, Postharvest Biol. Tec. 60 (2011) 130–135, DOI: 10.1016/j.postharvbio.2010.12.006.

[24] X. Luo, Z. Ye, H. Xu, D. Zhang, S. Bai, Y. Ying, Robustness improvement of NIR-based determination of soluble solids in apple fruit by local calibration, Postharvest Biol. Tec. 139 (2018) 82–90, DOI: 10.1016/j.postharvbio.2018.01.019.

[25] J.S. Shenk, M.O. Westerhaus, Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy, Monograph, NIRSystems, Inc., 12101 Tech Road, Silver Spring, MD 20904, USA, 1995.

[26] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777, DOI: 10.1366/0003702894202201.

[27] J.S. Shenk, M.O. Westerhaus, Routine Operation, Calibration, Development and Network System Management Manual, NIRSystems, Inc., 12101 Tech Road, Silver Spring, MD 20904, USA, 1995.

[28] J.S. Shenk, M.O. Westerhaus, Population definition, sample selection and calibration procedures for near infrared spectra and modified partial least squares regression, Crop Sci. 31 (1991) 469–474.

[29] ISI, The Complete Software Solution Using a Single Screen For Routine Analysis, Robust Calibrations And Networking, Manual, Foss NIRSystems/Tecator, Infrasoft International, LLC, Silver Spring MD, USA, 2000.

[30] F.E. Barton II., J.S. Shenk, M.O. Westerhaus, D.B. Funk, The development of near infrared wheat quality models by locally weighted

regressions, J. Near Infrared Spec. 8 (2000) 201–208, DOI: 10.1255/jnirs.280

[31] T. Naes, T. Isaksson, T. Fearn, A. Davies, A User-Friendly Guide to Multivariate Calibration and Classification, NIR Publications, Chichester, UK, 2002.

[32] J.S. Shenk, M.O. Westerhaus, Calibration the ISI way, in: A.M.C. Davies, P.C. Williams (Eds.), Near Infrared Spectroscopy: The Future Waves, NIR Publications, Chichester, UK, 1996, pp. 198–202.

[33] B.M. Nicolaï, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron, J. Lammertyn, Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review, Postharvest Biol. Tec. 46 (2007) 99–118, DOI: 10.1016/j.postharvbio.2007.06.024.

[34] H. Martens, M. Martens, Multivariate calibration: quality determination of wheat from high-speed NIR spectra, in: H. Martens, M. Martens, Multivariate Analysis Of Quality An Introduction, Wiley editorial, Chichester, UK, 2001, pp. 233-256.

[35] M.T. Sánchez, I. Torres, M.J. De la Haba, D. Pérez-Marín, First steps to predicting pulp colour in whole melons using near-infrared reflectance spectroscopy, Biosyst. 123 (2014) 12–18, DOI: 10.1016/j.biosystemseng.2014.04.010.

[36] J.S. Shenk, J. Workman, M.O. Westerhaus, Application of NIR spectroscopy to agricultural products, in: D.A. Burns, E.W. Ciurczac (Eds.), Handbook of Near Infrared Analysis, Marcel Dekker, Basel, 2008, pp. 347–386.

Table 6.2.1. Statistics for each set and parameter.

Parameter	Samples	Set	Number of samples	Range	Mean	SD	CV (%)
Soluble solid content (%)	Oranges	Calibration	283	6.80-15.30	10.73	1.91	17.80
		Validation	92	7.50-14.35	10.68	1.78	16.67
	Mandarins	Calibration	173	9.95-15.65	12.51	1.19	9.51
		Validation	56	9.95-15.00	12.58	1.07	8.51
	Oranges + mandarins	Calibration	456	6.80-15.65	11.41	1.88	16.48
		Validation	148	7.50-15.00	11.40	1.80	15.79
pH	Oranges	Calibration	283	3.01-4.15	3.69	0.21	5.69
		Validation	92	3.28-4.03	3.70	0.18	4.86
	Mandarins	Calibration	166	2.08-3.80	3.25	0.20	6.15
		Validation	55	2.86-3.69	3.26	0.21	6.44
	Oranges + mandarins	Calibration	449	2.08-4.15	3.52	0.30	8.52
		Validation	147	2.86-4.03	3.54	0.29	8.19
Titratable acidity (% citric acid)	Oranges	Calibration	282	0.36-1.21	0.62	0.14	22.58
		Validation	92	0.37-1.02	0.62	0.15	24.19
	Mandarins	Calibration	155	0.68-2.15	1.21	0.28	23.14
		Validation	50	0.79-1.77	1.89	0.27	14.29
	Oranges + mandarins	Calibration	437	0.36-2.15	0.83	0.34	40.96
		Validation	142	0.37-1.77	0.82	0.34	41.96
Maturity index (SSC/TA)	Oranges	Calibration	282	8.24-40.03	18.14	5.42	29.88
		Validation	92	8.55-35.79	18.55	6.02	32.45
	Mandarins	Calibration	155	5.41-17.27	10.86	2.32	21.36
		Validation	50	6.68-15.68	11.00	2.42	22.00
	Oranges + mandarins	Calibration	437	5.41-40.03	15.56	5.74	36.89
		Validation	142	6.68-35.79	15.59	6.21	39.83
BrimA index (%)	Oranges	Calibration	282	4.29-13.31	8.26	1.93	23.37
		Validation	92	4.63-12.22	8.22	1.98	24.09
	Mandarins	Calibration	155	2.93-10.33	7.70	1.42	18.44
		Validation	50	4.63-10.28	7.75	1.40	18.06
	Oranges + mandarins	Calibration	437	2.93-13.31	8.06	1.73	21.46
		Validation	142	4.63-12.22	8.05	1.81	22.48

Table 6.2.2. Validation statistics for predicting chemical quality parameters in Citrus fruits using non-linear (LOCAL) and linear (MPLS) regression algorithms and standard errors of laboratory (SEL).

Parameter	Set	LOCAL				GLOBAL			F	$F_{critical}$	SEL
		Settings	SEP	r^2_p	RPD_p	SEP	r^2_p	RPD_p			
Soluble solid content (%)	Oranges	100, 16, 4	0.80	0.81	2.23	0.97	0.75	1.84	1.47	1.40*	0.11
	Mandarins	140, 16, 4	0.71	0.57	1.51	0.84	0.39	1.27	1.40	1.43	0.07
	Oranges + mandarins	80, 14, 4	0.86	0.78	2.09	0.95	0.72	1.89	1.22	1.40	
pH	Oranges	100, 16, 4	0.16	0.25	1.13	0.18	0.15	1.00	1.27	1.40	0.02
	Mandarins	80, 16, 4	0.11	0.74	1.91	0.11	0.74	1.91	1.00	1.50	0.06
	Oranges + mandarins	100, 16, 4	0.15	0.72	1.93	0.17	0.64	1.71	1.28	1.36	
Titratable acidity (% citric acid)	Oranges	100, 15, 4	0.11	0.45	1.36	0.11	0.47	1.36	1.00	1.40	0.004
	Mandarins	100, 15, 4	0.13	0.76	2.08	0.18	0.65	1.50	1.92	1.48*	0.020
	Oranges + mandarins	80, 14, 4	0.14	0.84	2.43	0.18	0.75	1.89	1.65	1.40*	

* Values with significant differences ($P < 0.05$).

Table 6.2.2. Validation statistics for predicting chemical quality parameters in Citrus fruits using non-linear (LOCAL) and linear (MPLS) regression algorithms and standard errors of laboratory (SEL) (continuation).

Parameter	Set	LOCAL				GLOBAL			F	$F_{critical}$	SEL
		Settings	SEP	r^2_p	RPD_p	SEP	r^2_p	RPD_p			
Maturity index (SSC/TA)	Oranges	140, 16, 4	3.56	0.65	1.69	3.70	0.67	1.63	108	1.36	0.13
	Mandarins	100, 16, 4	1.13	0.79	2.14	1.38	0.68	1.75	1.49	1.48*	0.15
	Oranges + mandarins	160, 16, 4	2.98	0.77	2.08	3.52	0.72	1.76	1.40	1.31*	
BrimA index (%)	Oranges	100, 15, 4	0.85	0.82	2.33	0.89	0.80	2.22	1.10	1.40	0.11
	Mandarins	140, 16, 4	0.70	0.75	2.00	0.79	0.68	1.77	1.27	1.45	0.10
	Oranges + mandarins	140, 14, 4	0.84	0.78	2.15	0.94	0.73	1.93	1.25	1.32	

* Values with significant differences ($P < 0.05$).

Figure 6.2.1. Population distribution of chemical quality parameters for oranges (O), mandarins (M) and oranges and mandarins together (O + M).

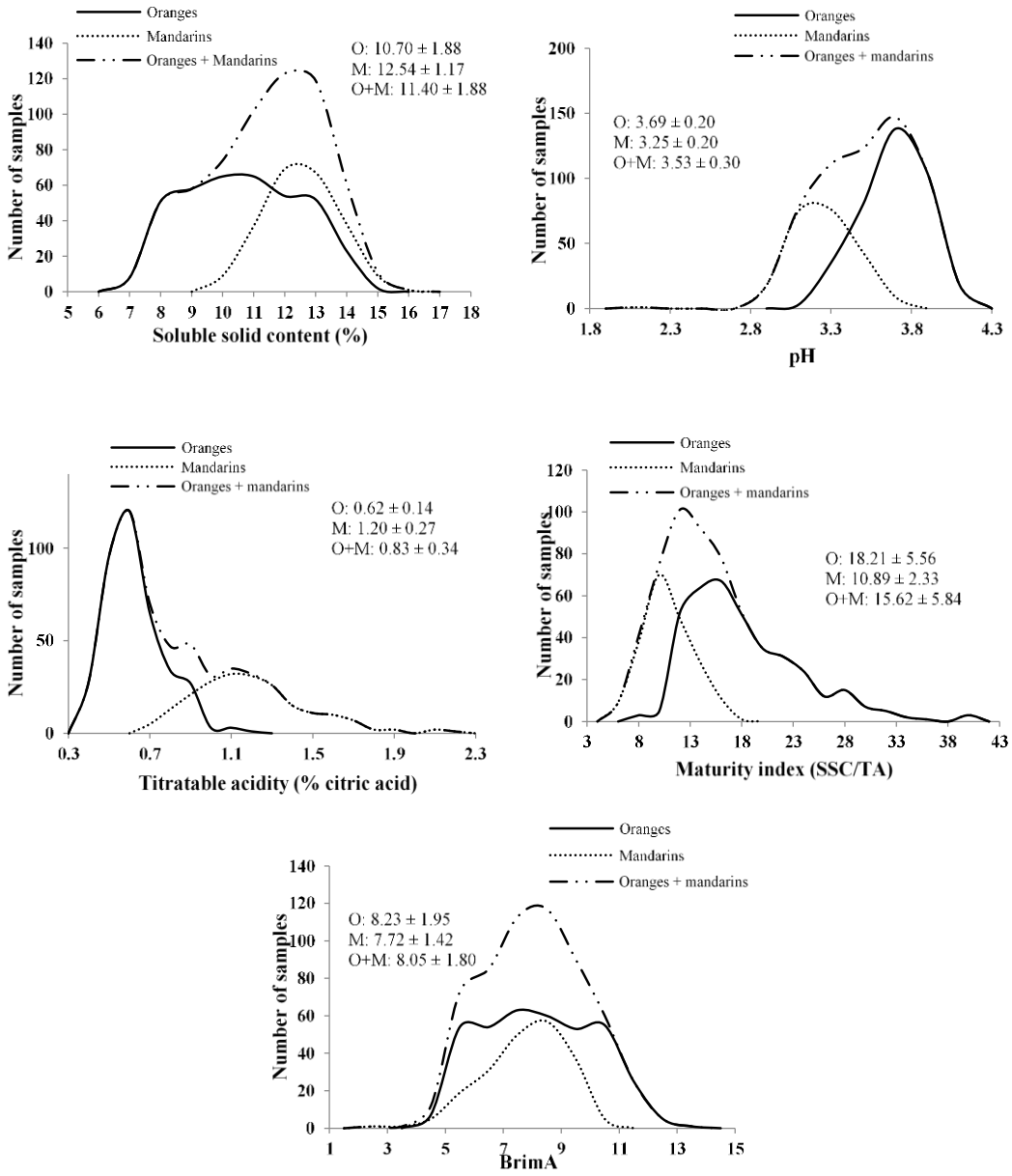
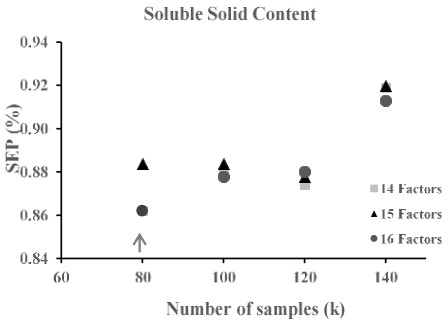
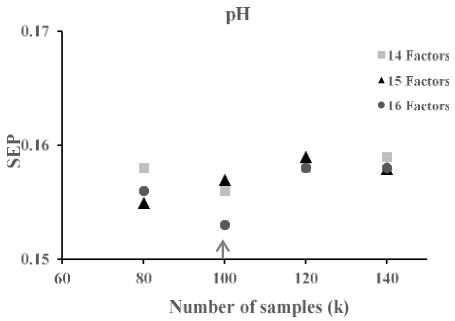


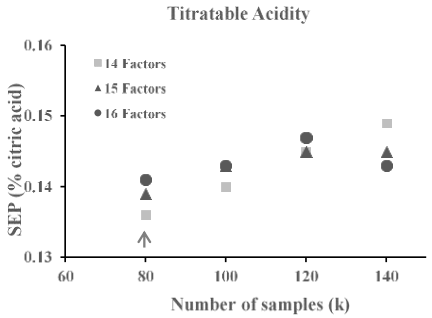
Figure 6.2.2. SEP values obtained for the prediction of chemical quality parameters in the set composed of intact oranges and mandarins using the LOCAL algorithm.



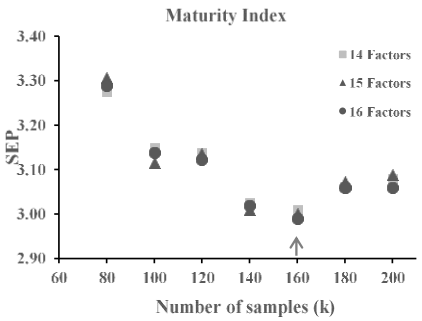
Mathematical treatment: 2,5,5,1



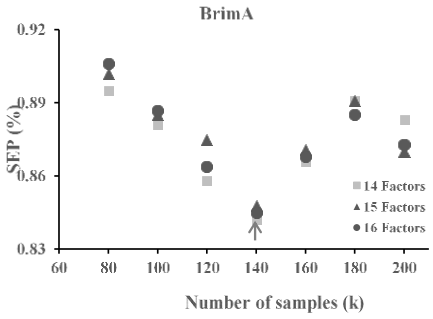
Mathematical treatment: 1,10,5,1



Mathematical treatment: 1,5,5,1

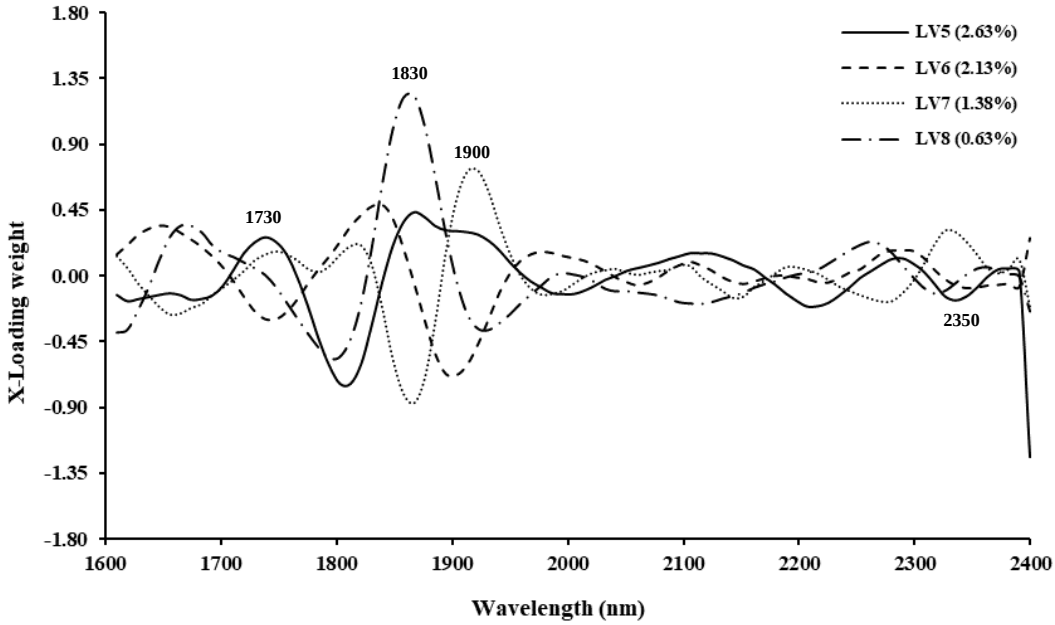


Mathematical treatment: 1,10,5,1



Mathematical treatment: 1,5,5,1

Figure 6.2.3. Loadings for BrimA.



Chapter 7

Chapter 7. SETTING UP A METHODOLOGY FOR THE ON-TREE ESTIMATION OF GREEN CITRUS FRUIT YIELD USING HYPERSPECTRAL IMAGING

Chapter 7.1.

Setting up a methodology to distinguish between green oranges and leaves using hyperspectral imaging

Irina Torres[a], María-Teresa Sánchez[a], Byoung-Kwan Cho[b], Ana Garrido-Varo[c], Dolores Pérez-Marín[c]

[a] Department of Bromatology and Food Technology, College of Agriculture and Forestry Engineering (ETSIAM), University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.
[b] Department of Biosystems Machinery Engineering, College of Agricultural and Life Science, Chungnam National University, 220 Gung-dong, Yuseong-gu, Daejeon 305-764, Republic of Korea
[c] Department of Animal Production, College of Agriculture and Forestry Engineering (ETSIAM), University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.

Computers and Electronics in Agriculture167, 105070 (2019)

Abstract

The estimation of green citrus fruit yield is a key parameter for growers and the industry. The early estimation of orange yield at the immature green stage could influence the future market price and allow producers to plan the harvest in advance, thus reducing costs. This research can be considered as a preliminary step for designing low-cost spectral cameras capable of being mounted on unmanned aerial vehicles (UAVs) to estimate orange yield and defects. Images were acquired from oranges and leaves from an orchard in Jeju island (Jeju, Republic of Korea), using two hyperspectral reflectance imaging systems, one working in the range 400–1000 nm (visible/near infrared, Vis/NIR) and the other between 900–2500 nm (short-wave infrared, SWIR). The main objective of the research was to set up a methodology to select the relevant bands - from the two spectral ranges studied - to distinguish between green oranges and leaves and to detect defects, which will allow citrus yield to be estimated. Analysis of variance (ANOVA) and principal component analysis (PCA) were used to select the key wavelengths for this purpose; next, a band ratio coupled with a simple thresholding method was applied. This study showed that the Vis/NIR hyperspectral imaging correctly classified 96.97% and 92.93% of the pixels, respectively, to distinguish between green oranges and leaves and to detect defects, while with the SWIR system, the percentage of pixels correctly classified for these two objectives were 74.79% and 89.31%, respectively. These results confirm that it is possible to use a low number of wavelengths to estimate harvest yield in oranges, which could pave the way for the future development of low-cost and low-weight equipment for the detection of green and sound fruit.

Keywords: Orange; Harvest yield; Defect detection; Hyperspectral and multispectral imaging

7.1.1. Introduction

The citrus sector is one of the most dynamic and important agricultural sectors. This sector is of considerable economic value to the countries of the Mediterranean Basin, China, Brazil, the United States and Southeast Asia (Republic of Korea), the main producers of citrus fruit, in general, and of oranges in particular (FAO, 2017).

In the light of the economic importance of the orange on the international market, it is of particular interest to obtain an estimate of the crop yield prior to harvesting, which usually takes place when the fruits are of the same green colour as the leaves (Obenland et al., 2009). In oranges, the fruits often reach physiological maturity and present excellent eating quality while the peel is still green. Post-harvest de-greening practices are used to speed up the fruit colour change and to make the fruit more acceptable for marketing, since a shiny, yellow peel is what the market demands (Porat, 2008).

Therefore, tools are needed to identify these green fruits on-tree, to make possibly decision on harvesting and to optimize the process, so that fruits of the highest quality are picked in keeping with their subsequent industrial use.

