



UNIVERSIDAD DE JAÉN

**ESCUELA POLITÉCNICA SUPERIOR
DE JAÉN**
**DEPARTAMENTO DE INGENIERÍA
ELECTRÓNICA Y AUTOMÁTICA**

TESIS DOCTORAL

**APPLICATION OF NON-CONVENTIONAL
SENSOR TECHNOLOGIES TO THE
AUTOMATIC MONITORING AND CONTROL
OF THE VIRGIN OLIVE OIL
PRODUCTION PROCESS**

**PRESENTADA POR:
DIEGO MANUEL MARTÍNEZ GILA**

**DIRIGIDA POR:
DR. D. JUAN GÓMEZ ORTEGA
DR. D. JAVIER GÁMEZ GARCÍA**

JAÉN, 13 DE JULIO DE 2016

ISBN 978-84-9159-059-0

Doctoral Thesis

APPLICATION OF NON-CONVENTIONAL SENSOR TECHNOLOGIES
TO THE AUTOMATIC MONITORING AND CONTROL OF THE VIRGIN
OLIVE OIL PRODUCTION PROCESS

Author:

D. Diego Manuel Martínez Gila

Directors:

Dr. D. Juan Gómez Ortega

Dr. D. Javier Gámez García

Programa de Doctorado en Tecnologías de la Información y la Comunicación (R.D. 99/2011)
Escuela Politécnica Superior de Jaén
Universidad de Jaén

2016

Tesis Doctoral: APPLICATION OF NON-CONVENTIONAL SENSOR TECHNOLOGIES
TO THE AUTOMATIC MONITORING AND CONTROL OF THE VIRGIN
OLIVE OIL PRODUCTION PROCESS

Autor: D. Diego Manuel Martínez Gila
Directores: Dr. D. Juan Gómez Ortega
Dr. D. Javier Gámez García

El tribunal compuesto por los siguientes doctores:

Presidente:

Vocal:

Secretaria:

acuerdan otorgarle la calificación de:

La Secretaria del Tribunal

Fecha:

Agradecimientos

En primer lugar me gustaría expresar mi más sincero agradecimiento a mis dos directores de tesis, Dr. Juan Gómez Ortega y Dr. Javier Gámez García, por darme la oportunidad de trabajar durante todos estos años en el Grupo de Robótica, Automática y Visión por Computador (GRAV) de la Universidad de Jaén.

Además me gustaría agradecer a Pablo Cano Marchal, a Julio Beltrán Ortega, a Óscar Cáceres Moreno y a Daniel Aguilera Puerto su colaboración en las diferentes tareas relacionadas con la recogida de muestras, con la experimentación llevada a cabo en planta y con la documentación de los resultados de investigación. También hago extensible mi agradecimiento a los profesores Alejandro Sánchez García, Silvia Satorres Martínez, Elisabet Estévez Estévez, Alonso Ruano Ruano y Jesús de la Casa Cárdenas, y al resto de compañeros del grupo de investigación, incluyendo a los profesores Fabio Mencarelli y Andrea Bellincontro que tan bien me atendieron durante mi estancia en Italia.

Por otro lado transmitir mi agradecimiento al laboratorio de análisis CM Europa por facilitar el suministro de muestras y sus análisis, y a la cooperativas San Francisco de Asís de Montefrío y San Bartolomé de Torredelcampo por ceder parte de su maquinaria para tareas de experimentación durante la campaña.

Para concluir, me gustaría dar las gracias a toda mi familia, a mis padres Diego y María José y a mi hermana María José porque sin su apoyo todo el tiempo que se le ha dedicado al desarrollo de esta tesis no se hubiese sobrellevado del mismo modo.

Diego Manuel Martínez Gila
Jaén, 2016

A mi familia

List of Figures

1.1.	Possibilities for the automatic control of the virgin olive oil production process according to the different sub-processes of the production process; see Chapter 2 for more information	2
1.2.	Possibilities for the automatic control of the virgin olive oil production process (continuation)	3
1.3.	Linkage between quality, yield and energy efficiency	4
1.4.	Logotype of the projects	6
2.1.	Olive fruits reception in an industrial plant of olive oil production	14
2.2.	Schematic overview of process control strategies in food manufacturing (obtained from the reference [191])	15
2.3.	Changes in free acidity and bitterness of Cornicabra virgin olive oil extracted at different stages of olive maturity (obtained from the reference [172])	17
2.4.	Farmer discharging his olive fruits into the hopper	18
2.5.	Virgin olive oil production process from olive reception to oil storage	19
2.6.	Picture of the inside of the thermomixer	20
2.7.	Horizontal decanter machine	21
2.8.	Olive oil cellar	22
2.9.	Model of infrared active molecule as a vibrating dipole between two atoms (From [202])	24
2.10.	Block diagram that shows the generic structure of a Near Infrared (NIR) device	26
2.11.	Halogen lamp housing, from which light is delivered through an optical fiber	26
2.12.	Michelson Interferometer concept (a) and NIR device based on this technology (b)	27
2.13.	Diffraction grating concept (a) and spectrograph device based on this technology (b)	28
2.14.	Hypercube	29
2.15.	Measurement methodologies: reflectance (a), transmittance (b) and transfectance (c)	30
2.16.	Sample turn table accessory of the spectrometer used in [38]	32
2.17.	Experimental setup used in [171]. CORONA 45 Vis-NIR (left) and Foss NIRSystem 6500 (right) instruments	33
2.18.	Schematic diagram of localization of the sensor NIR in the last phase of the olive oil extraction process used in [135]: (1) oil from horizontal centrifuge decanter; (2) vertical centrifuge for oil clarification; (3) tank for oil sedimentation; (4) continuous oil-weigher; (5) oil to storage; (6) NIR equipment	37
2.19.	Schematic decanter and localization for reflectance probe used in [108]. R_2 is the decanter radius; R_1 is the level output oil; L is the length of horizontal zone of decanter	44

2.20.	System obtained from [92] and designed for this work in order to take Raman measurements	45
2.21.	Schematic diagram of localisation of the sensor AOTF-NIR used in [2]	48
2.22.	Structure of an image analysis system	49
2.23.	The camera takes images which are acquired and processed by the machine vision system. This system sends the parameters of each olive to the control system which decides which class the olive belongs to and when it should be ejected (Figure obtained from [54])	51
2.24.	Damage types identified in olives by expert technicians at processing companies: (a) undamaged olives; (b) 'serpeta'; (c) 'granizo'; (d) 'rehús'; (f) wrinkled; (g) purple olive; (e) and (h) others. (Figure obtained from [165])	52
2.25.	Positioning of the CMOS camera, camera mode, spectral response and optical (Figure obtained from [31])	54
2.26.	Image acquisition set-up, composed by the illumination ring and the camera (top). In the bottom of the image is the porcelain basin containing the olive oil sample. (Figure obtained from [133])	58
2.27.	(a) System used in the acquisition of images of the edible vegetable oil samples where (1) is the notebook; (2) is a peristaltic pump; (3) is the fluorescent lamp; (4) is the closed compartment. (b) Detailed compartment where (5) is a webcam and (6) is the flow cell. (Figure obtained from [188])	59
3.1.	Vision system configured. In the figure, label 1 indicates the InGaAs hyperspectral camera, label 2 the illumination system used, label 3 the conveyor belt, label 4 the infrared sensor and label 5 indicates the variable speed drive	64
3.2.	a) Image of an olive oil sample contained in a basin. The red line denotes the piece of sample to analyze. The hyperspectral image for this line is shown in the Figure b. b) Image (320x256) captured by hyperspectral camera. Discontinuous lines emphasize the white reference (white paper), dark reference (conveyor belt) and olive oil reflectance information	65
3.3.	Spectrogram without any preprocessing of an olive oil hyperspectral image	66
3.4.	a) Components selected by GA method, marked with x marks. b) Regression fitted by cross validation.	69
3.5.	a) Components selected by SPA method, pointed with x marks. b) Regression fitted by cross validation.	70
3.6.	a) Components selected by SPA method, pointed with x marks. b) Regression fitted by cross validation.	71
4.1.	Work flow followed for the proposed approach	77
4.2.	Olive oil sampling process into the factory	78
4.3.	Experimental setup where label 1 is the NIR sensor, label 2 is the light source and label 3 is the transreflectance probe	79
4.4.	Olive oil sample spectra for the three groups: S_1 in red, S_2 in green and S_3 in blue.	81
4.5.	Blue line represents the CV coefficients for each wavelength; Red line is the CV mean value	86
4.6.	Figure shows the selected wavelengths after the filtering step (red marks) and the spectra of the whole dataset for each preprocessing algorithms: (a) without preprocessing, (b) Standard Normal Variate, (c) Multiplicative Scatter Correction and (d) Savitzky-Golay approach	87
4.7.	Predicted values versus target values when the built linear model is applied. Figure (a) represents the results for calibration and (b) for validation	88
4.8.	Blue and red lines represent the RMSEV and RMSEC values respectively for each windows size when Savitzky-Golay preprocessing algorithm is employed	89
4.9.	Figures (a) and (b) show the $RMSEC$ and $RMSEV$ values when different light intensities were employed and different number of acquisitions per sample were averaged respectively	90

5.1.	Malaxing of olive paste inside the thermomixer	94
5.2.	Block diagram of the control system for the malaxing paste state	96
5.3.	Image acquisition set-up. Subfigure (a) is composed by the vision camera and the illumination lamp which are installed in the thermomixer. At the left of this subfigure is the cabinet that houses the computer PC and the electronic associated. Subfigure (b) presents an example of image acquired by the infrared sensor. Subfigure (c) displays the same scene in the visible colour channel	98
5.4.	Image setup for at-line acquisitions. Subfigure (a) is composed by the vision camera, two halogen lamps and the conveyor belt. At the left of this subfigure appears the PC which handles the experimental setup. To the left of this subfigure is the Petri dish with a sample of paste below the vision device. Subfigure (b) displays the acquired images in the infrared (left) and colour (right) channels	99
5.5.	These figures show the olive paste adhered to one of the blades of the thermomixer in the red channel (a), green channel (b), blue channel (c) and infrared channel (d). In the last one the olive paste is more contrasted	100
5.6.	Infrared images of the surface of the olive paste inside the thermomixer. The image (a) is an example of paste without floating oil, and the image (b) shows the olive paste flooded by oil. Note their different tones	101
5.7.	Fuzzy Rule-Based System. Viscosity, floating oil and granularity (V , FO and G respectively) are the inputs. Malaxing time, malaxing temperature, coadjuvant addition and speed of the shovels (Mt , MT , C and S respectively) represents the outputs	103
5.8.	Input fuzzy membership functions	103
5.9.	Input fuzzy membership functions	104
5.10.	Viscosity index for different channels versus malaxing time	106
5.11.	Floating oil index for different channels versus malaxing time	106
5.12.	Homogeneity parameter in the green channel versus malaxing time	107
5.13.	Inputs values for the controller where H_G is the textural parameter homogeneity in the green channel, $I_{IR}(FO)$ is the intensity value of the infrared channel for floating oil and $I_{IR}(V)$ is the same value for viscosity	108
5.14.	Outputs for different olive paste conditions	109
6.1.	An example of commercial classification machine for table olives. Figure obtained from www.multiscan.eu	112
6.2.	At-line acquisition system setup. Subfigure (a) shows the setup as sketch and Subfigure (b) is the real setup. On the other hand, Subfigure (c) presents an example of tree olives and Subfigure d) is an example of soil olives	114
6.3.	Difference in brightness with different light intensity	115
6.4.	Figure a) shows the different components of the computer vision system; Figure b) is a real image of the vision system installed over the conveyor belt; Figure c) presents an example of tree olives; Figure d) is an example of soil olives	116
6.5.	Olives coming from different locations and different image analysis. The first row shows the original image of the samples (RGB); the second row presents the result after applying the textural filter (I) and the third and fourth rows show the same samples where the brownish colour has been detected (C_{rg} and C_{rb} respectively)	118
6.6.	Proposed FDA classifier diagram	121
6.7.	Neuronal network diagram where P is the input vector; IW is the input weights; OW is the output weights; b_1 and b_2 are the bias; f_1 and f_2 are the sigmoid functions and C is the classification result	121

6.8.	This figure shows the normalized image histograms after applying the textural filter (a), channel R minus channel G operation (b) and channel R minus channel B operation (c)	122
6.9.	This figure shows the statistical F – values and p – values between the aforementioned two classes of olives for the histograms in Figure 6.8	123
6.10.	Feature vectors of the whole dataset after filtering by ANOVA test	123
6.11.	All the dataset is projected in the new 3D space which is created by three principal components PC1, PC2 y PC3	124
6.12.	a) Lineal classification results without ANOVA component filter; b) Lineal classification results with ANOVA component filter; c) ANN classification results without ANOVA component filter; d) ANN classification results with ANOVA component filter	126
6.13.	Concept of machine developed by GRAV	128
A.1.	Posibilidades de automatización de las fases del proceso	141
A.2.	Posibilidades de automatización de las fases del proceso (continuación)	142
A.3.	Relación entre calidad, agotamiento y eficiencia energética	143
A.4.	Logotipo de los proyectos	146

List of Tables

2.1.	NIR detectors (adapted from [143])	28
2.2.	Research works about infrared spectroscopy applied to olive fruits	34
2.3.	Research works about infrared spectroscopy applied to olive fruits (continuation)	35
2.4.	Research works about infrared spectroscopy applied to olive oil	39
2.5.	Research works about infrared spectroscopy applied to olive oil (continuation I)	40
2.6.	Research works about infrared spectroscopy applied to olive oil (continuation II)	41
2.7.	Research works about infrared spectroscopy applied to olive oil (continuation III)	42
2.8.	Research works about infrared spectroscopy applied to olive oil (continuation IV)	43
2.9.	Research works about infrared spectroscopy applied to olive slurries	46
2.10.	Research works about infrared spectroscopy applied to olive slurries (continuation)	47
2.11.	Research works about computer vision applied to olive fruit	55
2.12.	Research works about computer vision applied to olive fruit (continuation I)	56
2.13.	Research works about computer vision applied to olive fruit (continuation II)	57
2.14.	Research works about computer vision applied to olive oil	60
3.1.	Reference parameters from laboratory	63
3.2.	Results for free acidity estimation	69
3.3.	Results for peroxide index estimation	70
3.4.	Results for moisture estimation	71
4.1.	Different groups of spectra considered	80
4.2.	Experimental results	88
5.1.	Textural parameters	107
6.1.	Classification results in laboratory	125
6.2.	Classification results in oil mill	127

Resumen

En los últimos años la producción de aceite de oliva ha experimentado un crecimiento desde el punto de vista cualitativo y cuantitativo, en parte gracias a la transferencia de conocimiento desde los centros de investigación hacia la planta de producción. Sin embargo, el grado de automatización del sistema productivo se encuentra de momento en una etapa inicial del recorrido que supone ponerse al nivel de otras industrias agroalimentarias. En este contexto, el trabajo que se desarrolla en la presente Tesis Doctoral tiene como objetivo profundizar en la aplicación de dos tecnologías no invasivas, espectroscopía en el infrarrojo cercano y visión por computador, para adquirir información a nivel de proceso, y en sus distintas etapas, y fundamentalmente con dos fines: (1) ayudar a determinar de forma cualitativa qué está ocurriendo en el proceso desde el punto de vista físico-químico y organoléptico, y (2) poder utilizar la información adquirida en estructuras de control automático de nivel local (subprocesos), global (proceso) o incluso en la toma de decisiones sobre el modo de operación de la planta como es el caso de la planificación de la producción.

Abstract

In recent years the production of virgin olive oil has experienced positive growth from qualitative and quantitative points of view, thanks to the transference of knowledge from the research centres to the production plant. However, the degree of automation of the virgin olive oil production process is at an earlier stage than other agronomic industries. In this context, the work developed in this doctoral thesis is seeking to deepen the application of two non-invasive technologies, near infrared spectroscopy and computer vision, in order to acquire information at process level, in different phases, and fundamentally with two specific purposes: (1) to assess in a qualitative form what is happening in the process from a physicochemical and organoleptic point of view, and (2) to use the acquired information in order to feed the automatic control structures at local level (sub-process), global level (process) or even in the decision making process about the operational mode of the production plant — i.e. the production planning —.

Contents

<i>List of Figures</i>	V
<i>List of Tables</i>	IX
<i>Resumen</i>	XI
<i>Abstract</i>	XIII
1. Introduction	1
1.1. Scientific context	5
1.2. Hypothesis and objectives	7
1.3. Contributions	8
1.4. Document structure	10
2. Bibliographical review	13
2.1. Introduction	13
2.2. Olive oil quality	16
2.3. Virgin olive oil production process	18
2.4. Infrared spectroscopy	23
2.4.1. Principles of operation	23
2.4.2. Research based on near infrared spectroscopy over olive fruits	30
2.4.3. Near infrared spectroscopy applied to olive oil	33
2.4.4. Near infrared spectroscopy related to olive pastes and pomaces	38
2.5. Computer vision	48
2.5.1. Principles of operation	48
2.5.2. Computer vision applied to olive fruit	50
2.5.3. Computer vision applied to olive oil	54
3. System based on hyperspectral information to estimate acidity, moisture and peroxides in olive oil samples	61
3.1. Introduction	61
3.2. Materials and methods	63
3.2.1. Olive oil samples	63
3.2.2. Analytical methods	63
3.2.3. Experimental setup	63

3.2.4. Hyperspectral images preprocessing	65
3.2.5. Wavelength selection methods	67
3.3. Results and discussion	68
3.4. Conclusions and future works	71
4. Non-invasive methodology to estimate polyphenol content in EVOO based on NIR and wavelength selection	73
4.1. Introduction	73
4.2. Problem description and earlier works	75
4.3. Materials and methods	76
4.3.1. Extra virgin olive oil samples	77
4.3.2. Experimental setup	78
4.3.3. Spectra acquisition procedure	79
4.3.4. Spectra preprocessing	81
4.3.5. Spectra filtering	83
4.3.6. Dimensionality reduction with regression models	84
4.4. Experimental results and discussions	86
4.5. Conclusions and future works	90
5. Expert system for monitoring the malaxing state of the olive paste based on computer vision	93
5.1. Introduction	93
5.2. Problem description	95
5.3. Materials and methods	97
5.3.1. Experimental setup	97
5.3.2. Viscosity determination	98
5.3.3. Floating oil determination	100
5.3.4. Olive paste granularity determination	100
5.3.5. Fuzzy controller design	102
5.4. Experimental results and discussions	105
5.5. Conclusions and future works	108
6. Sorting olive batches for the milling process using image processing	111
6.1. Introduction	111
6.2. Materials and methods	113
6.2.1. Experimental set-up	114
6.2.2. Olive fruit samples	117
6.2.3. Image analysis	117
6.2.4. Feature vector	119
6.2.5. Fisher Discriminant Analysis	120
6.2.6. Artificial Neuronal Network	121
6.3. Experimental results and discussions	122
6.4. Conclusions and future works	127
7. Results and discussion	129
8. Conclusions and future works	133
8.1. Summary of conclusions	134
8.2. Future researches	136