Currently, the indexes used to determine quality in oranges are colour intensity and uniformity, firmness, size, shape, quality of flavour, lack of decay and lack of defects including physical damage (abrasions and bruising), skin blemishes and discoloration, as well as insect damage (Arpaia and Kader, 1999): this last quality index is one of the most influential factors on yield in fresh oranges (Leemans and Destain, 2004; Li et al., 2011).

These days, the large-scale measurement of these production and quality indicators still constitutes a major difficulty for producers. On the one hand, it is believed that drones with a sensing unit may be used in the near future to achieve this aim due to the rise in their popularity for agricultural

applications; on the other, hyperspectral imaging (HSI) and multispectral imaging (MSI) are two emerging techniques which could be used for this purpose, due to their ability to acquire both spectral and spatial information and thus assess quality indexes in agricultural products (Dale et al., 2013).

As far as the present authors are aware, only two published studies have used HSI technology to distinguish between green oranges and leaves on-tree. In the first, Kane and Lee (2007) used an InGaAs camera with a spectral range of 900 to 1700 nm to identify green oranges in the field. They applied a combination of three different spectral band images to identify the green orange and achieved an accuracy of 84.5% in terms of correctly classified pixels. In the other, Okamoto and Lee (2009) using a CCD hyperspectral camera in the 369–1042 nm range applied stepwise forward variable selection method and linear discriminant analysis was then developed with selected variables - between 10 and 14 - to identify green citrus fruits in the field at different stages, with detection accuracies for complete fruit ranging between 80 to 89%.

As regards detecting external defects in oranges, a number of studies have been published to date, all of which were carried out under laboratory conditions, with the aim of selecting the optimal wavelengths for this application with a multispectral system (Li et al. 2011; Bulanon et al., 2013; Lorente et al., 2013; Li et al., 2016).

In addition, the recent development of small-sized hyperspectral and multispectral sensors has enabled them to be attached to UAVs in order to obtain images with a high spectral and spatial resolution (Dale et al., 2013). The use of multispectral instead of hyperspectral cameras would reduce the cost of the system and would speed up the data analysis (Kim et al., 2011), although it would be necessary to make a prior selection of the optimal spectral bands for each particular objective.

Likewise, it is important to consider that the HSI measurement devices can affect the quality of the image data, so the selection of the instrument and the wavebands play an important role in optimising the performance of the application (Kim et al., 2011). Most of the research carried out using multi and hyperspectral cameras to measure quality attributes (sweetness and acidity, firmness, stage of maturity or detection of defects) in fruit and vegetables have been carried out in the Vis/NIR region (400–1000 nm) (Kim et al., 2011; Li et al., 2018), although other studies have also used instruments in the SWIR range (1000–2500 nm) (Gowen et al., 2007). This spectral range is of particular interest to the citrus sector, since it not only allows us to differentiate between specific areas of the fruit (i.e. for detecting defects), but also includes the most suitable wavelengths for measuring the chemical parameters, such as soluble solid content and acidity (Williams, 2001).

Thus, given the numerous options available in terms of equipment, the successful implementation of hyperspectral or multispectral reflectance technology *in-situ* for crop yield estimation requires the instrument to be selected previously and the correct waveband combination to be found for this specific application. Although this aspect is hugely relevant when designing low-cost and low-weight cameras, to our knowledge there are no reports in the literature regarding the comparison between Vis/NIR and SWIR HSI systems to identify green citrus fruits and to pick out defective ones.

The objective of this work was to evaluate – at the laboratory scale – two line-scan hyperspectral reflectance imaging systems working in the Vis/NIR and SWIR ranges, respectively, to estimate the crop yield of oranges based on the distinction between green oranges and leaves and the detection of external defects (abrasions and bruising, skin blemishes and discoloration) in the oranges. The performance of these two systems was also compared and the optimal wavelengths were selected for the further development of a low-cost and low-weight MSI system which could be mounted on drones.

7.1.2. Material and methods

7.1.2.1. Sampling

The oranges and leaves were obtained from an orchard in Jeju island (Jeju, Republic of Korea) in autumn 2017.

For the first of the objectives – to differentiate green oranges from leaves – 20 leaves and 24 green oranges were placed in plastic bags and immediately taken to the laboratory. In the case of the leaves, wet tissues were put into the bags to maintain the moisture content of the leaves. Once in the laboratory, images of the leaves and green oranges were taken. Next, a set made up of 5 samples of green oranges together with their leaves was measured to validate the model obtained.

For the second objective – to identify oranges with defects – a total of 20 oranges with some external defects (abrasions and bruising, skin blemishes and discoloration) were measured.

Although the number of samples available in this study could appear to be limited, it is enough to evaluate the potential of the developed methodology.

7.1.2.2. Hyperspectral imaging systems and image acquisition

Two laboratory-based push-broom Vis/NIR and SWIR systems were used to obtain the hyperspectral images of oranges and leaves. Details of both imaging systems are given in Table 1.

The Vis/NIR system was made up of an Electron Multiplying Charge Coupled Device (EMCCD) camera (Luca R DL-604M, 14-bit, Andor Technology, South Windsor, CT, USA), a C-mount objective lens (F1.4 28-mm compact lens, Schneider Optics, Hauppauge, NY, USA), a line scan imaging spectrograph (Vis/NIR, Headwall photonics, Fitchburg, MA, USA) and halogen light sources (4 x 2 sets of 100 W) at a 45° angle to the sample. Next,

in order to reduce specular reflection, the system was equipped with a glass rotating polarizer available for Vis/NIR region. The nominal reflectance range was approximately 400 to 1000 nm, with a spectral resolution of 4.7 nm. In order to capture the actual shape of samples by keeping the pixels shape nearly square, the number of lines was set at 1200 and 1080, with the distance between the lines set at 0.250 and 0.278 mm for the oranges and the leaves, respectively. Spectral data was stored on a 1200 x 502 x 128 hypercube for the oranges and a 1080 x 502 x 128 hypercube for the leaves.

For the SWIR system, a 384 x 256 pixel InGaAs camera (MCT, Headwall Photonics, Fitchburg, MA, USA) with spectrograph (Headwall Photonics, Fitchburg MA, USA) and C-mount 1.4/25 mm focal length lens (Navitar, Inc., Rochester, NY, USA) was used to collect images over a wavelength range of 900 to 2500 nm with 6 nm spectral resolution. The illumination for reflectance imaging was provided by six tungsten halogen lamps (100 W) connected by optical fibres and set up at a 45º angle. Line-by-line images were collected by a conveyor unit enable to cover the spatial shape of the samples; it was set to move at a 0.328 mm/scan for oranges and a 0.364 mm/scan for leaves.

To obtain the imaging using each hyperspectral system, 3 oranges were placed on a tray (30 x 12.5 cm) in a single row, while for the leaves, a batch of four samples on a 30 x 14 cm tray was imaged with a single take. The sound, green orange samples were arranged so that the stems pointed upwards, whereas the defective oranges were manually arranged to present the defects for imaging. As for the leaves, these were arranged with the adaxial side facing upwards.

Tablet movement was controlled by the step interval and the number of steps. Visual Basic 6.0 (Microsoft, Seattle, WA, USA) was used to run the HSI system and to control both conveyor and motor speed (0.25 mm/s for Vis/NIR and 8 mm/s for SWIR), instantaneous field of view (IFOV) and exposure time

(Table 7.1.1). The two-dimensional spectral and spatial data were captured by the EMCCD and InGaAs cameras and stored in raw format as a 3D hypercube (two spatial and one spectral dimension), which comprised each spatial location and spectral information at each wavelength (λ).

Due to the heterogeneous intensities of the light source across the whole wavelength range, reflectance calibration was performed before each measurement by taking dark and white reference images. The dark reference was obtained by covering the camera lens with a black cap (0% reflectance) and the white reference by using white Teflon (99% reflectance). The reflectance value (R) of the raw images (R_0) was calculated using the dark (D) and white (W) reference and taking into account the correction factor for the reference panel (C) as in following equation (Kim et al., 2001):

$$R = \frac{R_0 - D_i}{W_i - D_i} \; x \; C_i$$

where *i* is the pixel index (i = 1, 2, 3, …, n) and *n* is the total number of pixels and the correction factor (C) of 1 was used.

7.1.2.3. Image processing and analysis

The steps followed in the full procedure of image processing and analysis for both instruments are shown in Figure 7.1.1. To achieve this aim, Matlab software (version 2015a, The Mathworks, Natick, MA, USA) was used. In the case of the SWIR system, the spectral range used was 900 to 1900 nm due to the fact that there was no signal beyond this wavelength.

After reflectance correction, the region of interests (ROIs) for sound green oranges and leaves were manually selected using a reflectance image. The spectra of all pixels in each ROI was extracted and averaged to obtain the mean intensity value for each wavelength.

For the identification of oranges, the selection of the most significant wavelengths for the distinction of green oranges and leaves was based on F-values of the analysis of variance (ANOVA) between the two groups (i.e. green oranges *versus* leaves). The higher the F-value, the more statistically significant the mean separation between groups (Neter et al., 1996; Cho et al., 2013). In the case that only one wavelength could be extracted from the ANOVA analysis, PCA was also used to determine the other wavelength needed to obtain the ratio. Thus, the wavebands which presented the greatest difference were used in the application of the ratio image.

For the detection of defects, PCA was used for all the hyperspectral data sets, including the spatial and spectral dimensions. This algorithm reduces the spectral dimensionality, since it converts the huge amount of data from the hypercube into a limited set of scores and loadings. In this work, the PC images and the loading vectors for the first three principal components (PC1, PC2, PC3) were used to select robust wavelengths for the proposed objectives; for this goal, the mean centre was performed as a pre-processing method (Wise et al., 2006).

Prior to using the PCA, a binary mask image was generated in order to avoid interferences from background that could decrease the accuracy of the method. To achieve this, the images at wavelengths 712.5 nm and 1065.11 nm were used for the segmentation for the Vis/NIR and SWIR hyperspectral images, respectively, since they showed the best contrast between the sample and the background. The background was removed by setting a simple threshold value ($R < 0.045$ for Vis/NIR and $R < 0.073$ SWIR) for these wavelengths, respectively, which was then applied to all the hypercubes.

Because the band-ratio can enhance the contrast between different regions (Vargas et al., 2005), two different band-ratios were used to distinguish between the green oranges and the leaves and to detect external defects in the oranges. The two-band ratio was performed as the following equation:

$$Q_{t/k} = \frac{R_t}{R_k}$$

where $Q_{t/k}$ represents a quotient of spectral reflectances, and R_t and R_k are the reflectance intensities at t nm and k nm.

The frequency histograms of the ratio values were recorded in order to select the optimal threshold values. Finally, the accuracy of the models was calculated as the percentage of correctly classified pixels.

To ensure the robustness of the models developed, their external validation was carried out. To achieve this, for differentiating between the leaves and the green oranges, the 5 samples of oranges attached to leaves were used, while to detecting the defects, 25% of the total defective samples (i.e. 5 fruits not used to develop the model) were randomly selected for validation.

7.1.3. Results and discussion

7.1.3.1. Spectral analysis

Representative mean reflectance spectra of green oranges and leaves after normalization, calculated from the pixel values of the ROIs for each system, are shown in Figure 7.1.2.

Figure 7.1.2(a) shows the mean spectra obtained using the Vis/NIR system in the spectral region 400–1000 nm. Although the spectral patterns for both – green oranges and leaves- were fairly similar, in the green-yellow area of the spectrum (500~600 nm) the average reflectance obtained from the green oranges samples was higher than that obtained from the leaves. This makes sense, since in the visible range, the dominant process taking place is pigment absorption; in particular, the peak around 530 nm is due to β-carotene (Keşan et al., 2016). In addition, one dominant spectral feature observed in both spectra is the absorption of chlorophyll *a* at approximately 680 nm (Cho et al., 2013),

whereas in the range between 900 and 1000 nm, corresponding to the water band (Williams, 2001), the oranges display lower intensity than the leaves.

The mean spectra obtained using the SWIR system in the spectral region 900–1900 is shown in Figure 7.1.2(b). As in the case of the Vis/NIR system, the characteristic shape of both spectra is very similar, with peaks and valleys in the same wavelengths. Here, the greatest difference in reflectance values occurred mainly around 1100 nm and 1400-1600 nm. These values were related to the molecular vibrations corresponding to the C–H and O–H bonds (Williams, 2001).

7.1.3.2. Optimal wavelengths selection to distinguish between green oranges and leaves

Figure 7.1.3 shows F-values for the ANOVA analysis between the two groups (i.e. green oranges *versus* leaves) for each wavelength and the two systems tested.

With the Vis/NIR system, the F-values of each waveband in the range 400–1000 nm obtained from the ANOVA analysis for distinguishing oranges and leaves are displayed in Figure 7.1.3(a). The highest F values were obtained for the bands 941.7 nm and 951.2 nm. Since these two bands were fairly close to each other, the band that yielded the highest F-value was selected; in this case, the band 951.2 nm was chosen as one of the dominant bands. This wavelength is very close to one of those chosen (967.2 nm) by Okamoto and Lee (2009) to fulfil this objective using a sample group consisting of 3 varieties of green oranges. This band corresponds to water absorption, which is the main component in oranges (Williams, 2001).

Since only one wavelength could be extracted from the ANOVA analysis, PCA had to be used to determine the other wavelength needed to obtain the ratio.

It was visually determined that the PC2 image (with 1.07% of the explained variance) appeared to provide the best discrimination between green oranges and leaves. The PC2 weighting coefficients showed high positive values in the red region and negative values in the NIR region related to the O-H bond (Williams, 2001); the 698.2 nm wavelength, related to the absorption of chlorophyll *a*, was taken as the maximum, dominant wavelength (Cho et al., 2013).

Based on these results, the 698.2 nm and 951.2 nm bands, obtained from the PCA and ANOVA analyses, respectively, were selected for the MSI design. The raw and binary images, as well as the image obtained after the application of band ratio $R_{\lambda 698}/ R_{\lambda 951}$ and its corresponding frequency histogram, are shown in Figure 7.1.4.

To obtain the global classification capacity of the model, this band ratio was applied to the validation set. The results indicated that for the $R_{\lambda 698}/ R_{\lambda 951}$ band ratio, the highest classification accuracy (96.97%) for oranges *versus* leaves was obtained using a threshold value of 1.00, which was obtained from the frequency histogram (Figure 7.1.4(d)), in which two clearly differentiated populations can be observed, with an overlap between the intensity values of 0.90 and 1.00.

In the same way as in the Vis/NIR system, when using the SWIR system, the dominant bands were selected from the F-values of ANOVA between the groups of green oranges and leaves (Figure 7.1.3(b)). The results showed that the wavelengths which gave rise to a more significant separation were 1165, 1259 and 1471 nm.

Since the purpose of this study was to minimize the number of spectral bands so that the measurement system to be developed would be as light weight, simple and economical as possible, only two of these wavelengths were selected, those whose ratio provided the greatest differentiation between leaves

and fruits. Thus, from all the two-band combinations available, the ratio of wavebands at 1165 and 1471 nm ($R_{\lambda 1165}$/ $R_{\lambda 1471}$), which corresponded to the molecular bonds C-H and the molecular vibrations caused by O-H bonds**,** respectively (Williams, 2001), produced the clearest separation, as shown in Figure 7.1.5. These wavelengths coincide with two of those selected by Kane and Lee (2007) for on-tree green citrus fruit identification**,** which is the only work with this objective found in the literature.

The validation results showed that by applying a simple threshold value (2.60), obtained from the frequency histogram (figure not shown), the classification accuracy was about 74.79%.

The accuracy obtained using the SWIR system was around 22% lower than that yielded with the Vis/NIR system. In Figure 7.1.5, it can be seen that when applying the band-ratio to the image with the SWIR equipment, not only is there a less distinct separation between the leaf and the green orange than with the Vis/NIR system (Figure 7.1.4(c)), but there is also an area in the centre of the fruit with less intensity which corresponds to the specular reflection caused by the geometry of the orange (Garrido-Novell et al., 2012). With the Vis/NIR system, this difference was reduced by using the polarizer, whose function is to reduce specular reflection in samples with curved or shiny surfaces; however, this was not possible with the SWIR system, as this accessory is not easily available for the SWIR spectral range. Thus, the problem is not the range of the camera used, but the lack of a polarizer to reduce the effect of specular reflection on the fruit.

7.1.3.3. Optimal wavelengths selection to detect external defects in oranges

Figure 7.1.6 shows the score images and loadings plots for the first three PCs obtained in the calibration set of hyperspectral images of defective oranges after background removal using the Vis/NIR system. The first three PCs accounted for 99.92% of the explained variance, and it was found that the PCs above the third did not provide any useful information for detecting

defects (data not shown). In the image corresponding to PC1 (96.37%), the areas of the oranges which especially stood out are those where, due to the spherical shape of fruit, higher intensities were produced because they were closer to the camera (Lee et al., 2008). Subsequent PC images represent other features ordered according to variations in spectral responses. In general terms, PC2 images exhibit the greatest contract between the sound and defective areas of the oranges, and so appear to have high discrimination power for identifying the defective areas**.** In PC3, the defects can be observed, and the stem of the fruit can clearly be seen, although no visual differences between these two features were recognized.

In addition, Figure 7.1.6 also shows the loadings for the PC images (PC1–PC3) obtained from the hyperspectral images across the Vis/NIR region. The peaks and valleys show the dominant wavelengths, with maximums of 760 nm observed in PC1, 679 nm in PC2 and 665 nm in PC3, and minimums around 755 nm and 693 nm in PC2 and PC3, respectively, with no minimum of note observed in PC1. In view of these results, it can be stated that within the visible spectrum range, the red region and, in particular, those wavelengths related to the absorption of chlorophyll a, are predominant (Martínez-Valdivieso et al., 2014; Garrido et al., 2016).

Based on the visual aspect, PC2 is the component which seems to provide the best detection of defective areas in oranges. Thus, based on the loading plot obtained for this PC, the two most powerful spectral bands (679 and 755 nm) in PC2 were selected.

The resultant band ratio ($R_{\lambda 679}$/ $R_{\lambda 755}$) was applied to the reflectance images, with which the contrast between the sound surface and defects was more noticeable. After the application of the mask, the threshold was established to isolate the defective surface.

However, given the high level of heterogeneity present in the samples, when the validation of the model was carried out, the threshold value with which the highest accuracy was reached was not the same for all the samples. As a result, to find an optimal threshold value for separating sound from defective areas in oranges, the classification accuracy was calculated with threshold values within the range 0.23–0.35 in an increment of 0.02.

Figure 7.1.7 shows the classification accuracy as a function of the threshold value established. After analysing the results shown in Figure 7.1.7, it can be concluded that the highest accuracy (92.93% of the correctly classified pixels) was reached with a threshold value of 0.35. These results were similar to those obtained by Li et al. (2011), who selected bands 630 and 687 nm by analysing the principal components to detect 9 types of defects in 'Navel' oranges, and, after applying the ratio, obtained a precision of 98.2% in terms of correctly classified pixels.

To discriminate between defective and sound areas using the SWIR system, the optimal wavebands were investigated using the same methodology used with the Vis/NIR system. Thus, from the loading plot for PC2, the 1206 and 1518 nm wavebands were selected, which are related with C–H y O–H absorptions, respectively (Williams, 2001). The ratio image ($R_{\lambda 1206}$/ $R_{\lambda 1518}$) was created, using 1206 nm and 1518 nm images.

According to Figure 7.1.7, which shows the accuracy obtained for each threshold value, the threshold value that yielded the best classification accuracy (89.31%) for the validation set was 0.29.

For this second objective, the difference in the accuracy obtained by both systems was not as great as in the differentiation between green leaves and oranges, and, the Vis/NIR system enabled to obtain the model with the greatest accuracy.

7.1.4. Conclusions

The results obtained in this study indicate the feasibility of using HSI technology to measure crop yield in oranges. The HSI systems can also potentially be developed further as a low-cost multi-spectral imaging system using the key wavelengths identified with the PCA method together with a simple ANOVA analysis from the calibration sets. Four wavelengths (679, 698, 755 and 951 nm) could potentially be implemented as MSI systems to differentiate green oranges from leaves and to detect orange peel defects, respectively, in the Vis/NIR system, and four wavelengths (1165, 1206, 1471 and 1518 nm) could also be used for the same purpose using the SWIR device.

For the two objectives proposed in this study (identification of fruits and detection of defects) in green orange, after using a two-band ratio coupled with a simple threshold method, a comparison of the two hyperspectral devices produced a better classification performance with the Vis/NIR system than with the SWIR system, with an accuracy of 96.97% when distinguishing between green oranges and leaves and an accuracy of 92.93% when detecting defects. Therefore, it could be concluded that Vis/NIR was the most suitable system for this application, with the added advantage of the equipment being more economical than the SWIR. However, it must be highlighted that the use of the polarizer with Vis/NIR system improved the signal reducing the specular reflection in samples, while for the SWIR system this accessory is not easily available. In addition, it must be added that if, as well as estimating the crop yield, certain chemical quality parameters in oranges also need to be measured simultaneously, it would be of great interest to incorporate a band related to the absorption of water or glucides, which would require the use of the spectral range of the SWIR system.

This work can be considered as a feasibility study and further studies are needed for in field application of these systems. In this study samples were

measured in laboratory conditions using halogen lights and for remote sensing, under sun-light illumination, other factors must be taken into account.

Acknowledgements

This work was carried out under the bilateral (Spain-Republic of Korea) EUREKA R & D project entitled 'Development of ICT fusion smart farm technology for the intelligent production and distribution of oranges, INTELLIGENT-CITRUS'. The authors wish to express their gratitude to the Spanish Ministry of Education, Culture and Sports for the support offered to Irina Torres Rodríguez through the Training Programme for Academic Staff (FPU).

References

Arpaia, M.L., Kader, A.A., 1999. Orange. Recommendations for Maintaining Postharvest Quality. Perishables Handling #98. May. Postharvest Technology Center, University of California, Davis, USA.

Bulanon, D.M., Burks, T.F., Kim, D.G., Ritenour, M.A., 2013. Citrus black spot detection using hyperspectral image analysis. Agric. Eng. Int.: CIGR J. 15, 171–180. https://doi.org/10.3965/j.ijabe.20140706.004.

Cho, B., Kim, M.S., Baek, I.S., Kim, D.Y., Lee, W.H., Kim, J., Bae, H., Kim, Y.S., 2013. Detection of cuticle defects on cherry tomatoes using hyperspectral fluorescence imagery. Postharvest Biol. Technol. 76, 40–49. http://dx.doi.org/10.1016/j.postharvbio.2012.09.002.

Dale, L.M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., Fernández-Pierna, J.A., 2013. Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: A review. Appl. Spectrosc. Rev. 48, 142–159. https://doi.org/10.1080/05704928.2012.705800.

FAO, 2017. Citrus Fruit Fresh and Processed Statistical Bulletin 2016. Food and Agriculture Organization of The United Nations, Rome, Italy.

Garrido, Y., Tudela, J.A., Hernández, J.A., Gil, M.I., 2016. Modified atmosphere generated during storage under light conditions is the main factor responsible for the quality changes of baby spinach. Postharvest Biol. Technol. 114, 45–53. https://doi.org/10.1016/j.postharvbio.2015.12.001.

Garrido-Novell, C., Pérez-Marín, D.C., Amigo, J.M., Fernández-Novales, J., Guerrero, J.E., Garrido-Varo, A., 2012. Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras. J. Food Eng. 113, 281–288. https://doi.org/10.1016/j.jfoodeng.2012.05.038.

Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G., Frias, J.M., 2007. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. Trends Food Sci. Tech. 18, 590–598. https://doi.org/10.1016/j.tifs.2007.06.001.

Kane, K.E., Lee, W.S., 2007.Multispectral imaging for in-field green citrus identification. ASABE Paper No. 073025. ASABE, St. Joseph, MI, USA. https://doi.org/10.13031/2013.23169.

Keşan, G., Litvin, R., Bina, D., Durchan, M., Slouf, V., Polivka, T., 2016. Efficient light-harvesting using non-carbonyl carotenoids: Energy transfer dynamics in the VCP complex from *Nannochloropsis oceanica*. Biochim. Biophys. Acta Bioenerg. 1857, 370–379. https://doi.org/10.1016/j.bbabio.2015.12.011.

Kim, M.S., Chen, Y.R., Mehl, P.M., 2001. Hyperspectral reflectance and fluorescence imaging system for food quality and safety. T. ASAE 44, 721–729. https://doi.org/10.13031/2013.6099.

Kim, M.S., Chao, K., Chan, D.E., Jun, W., Lefcourt, A.M., Delwiche, S.R., Kang, S., Lee, K., 2011. Line-scan hyperspectral imaging platform for agro-food safety and quality evaluation: system enhancement and characterization. T. ASABE 54, 703–711. https://doi.org/10.13031/2013.36473.

Lee, K., Kang, S., Delwiche, S.R., Kim, M.S., Noh, S., 2008. Correlation analysis of hyperspectral imagery for multispectral wavelength

selection for detection of defects on apples. Sens. Instrumen. Food Qual. 2, 90–96. https://doi.org/10.1007/s11694-008-9046-0.

Leemans, V., Destain, M.F., 2004. A real-time grading method of apples based on features extracted from defects. J. Food Eng. 61, 83–89. https://doi.org/10.1016/S0260-8774(03)00189-4.

Li, J., Rao, X., Ying, Y., 2011. Detection of common defects on oranges using hyperspectral reflectance imaging. Comput. Electron. Agric. 78, 38–48. https://doi.org/10.1016/j.compag.2011.05.010.

Li, J., Huang, W., Tian, X., Wang, C., Fan, S., Zhao, C., 2016. Fast detection and visualization of early decay in citrus using Vis-NIR hyperspectral imaging. Comput. Electron. Agric. 127, 582–592. https://doi.org/10.1016/j.compag.2016.07.016.

Li, X., Li, R., Wang, M., Liu, Y., Zhang, B., Zhou, J., 2018. Hyperspectral imaging and their applications in the nondestructive quality assessment of fruits and vegetables. In: Luna-Maldonado, A.L., Rodríguez-Fuentes, H., Vidales-Contreras, J.A. (Eds.), Hyperspectral Imaging in Agriculture, Food and Environment. IntechOpen Limited, London, UK, pp. 27–63. https://doi.org/10.5772/intechopen.72250.

Lorente, D., Blasco, J., Serrano, A.J., Soria-Olivas, E., Aleixos, N., Gómez-Sanchís, J., 2013. Comparison of ROC feature selection method for the detection of decay in citrus fruit using hyperspectral images. Food Bioprocess Technol. 6, 3613–3619. https://doi.org/10.1007/s11947-012-0951-1.

Martínez-Valdivieso, D., Font, R., Blanco-Díaz, M.T., Moreno-Rojas, J.M., Gómez, P., Alonso-Moraga, A., Del Río-Celestino, M., 2014. Application of near-infrared reflectance spectroscopy for predicting carotenoid content in summer squash fruit. Comput. Electron. Agric. 108, 71–79. http://dx.doi.org/10.1016/j.compag.2014.07.003.

Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W., 1996. Applied Linear Statistical Models, fourth ed. McGraw-Hill, Columbus, USA.

Obenland, D., Collin, S., Mackey, B., Sievert, J., Fjeld, K., Arpaia, M.L., 2009 Determinants of flavour acceptability during the maturation of navel

oranges, Postharvest Biol. Tecnol. 52, 156–163. https://doi.org/10.1016/j.postharvbio.2009.01.005.

Okamoto, H., Lee, W.S., 2009. Green citrus detection using hyperspectral imaging. Comput. Electron. Agric. 66, 201–208. https://doi.org/10.1016/j.compag.2009.02.004.

Porat, R., 2008. Degreening of citrus. Tree Forest. Sci. Biotechnol. 2, 71–76.

Vargas, A.M., Kim, M.S., Tao, Y., Lefcourt, A.M., Chen, Y.R., Luo, Y., Song, Y., Buchanan, R., 2005. Defection of fecal contamination on cantaloupes using hyperspectral fluorescence imagery. J. Food Sci. 70, 471–476. https://doi.org/10.1111/j.1365-2621.2005.tb11517.x.

Williams, P.C., 2001. Implementation of near-infrared technology. In: Williams, P.C., Norris, K.H. (Eds.), Near Infrared Technology in the Agricultural and Food Industries. AACC Inc., St. Paul, Minnesota, pp. 145–169.

Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, W., Koch, R.S., 2006. PLS_ToolBox 4.0. Manual for use with MATLAB (TM) [Computer software]. Eigenvector Research, Inc., Wenatchee, WA, USA.

Table 7.1.1. Details of the two hyperspectral cameras used.

Wavelength range	**Vis/NIR (400–1000 nm)**	**SWIR (900–2500 nm)**
Manufacture	Andor Technology (South Windsor, CT, USA)	Headwall Photonics (Fitchburg MA, USA)
Sensor	EMCCD	InGaAs
Bit depth	14 bits	12 bits
Spatial resolution	8 µm	24 µm
Number of bands	128	275
Spectral resolution	~ 4.7 nm	~ 6 nm
Illumination	Eight 100 W tungsten halogen lamps	Six 100 W tungsten halogen lamps
Exposure time	10 ms	50 ms

Figure 7.1.1. Comprehensive flow for data analysis.

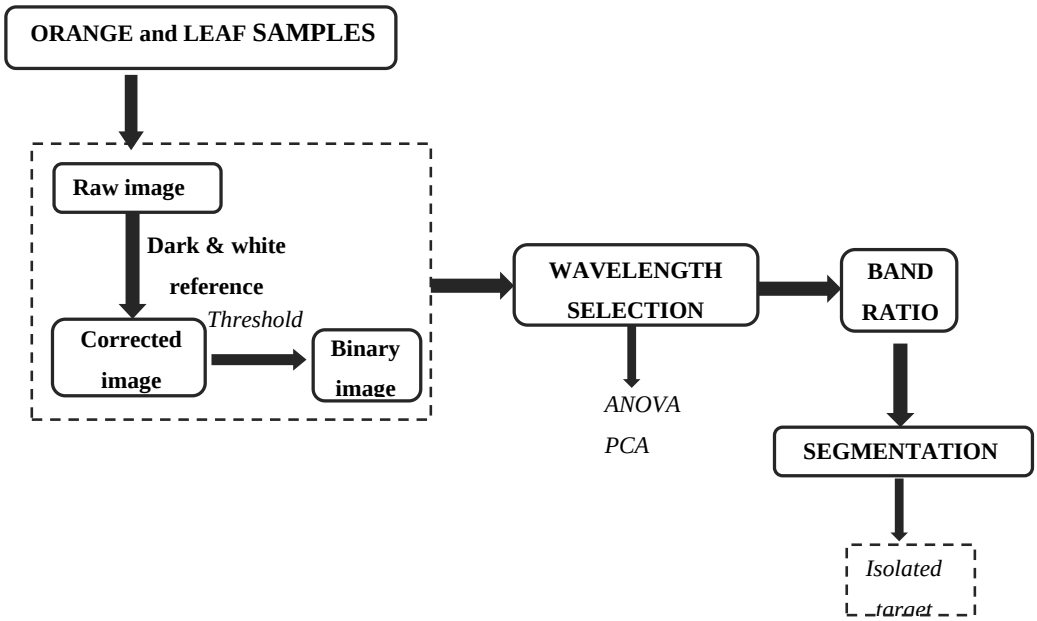


Figure 7.1.2. Spectral features for leaves and green oranges obtained using the Vis/NIR (a) and SWIR (b) hyperspectral imaging systems.

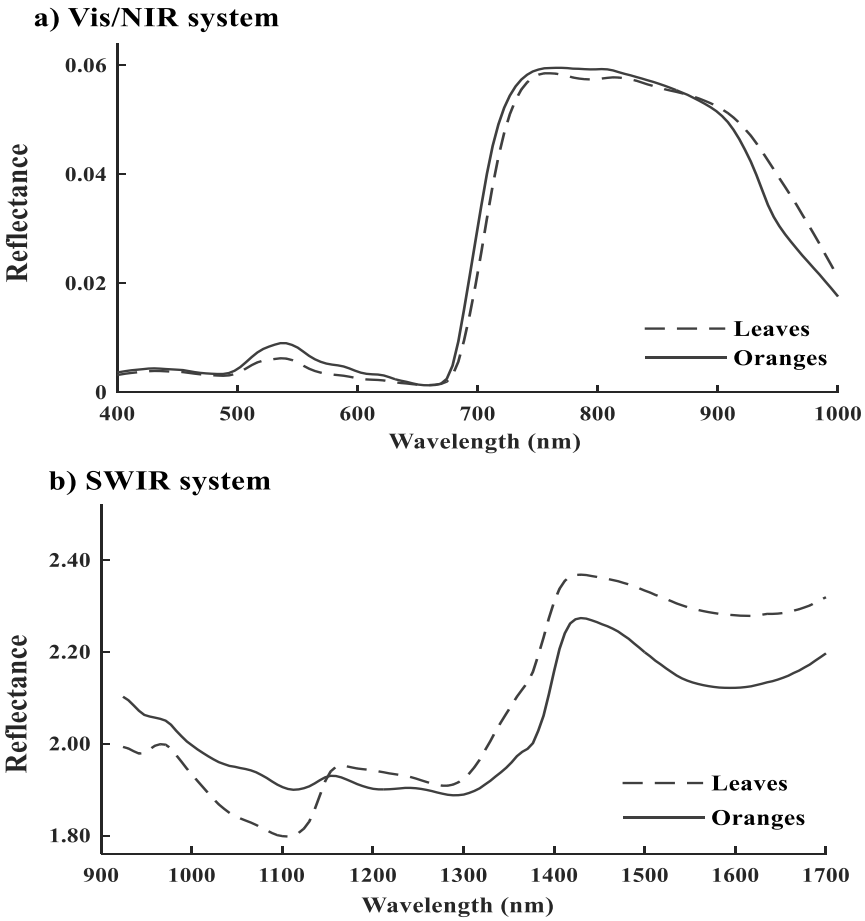


Figure 7.1.3. F-values obtained for the distinction between green oranges and leaves using the Vis/NIR (a) and SWIR (b) hyperspectral imaging systems. Loading plots for the first three principal components for the Vis/NIR (c) and SWIR (d) data sets.

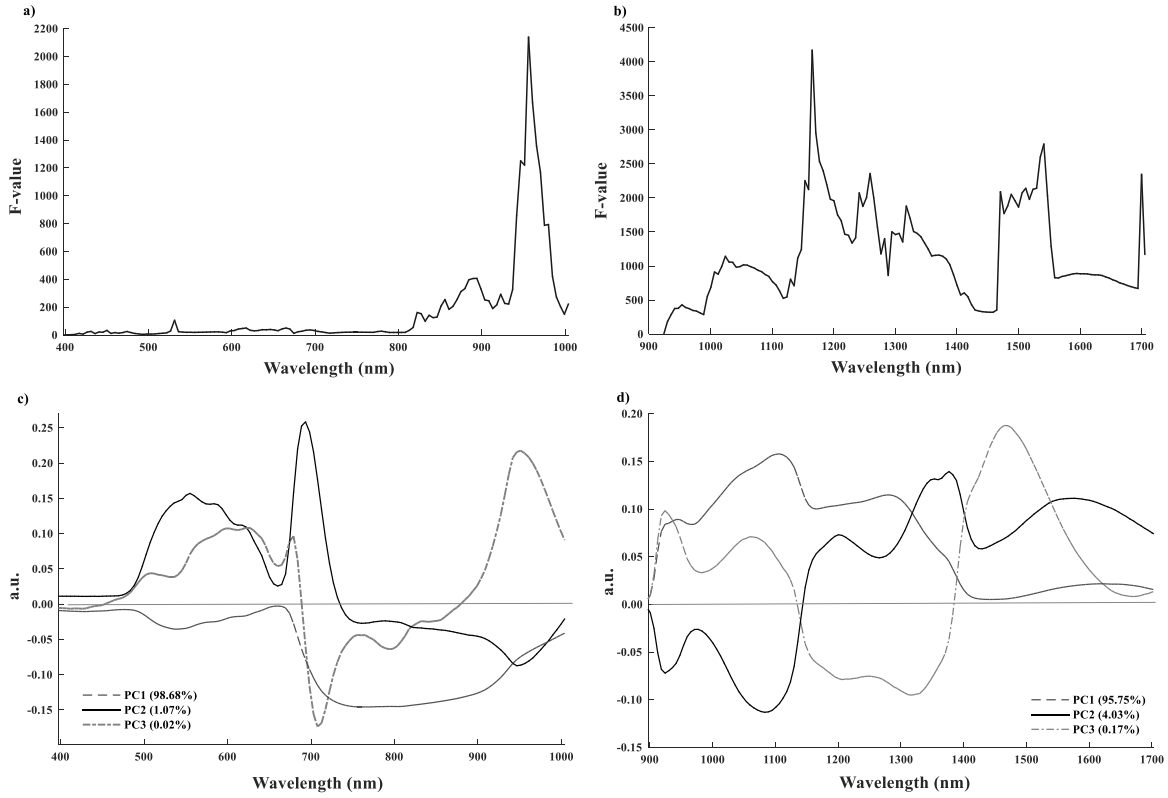


Figure 7.1.4. (a) Reflectance image, (b) Binary image (mask), (c) Ratio image, (d) Histogram of frequencies.

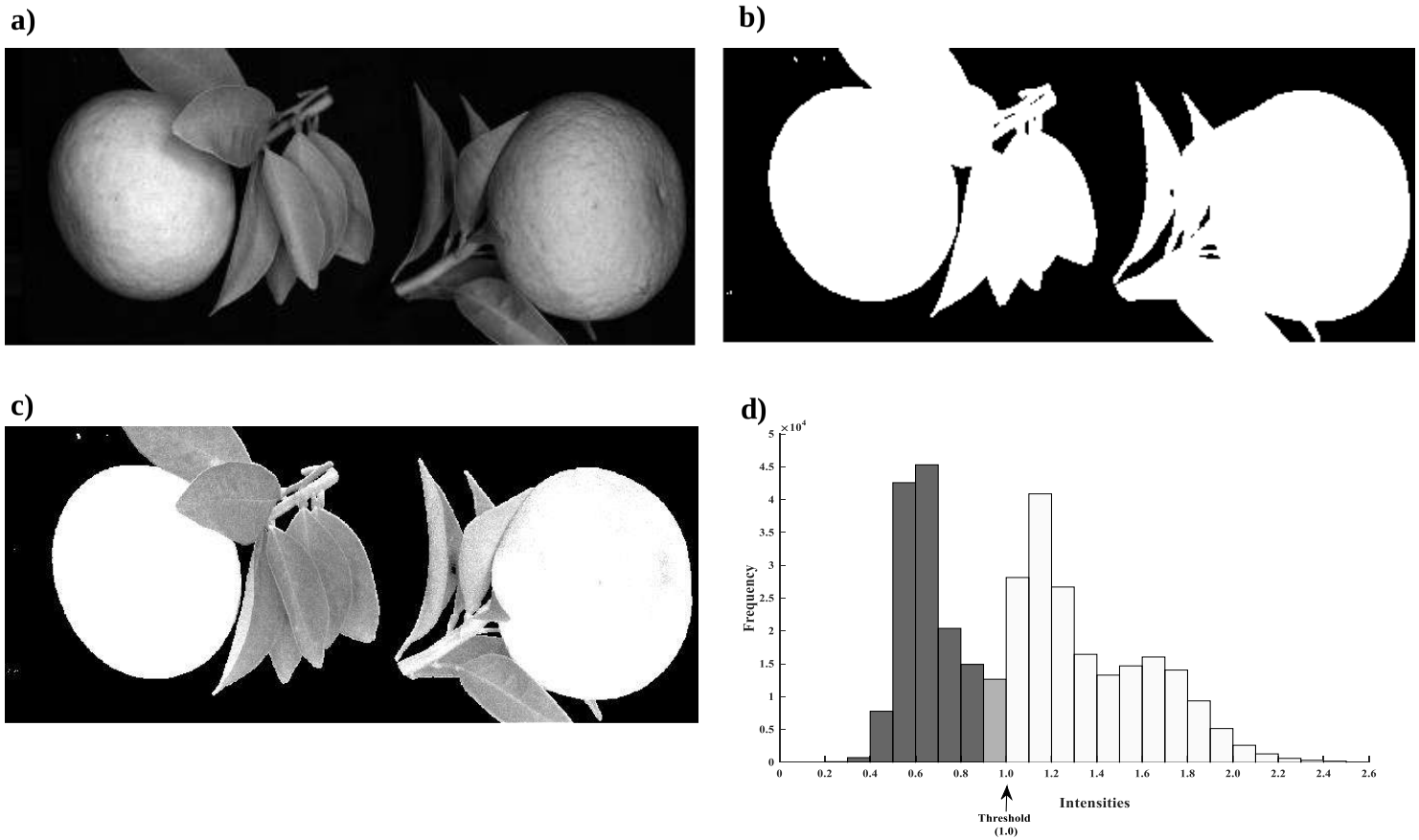


Figure 7.1.5. Band ratio image ($R_{\lambda 1165}$/ $R_{\lambda 1471}$) for the differentiation between green oranges and leaves using the SWIR system.