Appendix A. Documentación en castellano	137
A.1. Título	137
A.2. Índice	137
A.3. Introducción	140
A.3.1. Contexto científico	144
A.3.2. Hipótesis y objetivos	145
A.3.3. Contribuciones	147
A.3.4. Estructura del documento	150
A.4. Resumen	151
A.5. Conclusiones generales y líneas futuras	151
A.5.1. Resumen de conclusiones	152
A.5.2. Líneas futuras de investigación	154
<i>Bibliography</i>	157
<i>Glossary</i>	179

1 Introduction

Contents

1.1. Scientific context	5
1.2. Hypothesis and objectives	7
1.3. Contributions	8
1.4. Document structure	10

The production of virgin olive oil is an important economic activity carried out in more than thirty countries and it has experimented a significant evolution during the last years from the quality and quantity points of view [105]. Two-phases extraction systems have been extensively installed and it has allowed this industry to increase the production capacities and, at the same time, reduce the personnel needed for the plant operation. As indicative data, the quantity of olive oil produced in the last campaign was higher than 2.4 millions of tons [105]. The research works, the transference of good practices and the necessity of producing with quality in order to put in value the product, have increased the middle level of the olive oil quality [189]. Furthermore, in a growing competition and demanding market as regards price and quality of the virgin olive oil, the use of technology in the production process could be useful as a competitive advantage over the competitors. The automation and control systems applied to olive mills can be considered as one of the leading technological progresses which confer a competitive advantage to the production plant [43].

Automating processes involve the automatic control of the main parameters which modulate the operation of the process [175]. This fact can be achieved by the installation of sensors and actuators into the production plant, as well as control devices and data acquisition systems. Sensors are useful to measure the process variables at any time and to send electronically this information to the control

devices. In the case of Virgin Olive Oil Production Process (VOOPP) automation these sensors are typically temperature probes and flow meters to determinate parameters such as temperature and flow of different products (olive paste, water, olive oil and pomace), and ammeters and encoders in order to measure the power consumption and the rotating speed of the engines which are involved in the oil mill. Therefore, VOOPP is a complex process composed by different interconnected stages and it offers a wide range of possibilities for automatic control. Figure 1.1 and Figure 1.2 summarise these choices for each step of the process.



Figure 1.1 Possibilities for the automatic control of the virgin olive oil production process according to the different sub-processes of the production process; see Chapter 2 for more information.

On the other hand, at scientific level national and international research groups have evaluated the feasibility of using non invasive technologies to estimate olive and oil properties. Recently, the most studied technologies are Near Infrared Spectroscopy (NIRS), computer vision, electronic noses (e-nose) and electronic tongues (e-tongue). Whereas the applications based on computer vision have been mainly focused on characterising olive fresh by means of detecting external defects over the olive skin and assessing the maturity index of olive batches, the rest of applications have been employed to detect adulterated oil mixed with olive oil, predict the geographical origin and estimate chemical and organoleptic parameters related to the olive oil quality.

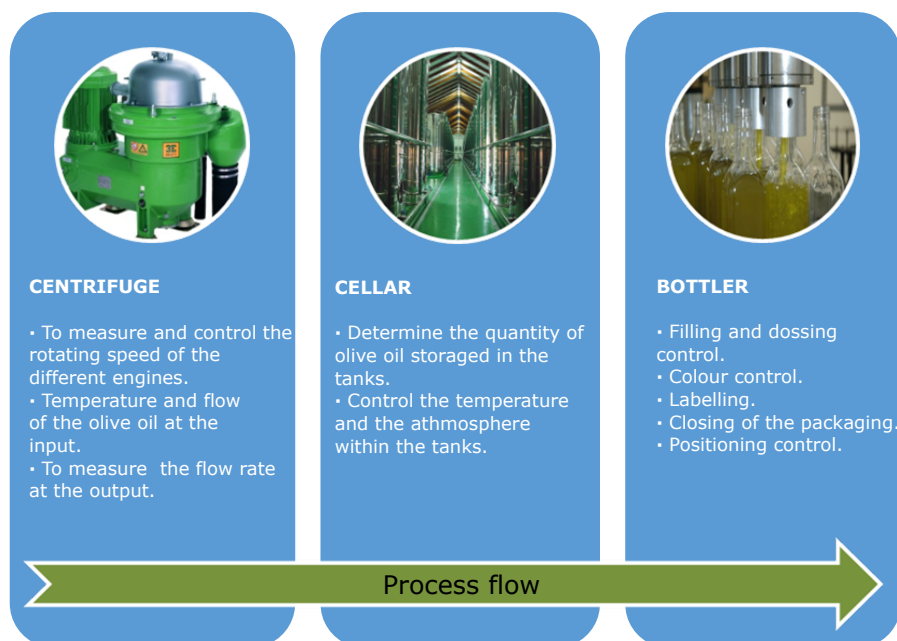


Figure 1.2 Possibilities for the automatic control of the virgin olive oil production process (continuation).

However, in addition to the sensors, it is necessary to install electronically controlled actuators in order to modify the operation mode of the process. Usually, in the VOOPP these actuators are electro valves which directly regulate different flows and indirectly they are use to modify the temperature different fluids. Also, there are variable frequency drivers to regulate the rotating speed of the engines and pumps.

Sensors and actuators are controlled by a control software installed in the programmable control system (PLC) or personal computer (PC), and it allows the system set the operation parameters in an automatic or manual form. According to the information acquired by the sensors and the consigned set points, this program computes the rotating speed of the engines and the electro valves state in order to reach the desired operation mode.

The automatic control could be interesting to reach the objective of operation. This is a weighted combination between different objectives such as yield, quality and energy efficiency. Figure 1.3 shows the relation among these objectives. Normally, it is not possible to reach the maximum value for each objective because the presence of inverse linkages. For example, in general terms it is not possible to obtain at the same time an olive oil with the highest quality level attainable and with the highest

extraction performance. The works presented in this doctoral thesis are focused on determining in line the real operation point of the process.

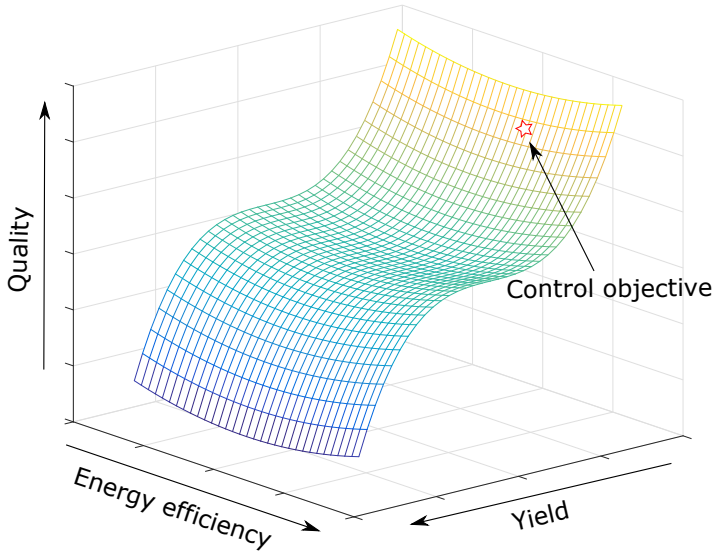


Figure 1.3 Linkage between quality, yield and energy efficiency.

Nowadays the automation level in the oil production factories is lower than in other agrofood industries [25]. In most cases, the master miller is the person with the role of taking decisions in the oil mill yard (e.g. he supervises the olive quality at the input of the process and then he selects the proper hopper) and he adjusts manually the process variables at the oil extraction facility by means of his own expertise (e.g. he regulates the water addition to the olive paste on the basis of its visual appearance). Furthermore, olives and olive oil are sampled at the input and output of the production process and the samples are analysed by external laboratories in accordance with the European Normative [66]. Usually this procedure takes a long time and then it is not possible to reconfigure, if necessary, the process variables according to the laboratory analysis (e.g. water addition modifies the polyphenol profile into the olive oil). On the other hand these laboratory techniques are tediously slow, destructive and they requires qualified personnel. At this point the non-invasive process analytical technologies could be an useful tool for acquiring information about the product within the process and feeding the upper level control loops.

The final purpose of this doctoral thesis is to provide added value to the VOOPP by

means of non-invasive sensorial technologies which allow to infer the properties of the product at its different production steps. These properties are useful to characterize the elaborated product and classify it automatically according to qualities or to take decision in order to correct deviations in the elaboration objective.