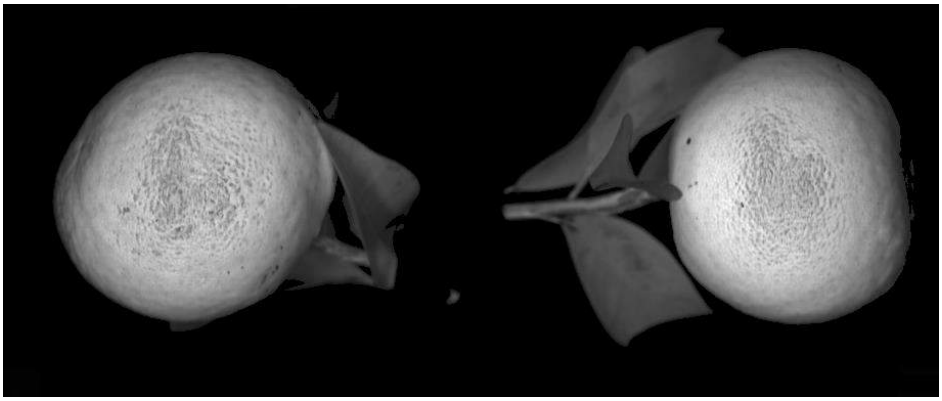
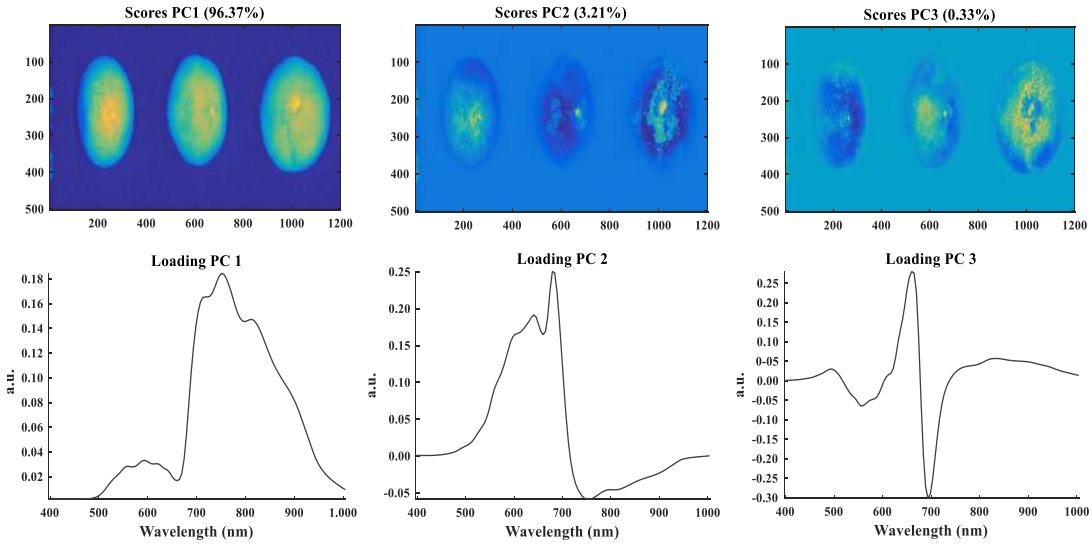
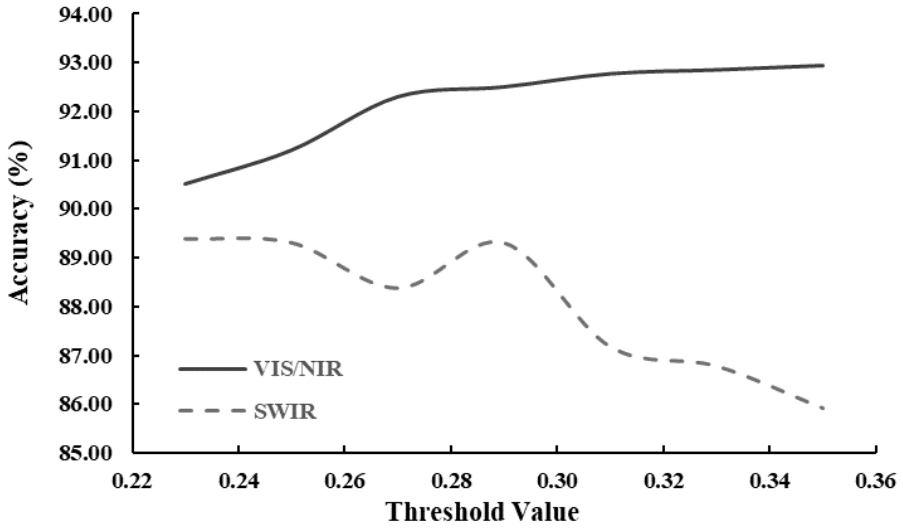


Figure 7.1.6. PCA score images and loadings plot for the first three



principal components using the Vis/NIR system.

Figure 7.1.7. Accuracy (%) for each threshold value applied using the Vis/NIR and SWIR based generated ratio images.



Chapter 8

Chapter 8. CONCLUSIONS

The main conclusions reached, based on the research carried out, the proposed strategies and the results obtained during the development of this PhD Dissertation, are:

a) Conclusions related to the feasibility of NIRS and HSI technologies for characterization, authentication and quality & safety assurance of horticultural products during the ripening process in the field and in the handling and sorting lines at industrial level.

1. Near infrared spectroscopy has demonstrated to be a promising tool for the high-speed on-site postharvest quality assessment in Raf tomatoes, allowing the determination of ripeness based on visual appearance and taste and enabling its incorporation into the grading systems. [*This conclusion was reached in the article: 'Fast and accurate quality assessment of Raf tomatoes using NIRS technology'. Postharvest Biology and Technology 107, 9–15 (2015)*].

2. The *in situ* quality determination in vegetables – summer squashes and spinach leaves – using handheld NIRS instruments provides the opportunity to take decisions regarding to the optimum harvest time and to stablish NIRS technology as a routine analysis to guarantee the quality of vegetables in the field. [*This conclusion was met in the research articles: 'Use of NIRS technology for on-vine measurement of nitrate content and other internal quality parameters in intact summer squash for baby food production'. Postharvest Biology and Technology 125, 122–128 (2017); 'Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors'. Postharvest Biology and Technology 154, 21–30 (2019); 'Pre-harvest screening on-vine of spinach quality and safety using*

NIRS technology'. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 207, 242–250 (2019)].

3. The results obtained showed the possibility to establish and classify vegetables, such as summer squashes and spinach leaves, by their final destination according to their nitrate content, measured both in the field or at the industrial level, confirming the feasibility of the NIRS technology for the safety assessment of horticultural products. [*This conclusion was reached in the research articles: 'Use of NIRS technology for on-vine measurement of nitrate content and other internal quality parameters in intact summer squash for baby food production'. Postharvest Biology and Technology 125, 122–128 (2017); 'Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors'. Postharvest Biology and Technology 154, 21–30 (2019); 'Pre-harvest screening on-vine of spinach quality and safety using NIRS technology'. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 207, 242–250 (2019)*].

4. The results showed that NIR spectroscopy can be successfully used for predicting the growing systems used in bell pepper production, which is of particular value to guarantee the authentication of outdoor-grown peppers. Additionally, the results showed that NIR spectroscopy can be used simultaneously as a rapid preliminary screening technique to measure quality. [*This conclusion was reached in the research article: 'Rapid, simultaneous, and in situ authentication and quality assessment of intact bell peppers using near-infrared spectroscopy technology'. Journal of the Science of Food and Agriculture 99, 1613–1622 (2019)*].

5. The models developed for the instantaneous and non-destructive prediction of relative water content in olive leaves measured on-tree using a portable NIRS sensor provide information related to the hydration status of the olive trees, enabling to make real time irrigation decisions in order to maintain the crop yield and to optimize the water resources management. [*This conclusion is met in the article: 'Irrigation decision support based on leaf relative water content determination in olive grove using near infrared spectroscopy'. Biosystems Engineering 180, 50–58 (2019)*].

6. NIRS technology by means of the application of non-linear regression techniques (LOCAL algorithm), enables the development of universal quality-prediction models for different species of the same gender, without being necessary to make specific models for each fruit and making easy the recalibration by adding other species. [*This conclusion is reached in the articles: 'Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors'. Biosystems Engineering 153, 140–148 (2017); 'LOCAL regression applied to a citrus multispecies library to assess chemical quality parameters using near infrared spectroscopy'. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 217, 206–214 (2019)*].

7. The hyperspectral imaging technique can be used to estimate the crop yield in citrus orchards prior to the harvest time, which takes place when fruits present the same green color as leaves. PCA and ANOVA were useful methods for the selection of the optimal wavelengths to develop the two-band ratio used for the distinction between green oranges and leaves and the detection of defects in oranges. [*This conclusion is met in the research article: 'Setting up a methodology to distinguish between green oranges and leaves using hyperspectral imaging'. Computers and Electronics in Agriculture (under review)*].

8. A sufficiently large and highly representative sample database is necessary to make the quantification of quality and safety parameters in horticultural products more robust. Over the coming years, further studies will be needed in order to improve the calibration specificity and accuracy of the models here obtained. [*This conclusion is reached in the research articles: 'Pre-harvest screening on-vine of spinach quality and safety using NIRS technology'. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 207, 242–250 (2019); 'Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors'. Postharvest Biology and Technology 154, 21–30 (2019); 'Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors'. Biosystems Engineering 153, 140–148 (2017); 'LOCAL regression applied to a citrus multispecies library to assess chemical quality parameters using near infrared spectroscopy'. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 217, 206–214 (2019)*].

b) Conclusions related with the NIRS instrumentation and the hyperspectral imaging technologies tested in this PhD dissertation and their practical applicability for *in situ* quality and safety determinations in the field and in the industrial processing of fruits and vegetables.

9. The portable manual NIR instruments here tested, could be promising tools for their use by the growers and industry at any time in the food supply chain (from the field to the dinner table) to characterize and authenticate fruits and vegetables and also for a selective harvest at the optimum time in order to obtain a product of the highest quality which is intended both for fresh consumption and for the processing industry. However, it is highly important the optimization of the analysis methodology applied with these instruments for each specific product analysed prior to their final incorporation for this purpose. [*This conclusion is met in the research articles: 'Developing universal models for the prediction of physical quality in citrus fruits analysed on-tree using portable NIRS sensors'. Biosystems Engineering 153, 140–148 (2017); 'LOCAL regression applied to a citrus multispecies library to assess chemical quality parameters using near infrared spectroscopy'. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 217, 206–214 (2019); 'Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors'. Postharvest Biology and Technology 154, 21–30 (2019)*].

10. For the MicroNIR™ 1700, the comparison between two different analysis modes (static *vs* dynamic) confirmed that taking of point spectra (static mode) was the most suitable way of analysis to measure both the quality and safety parameters in summer squash. [*This conclusion is met in the research article: 'Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors'. Postharvest Biology and Technology 154, 21–30 (2019)*].

11. The incorporation of the FT-NIR instrument Matrix-F, suitable for online analysis in the sorting lines at industrial level, requires the optimization of the analysis procedure, taking special attention to the distance between the measurement head and the conveyor belt due to the different fruit sizes tested. From this optimization process, it was also determined that the measurement of a single spectrum taken in the sorting lines would be enough to guarantee the quality and safety of the horticultural products, enabling the incorporation of the NIRS technology in the classification process in the industry. [*This objective was reached in the research article: 'Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors'. Postharvest Biology and Technology 154, 21–30 (2019)*].

12. Results obtained using two hyperspectral imaging systems indicated that initially Vis/NIR showed to be the most suitable system for identification of fruits and detection of defects in green oranges on-tree. However, it must be stressed that the use of the polarizer enabled to improve the signal, reducing the specular reflection in samples, while for the SWIR system this accessory is not currently easily available. In addition, it must be added that if, as well as estimating the crop yield, certain chemical quality parameters in oranges are also needed to be measured simultaneously, it would be of great interest to incorporate a band related to the absorption of water or glucides, which would require the use of the spectral range of the SWIR system. [*This conclusion is met in the research article: 'Setting up a methodology to distinguish between green oranges and leaves using hyperspectral imaging'. Computers and Electronics in Agriculture (under review)*]

Chapter 9

Chapter 9. FINAL CONSIDERATIONS AND RECOMMENDATIONS FOR FUTURE R&D&I WORKS

The results obtained in this PhD dissertation show the feasibility of using NIR sensors for quality and safety assessment in horticultural products along the supply chain. The studies carried out are highly relevant for the horticultural sector because they can be considered as simulation studies in real conditions, paving the way for the routine *in situ* incorporation of this technology.

Additionally, NIRS analysis methodology has been fine-tuned and optimized for the measurement of quality and safety parameters in horticultural products. While, it is true that in recent years there has been an increase in the number of NIRS applications in the horticultural sector, however, more efforts must be put in order to consider NIR spectroscopy as an official analytical method for the determination of quality and safety parameters in fruits and vegetables.

This chapter of this PhD dissertation is concerned with future research works that should be carried out for the optimization and consolidation of NIRS sensors for the characterization, authentication and also for the quality and safety assessment of horticultural products.

- ✓ Fine-tuning and validation of a NIRS analysis methodology to measure quality and safety parameters in horticultural products both *in situ*/on the plant and on-tree under non-control environmental conditions, and online, in the industry.

- ✓ Intensification in the sampling sets used in this PhD dissertation in order to increase the robustness of the models here developed for their routine use in the field and at industrial level.

- ✓ Due to the development of a new generation of portable, compact and extremely light-weight NIRS instruments, ideally suited for use in the field and for taking *in situ* measurements, it is of great importance to have ready a transfer protocol between databases obtained with different NIR instruments in previous research works, so that spectral databases can be expanding and updating easily for the new equipments.

- ✓ Multispecies spectral libraries for different fruits and vegetables genus should be created, so that universal models could be easily obtained for all the fruit and vegetable genus. This would facilitate the updating of the models and the management of the spectral libraries.

- ✓ It could be highly relevant for the success of NIRS technology the collaborative creation of spectral databases of great heterogeneity, which would incorporate all kind of fruits and vegetables and different growing seasons and cultural practices. Working in this way, reference data based on the NIR spectral fingerprint, could be obtained in any part of the world, without the development of a specific NIR model for this product on purpose.

- ✓ These days, the large-scale measurement of production and quality indicators in fruits and vegetables still constitutes a major difficulty for producers and industry. In this PhD dissertation, the feasibility of using HSI and MSI systems for this purpose have been demonstrated. However, this work can be can be considered as a preliminary step for designing low-cost spectral cameras capable of being mounted on unmanned aerial vehicles (UAVs) to estimate yield and quality in horticultural products. Thus, further studies are needed related to the spectral signal, the best processing options, the optimization of the analysis with different devices and accessories to increase the signal quality for each particular application, paving the way for the future development of low-cost and low-weight equipment for in field application of these systems.

Chapter 10

Chapter 10. REFERENCES

Bruhn, C.M. 2002. Consumer issues in quality and safety: Definition and evaluation for fresh horticultural crops. In: Kader, A.A. (Ed.), Postharvest Technology of Horticultural Crops, 3rd edition. Division of Agriculture and Natural Resources, University of California, Oakland, CA, pp. 31–37.

Dale, L.M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., Fernández-Pierna, J.A. 2013. Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: A review. Applied Spectroscopy Reviews. 48, 142–159.

Elia, A., Santamaria, P., Serio, F. 1998. Nitrogen nutrition, yield and quality of spinach. Journal of the Science of Food and Agriculture 76, 341–346.

Kader, A.A. 2002. Quality and safety factors: Definition and evaluation for fresh horticultural crops. In: Kader, A.A. (Ed.), Postharvest Technology of Horticultural Crops, 3rd edition. Division of Agriculture and Natural Resources, University of California, Oakland, CA, pp. 279–285.

Magwaza, L.S., Opara, U.L., Nieuwoudt, H., Cronje, P.J.R., Saeys, W., Nicolaï, B. 2012. NIR spectroscopy applications for internal and external quality analysis of citrus fruit – a review. Food and Bioprocess Technology 5, 425–444.

Meca, D., Céspedes, A. 2015. Varietal trends in protected horticulture. Vida Rural 406, 54–58 (in Spanish).

Nicolaï, 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 Biology and Technology 46, 99–118.

Obenland, D., Collin, S., Mackey, B., Sievert, J., Fjeld, K., Arpaia, M.L. 2009. Determinants of flavour acceptability during the maturation of navel oranges. Postharvest Biology and Technology 52, 156–163.

Official Journal of the European Union (OJEU). 2011. Commission Regulation (EC) No 1258/2011 of 2 December 2011 Amending Regulation (EC) No 1881/2006 As Regards Maximum Levels for Nitrates in Foodstuffs. OJ L 320/15-17, 3.12.2011.

Pasquini, C. 2018. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. Analytica Chimica Acta 1026, 8–36.

Sánchez, M.T., Pérez-Marín, D. 2011. Nondestructive measurement of fruit quality by NIR spectroscopy. In: Vázquez, M., Ramírez, J.A. (Eds.), Advances in Post-Harvest Treatments and Fruit Quality and Safety. Nova Science Publishers Inc., Hauppauge, NY, pp. 101–163.

Watkins, C.B. 2008. Postharvest ripening regulation and innovation in storage technology. Acta Horticulturae 796, 51–58.

Yan, H., Siesler, H.W. 2018. Hand-held near-infrared spectrometers: State-of-the-art instrumentation and practical applications. NIR News 29, 1–15.

Zhang, Y., Luo, L., Li, J., Li, S., Qu, W., Ma, H., Ye, X. 2017. *In situ* and real-time monitoring of enzymatic process of wheat gluten by miniature fiber NIR spectrometer. Food Research International 99, 147–154.

Appendixes

Appendix I. Participation in Congresses, Conferences and Symposia

I.1. VIII Congreso Español de Ingeniería de Alimentos (CYTA/CESIA 2015, Badajoz, Spain)

MONITORIZACIÓN DE LA CALIDAD Y DE LA MADUREZ DE NARANJAS ANALIZADAS EN ÁRBOL MEDIANTE TECNOLOGÍA NIRS

M.T. Sánchez[1*], M.J. De la Haba[1], I. Torres[1], D. Pérez-Marín[2]
[1]Departamento de Bromatología y Tecnología de los Alimentos. ETSIAM. Universidad de Córdoba. teresa.sanchez@uco.es
[2]Departamento de Producción Animal. ETSIAM. Universidad de Córdoba

INTRODUCCIÓN

Para el sector citrícola, sería de gran interés el poder evaluar la calidad de las naranjas durante el proceso de maduración en árbol y a su entrada en la industria, empleando una tecnología rápida, precisa, de bajo coste y no destructiva, que clasifique la naranja por su calidad.

La Espectroscopía de Reflectancia en el Infrarrojo Cercano (NIRS), hoy en día, es una de las alternativas idóneas como solución a las exigencias de control de calidad de la industria citrícola.

OBJETIVO

Establecer un sistema de apoyo al sector, basado en la aplicación de la huella espectral NIRS, para el control de calidad del fruto en campo, asegurando su recolección en el momento óptimo en función de su destino final.

MATERIAL y MÉTODOS

Material vegetal

Naranjas:

- 126 naranjas (*Citrus sinensis*. (L.) Osbeck var. 'Powell Summer Navel').

Análisis de referencia

- Peso del fruto (g).
- Diámetro ecuatorial (mm).
- Firmeza (N). Método de Magness-Taylor.
- Contenido en sólidos solubles totales (%). Refractometría.
- Acidez titulable (% ácido cítrico). Titración.

Análisis NIRS

- Espectrofotómetro MEMS:
  - Phazir 2400, Polychromix, Inc., Wilmington, MA, EE.UU. Reflectancia. 1600-2400 nm.
- Producto intacto.

Desarrollo de calibraciones

- Método de regresión: MPLS.
- Corrección de radiación difusa: SNV + DT.
- Derivadas: 1,5,5,1; 1,10,5,1; 2,5,5,1; 2,10,5,1.
- Software: WINISI II, versión 1.5.