1.1 Scientific context

The studies presented in this doctoral thesis have been aligned with the research line related to the application of the automatic control to the olive growing and oil industry. This line is being developed by the Robotics, Automation and Computer Vision Research Group (GRAV) belonging to the University of Jaén, TEP-237 (grav.ujaen.es). Furthermore, the works carried out by the doctoral student has been supported by different projects. In chronological order, they are:

- **AGR-6616:** *"Modelado y optimización del proceso de elaboración de aceite de oliva virgen desde el punto de vista de la calidad, productividad y eficiencia energética desarrollando metodologías analíticas y de control avanzadas"* (Modelling and optimization of the virgin olive oil production process from the point of view of the quality, yield and energy efficiency by means of developing analytic and control advanced methodologies)
- **NUTRAOLEUM:** *"Estudio de nuevos procesos en la industrial almazarera andaluza para la elaboración de nuevos productos de alto valor biológico con aplicación en salud humana"* (New processes in the olive oil industry in order to produce new products of high biological value for human health)
- **CARDIOLIVE:** *"Desarrollo de suplementos alimenticios a partir de los productos del olivar dirigidos a la protección cardiovascular humana"* (Development of food supplements based on olive products and focused on the human cardiovascular protection)

The first project, AGR-6616, was started in 2011 by means of the Excellence Projects program funded by the Government of Andalusia. The objectives of this project were three: Characterizing the VOOPP in order to regulate the quality of the olive oil elaborated; Determining the right control reference according to the production conditions; Development and implementation of multiobjective and multivariate control strategies within the VOOPP. Since then, the GRAV resources related to this research line have increased significantly and nowadays the group has several resources available such as an experimental oil mill in GEOLIT Technological Park,

analytical non-invasive technologies (e.g. NIR device and electronic nose), vision cameras and lighting systems in the visible and infrared spectrum.

The second project, NUTRAOLEUM, was funded in 2013 by the FEDER Innterconecta program (www.cdti.es). This project covered different areas of knowledge, from the agronomic field to clinical trials. The final objective was to elaborate olive oil with high content of nutritional components. The experiments were performed with different varieties of olives and different process variables in a real olive mill. The olive oil with the highest content of the studied components were introduced into the diet of selected people to evaluate its influence on the metabolic syndrome.

The third project, CARDIOLIVE, was started in 2015 also funded by the FEDER Innterconecta program (www.cdti.es) and it is currently in progress. It has two objectives: The first one is to produce a nutraceutical products with the main nutritional components extracted from the olive oil; The second objective is to research new extraction systems in order to enhance the presence of these components.

On the other hand, the experimentation works performed to develop the studies presented in this doctoral thesis have been partly carried out by means of collaborative agreements between GRAV and different oil mills in the province of Jaén (eg, Almazara San Bartolomé from Oleocampo group, Torredelcampo, Jaén) and also with laboratories for chemical analysis of oil (eg, CM Europa, Martos, Jaén). Moreover, in the context of the aforementioned projects, different technological transference contracts were signed between GRAV and the company GEA Westfalia Separator Ibérica S.A. In addition, they have allowed to perform different experimentation works with the aims of validating all the technological approaches presented by this doctoral thesis in the olive oil cooperative San Francisco de Asís S.C.A. (Montefrío, Granada).



Figure 1.4 Logotype of the projects.

1.2 Hypothesis and objectives

Within the context of the agro-alimentary sector related to the virgin olive oil extraction process, the central hypothesis of this doctoral thesis is the introduction of non-invasive and advanced sensors into the process in order to know the elaboration point and to improve the VOOPP in terms of quality, quantity and nutritional value. This, together with automatic control technologies, will entail a scientific advance for the production of VOO.

On that assumption, the research studies carried out in this doctoral thesis have been focused on four main objectives; they are, in turn, divided into specific objectives. These objectives have been enumerated as follows:

1. Evaluating the feasibility of using hyperspectral imaging as automatic and non-contact technology for the in line characterization and quality control of olive oil.
 - a) Development of predictive methodologies in order to assess the quality of the olive oil rapidly and in a non-destructively way by means of the estimation of moisture, peroxide index and free acidity according to hyperspectral information.
 - b) Determination of the optimal wavelengths which are better correlated with the considered olive oil parameters.
2. Development of predictive models based on Near Infrared Spectroscopy to automatically determine the concentration of polyphenol content in olive oil samples.
 - a) Evaluation of two approaches of wavelength selection with the objective of minimizing the error of prediction identifying the most correlated bands with polyphenol content.
 - b) Determining how the number of acquisition per sample and the lighting system affect the prediction results.
3. Assisting to the master miller in order to regulate the technological variables which are involved within the malaxing process.
 - a) Developing a software sensor to assess the malaxing state of the olive paste according to different parameters such as floating oil, granularity and viscosity.
 - b) Assisting the master miller to characterize the olive paste continuously and recommending set points according to the olive paste malaxing state and

the master miller knowledge.

4. Characterizing olive fruits at the input of the olive oil production process by means of computer vision.
 - a) Extracting and identifying discriminant features between sound and defective olives.
 - b) Classifying olive batches according to the harvesting method automatically.

1.3 Contributions

This section shows the works carried out according to the objectives of this doctoral thesis. They have been presented as contributions in journals together with national and international conferences. These contributions can be organized into two main blocks. The block number one contains the contributions where the non-invasive infrared spectroscopy technology have been evaluated to determine the operation state of the process. Consecutively, the second block shows the contributions related to the results obtained by means of the use of computer vision as non-invasive technology within the VOOPP.

1. Contributions related to the application of non invasive technologies in order to determine automatically olive oil quality parameters, polyphenol content, moisture and presence of impurities:
 - [138] **Martínez Gila, D.M.**, Cano Marchal, P., Gámez García, J., Gómez Ortega, J., *On-line system based on hyperspectral information to estimate acidity, moisture and peroxides in olive oil samples*. In Journal of Computers and Electronics in Agriculture, 2015, 116:1–7.
DOI: 10.1016/j.compag.2015.06.002
 - [133] Cano Marchal, P., **Martínez Gila, D.M.**, Gámez García, J., Gómez Ortega, J., *Expert system based on computer vision to estimate the content of impurities in olive oil samples*. In Journal of Food Engineering, Elsevier Ltd, 2013, 119(2):220–228.
DOI: 10.1016/j.jfoodeng.2013.05.032
 - [154] Beltran Ortega, J., **Martínez Gila, D.M.**, Aguilera Puerto, D., Gámez García, J., Gómez Ortega, J., *Novel Technologies for Monitoring the In-Line Quality of Virgin Olive Oil during Manufacturing and Storage*. Journal of Science in Food and Agriculture, 2016. Unedited article.
DOI: 10.1002/jsfa.7733

- [140] **Martínez Gila, D.M.**, Cano Marchal, P.C., Gámez García, J., Gómez Ortega, J., *Non-invasive methodology to estimate polyphenol content in EVOO based on Stepwise Multilinear Regression*. In Journal of Computers and Electronics in Agriculture, 2016, Under review.
 - [81] **Martínez Gila, D.M.**, Cano Marchal, P., Gámez García, J., Gómez Ortega, J., *Hyperspectral imaging for determination of some quality parameters for olive oil*. In Proceedings of IEEE 18 International Conference on Automation and Computing (ICAC), 2012.
 - [80] **Martínez Gila, D.M.**, Beltran Ortega, J., Gámez García, J., Gómez Ortega, J., *Rapid quantification of total polyphenol content in EVOO using NIR sensor with wavelength selection and FS-MLR*. In Proceedings of IEEE International Conference on Imaging Systems and Techniques (IST), 2015. DOI: 10.1109/IST.2015.7294542
 - [134] Cano Marchal, P., **Martínez Gila, D.M.**, Gámez García, J., Gómez Ortega, J., *Application of computer vision and Support Vector Machines to estimate the content of impurities in olive oil samples*. In Proceedings of IEEE 18 International Conference on Automation and Computing (ICAC), 2012.
 - [139] **Martínez Gila, D.M.**, Beltran Ortega, J., Gámez García, J., Gómez Ortega, J., *Tecnología no invasiva para la determinación automática del contenido total de polifenoles en el aceite de oliva virgen*. In Proceedings of XXXVI Jornadas de Automática, 2015.
2. Contributions related to the use of computer vision to characterize and classify olive fruit at the input of the production process and control the malaxing state of the olive paste at the kneading step:
- [160] Aguilera Puerto, D., **Martínez Gila, D.M.**, Gámez García, J., Gómez Ortega, J., *Sorting Olive Batches for the Milling Process Using Image Processing*. In Journal of Sensors, 2015, 15(7):15738–15754. DOI: 10.3390/s150715738
 - [82] **Martínez Gila, D.M.**, Cano Marchal, P., Gámez García, J., Gómez Ortega, J., *Control System of the Malaxing State for the Olive Paste Based on Computer Vision and Fuzzy Logic*. In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2013, pp. 1951–1956. DOI: 10.1109/SMC.2013.335
 - [137] **Martínez Gila, D.M.**, Cano Marchal, P., Gámez García, J., Gómez Ortega, J., *Extractability control for the virgin olive oil elaboration process*.

In Proceedings of IEEE International Conference on Industrial Technology (ICIT), 2015, pp. 240–245.