RESULTADOS

Tabla 1. Estadísticos del colectivo de calibración: rango, media, desviación típica (DT) y coeficiente de variación (CV)

Parámetro	Rango	Media	DT	CV (%)
Peso (g)	170,00-598,30	306,24	67,34	21,99
Diámetro ecuatorial (mm)	69,24-108,34	83,02	6,98	8,41
Firmeza (N)	5,34-75,17	21,44	14,17	66,09
Contenido en sólidos solubles totales (%)	9,35-14,8	12,13	1,28	10,53
Acidez titulable (% ácido cítrico)	0,36-1,05	0,59	0,15	26,01

Tabla 2. Estadísticos de validación cruzada para la predicción de parámetros de calidad de naranjas intactas empleando la regresión MPLS

Parámetro	Tratamiento matemático	N	r^2	ETVC	RPD	RER	CV
Peso (g)	1,5,5,1	123	0,47	42,30	1,37	5,90	14,01
Diámetro ecuatorial (mm)	2,10,5,1	125	0,59	4,11	1,57	7,46	4,97
Firmeza (N)	2,10,5,1	113	0,42	10,03	1,30	5,46	48,43
Contenido en sólidos solubles totales (%)	2,10,5,1	126	0,62	0,79	1,61	6,94	6,47
Acidez titulable (% ácido cítrico)	1,5,5,1	124	0,09	0,14	1,05	4,06	23,78

CONCLUSIONES

Los resultados obtenidos confirman que con un mismo instrumento es posible determinar de forma simultánea un elevado número de parámetros de calidad tanto interna como externa en naranjas intactas en el propio árbol. Asimismo, el instrumento Phazir 2400 puede ser considerado como una alternativa de futuro para la determinación del momento óptimo de cosecha de los frutos.

Esta investigación ha sido financiada por la Junta de Andalucía (P. Excelencia No. 3713) y por la Unión Europa Proyecto IRRIQUAL (FP6-FOOD-CT-2006-023120).

INCORPORACIÓN DE SENSORES NIRS PARA LA CATEGORIZACIÓN ON-LINE DE MANDARINAS EN LA INDUSTRIA

M.J. De la Haba[1*], D. Pérez-Marín[2], I. Torres[1], M.T. Sánchez[1]
[1]Departamento de Bromatología y Tecnología de los Alimentos. ETSIAM. Universidad de Córdoba. bt1hacem@uco.es
[2]Departamento de Producción Animal. ETSIAM. Universidad de Córdoba

INTRODUCCIÓN

La industria citrícola está demandando la incorporación de sensores no destructivos, rápidos, precisos y de bajo coste, en las líneas de clasificación de producto en fresco, que permitan el establecimiento de la calidad del mismo y su categorización en función de parámetros de calidad externa e interna.

La Espectroscopía de Reflectancia en el Infrarrojo Cercano (NIRS) se presenta como una tecnología alternativa, frente a los métodos tradicionales de análisis, para hacer frente a dicho requerimiento.

OBJETIVO

El objetivo del presente trabajo es analizar la viabilidad de incorporar la tecnología NIRS en las líneas de manipulación de mandarina en fresco para realizar la clasificación del producto intacto de forma precisa y en segundos.

MATERIALES Y MÉTODOS

Material vegetal:

- 107 mandarinas (var. "Clemevilla").
- Las muestras fueron almacenadas en condiciones de refrigeración (5ºC y 90% H.R.) antes de realizar los análisis de laboratorio.

Análisis de referencia:

- Peso del fruto (g).
- Diámetro ecuatorial (mm).
- Firmeza (N). Método de Magness-Taylor.
- Contenido en sólidos solubles totales (%). Refractometría.
- Acidez titulable (% ácido cítrico). Titración.

Análisis NIRS:

Espectrofotómetro de red de diodos Perten DA-7000 (Perten Instruments North America Inc., Springfield, IL, EE.UU.).

- Rango espectral: 400-1700 nm.
- Modo de análisis: Reflectancia.

- Desarrollo de calibraciones:
  - Método de regresión: MPLS.
  - Corrección de *scatter*: SNV+DT y sin corregir (*none*).
  - Derivadas: 1,5,5,1; 2,5,5,1; 1,10,5,1; 2,10,5,1.
  - Software: WinISI II, versión 1.5.

RESULTADOS Y DISCUSIÓN

Tabla 1. Rango, media, desviación típica (DT) y coeficiente de variación (CV) del colectivo de calibración.

Parámetro	Rango	Media	DT	CV (%)
Peso (g)	71,09-217,22	136,24	34,02	24,97
Diámetro ecuatorial (mm)	42,38-68,10	55,08	5,61	10,19
Firmeza (N)	2,07-30,71	8,21	4,80	58,50
Contenido en sólidos solubles totales (%)	9,95-15,45	12,79	1,13	8,81
Acidez titulable (% ácido cítrico)	0,69-2,06	1,16	0,27	23,42

Tabla 2. Estadísticos de calibración para la predicción de parámetros de calidad externa e interna en mandarinas.

Parámetro	Media	DT	ETVC	r^2	CV	RPD
Peso (g)	136,24	34,29	16,84	0,76	12,36	2,04
Diámetro ecuatorial (mm)	66,50	5,72	1,49	0,93	2,24	3,85
Firmeza (N)	6,81	2,68	1,98	0,48	28,27	1,39
Contenido en sólidos solubles totales (%)	12,85	1,04	0,63	0,64	4,88	1,66
Acidez titulable (% ácido cítrico)	1,13	0,24	0,12	0,77	10,24	2,10

CONCLUSIÓN

Los resultados obtenidos muestran la viabilidad de incorporar sensores NIRS en las líneas de clasificación de cítricos para la determinación simultánea de parámetros de calidad en frutos intactos y su categorización. El instrumento de red de diodos testado ha resultado ser muy adecuado para tal fin.

Esta investigación ha sido financiada por la Junta de Andalucía (R. Excelencia No. 3713) y por la Unión Europa Proyecto IRRIQUAL (FP6-FOOD-CT-2006-023120).

I.2. 17th International Conference on Near Infrared Spectroscopy (Foz do Iguassu, Brazil)

NEAR INFRARED SPECTROSCOPY FOR DISCRIMINATION BETWEEN EXTRA VIRGIN OIL AND LAMPANTE OLIVE OIL

I. Torres[1], M.T. Sánchez[1], M.J. De la Haba[1], D. Pérez-Marín[2], J.E. Guerrero-Ginel[2], A. Garrido-Varo[2,*]

[1]Department of Food Science and Food Technology. University of Córdoba (Spain)
[2]Department of Animal Production. University of Córdoba (Spain). *pa1gavaa@uco.es

INTRODUCTION

The International Olive Oil Council (IOC) defines three categories of olive oils (Extra Virgin, Virgin and Lampante) [1]. Extra Virgin Olive Oil is the most expensive and healthy promoting of all the categories of olive oils while Lampante Olive Oil is intended for refining or for technical use, not being possible its direct use for human consumption.

Despite of the huge amount of efforts invested in research about methods for determining quality, the adulteration of Extra Virgin Olive Oil with other low quality oils remains a major international problem and of enormous media impact. An exhaustive analytical control must be carried out on a large volume of olive oils traded at world level to avoid recurrent fraud episodes. Due to the fact that Extra Virgin Olive Oil can not be categorized by a single parameter, NIRS technology could be one of the approaches best suited to increase significantly the number of samples inspected per year.

OBJECTIVE

To evaluate the potential of NIRS for the discrimination between Extra Virgin and Lampante Virgin Olive oils.

MATERIAL AND METHODS

Sampling

- A total of 299 Virgin Olive Oil samples:
  - ✓ 189 Extra Virgin Olive Oil (EVOO) samples.
  - ✓ 110 Lampante Virgin Olive Oil (LVOO) samples.
- Samples were received in dark green bottles of 250 ml capacity.

Reference analysis

- Organoleptic evaluation: Panel test [2].
- Chemical analysis: (free acidity, peroxide index, absorbency in ultra-violet (K_{232} and K_{270}) and fatty acid ethyl esters) [1, 3].

NIRS analysis

- FNS-6500 SY-II scanning monochromator (FOSS NIRSystems, Silver Spring, MD, USA).
  - ✓ Spectral range: 400-2500 nm.
  - ✓ Transflectance mode.

Discriminant models development

- PLS Discriminant Analysis (PLS-DA) [4].
  - ✓ Scatter correction: SNV + DT.
  - ✓ Derivatives: 1,5,5,1; 2,5,5,1; 1,10,5,1; 2,10,5,1.
  - ✓ Software: WinISI v. 1.50.

REFERENCES

[1] International Olive Council, 2015. COI/T.15/NC No 3/Rev. 9. June 2015. Available online: http://www.internationaloliveoil.org/estaticos/view/222-standards.
[2] Jiménez-Herrera, B., Carpio-Dueñas, A. 2008. Instituto de Investigación y Formación Agraria y Pesquera. Consejería de Agricultura y Pesca. Junta de Andalucía. Sevilla. España
[3] Official Journal of The European Union (ODUE). 2013. Commission Implementing Regulation (EU) No 1348/2013 of 16 December 2013 amending Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. Official Journal L 338, 17 of December 2013, pp. 31–67.
[4] Naes, T., Isaksson, T., Fearn, T., Davies, A. 2002. NIR Publications, Chichester, UK.

RESULTS AND DISCUSSION

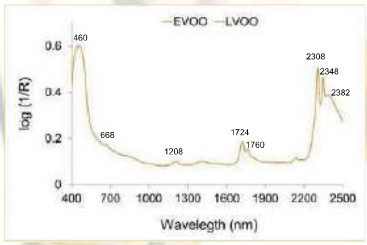


Figure 1. Mean NIR spectra of the Extra Virgin and Lampante Virgin Olive Oils.

Table 1. Classification results for the training set during the development of the best model "1,10,5,1".

Belonging to…	Classified as…		% Correctly classified training samples
	EVOO	LVOO	
EVOO (N=140)	130	10	92.86
LVOO (N=75)	15	60	80.00
% Correctly classified model: 88.37			

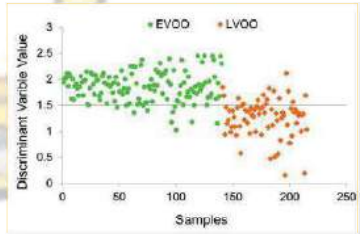


Figure 2. Discriminant Variable Values for the training set get used to discriminate between EVOO and LVOO.

CONCLUSIONS

The results obtained demonstrate that NIRS technology could be a first-line screening analytical method for inspection laboratories which later only will need to analyze for the most costly official methods the samples detected as "uncertain".

This research was funded by the Project 12013018. Thanks to the Olive Oil Interprofessional and the Agrarian Arbitrary Laboratory for providing samples

 DOI: 10.17648/NIR-2015-34359

NEAR INFRARED SPECTROSCOPY FOR DISCRIMINATION BETWEEN EXTRA VIRGIN OIL AND LAMPANTE OLIVE OIL

I. Torres[1], M.T. Sánchez[1], M.J. De la Haba[1], D. Pérez-Marín[2]
J.E. Guerrero-Ginel[2], A. Garrido- Varo[2]*

[1] Faculty of Agriculture and Forestry Engineering. Department of Bromatology and Food Technology. University of Cordoba. Campus Rabanales. Ctra. Nacional IV – Km 396. 14071 Cordoba. Spain.
[2] Faculty of Agriculture and Forestry Engineering. Department of Animal Production. University of Cordoba. Campus Rabanales. Ctra. Nacional IV – Km 396. 14071 Cordoba. Spain.

Corresponding author: pa1gavaa@uco.es

ABSTRACT

NIR spectroscopy has been used as a non destructive analytical technique to classify the different classes of Olive Oils attached to the official method named "Panel Test". In particular, the present work is aimed to evaluate the potential of the NIR spectroscopy for the discrimination between extra virgin and lampante olive oils. A total of 299 olive oil samples, all of them provided by official inspectors, were used. 189 samples were authenticated by inspectors as Extra Virgin Olive Oil, and 110 samples as Lampante Olive Oil. For collecting NIR spectra, the FNS- 6500 SY-II scanning monochromator (spectral range 400-2500 nm) was used. Samples were scanned using a folded-transmission gold reflector cup, diameter 3.75 cm, with a pathlength of 0.1 mm, in the transflectance mode. Discriminant models were constructed to classify olive oils by category, using PLS discriminant analysis (PLS-DA). The best models developed correctly classified 92.86% of the samples as EVOO and 80.00% as LOO. The results obtained demonstrate that NIRS technology could be a first-line screening analytical method for inspection laboratories which latter on will only need to analyze for the most costly and destructive official methods, the samples detected as "uncertain".

KEYWORDS: NIR Spectroscopy; Olive Oil; Categorization; Authentication; Adulteration.

INTRODUCTION

The International Olive Council (IOC) has defined standards that can be applied to all the olive oils and olive-pomace oils that are the object of international trade. The IOC recognizes three categories of Virgin olive oils (Extra Virgin, Virgin, and Lampante)[1]. Extra Virgin and Virgin Olive Oils (EVOO and VOO) are the most expensive and health promoting of all the categories of olive oils, while Lampante Olive Oil (LOO) is intended for refining or for technical use, not being possible its direct use for human consumption. Despite of the huge amount of efforts invested in research about methods for determining quality, purity and authenticity of Virgin Olive oils and that international standards are used in official inspections by the different producers and importer countries, the adulteration of Extra Virgin Olive oil with other low grade oils remains a major international problem and of enormous media impact. The main reason for recurrent fraud episodes is that inspection bodies are lacking of a suitable and affordable methodology that can help to increase significantly the number of samples inspected per year, providing more evidences than now and exhaustive analytical control is carried out on a large volume of Olive oils traded at world level. In that sense, NIRS technology could be one of the approaches best suited to these requirements.

MATERIALS AND METHODS

Samples

A total of 299 samples of olive oil from different categories, Extra Virgin Olive Oil (EVOO) (N=189) and Lampante Olive Oil (LOO) (N=110) classified according to the Regulation (EU) No 1348/2013[2], were used in this work. Samples were received in dark green bottles of 250 ml capacity and were storage at 5ºC and 85% relative humidity. Prior to the NIR analysis, olive oil samples were left at the laboratory temperature in order to allow the sample temperature to stabilize.

DOI: 10.17648/NIR-2015-34359

Spectra Collection

Spectra were collected in transflectance mode (log (1/R)) using a Foss NIRSystems 6500 SY-II (FNS-II) scanning monochromator (FOSS NIRSystems, Silver Spring, MD, USA). A folded-transmission gold reflector cup, diameter 3.75 cm, with a pathlength of 0.1 mm was used. This instrument operates in the 400-2500 nm range, with a 2nm scanning interval. Two spectral measurements were made on each sample and these were averaged to provide a mean spectrum. NIR spectra data were collected using the WinISI II software package version 1.50 (Infrasoft International, Port Matilda, PA, USA).

NIRS classification models

Discriminant models were constructed to classify the olive oil samples by category, using PLS discriminant analysis (PLS-DA) for supervised classification[3]. Specifically, the PLS2 algorithm was applied, using the "Discriminant Equations" option in the WinISI II software.

Prior to developing the PLS2-DA models, the CENTER algorithm was applied to calculate the centre of the population and the distance of samples (spectra) from that centre in a *n*-dimensional space, using the Mahalanobis distance (GH); samples with a GH greater than 3 were considered outliers or anomalous spectra and were eliminated.

The discriminant model was obtained using the Standard Normal Variate (SNV) and Detrending (DT) for scatter correction. Furthermore, four derivative mathematical treatments were tested in the development of NIR models (1,5,5,1; 2,5,5,1; 1,10,5,1; 2,10,5,1).

The precision of the model obtained was evaluated using the percentage of correctly-classified samples.

RESULTS AND DISCUSSION

Spectral features

Typical log (1/R) spectra for both categories, Extra Virgin Olive Oil and Lampante Olive Oil, together with the most relevant absorption bands are shown in Figure 1. It can be observed that the shape of the spectra in both categories is very similar.

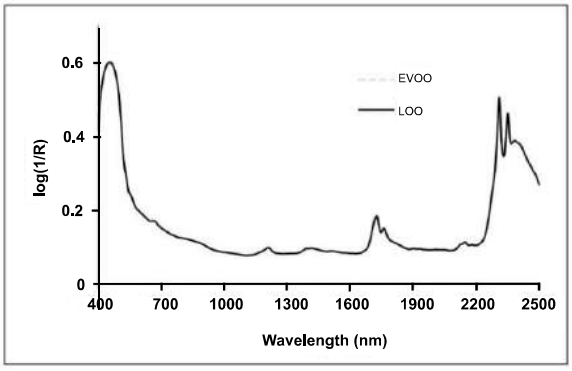


Figure 1. Typical log (1/R) spectra for Extra Virgin and Lampante Olive Oils.

In the visible region of the spectrum, absorbance spectra measured on both categories show peaks occurring at positions (420-460 nm and 668 nm) respectively, these are indicative of the presence of red pigments (carotenoids and anthocyanins) and of abundant chlorophyll. Strong absorption by chlorophyll a was evident at 680 nm, with a shoulder at 630 nm due to absorption by chlorophyll *b*[4]. In addition, red pigments (carotenoids and anthocyanins) have a typical absorption band in the 490-550 nm region of the visible spectrum.[5]

DOI: 10.17648/NIR-2015-34359

In the NIR region, the main absorption peaks are observed at 1208, 1400, 1724, 1760, 2144, 2308, 2348 and 2382 nm. The bands at 1208 nm correspond to a weak second overtone due to C-H stretching vibration, whereas at 1400, 1724 and 1760 nm are assigned to C-H stretch first overtones of the methyl, methylene and ethylene groups. Moreover, bands at 2144 are related to the absorption by N-H bands while at 2308, 2348 y 2382 nm are combinations involving C-H stretching with other vibrational modes.[6-9]

Discrimination between EVOO and LOO

The training set of 299 samples of olive oil available for this study was built with 140 and 75 EVOO and LOO samples, respectively.

The most accurate discriminant model (Table 1) was obtained using the first derivative (1,5,5,1) of log (1/R) with scatter correction (SNV and DT).

Table 1. Classification results for the training set during the development of Model "1,5,5,1".

Belonging to...	Classified as...		% Correctly classified training samples
	EVOO	LOO	
EVOO (N=140)	130	10	92.86
LOO (N=75)	15	60	80.00
% Correctly classified model: 88.37			

As shown in Table 1, 130 out of the 140 samples (92.86%) in the EVOO category and 60 out of the 75 samples (80%) in the LOO category were correctly classified.

In the model developed using PLS2, a discrimination value of between 1 and 2 was assigned to each sample for each class (EVOO vs LOO); thus, the sum of the two call values should equal 3. A value of 2 perfectly reflects membership of a given class; so the value for non-membership of the other class must be 1; a value of 1.5 for both classes means that assignment could go either way.[10] Several authors[11,12] recommend taking 1.5 as the cut-off value for classification using the discriminant strategy, i.e. a value above 1.5 signifies membership of the class in question, while a value of below 1.5 signifies non-membership of that class. Figure 2 shows discriminant variable values for the training set samples assigning a value of 2.0 to the EVOO.

As Figure 2 shows, the ten misclassified EVOO samples displayed values below 1.5, indicating discriminant variable values of close to 1.0 for the LOO and thus classified as Lampante Olive Oil. The fifteen misclassified LOO samples displayed values over the cut-off limit of 1.5 and were thus classified as EVOO.

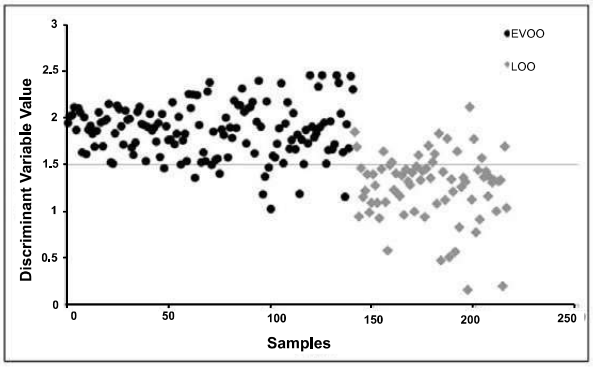


Figure 2. Discriminant Variable Values for the training set get used to discriminate between EVOO and LOO.

DOI: 10.17648/NIR-2015-34359

CONCLUSION

The results obtained demonstrate that NIRS technology provides a non-destructive means of screening for inspection laboratories allowing a larger number of samples to be tested. Only "uncertain" samples may then be analysed using more sophisticated – though slower and more costly – official methods.

Acknowledgments

This research was funded by the Project 12013018. Thanks to the Olive Oil Interprofessional and the Agrarian Arbitrary Laboratory for providing samples and Ms. Ma C. Fernandez for the technical assistance.

References

1. International Olive Council. 2015. COI/T.15/NC No 3/Rev. 9. June 2015. Available online: http://www.internationaloliveoil.org/estaticos/view/222-standards.
2. Official Journal of The European Union (ODUE). 2013. Commission Implementing Regulation (EU) No 1348/2013 of 16 December 2013 amending Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. Official Journal L 338, 17 of December 2013, pp. 31–67.
3. Naes, T., Isaksson, T., Fearn, T., Davies, A. 2002. NIR Publications, Chichester, UK.
4. McGlone, V.A., Martinsen, P.J., Clark, C.J., Jordan, R.B. 2005. Postharvest Biology and Technology 37, 142-151. 5. Strayer, L. 1995. Biochemistry. 4th ed. W.H. Freeman and Company/Worth Publishers. New York, USA.
6. Murray I. 1987. Int. NIR/NIT Conference. Akadémiai Diadó, Budapest, Hungary.
7. Downey, G., McIntyre, P., Davies A.N. 2002. Journal of Agricultural and Food Chemistry 50, 5520-5525.
8. Hourant, P., Baeten, V., Morales, M.T., Meurens, M., Aparicio, R. 2000. Applied Spectroscopy 54, 1168-1174.
9. Garrido-Varo, A., García-Olmo, J., Pérez-Marín, D.C. 2004. Near-Infrared Spectroscopy in Agriculture, ASA, CSSA, SSSA. Madison, US.
10. Murray, I., Aucott, L.S., Pike, H.I. 2001. Journal of Near Infrared Spectroscopy, 9, 297-311.
11. McElhinney, J., Downey G., Fearn, T. 2000. Proceedings of the 9th International Conference on Near Infrared Spectroscopy. NIR Publications, Chichester, UK.
12. Barnes, R.J., Dhanoa, M.S., Lister, S.J. 1989. Applied Spectroscopy 43, 772-777.

I.3. VIII International Postharvest Symposium (Cartagena, Murcia, Spain)

In-situ determination of external quality parameters in intact summer squash using near-infrared reflectance spectroscopy

M.T. Sánchez[1,a], I. Torres[1], B. Gil[1], D. Pérez-Marín[2], A. Garrido-Varo[2] and M.J. De la Haba[1]

[1]Department of Food Science and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain; [2]Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.

Abstract

Near-infrared reflectance (NIR) spectroscopy was used for the non-destructive measurement of external quality parameters (weight, length, equatorial diameter and colour (a*)) in intact summer squash using a handheld MEMS-based NIR spectrophotometer (spectral range 1600-2400 nm). A number of 156 summer squashes were analysed by traditional physical methods, and used to build calibration models by applying different spectral signal pretreatments and MPLS regression. The models obtained yielded promising results for in-situ external quality measurements, particularly for morphological-related parameters (r^2>0.8). These results confirm that changes in intact summer squashes quality parameters during on-vine ripening can be measured non-invasively using the new generation of portable handheld MEMS-based NIRS sensors, allowing the ripening process to be charted in individual fruits, not only in terms of visual appearance, but also in terms of morphological-related features, this being a major step towards the selective harvesting of summer squashes depending on their final destination.

Keywords: summer squash, NIR spectroscopy, morphological parameters, colour

INTRODUCTION

Appearance is one of the key characteristics governing vegetable acceptability and consumption (Abbott, 1999). The relative importance of this quality parameter depends on the consumer demands and preferences, and is influenced by a variety of factors ranging from agronomic to personal considerations (Drewnowski, 1997).

In summer squash, external appearance is one of the sensory properties most influencing consumer acceptance or rejection. Attributes such as size, shape uniformity, and green surface colour are considered the main external quality factors in this product (Kader, 2002).

Appearance analysis is thus a key element in summer squash quality control. However, the development of a non-invasive instrumental method for appearance determination in intact vegetables, permitting speedy, accurate and reproducible results, has not proved an easy task.

Near infrared (NIR) spectroscopy has shown considerable potential for the non-destructive measurement of external attributes in fruits and vegetables (Nicolaï et al., 2007; Sánchez and Pérez-Marín, 2011). In recent years, moreover, the development of handheld near-infrared devices has enhanced the potential of NIR spectroscopy for the in situ monitoring and analysis of the fruit ripening process (Pérez-Marín et al., 2009; Sánchez et al., 2011, 2013a, b).

Although a few number of papers address the application of NIRS to quality control in summer squash (e.g., Barnaba et al., 2012; Blanco-Díaz et al., 2014; Martínez-Valdivieso et al., 2014a, b), no studies have focused specifically on the use of this technology for appearance determination with whole intact summer squashes, perhaps due to the complexities involved.

[a]E-mail: bt1sapim@uco.es

Acta Hortic. 1194. ISHS 2018. DOI 10.17660/ActaHortic.2018.1194.178
Proc. VIII International Postharvest Symposium:
Enhancing Supply Chain and Consumer Benefits – Ethical and Technological Issues
Eds.: F. Artés-Hernández et al.

The aim of this work was to investigate the feasibility of NIRS technology for predicting major external quality parameters of intact summer squash using a low cost miniaturised, handheld, near-infrared device based on MEMS technology.

MATERIALS AND METHODS

Sampling

156 summer squashes harvested at commercial stage in the Province of Cordoba (Spain) were evaluated. Harvested summer squashes were kept in refrigerated storage at 5°C and 90% RH until the following day, when laboratory testing was performed. Prior to each test, summer squashes were allowed to reach room temperature. All tests were performed at 20°C.

NIR spectrum acquisition

Spectra were collected on summer squashes in reflectance mode (Log 1/R) using a handheld MEMS spectrometer (Phazir 2400, Polychromix, Inc., Wilmington, MA, USA). The instrument scans at non-constant intervals of approximately 8 nm (pixel resolution 8 nm, optical resolution 12 nm), across the range of NIR wavelengths 1600-2400 nm.

Using this instrument, four spectral measurements were made on each summer squash. The first was made on the equatorial part of the fruit, and the fruit was then turned through 90° for each successive measurement. The four spectra were averaged to provide a mean spectrum for each fruit.

Determination of external quality parameters

Summer squashes were individually weighed on an electronic balance (0-1000±0.01 g; model P1000 N, Metter-Toledo, GmbH, Greifensee, Switzerland). Length was measured using a measuring tape and equatorial diameter was then measured using a digital precision calibrator (0-300±0.01 mm; Comecta, Barcelona, Spain). Fruit colour was evaluated with a colorimeter (Minolta, Model CR-400, Osaka, Japan) by measuring the parameter a* in the equatorial zone of the fruit (CIE, 2004).

Quantitative calibrations: sets and data processing

The WinISI II version 1.50 software package was used for the chemometric treatment of data (ISI, 2000).

Prior to the development of NIRS calibrations, the structure and spectral variability of the sample population were determined following the method recommended by Shenk and Westerhaus (1991), using the CENTER algorithm. Then, having ordered the sample set by spectral distances (from smallest to greatest distance to the centre) the outlier spectra samples (GH>3) were eliminated (N=6); the approximately 25% of samples forming the validation set were selected, by taking one of every four samples in the initial set.

Prediction equations were obtained using the MPLS regression method with cross-validation. Different mathematical treatments were evaluated for scatter correction, including SNV and DT methods (Barnes et al., 1989). Moreover, four derivate mathematical treatments were tested in the development of NIRS calibrations: 1,5,5,1; 2,5,5,1; 1,10,5,1 and 2,10,5,1 where the first digit is the number of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing, and the fourth the second smoothing (Shenk and Westerhaus, 1995).

The statistics used to select the best equations were: standard error of calibration (SEC), coefficient of determination of calibration (R^2), standard error of cross-validation (SECV), coefficient of determination for cross-validation (r^2), RPD or ratio of the standard deviation of the original data (SD) to SECV, and the CV (coefficient of variation) or ratio between the SECV and the mean value of the reference data of the studied parameter for the training set. This statistic enables SECV to be standardized, facilitating the comparison of the results obtained with sets of different means (Williams, 2001).

The best-fitting equations, as selected by statistical criteria, were subsequently

evaluated by external validation, a procedure determining the predictive ability of an equation based on a sample set, which has not been used in the calibration procedures, following the protocol outlined by Windham et al. (1989).

RESULTS AND DISCUSSION

Descriptive data for NIR calibrations and validations sets

Values for range, mean, standard deviation (SD) and coefficient of variation (CV) for each of the parameters analysed using the calibration and validation sets after application of the CENTER algorithm and removal of spectral outliers are shown in Table 1, together with the number of samples for each parameter that was measured for the product tested.

Table 1. Statistical analysis of calibration and validation sets: number of samples (*N*), data range, mean, standard deviation (SD), and coefficient of variation (CV).

Parameter	Set	N	Range	Mean	SD	CV (%)
Weight (g)	Calibration	116	156.00-1247.20	549.89	310.36	56.44
	Validation	34	169.30-1209.17	549.83	266.08	48.39
Length (mm)	Calibration	116	16.00-38.00	26.45	5.65	21.36
	Validation	34	18.00-38.00	26.70	4.90	18.35
Equatorial diameter (mm)	Calibration	116	38.00-81.00	55.52	12.39	22.32
	Validation	34	39.00-76.00	55.81	11.29	20.23
a*	Calibration	116	-13.44-(-1.70)	-6.15	2.52	40.97
	Validation	34	-9.14-(-1.97)	-5.06	1.91	37.75

It should be stressed that structured selection using only spectral information treatment algorithms such as CENTER proved to be adequate and useful, since the calibration and validation sets displayed similar values for mean, range and standard deviation for all studied parameters, and that ranges for the validation set lay within the range recorded for the calibration set. All studied parameters covered a wide range of values as evidenced by the CV values, mainly weight (CV=56.44%) and a* (CV=40.97%).

Prediction of external quality parameters using MPLS regression and NIR spectra

The best calibration models for external quality parameter (weight, length, equatorial diameter, and colour (a*)), using the combination of signal pretreatments yielding the best results, are shown in Table 2.

Table 2. MPLS regression statistics for NIR-based models for predicting external quality parameters in intact summer squashes.

Parameter	Math treatment	N	Range	Mean	SD	SECV	r^2	CV (%)	RPD
Weight (g)	2,10,5,1	105	156.00-1170.54	547.05	301.03	104.68	0.88	19.14	2.88
Length (mm)	2,5,5,1	112	17.00-38.00	26.64	5.59	2.48	0.81	9.30	2.26
Equatorial diameter (mm)	2,10,5,1	112	38.00-81.00	55.53	12.53	5.19	0.83	9.34	2.42
a*	2,5,5,1	106	-12.01-(-1.70)	-5.79	2.24	1.03	0.79	17.77	2.18

The calibration model displaying the greatest predictive capacity for the weight parameter was obtained using D2 log (1/R) (r^2=0.88; SECV=104.68 g), which enabled good quantification of intact summer squash weight values (Shenk and Westerhaus, 1996).

Whilst no research published to date has examined the correlation between spectroscopic data and external quality parameters in intact summer squashes, a number of studies have addressed the use of NIR spectroscopy for predicting external parameters such as weight. Pérez-Marín et al. (2009) tested the same MEMS-based device used here (Phazir-

2400) for predicting weight in intact nectarines (RPD=1.40; CV=17.30%) while Sánchez et al. (2013a, b) used the same instrument for predicting weight in mandarins (RPD=1.43; CV=18.52%) and in intact oranges (RPD=1.37; CV=14.01%). The results obtained by these authors were slightly poorer than the results obtained in this research, although important differences in the characteristics of the peel between citrus fruits and summer quashes should be taken into account.

As Table 2 shows, good predictive ability (r^2=0.81 and 0.83; SECV=2.48 and 5.19 mm) was recorded for the measurement of length and equatorial diameter (Shenk and Westerhaus, 1996). For this MEMS instrument, the major absorption peak at around 1920-1930 nm was mainly related to water absorption, as was the peak at around 2200 nm; this was to be expected, since summer squash is around 90% water (Osborne et al., 1993). This was achieved using a rapid, non-destructive handheld sensor in intact summer squashes, thus providing the growers and the industry with an instant response about summer squash harvesting. No references have been found in the literature to the measurement of these parameters in intact squashes using NIRS technology. However, Pérez-Marín et al. (2009) and Sánchez et al. (2011), using the same MEMS-based device used here, reported poorer results for the prediction of equatorial diameter (RPD=1.31 and 1.45) in intact nectarines analysed on-tree and during postharvest storage. For citrus fruits, Sánchez et al. (2013a, b) obtained RPD values for the equatorial diameter ranging from 1.46 (intact mandarin) to 1.57 (intact orange) while in the case of the axial diameter the values of the RPD statistic were comprised between 1.42 (intact orange) and 1.59 (intact mandarin), inferior to the ones here obtained.

Models constructed for a* parameter displayed good predictive capacity within the limits established by Shenk and Westerhaus (1996). It should also be stressed that the MEMS instrument enables this colour parameter to be measured on-site, which is particularly useful for the squash handling industry.

Although no references have been found in the literature related to the prediction of this colour parameter in intact summer squashes, the RPD value recorded here was higher than this of 1.36 reported by Torres et al., (2015) for a* prediction in Raf tomatoes at varying degrees of ripeness with the same handheld spectrophotometer used here.

Validation statistics for the prediction of these parameters in intact summer squashes are shown in Table 3. Examination of the validation statistics shown in Table 3 and the calibration statistics (Table 2) revealed similar SECV and SEP values for NIRS prediction of the external parameters analysed, suggesting that application of these models was feasible.

Table 3. Validation statistics for predicting external quality parameters in intact summer squash using the MPLS regression method.

Parameter	Validation statistics				Control limits	
	r^2	SEP	SEP (c)	Bias	SEP (c)	Bias
Weight (g)	0.65	158.78	157.10[1]	-35.47	124.18	±57.31
Length (mm)	0.61	3.09	3.12[1]	-0.30	2.70	±1.25
Equatorial diameter (mm)	0.64	6.77	6.75[1]	-1.28	6.25	±2.89
a*	0.73	1.07	1.07	0.19	1.13	0.52

[1]Values exceeding control limits recommended by Windham et al. (1989).

In terms of the validation protocol recommended by Windham et al. (1989) for the routine implementation of NIRS prediction models for all the parameters tested, r^2 results always attained recommended minimum value (r^2>0.6) and the bias were below the confidence limits; however the SEP(c) were above the control limits established. These results indicate that the NIRS models constructed here can be considered as a first step in the fine-tuning of NIRS technology for the monitoring on-vine of external quality parameters in summer squashes.

CONCLUSIONS

The results obtained show that near infrared reflectance spectroscopy combined with multivariate analysis is a very useful and promising tool for determining external appearance in intact summer squashes and deciding on the optimum time for harvesting. It must be highlighted that the results obtained here with the analysis of summer squashes in intact form, requiring no previous sample preparation, should be considered a first step in the tuning of NIRS technology for on-site control purposes in the field using the latest-generation, hand-held compact MEMS instruments.

ACKNOWLEDGEMENTS

This research was funded by Gelagri Ibérica S.L. under the Research Project 'Characterisation of summer squashes cultivated in Santaella (Cordoba)'.

Literature cited

Abbott, J.A. (1999). Quality measurement of fruits and vegetables. Postharvest Biol. Technol. *15* (*3*), 207–225 https://doi.org/10.1016/S0925-5214(98)00086-6.

Barnaba, F.E., Bellicontro, A., and Mencarelli, F. (2012). NIR-AOTF spectroscopy: innovative technique to predict dry matter and the harvest time of zucchini. Ind. Aliment. *51*, 30–35.

Barnes, R.J., Dhanoa, M.S., and Lister, S.J. (1989). Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra. Appl. Spectrosc. *43* (*5*), 772–777 https://doi.org/10.1366/0003702894202201.

Blanco-Díaz, M.T., Del Río-Celestino, M., Martínez-Valdivieso, D., and Font, R. (2014). Use of visible and near-infrared spectroscopy for predicting antioxidant compounds in summer squash (*Cucurbita pepo* ssp *pepo*). Food Chem *164*, 301–308 https://doi.org/10.1016/j.foodchem.2014.05.019. PubMed

CIE. (2004). Colorimetry, 3rd edn (Vienna, Austria: Commission Internationale de L'eclairage).

Drewnowski, A. (1997). Taste preferences and food intake. Annu. Rev. Nutr. *17* (*1*), 237–253 https://doi.org/10.1146/annurev.nutr.17.1.237. PubMed

ISI. (2000). The Complete Software Solution Using a Single Screen for Routine Analysis, Robust Calibrations, and Networking. Manual, FOSS NIRSystems/TECATOR (Silver Spring, MD, USA: Infrasoft International, LLC.).

Kader, A.A. (2002) Quality and safety factors: Definition and evaluation for fresh horticultural crops. In Postharvest Technology of Horticultural Crops, 3rd edn, A.A. Kader, ed. (Oakland, California: University of California, Division of Agriculture and Natural Resources), p.279–285.

Martínez-Valdivieso, D., Font, R., Gómez, P., Blanco-Díaz, T., and Del Río-Celestino, M. (2014a). Determining the mineral composition in Cucurbita pepo fruit using near infrared reflectance spectroscopy. J. Sci. Food Agric. *94* (*15*), 3171–3180 https://doi.org/10.1002/jsfa.6667. PubMed

Martínez-Valdivieso, D., Font, R., Blanco-Díaz, M.T., Moreno-Rojas, J.M., Gómez, P., Alonso-Moraga, A., and Del Río-Celestino, M. (2014b). Application of near-infrared reflectance spectroscopy for predicting carotenoid content in summer squash fruit. Comput. Electron. Agric. *108*, 71–79 https://doi.org/10.1016/j.compag.2014.07.003.

Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., and Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. Postharvest Biol. Technol. *46* (*2*), 99–118 https://doi.org/10.1016/j.postharvbio.2007.06.024.

Osborne, B.G., Fearn, T., and Hindle, P.H. (1993). Practical NIR spectroscopy with applications. In Food and Beverage Analysis, B.G. Osborne, T. Fearn, and P.H. Hindle, eds. (Essex, UK: Longman), p.11–35.

Pérez-Marín, D., Sánchez, M.T., Paz, P., Soriano, M.A., Guerrero, J.E., and Garrido-Varo, A. (2009). Non-destructive determination of quality parameters in nectarines during on-tree ripening and postharvest storage. Postharvest Biol. Technol. *52* (*2*), 180–188 https://doi.org/10.1016/j.postharvbio.2008.10.005.

Sánchez, M.T., and Pérez-Marín, D. (2011). Nondestructive measurement of fruit quality by NIR spectroscopy. In Advances in Post-Harvest Treatments and Fruit Quality and Safety, M. Vázquez, and J.A. Ramírez, eds. (Hauppauge, New York, USA: Nova Science Publishers, Inc.), p.101–163.

Sánchez, M.T., De la Haba, M.J., Guerrero, J.E., Garrido-Varo, A., and Pérez-Marín, D. (2011). Testing of a LOCAL approach for the prediction of quality parameters in intact nectarines using a portable NIRS instrument. Postharvest Biol. Technol. *60* (*2*), 130–135 https://doi.org/10.1016/j.postharvbio.2010.12.006.

Sánchez, M.T., De La Haba, M.J., and Pérez-Marín, D. (2013a). Internal and external quality assessment of mandarins on-tree and at harvest using a portable NIR spectrophotometer. Comput. Electron. Agric. *92*, 66–74

https://doi.org/10.1016/j.compag.2013.01.004.

Sánchez, M.T., De La Haba, M.J., Serrano, I., and Pérez-Marín, D. (2013b). Application of NIRS for non-destructive measurement of quality parameters in intact oranges during on-tree ripening and at harvest. Food Anal. Methods *6* (*3*), 826–837 https://doi.org/10.1007/s12161-012-9490-7.

Shenk, J.S., and Westerhaus, M.O. (1991). Population structuring of near infrared spectra and modified partial least squares regression. Crop Sci. *31* (*6*), 1548–1555 https://doi.org/10.2135/cropsci1991.0011183X003100060034x.

Shenk, J.S., and Westerhaus, M.O. (1995). Routine Operation, Calibration, Development and Network System Management Manual (Silver Spring, MD, USA: NIRSystems, Inc.).

Shenk, J.S., and Westerhaus, M.O. (1996). Calibration the ISI way. In Near Infrared Spectroscopy: the Future Waves, A.M.C. Davies, and P.C. Williams, eds. (Chichester, UK: NIR Publications), p.198–202.

Torres, I., Pérez-Marín, D., De La Haba, M.J., and Sánchez, M.T. (2015). Fast and accurate quality assessment of Raf tomatoes using NIRS technology. Postharvest Biol. Technol. *107*, 9–15 https://doi.org/10.1016/j.postharvbio.2015.04.004.

Williams, P.C. 2001. Implementation of near-infrared technology. In Near-Infrared Technology in the Agricultural and Food Industries, P.C. Williams, and K.H. Norris, eds. (St. Paul, Minnesota: AACC, Inc.), p.145–169.

Windham, W.R., Mertens, D.R., and Barton, F.E., II. (1989). Protocol for NIRS calibration: sample selection and equation development and validation. In Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality. Agriculture Handbook, Vol. 643, G.C. Martens, J. Shenk, and F.E. Barton, eds. (Washington, D.C., USA: Government Printing Office).