DOI: 10.1109/ICIT.2015.7125105

- [136] **Martínez Gila, D.M.**, Aguilera Puerto, D., Gámez García, J., Gómez Ortega, J., *Automatic classification of olives for oil production using computer vision*. In Proceedings of IEEE International Conference on Industrial Technology (ICIT), 2015, pp. 1651–1656.
DOI: 10.1109/ICIT.2015.7125334
- [149] Cáceres Moreno, O., **Martínez Gila, D.M.**, Aguilera Puerto, D., Gámez García, J., Gómez Ortega, J., *Automatic determination of peroxides and acidity of olive oil using machine vision in olive fruits before milling process*. In Proceedings of IEEE International Conference on Imaging Systems and Techniques (IST), 2015.
DOI: 10.1109/IST.2015.7294543
- [141] **Martínez Gila, D.M.**, Cano Marchal, P., Gámez García, J., Gómez Ortega, J., *Determinación del estado de batido de la pasta de aceituna empleando visión por computador*. In Proceedings of XXXIV Jornadas de Automática, 2013, pp. 826–33.
- [159] Aguilera Puerto, D., Cáceres Moreno, O., **Martínez Gila, D.M.**, Gámez García, J., Gómez Ortega, J., *Olive batches automatic classification in mill reception using computer vision*. In Proceedings of IEEE International Conference on Imaging Systems and Techniques (IST), 2015.
DOI: 10.1109/IST.2015.7294544
- [1] Aguilera Puerto, D., Cáceres Moreno, O., **Martínez Gila, D.M.**, Gámez García, J., Gómez Ortega, J., *Clasificación automática de lotes de aceitunas en almazara mediante visión por computador*. In Proceedings of XXXVI Jornadas de Automática, 2015, pp. 59–65.

1.4 Document structure

The present doctoral thesis has been structured in the format of a compendium of chapters as follows:

- **Chapter I** is the introduction chapter. It presents the context where this thesis has been developed, the objectives to be reached with the development of this work, the supporting contributions and the structure of the document.
- **Chapter II** is a complete bibliographical review. It describes the main olive oil

quality parameters and presents the different steps of the production process. Moreover, it contains a complete revision of the literature related to the use of non-invasive technologies within the olive oil production process.

- **Chapter III** exposes the feasibility of using a system based on hyperspectral information to estimate free acidity, peroxide index and moisture content in olive oil.
- **Chapter IV** shows the methodology employed to estimate polyphenol content in olive oil samples based on FT-NIR and wavelength selection.
- **Chapter V** presents the procedures carried out for monitoring the malaxing state of the olive paste and for adjusting the technological variables according to the olive paste appearance.
- **Chapter VI** describes the system developed for automatically sorting olives batches before the milling process and according to the harvesting method (soil or tree).
- **Chapter VII** presents a general discussion of the results obtained in the studies.
- Finally, **Chapter VIII** contains the conclusions which has been obtained from each research works and it lists different lines as future works.

2 Bibliographical review

Contents

2.1. Introduction	13
2.2. Olive oil quality	16
2.3. Virgin olive oil production process	18
2.4. Infrared spectroscopy	23
2.5. Computer vision	48

2.1 Introduction

Virgin olive oil is the juice of olive fruit obtained exclusively by mechanical or physical processes under controlled conditions. It is a product well valued by the consumers due to its exceptional organoleptic and health properties. The production of virgin olive oil is an important economic activity and originally it was developed in Mediterranean countries, with Spain as the largest producer (43.5% of world production in the last campaign). Nowadays it has been extended globally to the point that the total worldwide production of olive oil in 1993/1994 was 1825.0 thousand tons and twenty years later, in 2013/2014, the production expanded to 3270.5 thousand tons, according to the final balance published by International Olive Council [105]. This fact means a significant increase in production and it represents the importance of this sector.

The quality parameters of the olive oil are recommended by the International Olive Council and different countries employ these recommendations to draft their



Figure 2.1 Olive fruits reception in an industrial plant of olive oil production.

regulations. For example the European Normative number 1348/2013 [66] defines the values of the physical-chemical and the organoleptic olive oil parameters to classify it by quality, as well as methods of assessing these parameters.

The challenge of the olive oil industry is to reach the optimal quality of the olive oil in order to be more competitive in the international market. However, the reachable optimal quality is not the same during the harvest campaign and it is directly or indirectly related to the state of the olive fruits (maturity index, harvesting conditions, diseases, ...) and the process parameters (kneading time, temperature, ...) [53] [12].

Nowadays the values of the process variables are regulated according to the master miller knowledge and sometimes this person uses low-level control loops to adjust automatically these variables (e.g. the temperature in the thermomixer with a basic PID controller). Normally the quality assessment of the olive oil produced is carried out off-line by external chemical laboratories and the master miller gets the results with high latency. Therefore the automatic control of the VOOPP is at an early stage. This conclusion was reported by Cano Marchal et al. in 2011 [25].

One of the main olive oil research lines of the scientific community is to supply automatic systems to adjust the process variables at real time according to the objective of production [26]. It could be focused at same time on the quantity of olive oil produced, on the olive oil quality, on the batch traceability and on reducing the production costs. Therefore this is not an easy task and it relies on employing

the Process Analytical Technologies (PAT) principle.

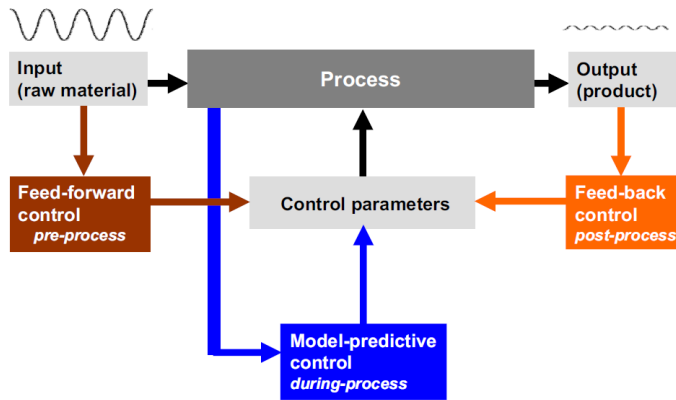


Figure 2.2 Schematic overview of process control strategies in food manufacturing (obtained from the reference [191]).

This principle was introduced originally in 2004 for the pharmaceutical industry and recently it has been adopted by the food industry [191]. Its approach involves multivariate data collection through in-line and on-line analytical techniques (such as near infrared spectroscopy, chemical sensors, ...) coupled to multivariate data analysis (chemometrics) for continuous learning and information [181]. It was defined by United States Food and Drug Administration (FDA) as systems for analysis and control of manufacturing processes based on real-time process monitoring of critical quality parameters and performance attributes of raw materials and in-process products, to assure acceptable end-product quality at the completion of the process [69]. In 2012, van den Berg et al. [191] defined the PAT concept as the principle that allows the change from post-problem or feed-back process control - where the end-product is tested and the process adjusted accordingly - to during-problem or model-predictive process control - where the larger variations in raw materials are compensated for by process adjustments during manufacturing to obtain the in-spec end-product (Figure 2.2). In 2015, Tajammal et al. [187] defined the current status of this approach applied to different process industries. Also authors focused on the dairy industry.

PAT principle includes the use of different non-invasive technologies that can be sorted according to their physical basis. Mainly two categories appear in the literature. While the first group is based on the quantity of electromagnetic radiation that is absorbed by the sample, the second category includes sensors that present

chemical affinity for certain molecular compounds.

The objective of this chapter is to review the state of the art of different technologies which have been employed in this PhD Thesis, such as infrared spectroscopy and computer vision applied to the virgin olive oil production process. It summarises the principal concepts and gathers the main research works that have appeared during recent years which have employed these technologies to monitor different parameters in the olive oil industry.

The chapter has been mainly organised in four sections. Section 2.2 shows the normalized parameters which determine the olive oil quality; Section 2.3 presents the different VOOPP phases; Section 2.4 deals with the infrared spectroscopy applied to VOOPP; Finally, Section 2.5 is focused on the use of computer vision in the same production process. At a secondary level, the second and third sections has been reclassified according to the virgin olive oil production phase where each referenced work is contextualized.

2.2 Olive oil quality

The quality of a product can be defined as the sum of features which allow it to be classified as better, equal or worse than other products of the same species. Particularly, the olive oil quality is determined by means of different values and methods according to the Regulation No 1348/2013 of the European Union (EU) Commission [66]. The main quality parameters are based on organoleptic features (sensory profile) and physicochemical parameters, such as acidity value, peroxide value and ultraviolet light absorption, among others [30].

- Free acidity degree

This chemical parameter is related to the quantity of free fatty acids contained in the olive oil. It is presented as grams of oleic acid per 100 grams of oil. Normally, sound olive fruits have 0% of free acidity. Then the percentage of free fatty acids depend on extrinsic factors such as plagues, delays between harvesting and extraction, bruised and damaged fruits, fungal diseases, maturity or adverse conditions during olive oil storage (Figure 2.3). Extra virgin olive oils have less than 0.8% of free acidity.