VIII International Postharvest Symposium
21-24 June 2016. Cartagena. Spain
http://www.postharvest2016.org
ISHS
Universidad Politécnica de Cartagena
Reference: VII.II.S3.2.P1

In-situ prediction of internal quality parameters in intact bell peppers using NIRS technology

M.J. De la Haba[1], I. Torres[1], A. Chamorro[1], A. Garrido-Varo[2], D. Pérez-Marín[2], and M.T. Sánchez[1]

[1]Department of Food Science and Food Technology ; [2]Department of Animal Production.
University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.

Introduction

Sugar content, acidity and dry matter content are indicators of bell pepper quality of key importance determination, for consumer acceptance. These parameters are routinely evaluated using analytical methods, which involve the destruction of the product. Producers and the industry would benefit greatly from a rapid, accurate, economical and above all non-destructive technology, enabling peppers to be classified by internal quality, particularly if this method could be implemented *in-situ*, in the field.

Objectives

To assess the potential of Near Infrared Reflectance Spectroscopy (NIRS) for determining total soluble solids content, titratable acidity and dry matter content in intact bell peppers using a handheld MEMS- based NIR spectrophotometer.

Materials and Methods

Sampling:

- 147 bell peppers *(Capsicum annum* L.*)* were harvested at comercial stage in Murcia (Spain).

Reference analysis:

- Soluble solids content (SSC) (%): Refractometry.
- Titratable acidity (TA) (% citric acid): Titration.
- Dry matter (DM) (% fw): AOAC (2000).

NIR analysis:

- Phazir 2400, Polychromix, Inc., Wilmington, MA, USA.
- Reflectance: 1600-2400 nm.
- Intact product.

Calibration development:

- Regression procedure: MPLS.
- Scatter correction: SNV + DT.
- Derivatives: 1,5,5,1; 2,5,5,1.
- Software: WinISI II, ver. 1.50.

Results

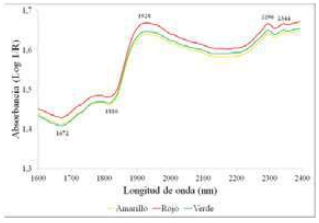


Table 1. MPLS regression statistics for NIR-based models for predicting internal quality parameters in intact bell peppers (N = 147).

	Mean	SD	SECV	r^2	RPD
SSC (%)	6.58	1.33	0.89	0.55	1.49
TA (% citric acid)	0.21	0.06	0.04	0.54	1.50
DM (% fw)	6.71	1.06	0.65	0.63	1.64

Conclusions

The handheld spectrometer tested enabled intact bell peppers to be screened in situ for internal quality parameters, allowing increased sampling of each batch on vine and at harvest, and thus ensuring a more precise and accurate guarantee of specific quality.

This research was funded by Gelagri Ibérica S.L. under the Research Project 'Characterisation of bell pepper cultivated in Santaella (Cordoba)').

I.4. International Diffuse Reflectance Conference (Chambersburg, Pennsylvania, USA)

LOCAL ALGORITHM FOR THE PREDICTION OF OPTIMUM HARVEST DATES IN FRUITS OF DIFFERENT GENUS

I. Torres[1*], D. Pérez-Marín[2], A. Garrido-Varo[2,*], M.J. De la Haba[1], M.T. Sánchez[1]

[1]Department of Food Science and Food Technology. University of Cordoba (Spain). *g72toroi@uco.es
[2]Department of Animal Production. University of Cordoba (Spain).

IDRC 2016

1. INTRODUCTION

The fruit sector needs a rapid, economical and non-destructive technology for monitoring the internal quality of fruits during on-tree and on-vine ripening process. Furthermore, the development of a robust and accurate universal mathematic model for this propose based on heterogeneous samples is demanded for producers and the industry.

Handheld near infrared reflectance (NIR) instruments are suitable for using them in the field and could be able to measure *in-situ* the main quality parameters in different fruit genus.

2. OBJECTIVE

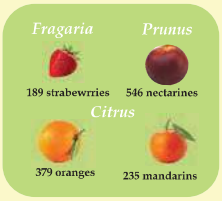
To evaluate the feasibility of LOCAL algorithm for the prediction of internal quality parameters in different fruit genus.

3. MATERIAL and METHODS

Sampling

Number of samples: 1,349

- 3 different genus
- 4 different fruits



NIRS analysis

Instrument: Phazir 2400

- MEMS technology
- Spectral range: 1600-2400 nm
- Reflectance mode

Reference analysis

- Soluble solid content (°Brix). Refractometry.
- Titratable acidity (% citric acid). Titration.

LOCAL algorithm

Software: WinISI II, version 1.50

- Number of calibration samples: *80, 100, 120, 140*
- PLS factors: *14, 15, 16*
- Derivatives: *1,5,5,1* and *2,5,5,1*

4. RESULTS and DISCUSSION

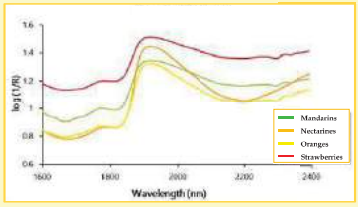


Figure 1. Mean NIR spectrum for mandarins, oranges, nectarines and strawberries.

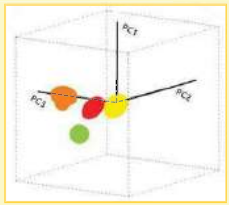


Figure 2. GH distance to the center of the population in the three main PC.

Table 1. Validation statistics for the prediction of soluble solid content (SSC) and titratable acidy (TA) in fruits of different genus using Local algorithm.

Parameter	Selected samples	Math treatment	$SEP_{(C)}$	r^2
SSC (°Brix)	80	2,5,5,1	0.88	0.89
TA (% citric acid)	120	1,5,5,1	0.13	0.80

5. CONCLUSIONS

These results confirm that NIRS technology, using a non-linear calibration method such as the LOCAL algorithm, could be an excellent tool for the on-tree and on-vine prediction of internal quality parameters in intact fruits. This could be the first attempt to develop an universal quality model for the prediction of the optimum harvest dates in fruits of different genus.

REFERENCES

Pérez-Marín, D., Garrido-Varo, A., & Guerrero, J. E. 2005. *Implementation of local algorithm with near-infrared spectroscopy for compliance assurance in compound feedingstuffs*. Applied Spectroscopy, 59, 69–77.

Pérez-Marín, D., Garrido-Varo, A., & Guerrero, J. E. 2007. *Non-linear methods in NIRS quantitative analysis*. Talanta, 72, 28–42.

I.5. V Congreso Científico de Investigadores en Formación de la Universidad de Córdoba (Córdoba, Spain)

V Congreso Científico de Investigadores en Formación de la Universidad de Córdoba

SENSORES ESPECTROSCÓPICOS NIRS PARA LA IDENTIFICACIÓN, CARACTERIZACIÓN Y MEJORA VARIETAL DE PRODUCTOS HORTOFRUTÍCOLAS

DOCTORADO EN INGENIERÍA AGRARIA, ALIMENTARIA, FORESTAL Y DEL DESARROLLO RURAL SOSTENIBLE

ALUMNA
Irina Torres Rodríguez

DIRECTORAS
Dra. Mª Teresa Sánchez Pineda de las Infantas
Dra. Dolores Pérez Marín

I.6. 18th International Conference on Near Infrared Spectroscopy (Copenhagen, Denmark)

Copenhagen, 14 June 2017

APPLICATION OF LOCAL REGRESSION METHODS TO NIR SPECTRA DATABASE OF CITRUS FRUIT FOR THE ASSESSMENT OF INTERNAL QUALITY

Authors:
Irina Torres
María-Teresa Sánchez
María-José de la Haba
Ana Garrido-Varo
Dolores Pérez-Marín

ICNIRS 2017 DENMARK COPENHAGEN 11-15 JUNE

EVALUATION OF NIRS FOR THE PREDICTION OF PHYSICAL AND SENSORIAL QUALITY PARAMETERS IN INTACT RAF TOMATO

María-Teresa Sánchez[1*], Irina Torres[1], María-José De la Haba[1], Ana Garrido-Varo[2], and Dolores Pérez-Marín[2]

[1] Department of Food Science and Food Technology. University of Cordoba, Spain. *teresa.sanchez@uco.es
[2] Department of Animal Production. University of Cordoba, Spain. dcperez@uco.es

INTRODUCTION

Raf tomato, which has a high market value, is a variety of tomato resistant to *Fusarium*. This variety is characterized by its irregular shape, with groove along the sides. Its colour ranges from deep green to red, with almost black-shoulder. In general, fruits are sweeter when the colour is darker.

Raf tomato harvest is typically determined by subjective measurements and little information exits on the changes in weight and sensory properties with ripeness.

Sensory parameters are routinely evaluated using methods which involve destruction of the piece of vegetable, only enabling quality control of a few samples per batch, rather than of each individual item.

The producers and the industry would benefit greatly from a rapid, economic and non-destructive method enabling Raf tomato to be classified by weight and sensory properties, particularly if this method could be built into the processing line.

OBJECTIVES

To assess the feasibility of Near Infrared Reflectance Spectroscopy (NIRS) for the measurement of external -weight and colour- and internal –taste- quality parameters in Raf tomatoes.

MATERIAL AND METHODS

SAMPLING

- 165 Raf tomatoes grown in the province of Almeria (Spain).

REFERENCE ANALYSIS

- Weight. Gravimetric method.
- External colour. Hedonic scale. Scored on a 6 to 1 scale, where: 6, red; 5, light red; 4, pink; 3, turning, 2, breaker; 1 green.
- Taste: Hedonic scale. Scored on a 5 to 1 scale, where: 1, very sweet; 2, sweet; 3, fairly sweet; 4, acid; 5, very acid.

NIRS ANALYSIS

- Intact product.
- Diode array NIRS instrument: PERTEN DA-7000
- Reflectance.
- Rotation module.
- Program WINISI II, version 1.5.
- Calibration development:
  - Regression Procedure: MPLS.
  - Spectral Range: 515-1650 nm.
  - Scatter Correction: SNV + DT.
  - Derivatives: 1,5,5,1; 2,5,5,1; 1,10,5,1; 2,10,5,1.

RESULTS

Table 1. Statistical analysis of the calibration and validation sets, i.e. data range, mean, standard deviation (SD) and coefficient of variation (CV).

Parameter	Set	Number	Range	Mean	SD	CV (%)
Weight (g)	Calibration	119	83.00-285.14	176.93	46.30	26.17
	Validation	40	101.61-240.04	151.40	34.10	22.52
External colour	Calibration	119	1.00-5.00	2.65	1.10	41.98
	Validation	40	1.00-5.00	2.53	1.17	46.25
Taste	Calibration	118	1.00-4.33	2.68	0.69	25.75
	Validation	40	1.33-4.00	2.46	0.72	29.27

Table 2. Calibration statistics for the best models obtained for the prediction of weight and sensory parameters in "Raf" tomatoes

Parameters	Mathematic treatment	Mean	SD	SECV	r^2_{cv}	RPD
Weight (g)	1,5,5,1	175.96	45.26	9.78	0.95	4.63
External colour	1,5,5,1	2.64	1.10	0.36	0.90	3.06
Taste	2,5,5,1	2.49	0.70	0.50	0.44	1.44

Table 3. Validation statistics for the best models obtained for the prediction of weight and sensory parameters in Raf tomatoes

Parameters	Mean		SD		r^2_p	SEP	SEP(c)	Bias
	R^a	P^b	R^a	P^b				
Weight (g)	152.58	151.56	33.70	35.04	0.87	12.73	12.86	1.02
External colour	2.53	2.44	1.17	1.10	0.91	0.36	0.36	0.09
Taste	2.46	2.52	0.72	0.56	0.37	0.58	0.58	-0.06

a, Reference values; b, NIRS predicted values.

CONCLUSIONS

These results confirm that NIRS technology could be used for determining quality parameters such as weight and colour in Raf tomato, in contrast to the taste.

REFERENCES

Flores, K., Sánchez, M.T., Pérez-Marín, D., Guerrero, J.E., Garrido-Varo, A., 2009. Feasibility in NIRS instruments for predicting internal quality in intact tomato. Journal of Food Engineering 91, 311-318.

Torres, I.; Pérez-Marín, D., De la Haba, M.J., Sánchez, M.T. 2015. Fast and accurate quality assessment of Raf tomatoes using NIRS technology. Postharvest Biology and Technology 107, 9-15.

This research was founded by the Andalusian Regional Government under the Research Excellence Program (Project N° 3713).

I.7. 5th Food Integrity Conference 'Assuring the Integrity of the Food Chain' (Nantes, France)

IN SITU ASSESSMENT OF SAFETY OF VEGETABLES BY MEASURING THE NITRATE CONTENT USING NEAR INFRARED TECHNOLOGY

I. Torres[1*], M.T. Sánchez[1], A. Garrido-Varo[2], D. Pérez-Marín[2]
[1]Department of Food Science and Food Technology. University of Cordoba (Spain) ([*]g72toroi@uco.es)
[2]Department of Animal Production. University of Cordoba (Spain)

1. INTRODUCTION

In response to the growing public concern about the presence of nitrates in foods, the European Union passed Commission Regulation (EC) No 1258/2011 of 2 December 2011, setting maximum levels for nitrates in vegetables, according to their final destination.(1)

Near Infrared (NIR) Spectroscopy by means of the use of portable instruments and in conjunction with multivariate analysis strategies, is an appropriate non-destructive technology for the *in situ* determination of the nitrate content in vegetables.(2,3)

2. OBJECTIVE

To develop robust and accurate models for the prediction of the nitrate content in summer squashes and spinach plants during on-vine ripening, with a view to optimizing harvesting times and enabling staggered harvesting by quality, thus allowing certain harvested squashes and spinach plants to be used in the production of baby foods.

3. MATERIAL AND METHODS

SAMPLING

- 157 summer squashes
- 128 spinach plants

NIRS ANALYSIS

- Instrument: Phazir 2400
- MEMS technology
- 1600 – 2400 nm

REFERENCE ANALYSIS

- Nitrate content (mg/kg): Reflectometry.(4)

CHEMOMETRIC ANALYSIS

Software: WinISI II, v1.50

- MPLS regression
- Derivatives: 1,5,5,1; 2,5,5,1
- Scatter correction: SNV + DT

4. RESULTS

Table 1. Calibration statistics obtained for the prediction of nitrate content.

STATISTICS	PRODUCT	
	Summer squashes	Spinach leaves
RANGE	30.00 – 1074.00	109.50 – 5177.00
r^2_{cv}	0.73	0.41
SECV (mg/kg)	154.04	836.26
RPD	1.91	1.29

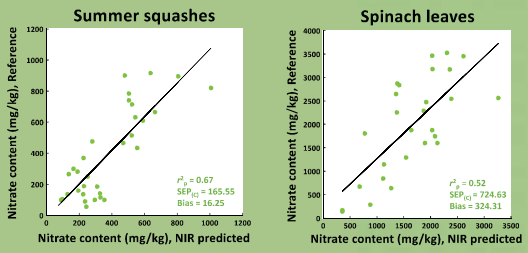


Figure 1. Reference and NIR predicted values for the nitrate content in summer squashes and spinach leaves.

5. CONCLUSIONS

This work is just a viability study, but the results showed that NIRS technology has a great potential for measuring nitrate content in intact summer squashes and spinach plants. It enables the industry to establish the final destination of the product according to the maximum levels of nitrate content stipulated by the EU legislation, allowing the selection of those vegetables that can be used for baby food production.

REFERENCES

(1) OJEU. 2011. Commission Regulation (EU) No 1258/2011 of 2 December 2011 amending Regulation (EC) No 1881/2006. OJ L 320/15 -17.
(2) Perez-Marín et al. 2019. Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy 207, 242-250.
(3) Sánchez et al. 2017. Postharvest Biology and Technology 125, 122-128.
(4) Thompson et al. 2009. Spanish Journal of Agriculture Research, 7, 200 – 211.

ACKNOWLEDGMENTS

This research was under the Research Projects 'Quality determination of summer squash grown on an open-air plantation in Santaella (Córdoba)' and 'Quality determination of spinach grown in Santaella (Córdoba)', funded by Gelagri Ibérica, S.L.

I.8. VI International Symposium on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain – Model IT (Molfetta, Italy)

Online postharvest assessment of quality in spinach plants using Near-Infrared Spectroscopy

I. Torres[1,a], D. Pérez-Marín[2], J.A. Entrenas[1], M.T. Sánchez[1]

[1]Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain; [2]Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.

Abstract

Dry matter (DM) and soluble solid content (SSC) are important attributes to assess the quality and freshness in spinach plants. Currently, these parameters are evaluated using time-consuming and destructive methods. However, over recent years there has been an increasing interest from the vegetable processing industry to incorporate Near Infrared (NIR) Spectroscopy as a non-destructive analytical technology to assess and classify vegetables by quality, which makes it necessary to simulate the working conditions of the industrial processes. The main challenge of this work was to optimize the NIR measurement process and to develop robust and accurate models by means of multivariate analysis strategies for the prediction of DM and SSC in spinaches. For this purpose, 128 spinach plants grown in open-air fields in the provinces of Cordoba and Seville (Spain) were online scanned in reflectance mode, using a Fourier Transform (FT)-NIR instrument (Matrix-F), working in the spectral range of 834–2,502.40 nm, equipped with a conveyor belt for the movement of the sample. For the construction of the calibration models for the parameters previously cited, modified partial least squares (MPLS) regression and various spectral signal pretreatments were tested. The results obtained showed that NIRS technology has a great potential for the rapid, accurate and non-destructive online determination of DM and SSC in the industry, with a view to guarantee the quality of the processed vegetables.

Keywords: Spinach plants, dry matter, soluble solid content, NIR instrument, online analysis.

INTRODUCTION

Spinach is a green, leafy vegetable with high water content and very low in carbohydrates and fats (Beis et al., 2002). After harvesting, due to the high perishable character of this vegetable, it is crucial to maintain high standard storage conditions in order to extend the shelf-life of this product.

In spinach plants, quality is the sum of the characteristics, attributes and properties that give this vegetable its food value. The degree of freshness usually associated with retention of good quality, usually declines during the ripening and subsequent storage due to water loss after harvest (Kader, 2002a). Thus, an important parameter to be measured in spinach as indicator of quality and turgidity is water content (Kader et al., 2002b). Additionally, SSC is also another key attribute to be assessed in order to establish the quality and freshness of spinach plants during their shelf-life.

NIR spectroscopy has demonstrated to be a potential tool for the non-destructive assessment of quality parameters in vegetables, which has been enhanced with the development of new instruments, ideally suited for the online analysis. This has led to a growing interest on the part of horticultural industries to incorporate this technique in their sorting lines as a routine analytical technique (Porep et al., 2015).

Sánchez et al. (2018) and Pérez-Marín et al. (2019) have demonstrated the potential of NIR spectroscopy to establish the quality of spinach plants at field level. However, no studies

have focused on the use of this technology for online applications in spinach with a view to measure quality parameters in the handling and sorting lines.

The aim of this work was to optimize the online NIR measurement process and to develop robust and accurate models applying multivariate analysis strategies for the prediction of DM and SCC in spinach plants using an FT-NIR instrument.

MATERIALS AND METHODS

Sampling

128 spinach plants (*Spinacia oleracea* L., cv. 'Solomon', 'Novico', 'Meerkat' and 'Gorilla') grown in open-air plantations in the Provinces of Cordoba and Seville (Spain) were harvested between January and March 2017. Harvested spinach were kept in refrigerated storage at 4°C and 85% RH until the following day. Prior to each test, spinach plants were allowed to reach room temperature.

Before conducting NIR and reference analyses, samples were washed following Karaca and Velioglu (2014). For each spinach plant, one leaf was selected for the determination of DM and the remaining leaves (between 4 and 10) were used for measuring SSC. This distinction led to the creation of two different sets (C1 and C2), intended for the determination of DM and SSC, respectively.

NIR spectrum acquisition

Spectra were collected on spinach leaves in reflectance mode (Log 1/R) in the spectral range of 4000 to 12000 cm^{-1} (834–2,502.40 nm) using the FT-NIR spectrophotometer Matrix-F (Bruker Optik GmbH, Ettlingen, Germany). This equipment consists of a sensor head with two NIR light sources which illuminate a sampling spot of 10 mm in diameter, attached to the instrument via a 5 m-long fibre optic probe. Furthermore, the system was equipped with a conveyor belt to move the sample, whose speed was set at 15 kHz. An internal white reference was also collected every ten minutes.

The number of scans were 16 and two spectral measurements were made per each sample. In the case of the spinach leaves for DM determination, the NIR spectra were taken in static mode, without movement of the conveyor belt with the leaf placed on a methacrylate black plate. Because between 4 and 10 leaves were used for SSC determination, the spectra acquisition was carried out in dynamic mode, with displacement of the conveyor belt. In both cases, two spectra were recorded on each sample. The two spectra were averaged to provide a mean spectrum for each sample.

Reference data

DM (% fw) was determined gravimetrically by desiccation at 105°C for 24h (AOAC, 2000), and the final dry weight was calculated as a percentage of the initial wet weight.

SSC (ºBrix) was measured as the refractometer reading for the spinach juice, using a temperature-compensated digital Abbé-type refractometer (model B, Zeiss, Oberkochen, Würt, Germany).

Quantitative calibrations: sets and data processing

The WinISI II version 1.50 software package was used for the chemometric treatment of data (ISI, 2000).

Prior to the development of NIRS calibrations, the structure and spectral variability of the sample population were determined following the method recommended by Shenk and Westerhaus (1991), using the CENTER algorithm. Then, having ordered the sample sets by spectral distances (from smallest to greatest distance to the centre) the outlier spectral samples (GH > 4) were eliminated.

Prediction equations were obtained using the MPLS regression method with cross-validation. Different mathematical treatments were evaluated for scatter correction, including Standard Normal Variate (SNV) and De-trending (DT) (Barnes et al., 1989). Moreover, two derivative mathematical treatments were tested in the development of NIRS

calibrations: 1,5,5,1 and 2,5,5,1, where the first digit is the number of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing, and the fourth the second smoothing (Shenk and Westerhaus, 1995).

The statistics used to select the best equations were: standard error of cross-validation (SECV), coefficient of determination for cross-validation (r^2_{cv}), RPD or ratio of the standard deviation of the original data (SD) to SECV, and the coefficient of variation (CV) or ratio between the SECV and the mean value of the reference data of the studied parameter for the training set. These latter two statistics enable SECV to be standardized, facilitating the comparison of the results obtained with sets of different means (Williams, 2001).

RESULTS AND DISCUSSION

Descriptive data for NIR calibration sets

Values for range, mean, standard deviation and coefficient of variation for each of the parameters analysed after application of the CENTER algorithm and removal of spectral outliers are shown in Table 1, together with the number of samples for each measured parameter.

Table 1. Statistical analysis of the calibration sets: number of samples (N), data ranges, means, standard deviations (SD), and coefficients of variation (CV)

Parameter	N	Range	Mean	SD	CV (%)
Dry matter (% fw)	124	6.14–19.67	12.03	2.97	24.69
SSC (ºBrix)	128	1.34–14.25	8.37	2.63	31.42

For the set destinated to the prediction of DM, the spectral anomalous detected were 4, whereas no outliers were detected for the SSC set. A detailed analysis showed that one of the DM anomalous was associated with a small size leaf. No physicochemical explanation was found for the other three DM outliers, so it could be due to their lower spectral representation compared to the whole population.

In relation to the variability, the studied parameters displayed similar range of values for the coefficient of variation (CV = 24.53% for DM and CV = 31.42% for SSC), although for the SSC set, it was slightly higher. This fact could be explained due to the fact that all the spinach plants tested were collected at their commercial maturity stage.

Prediction of quality parameters using MPLS regression and NIR spectra

Table 2 shows the results of the best prediction models obtained for each parameter analysed (DM and SCC), using different pre-treatments of the spectral signal.

Table 2. MPLS regression statistics for NIR-based models for predicting quality parameters in spinach plants

Parameter	Math treatment	N	Mean	SD	SECV	r^2_{cv}	CV (%)	RPD_{cv}
Dry matter (% fw)	2,5,5,1	118	11.76	2.75	1.40	0.74	11.90	1.96
SSC (ºBrix)	1,5,5,1	118	8.80	1.91	0.73	0.86	8.30	2.62

For DM the calibration model showed a good predictive capacity in terms of the coefficient of regression and according to Shenk and Westerhaus (1996). The non-destructive measurement of this parameter online is, in fact, of great importance for the handling of the

postharvest crop. Conte et al. (2008) showed that DM values of around 10–12 % fw ensure a good resistance to handling and allow maintenance of visual quality at a high standard during storage.

The unique scientific article published on the measurement of DM in spinach plants using NIRS technology was published by Sánchez et al. (2018). The authors obtained the same results (RPD_{cv} = 1.96) than those obtained in this work, using a handheld MEMS-NIR spectrophotometer in the spectral range 1600-2400 nm, suited for *in-situ* analysis.

As Table 2 shows, good predictive ability was obtained for the measurement of SSC (Shenk and Westerhaus, 1996). Several works have already demonstrated the possibility of measuring this parameter using NIRS technology (Nicolaï et al., 2007; Sánchez and Pérez-Marín, 2011). No reports have been found regarding the online determination of SSC in spinach; however, Pérez-Marín et al. (2019) for the *in-situ* determination of this parameter, obtained models with slightly lower predictive capacity (RPD_{cv} = 2.53). It should be taken into consideration that the authors used a portable spectrophotometer (Phazir 2400) with a shorter spectral range (1,600–2,400 nm) than the one used in this work and a lower spectral resolution (8 nm for the Phazir 2400 *versus* 1.61 nm for the Matrix-F).

The regression coefficients for the best predictive models obtained for DM and SSC are shown in Fig. 1.

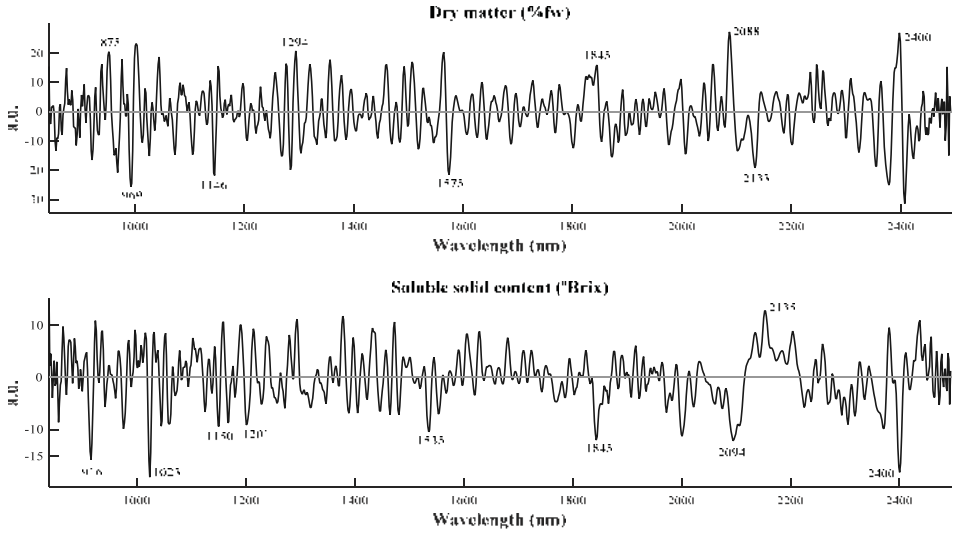


Figure 1. Regression coefficients for DM and SSC of intact spinach leaves.
*a.u. = arbitrary units

For both parameters, these regression coefficients show significant importance for the region around at 960 nm, which correspond to the water absorption (Shenk et al., 2008), the major component of this product. The absorbance regions 1,150–1,250 nm, 1,650–1,850 nm and 2,350–2,400 nm, which also have a great importance, could be attributed to the first overtone of C-H stretching (Shenk et al., 2008). Furthermore, these models also show variations in the 2,080–2,180 nm due to the O-H and N-H bonds, as well as the C=O vibration bands, corresponding to carbohydrates and proteins (Williams, 2001; Shenk et al., 2008).

CONCLUSIONS

The results obtained confirm that NIR technology has a great potential for the rapid, accurate and non-destructive online determination of DM and SSC in the horticultural

industry with a view to guarantee the quality of the processed vegetables. The FT-NIR device tested here should be considered for use in the routine, non-destructive analysis of spinach in the handling and sorting processing lines, enabling to provide consumers with high quality products.

ACKNOWLEDGEMENTS

This research was funded by Gelagri Ibérica, S.L. under the Research Project 'Quality determination of spinach grown in Santaella (Córdoba)'. The authors would also like to express their gratitude to the Spanish Ministry of Education, Culture and Sports for the support offered to Irina Torres through the Training Programme for Academic Staff (FPU).

REFERENCES

AOAC. (2000). Official Methods of Analysis of AOAC International. 17th ed. (Gaithersburg, MD, USA: AOAC).

Barnes, R.J., Dhanoa, M.S., and Lister, S.J. (1989). Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra. Appl. Spectrosc. 43, 772–777 http://dx.doi.org/10.1366/0003702894202201.

Beis, G.H., Siomos, A.S., and Dogras, C.C. (2002). Spinach composition as affected by leaf age and plant part. Acta Hortic. *57*, 653–658 http://dx.doi.org/10.17660/ActaHortic.2002.579.115.

Conte, A., Conversa, G., Scrocco, C., Brescia, I., Laverse, J., Eliba, A., and Nobile, M.A.D. (2008). Influence of growing periods on the quality of baby spinach leaves at harvest and during storage as minimally processed produce. Postharvest Biol. Technol. *50*, 190–196 https://doi.org/10.1016/j.postharvbio.2008.04.003.

ISI. (2000) The Complete Software Solution Using a Single Screen for Routine Analysis, Robust Calibrations, and Networking. Manual, FOSS NIRSystems/TECATOR. (Silver Spring, MD, USA: Infrasoft International, LLC.).

Kader, A.A. (2002a). Postharvest biology and technology: An overview. In: Postharvest Technology of Horticultural Crops, 3rd Edition, A.A. Kader, ed. (Oakland, California, USA: University of California, Division of Agriculture and Natural Resources), p. 39–47.

Kader, A.A. (2002b). Quality and safety factors: Definition and evaluation for fresh horticultural crops. In: In: Postharvest Technology of Horticultural Crops, 3rd Edition, A.A. Kader, ed. (Oakland, California, USA: University of California, Division of Agriculture and Natural Resources), p. 279–285.

Karaca, H., and Velioglu, Y.S. (2014). Effects of ozone treatments on microbial quality and some chemical properties of lettuce, spinach, and parsley. Postharvest Biol. Technol. *88*, 46–53 https://doi.org/10.1016/j.postharvbio.2013.09.003.

Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., and Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. Postharvest Biol. Technol. *46*, 99–118 http://dx.doi.org/10.1016/j.postharvbio.2007.06.024.

Pérez-Marín, D., Torres, I., Entrenas, J.A., Vega, M., and Sánchez, M.T. (2019). Pre-harvest screening on-vine of spinach quality and safety using NIRS technology. Spectrochim. Acta A Mol. Biomol. Spectrosc. *207*, 242–250. http://dx.doi.org/10.1016/j.saa.2018.09.035

Porep, J.U., Kammerer, D.R., and Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. Trends Food Sci. Tech. *46*, 211–230 https://doi.org/10.1016/j.tifs.2015.10.002

Sánchez, M.T., and Pérez-Marín, D. (2011) Nondestructive measurement of fruit quality by NIR spectroscopy. In: Advances in Post-Harvest Treatments and Fruit Quality and Safety, M. Vázquez and J.A. Ramírez, eds. (Hauppauge, New York, USA: Nova Science Publishers, Inc.), p. 101–163.

Sánchez, M.T., Entrenas, J.A., Torres, I., Vega, M., and Pérez-Marín, D. (2018). Monitoring texture and other quality parameters in spinach plants using NIR spectroscopy. Comput. Electron. Agr. *155*, 446–452 https://doi.org/10.1016/j.compag.2018.11.004

Shenk, J.S., and Westerhaus, M.O. (1991). Population structuring of near infrared spectra and modified partial least squares regression. Crop Sci. *31*, 1548–1555 http://dx.doi.org/10.2135/cropsci1991.0011183X003100060034x

Shenk, J.S., and Westerhaus, M.O. (1995) Routine Operation, Calibration, Development and Network System Management Manual. (12101 Tech Road, Silver Spring, MD 20904, USA: NIRSystems, Inc.).

Shenk, J.S., and Westerhaus, M.O. (1996) Calibration the ISI way. In: Near Infrared Spectroscopy: The Future Waves, A.M.C. Davies, and P.C. Williams, eds. (Chichester, UK: NIR Publications), p.198–202.

Shenk, J.S., Workman, J., and Westerhaus, M.O. (2008) Application of NIR spectroscopy to agricultural products. In: Handbook of Near Infrared Analysis, D.A. Burns, and E.W. Ciurczac, eds. (Basel, Switzerland: Marcel Dekker), p. 347–386.

Williams, P.C. (2001) Implementation of near-infrared technology. In: Near-Infrared Technology in the Agricultural and Food Industries, P.C. Williams, and K.H. Norris, eds. (St. Paul, Minnesota, USA: AACC, Inc.), p. 145–169.

I.9. 19th International Conference on Near Infrared Spectroscopy (Gold Coast, Queensland, Australia)

ABS-ID#

19th International Council for NIR Spectroscopy Meting, NIR2019
15-20 September 2019, Gold Coast, Australia

Routine safety control of vegetables at different stages of the production chain using Near Infrared Spectroscopy

Irina Torres[1,*], María-Teresa Sánchez[1], José-Antonio Entrenas[1], Ana Garrido-Varo[2], Dolores Pérez-Marín[2]

[a] Department of Bromatology and Food Technology, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.

[b] Department of Animal Production, University of Cordoba, Campus of Rabanales, 14071 Córdoba, Spain.

ABSTRACT

The demand in terms of safety in the horticultural sector stresses the need of the producers and the agri-food industry of using non-destructive analysis technologies, such as Near Infrared Spectroscopy (NIRS), which allows to measure the level of nitrates in a matter of seconds. This work is focused on the prediction of the nitrate content in intact summer squashes and spinach plants at different stages of the production chain using two different spectrophotometers, ideally suited ideally suited to analysing these horticultural products *in-situ* and on-line, respectively. For this purpose, 230 summer squashes and 199 spinach plants grown outdoors in the provinces of Cordoba and Seville (Spain) were scanned using two NIR instruments (MicroNIR™ 1700 and Matrix F). For the construction of the calibration models, modified partial least squares regression and various spectral signal pretreatments were tested. The results obtained showed that NIR technology can be used as a rapid, accurate and non-destructive technology throughout the production chain with a view to guarantee the safety of these two vegetables.

INTRODUCTION

Over recent years, consumers have become increasingly aware of the risk of the presence of nitrates in vegetables. The European Union is aware of this problem and has established maximum limits for nitrates content in vegetables such as summer squashes and spinach plants according to their final destination (OJEU, 2011). As a result of this, both producers and the agri-

food industry are increasingly anxious to provide consumers with assurances regarding the safety of these products. Near Infrared (NIR) Spectroscopy by means of the use of a new generation of instruments (portable and online) and in conjunction with the application of multivariate analysis strategies, is an appropriate non-destructive technology for the determination of the nitrate content in vegetables throughout the production chain. The main challenge of this work was to develop robust and accurate models for the *in-situ* and online prediction of the nitrate content in two vegetables: summer squashes and spinach plants using two new generations of NIRS instruments.

METHODS AND MATERIALS

230 summer squashes and 199 spinach plants — both grown in an open-air plantation — were scanned using two spectrophotometers: the first is the MicroNIR^{TM} 1700 (900–1700 nm, with an interval of 6.2 nm), a manual, portable instrument based on Linear Variable Filter (LVF) technology, ideally suited to analysing horticultural products *in situ*, and the other is the Matrix-F (834–2502 nm, with an interval of 1.61 nm), a Fourier Transform NIR device, suitable for online analysis.

For both vegetables, the NIR analysis using the portable instrument was made in static mode. On the one hand, in the case of the summer squashes, fruits were divided into three regions (stem, equatorial and stylar regions) and 4 spectra were taken at the centre of each of the selected regions, at 90º intervals, so that 12 spectra were obtained per summer squash. On the other hand, six spectral measurements were made on each spinach leaf; since between 4 and 10 leaves were used for the chemical analysis, spectra were taken for each leaf. In both cases, spectra were averaged to obtain a mean spectrum for each plant.

With the Matrix-F instrument, the analysis was carried out online as the product was moving on the conveyor belt. For summer squashes, the number of scans selected per fruit was 12, covering the entire length of the fruit as it moved on the belt. A total of 4 spectra per fruit were taken, rotating the fruit 90º after each measurement. For spinach plants, the number of scans were 16 and 2 spectra per sample were taken. The spectra acquired for each sample were averaged to obtain a mean spectrum.

Nitrate content (mg NO_3^-/kg) was measured by reflectometry following the procedure described in Sánchez et al., (2017) and Pérez-Marín et al., (2019).

Data analysis was performed using the WinISI software package ver. 1.50 (Infrasoft International LLC, Port Matilda, PA, USA). The CENTER algorithm was applied for the selection

of calibration and validation sets based on spectral information, using as pre-treatments for the scatter correction the standard normal variate and detrending (SND + DT) together with the first derivative (1,5,5,1); those samples with a GH value higher than 4 were considered outliers. Modified partial least squares (MPLS) regression was used to obtain NIR calibration models, using SNV + DT for scatter correction and different derivative mathematical (1,5,5,1; 2,5,5,1). Finally, once the best equations were selected according to the statistical criteria, the models were subjected to an external validation process following the protocols outlined by Windham et al., (1989).

RESULTS

The calibration statistics obtained for the prediction of nitrate content for each product analyzed are shown in Table 1. Results showed that NIRS technology has a great potential for measuring nitrate content in vegetables *in-situ*, in the field. Using the MicroNIR™ 1700, the predictive capacity of the models obtained for both, summer squashes and spinach plants allowed to distinguish between high, medium and low values for this parameter (Shenk and Westerhaus, 1996). It is shown that the recent development of applications using new manual, portable, light-weight equipment, such as the MicroNIR™ 1700 can be of great use to producers.

As regards Matrix-F, the results obtained for summer squashes are limited for routine use whereas in the case of spinach plants the model obtained allowed to differentiate between low and high reference values (Shenk and Westerhaus, 1996). The reason for this may be that, for parameters such as nitrates, which are found in very low concentrations in summer squash and spinach, the fact that no contact is made with the vegetable when taking the measurement may affect the results for minority parameters.

CONCLUSIONS

The results obtained shows that NIRS is a potential tool for the rapid, accurate and non-destructive measurement of nitrate content, with a view to guarantee the safety of vegetables for human consumption throughout the whole production chain, i.e., from farm to fork. However, since it is a complex and minor parameter with a huge range of values, a larger and highly representative set of samples is needed for the development of robust models.

REFERENCES

Official Journal of the European Union (OJEU). (2011). Commission Regulation (EC) No 1258/2011 of 2 December 2011 amending Regulation (EC) No 1881/2006 as regards maximum levels for nitrates in foodstuffs. OJ L 320/15-17, 3.12.2011.

Pérez-Marín, D., Torres, I., Entrenas, J.A., Vega, M., Sánchez, M.T. (2019). Pre-harvest screening on-vine of spinach quality and safety using NIRS technology. *Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy*, 207, 242-250.

Sánchez, M.T., Pérez-Marín, D., Torres, I., Gil, B., Garrido-Varo, A., De la Haba, M.J. (2017). Use of NIRS technology for on-vine measurement of nitrate content and other internal quality parameters in intact summer squash for baby food production. *Postharvest Biology and Technology,* 125, 122–128.

Shenk, J.S., Westerhaus, M.O. (1996). Calibration the ISI way. In: Davies, A.M.C., Williams, P.C. (Eds.), Near Infrared Spectroscopy: The Future Waves. NIR Publications, Chichester, West Sussex, UK, pp. 198–202.

Windham, W.R., Mertens, D.R., Barton II, F.E. (1989). Protocol for NIRS calibration: sample selection and equation development and validation. In: Martens, G.C., Shenk, J.S., Barton II, F.E. (Eds.), Near Infrared Spectroscopy (NIRS): Analysis of Forage Quality. Agriculture Handbook, vol. 643. USDA-ARS, US Government Printing Office, Washington, DC, USA, pp. 96–103.

Table 1. Calibration and validation statistics for the prediction of nitrate content (mg NO_3^-/kg) in summer squashes and spinach plants.

Product	Instrument	Cross-validation				External validation		
		Mean	SD	SECV	r^2_{cv}	Bias	$SEP_{(C)}$	r^2_p
Summer	MicroNIR™ 1700	327.02	246.98	154.06	0.61	7.94	172.45	0.54
squash	Matrix F	360.03	299.33	211.10	0.50	-33.64	231.30	0.40
Spinach	MicroNIR™ 1700	1394.47	860.90	601.72	0.51	-77.30	621.90	0.57
plant	Matrix F	1438.50	914.21	720.21	0.38	-283.92	716.79	0.30