- Peroxide index

Peroxides are the primary products of oxidation of olive oil. These products

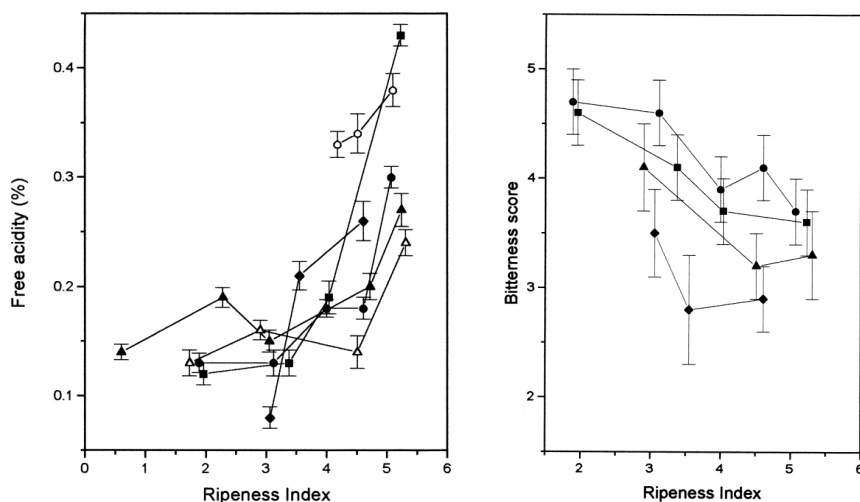


Figure 2.3 Changes in free acidity and bitterness of Cornicabra virgin olive oil extracted at different stages of olive maturity (obtained from the reference [172]).

appear when fats and oils come in contact with oxygen (auto-oxidation or photo-oxidation). The organoleptic defects could be increased by the oxidation products and they may adversely affect the nutritional value of the oil. Peroxide value is expressed as milliequivalents (m.e.q.) of active oxygen per kilogram of oil. In order to be extra virgin, olive oil must have less than 20 m.e.q. $O_2 \text{ kg}^{-1}$.

- Ultraviolet absorption

It is a spectrophotometric test which measures the absorption under ultraviolet light. The absorption is expressed as K (extinction coefficient) for the specified wavelength. K_{232} and K_{270} are measured at 232 and 270 nm respectively. Moreover, 264 and 274 nm are measured to calculate ΔK . It is an indicator of secondary oxidation products, especially in oils that have been refined.

- Sensory profile

It describes a set of organoleptic properties which can be identified and assessed by the human senses (odour, flavour and colour). These properties are strongly connected with the variety of olive fruits and the ripening index (Figure 2.3). The organoleptic assessment is carried out by normalized tasting panels.

2.3 Virgin olive oil production process

A modern virgin olive oil factory usually has three different zones which are the reception yard, the extraction facility and the oil cellar. The virgin olive oil production process starts within the reception yard where the farmers discharge their olive fruit batches into hoppers (Figure 2.4). Then the olives are conveyed to the weighting machine in order to give back to the farmer. Also the olives can be cleaned and/or washed depending on the state of the olive batch. After that the olives are crushed by the milling machine and the resulting paste is transported by pipelines to the extraction facility. Then the olive paste is preprocessed within the malaxing process. Later, the paste is pumped to the horizontal centrifugal machine where the olive oil is extracted from the olive paste. The remaining impurities are removed by the vertical centrifugal machine and finally the olive oil is stored into stainless steel tanks. Figure 2.5 shows the process flow.



Figure 2.4 Farmer discharging his olive fruits into the hopper.

The agronomical factors are directly correlated with the highest quality attainable of the olive oil produced. They are divided in intrinsic factors (variety or climatic conditions) and extrinsic factors, which can be controlled by the farmer (crop maintenance and the method or the moment to harvest) [147]. The most extended classification process is carried out at the reception yard where the operator classifies the olive batches according to the harvesting method (soil or tree).

However an automatic classification system according to different extrinsic factors is desirable to modulate the quality of the olive oil produced.

Olives are separated from leaves and branches through a blowing process before the weighting phase. Normally, if the olives come from the soil they are washed in order to remove pebbles and mud. Closed-loop washing system is not desirable due to the risk of microbiological cross-contamination and it decreases the grade of bitter, pungent and fruity attributes in final oils [194]. Finally, the olives are stored in big hoppers before milling. It is advisable to adjust the production rate in order to reduce the storage time. Long times could affect negatively to the oil quality [95] [195] [75] [116] [103].

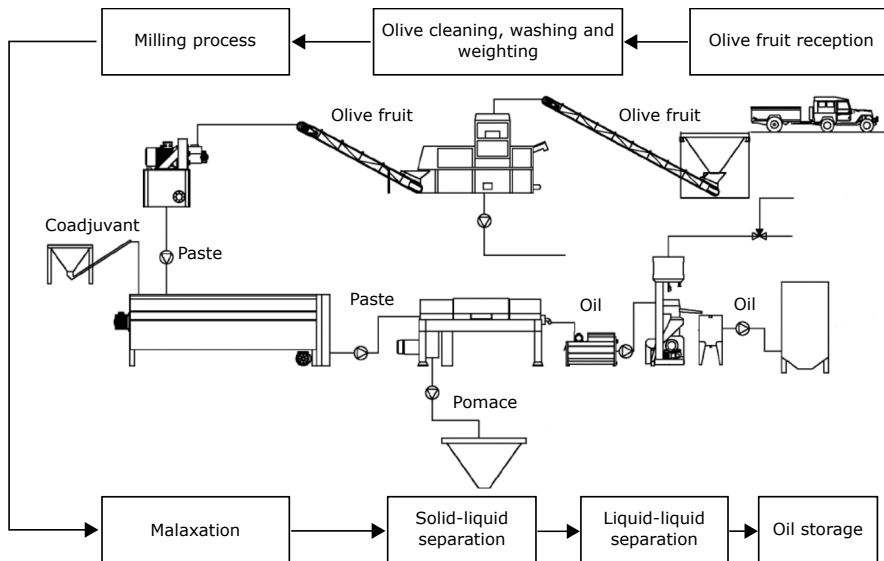


Figure 2.5 Virgin olive oil production process from olive reception to oil storage.

The olive crushing is the process through which the olive fruit turn into olive paste and it releases the oil droplets contained into the vegetable cell vacuoles. There are two main topologies of mills, stone mills and metal crushers. The first one was the standard model until the appearance of metal crushers. It has certain advantages such as the dilacerated effect that makes easier the malaxing process and it does not transmit heat to the paste. However it takes up a lot of space and the production performance is not very high. On the other hand, metal crushers have been engineered for higher productions, space saving and lower residence times to avoid fermentation processes [29]. Moreover, they usually provide fixed or mobile

grids with different hole diameters in order to regulate the pit particle size and the oil drop distribution in olive paste [125]. Different types of olive crushers are currently available, such as hammer mills, toothed crushers or blade cutters, and the main technological parameters involved are the grid hole size and the rotation speed. Furthermore, the use of the different crushing conditions may be useful to modulate the Virgin Olive Oil (VOO) content in minor compounds related to its overall quality [102] [27] [193].

The malaxing process is the second phase into the extraction facilities. It consists on prepare the paste before the extraction stage with the objective of achieving satisfactory yields of extraction. This process occurs inside the thermomixer and it favours the quantity of free oil by means of helping the small droplets of the oil formed during the milling phase to merge into large drops that can be easily separated through mechanical systems. Also it completes the crushing step by disrupting the oily cells which remain uncrushed. Typically this process takes place inside heated tanks with revolving shovels known as thermomixer (Figure 2.6), however other technologies have been recently studied in order to prepare the olive paste such as microwaves and ultrasounds [46]. Different technological parameters of operation can be controlled in the thermomixer, such as the time and temperature of malaxation, the speed of the shovels, the employment of technological coadjuvants and lately, the atmosphere of malaxation.



Figure 2.6 Picture of the inside of the thermomixer.

Furthermore, plenty of works have studied the linkage between the aforementioned parameters and the quality and quantity of the olive oil produced. Generally, increasing the malaxing time improves the oil extraction yield to a point where the quantity of oil extracted is constant (60 or 75 min) [163]. Also, whereas the polyphenol content increases with the malaxing time, the commercial qualitative parameters such as free fatty acids or peroxide values remain invariant. On the other hand, the malaxing temperature is directly correlated with the extraction yield up to 40°C, when the interaction between lipids, proteins and carbohydrates can culminate in the entrapment of olive in the olive paste, which result in decreased yields [115]. Moreover, there are no significant correlation between the temperature of malaxation and the free acidity, peroxide value and ultraviolet absorption characteristics [104], however the volatile compounds are negatively affected [5] and the phenolic compounds are positively correlated [104]. The employment of technological coadjuvants favours the oil extraction from difficult pastes. The most frequently employed coadjuvants are lukewarm water, talc and salt, and only the first one can affect negatively to the polyphenol level [32] [45]. Finally, the application of automatic systems to control the oxygen concentration during the malaxation process can extend the malaxing time without damaging the resulting oil and it can improve the organoleptic properties of the oil [173] [36].



Figure 2.7 Horizontal decanter machine.

After the malaxation process, the olive oil is separated from the olive paste by means of the decanting physical procedure. This process is continuously carried out

inside the decanter machine (Figure 2.7), that is a horizontal centrifugal machine which reduces the decanting time. Three-phase centrifugal decanters and two-phase centrifugal decanters are used in VOO industry. Whereas three phase decanters separate dry pomace, water and oil by injecting water, two-phase decanter separate oil and wet pomace and they can work without water addition. Different works have proven that the acidity and Ultraviolet (UV) indices are not affected by the number of phases, however, differences were observed in their antioxidant content and functional properties: compared to three-phase, olive oil produced by two-phase decanter exhibited significantly lower peroxide value, and significantly higher total phenolics [52] [107] [114]. Normally, the horizontal decanter has two concentric and mobile parts: the external drum and the internal endless screw [29]. Other works have evaluated the linkage between the quality of the extracted virgin olive oil and the differential speed between drum and endless screw. The obtained results showed that increasing the differential speed, the oils obtained were more pungent and richer in antioxidant compounds, whereas the oils extracted at lower drum speeds were more bitter and showed higher peroxide and UV absorption values and higher amounts of volatile compounds [28].

Typically, the extracted olive oil contains vegetable water and solid impurities that need to be removed before the storage process. This action is performed by the vertical centrifugal machine at high speeds with the drawback of possible oil heating during the residence time and marked loss of volatile aromatic compounds [142].



Figure 2.8 Olive oil cellar.

Finally the olive oil is stored in a cellar inside stainless steel tanks (Figure 2.8) before the bottling process. A cylindrical shape with a conical base is desirable in order to purge decanted solids and water. Furthermore, it is important to control main oxidation factors during storage, such as air contact, temperature and light [132]. Virgin olive oils stored at low temperatures maintained better quality parameters than oils stored at room temperature [22].

2.4 Infrared spectroscopy

Infrared vibrational spectroscopy is the method that allows to analyse the chemical properties of solid, liquid and gaseous samples of materials by describing the energy transfer between the incident light and matter. On the other hand, most cited infrared spectroscopic techniques into the literature are NIRS, Mid Infrared Spectroscopy (MIRS) and Raman [176]. The appearance of works dealing with the application of these technologies into the olive oil elaboration process has been increased lately. The majority of these papers have been published in the last five years showing the novelty of these techniques.

Their basic physical origin is the same, that is, spectra of chemical compounds can be observed as a consequence of molecular movements and vibrations. However, while Raman spectroscopy is a scattering technique, NIRS and MIRS are based on the absorption of electromagnetic radiation in the spectral bands 780-2500 nm and 2500-7500 nm respectively [144]. Usually NIRS and MIRS techniques operate with polychromatic light sources for the frequency range of interest. In Raman spectroscopy the sample is irradiated with monochromatic laser light whose frequency may vary from the Visible (VIS) to the NIR region. Then, the scattering happened on the sample is detected by the NIR sensor. All three techniques allow qualitative, quantitative non-invasive and non-destructive analysis without the addition of auxiliary chemical compounds. For all these reason, spectroscopic techniques are ideally for industrial quality control and process monitoring.

2.4.1 Principles of operation

NIRS and MIRS techniques are based on molecular absorption principle; briefly, when incoming photons of the incident light strike different molecules in a sample, two results may occur: the frequency of the incident light does not match one or more of the natural vibrational frequencies of the molecules into the sample, or the frequency does match one or more of the vibrational frequencies of the molecules

[184]. When there is a match between the frequency of the illumination energy and the natural vibrational frequency of a molecule in the sample, the molecule absorbs this energy, which in turn increases the vibrational amplitude of the absorbing dipoles.

Molecular bonds are fundamentally X-H, where X is the atom of carbon, nitrogen or oxygen and the molecular absorption are caused by the variations of the length of the bond (stretching movements) and variation or change in the angle of the bond (bending movements) [177] (Figure 2.9).

The energy and frequency of fundamental absorption bands can be predicted with relative accuracy by using the simple harmonic oscillator model. However it not predict band positions exactly because bonds are not true harmonic oscillators [146]. Fundamental frequencies of vibration occur into the Mid Infrared (MIR) spectral range and overtones, whereas combinations of the fundamental tones occur into de NIR spectral range [202]. Although NIR absorption bands are weaker than the corresponding fundamental MIR ones, this fact is an analytical advantage since it allows direct analysis of strongly absorbing matrices such as olive pastes and pomaces.

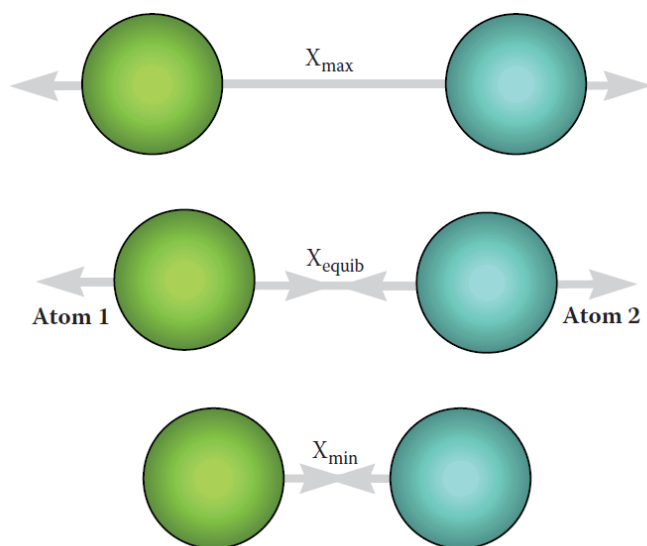


Figure 2.9 Model of infrared active molecule as a vibrating dipole between two atoms (From [202]).

The amplitude of the absorption at any particular wavelength depends on the

number of molecules encountered within the beam path of the measuring instrument and the molecular absorptivity. The relationship was described by Bouguer, Lambert and Beer and is known as Beer's law [200] [152]. This law describes the absorbance of an analyte at specific wavelength as the product of the molecular absorptivity for the wavelength (ϵ), the concentration (c) of the molecules in the measurement beam, and the pathlength (l) of the sample holder. It is often expressed as

$$A = \epsilon cl \quad (2.1)$$

where:

A : Absorbance.

ϵ : Molar absorptivity ($\text{liter} \cdot \text{mole}^{-1} \cdot \text{cm}^{-1}$).

c : Concentration of molecules in the spectrometer beam ($\text{mole} \cdot \text{liter}^{-1}$).

l : Pathlength of the beam (cm).

On the other hand the absorbance value can be formulated as the relationship between the transmitted or reflected energy with the incident energy for each wavelength of interest. In this case it is expressed as

$$A = -\log_{10} \left(\frac{I}{I_0} \right) \quad (2.2)$$

where:

I_0 : Initial energy incident to the sample.

I : Attenuated energy detected after sample interaction.

The absorbance values of the sample for each wavelength form the spectral response at the NIR spectral range. This response is named digital fingerprint or spectrum and it can be correlated with the physical and chemical information of the sample.

The spectrophotometer is the device used in infrared spectroscopy for the acquisition of spectra. Basically it consists of three components: light source, monochromator system and detector.

According to the location of the monochromator system with respect to the sample,

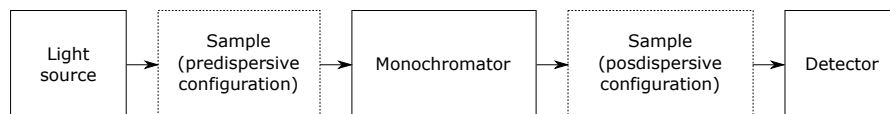


Figure 2.10 Block diagram that shows the generic structure of a NIR device.

NIR devices can be predispersive (before the sample) or posdispersive (after the sample). Figure 2.10 shows the light flow through the different components and the two aforementioned configurations.

The first component is the light source. It allows to create the electromagnetic radiation which strikes the sample. Light sources can be classified in two groups according to the bandwidth: broadband and narrowband.

The first group comprises heat sources where the light emission is reached when the current flows through a filament. Then an incandescent emission is generated. Halogen lamps with filament made of tungsten are the most common broadband illumination sources used in visible and near-infrared spectral regions and one of these has been employed in the works of this PhD Thesis (Figure 2.11).



Figure 2.11 Halogen lamp housing, from which light is delivered through an optical fiber.

The second group of light sources are solid state sources that emit light when electricity is applied to a semiconductor. They are non-heat sources and spectral band of emission depends on the semiconductor material (e.g. Indium Gallium Arsenide (InGaAs) emits up to 1600 nm). Light Emitting Diodes (LED) and Lasers can be included in this group.

Monochromator is the main component of any NIR device and it is used for dispersing broadband light (polychromatic light) into different wavelengths (monochromatic light). Nowadays there are numerous brands and models [201] of NIR devices and they are based on filter wheel encoders, diffraction gratings, Fourier Transform Near Infrared (FT-NIR), Acousto-Optic Tunable Filters (AOTF) and Liquid Crystal Tunable Filters (LCTF) [161]. The paragraphs below describe briefly the technologies which have been used in this PhD Thesis: FT-NIR and diffraction grating based technologies.

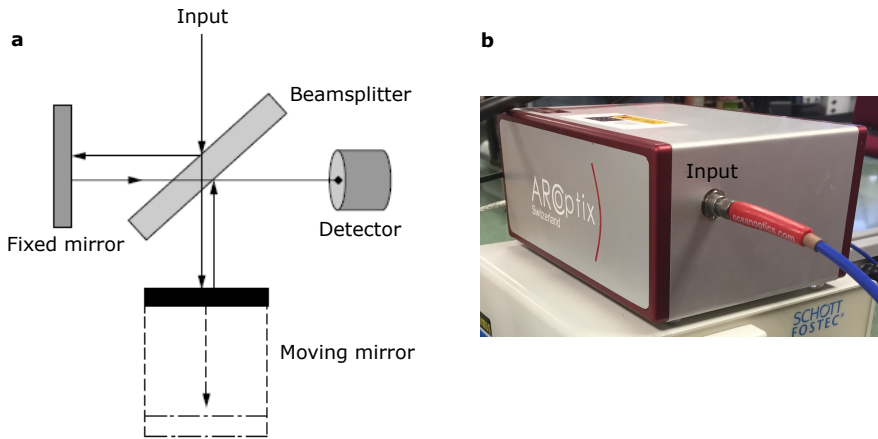


Figure 2.12 Michelson Interferometer concept (a) and NIR device based on this technology (b).

FT-NIR devices are founded on the principle of interferometry, that is, interferences between broadband lights can generate an interferogram that carries its spectral information. Then, the constituent frequencies of the broadband lights can be obtained through an inverse Fourier transform to the generated interferogram. Michelson interferometer is the widely used concept in commercial Fourier transform spectrometers in the infrared region (Figure 2.12). It consists of two perpendicularly disposed mirrors, one of which moves along its normal axis, and a beamsplitter. The beamsplitter divides the beam into two equal beams. One of them is reflected by the fixed mirror and the second one is reflected by the moving mirror. The beams reflected from the mirrors are recombined by the beamsplitter and directed to the detector [143].

On the other hand the second NIR technology employed is based on a diffraction grating. It is an array of elements separated by a distance comparable to the wavelength of the light under investigation. These elements transmit or reflect the

incident light to the detector. The physical concept is the spatial modulation of the refractive index, by which the phase and/or amplitude of the incident electromagnetic wave can be controlled [156]. Figure 2.13 shows the operating principle of this technology. The polychromatic incoming light is collimated by the first lens. Then, the diffraction grating disperses the collimated beams into different directions according to different wavelengths. Finally monochromatic beams are projected into the detector or diode array, through the final lens.

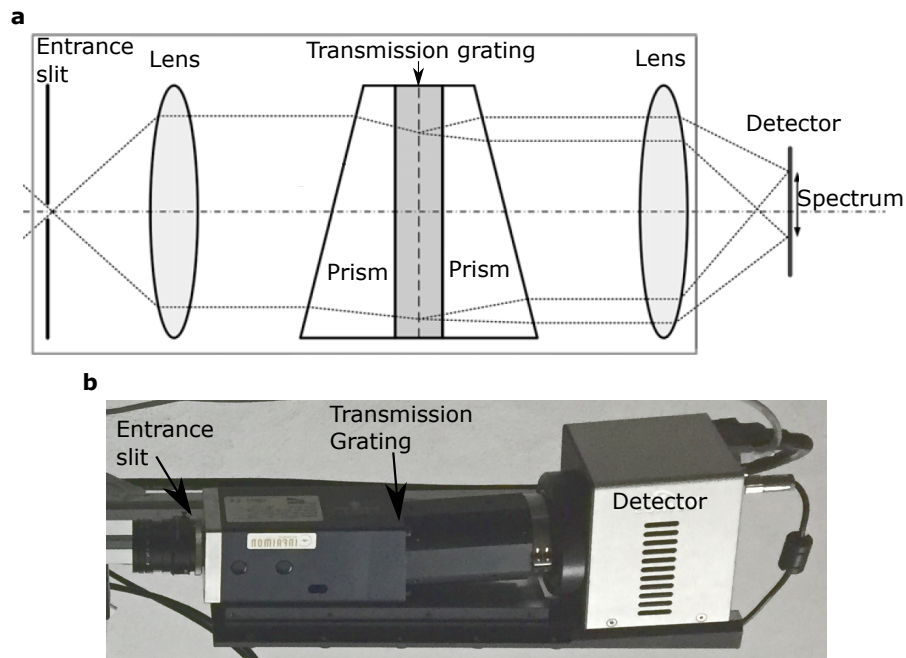


Figure 2.13 Diffraction grating concept (a) and spectrograph device based on this technology (b).

Table 2.1 NIR detectors (adapted from [143]).

Detector	Responsivity	Detectivity	Spectral range of measurement
Deuterated Triglycine Sulfate (DTGS)	Slow	Low	1250-5000 nm
Lead Selenide (PbSe)	Intermediate	Intermediate	1700-5500 nm
Lead Sulfide (PbS)	Slow	Intermediate	1100-3000 nm
Indium antimonide (InSb)	Fast	Very high	1800-6800 nm
Mercury Cadmium Telluride (MCT)	Fast	Very high	1000-17000 nm
Indium Gallium Arsenide (InGaAs)	Fast	High	900-1700 nm

NIR detectors have different features according to their source material [65]. Table 2.1 provides the most common detectors and their features. Also the number of photodiodes contained in the detector defines the NIR device operation mode. One photodiode as detector only allows to scan one point of the sample and one wavelength at the same time. However, one row of photodiodes can acquire one point of the sample but different wavelengths at the same time. Finally, the photodiode 2D array configuration acquires one line of spatial information of the sample and different wavelengths for each point of the line at the same time. This line-scanning operating mode is the basis of the hyperspectral systems. An hypercube (Figure 2.15) is formed when the sample is moved along the axis perpendicular to the acquisition line (e.g. on conveyor belt). This operation mode is also known as pushbroom method.

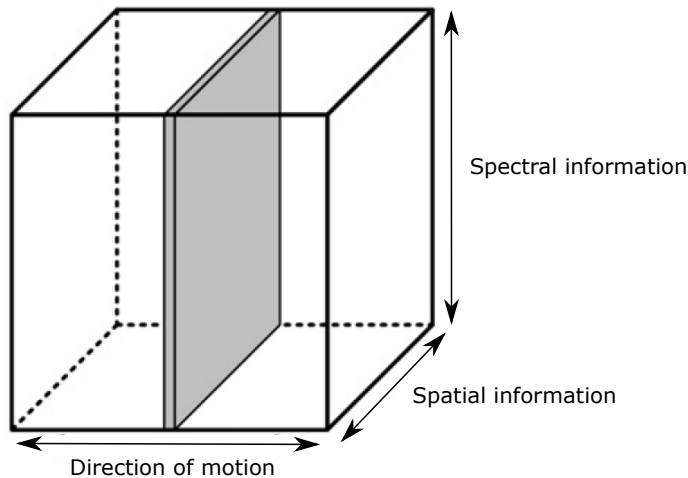


Figure 2.14 Hypercube.

Fundamentally, there are three operating modes to acquire the spectral information from the sample according to the position of detector with respect to the sample: transmittance, reflectance and transreflectance (or double transmittance) [119].

In the transmittance mode, the sample is located between the light source and the detector. The light beams passes through the sample and the light which has not been absorbed by the sample goes to the detector. This methodology is often employed with liquid and transparent or semitransparent materials. Also it can be applied to thin slices of solid samples.

Reflectance mode is employed with solid samples and it is based on acquiring the

light reflected by the sample after penetrating a few millimeters. Normally detector and light source are located in the same plane.

Transflectance mode is also known as double transmittance. In this mode the light beams passes through the sample to reach a specular surface (or reflectance pattern). The beam is reflected by the surface and it passes for a second time through the sample. Then the beam goes to the detector. This mode can be applied to liquid and semi-liquid samples.

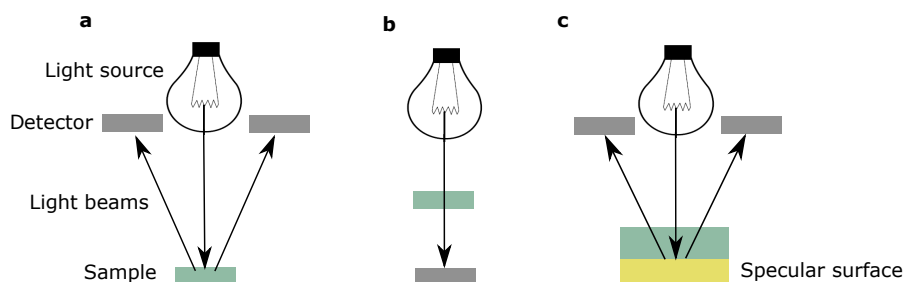


Figure 2.15 Measurement methodologies: reflectance (a), transmittance (b) and transflectance (c).

Infrared spectroscopy has been applied in the olive oil elaboration process to determinate the properties of the different products that appear in the process, such as olive fruits, slurries and olive oil. The objective of this section is to show the state of the art of the use of this technology with the aforementioned products. So it has been divided in three subsections: the first one introduces the spectroscopy techniques that appear in the literature that have been applied to olive fruits; the second talks about the infrared technologies employed with olive oil samples; finally, the last point presents the current state of this technology used with olive slurries.

Furthermore Table 2.2, Table 2.4 and Table 2.9 provide an extensive and up-to-date review (on March 2016) about the different research works related to the use of NIR technology in the olive oil industry.

2.4.2 Research based on near infrared spectroscopy over olive fruits

Infrared spectroscopy applied to olive fruit is related to the prediction of olive oil constituents at the beginning of the production process before the milling phase and on tree. The most cited in the literature is the oil content estimation on fresh olives, but there are others such as moisture, free acidity of the extracted oil or

fruit maturity index. Rapid determination of these parameters has been useful for breeding programs, assessing the optimal state of harvesting and predicting oil quality.

Breeding programs for olive trees try to combine several characteristics of different olive tree varieties in order to get specimens with the desired features (oil content, polyphenols content, resistance to diseases, etc). These programs are based on recurrent steps for several years. For each year a high number of samples must to be analysed to select the genotypes. In 2012, León Moreno [124] assessed the feasibility of portable NIR technology for the determination of oil and moisture content to be used as a decision support tool in olive breeding programs. In this work the author carried out two different experiments with samples collected in different selection steps (progenies and selection plots) of the breeding work and under different experimental conditions (on-tree and under laboratory conditions). The results show a correlation coefficient greater than 0.9 and standard errors of prediction lower than 2.21% in olive and moisture content. Similar works appear in previous research studies [123] [122].

As a general rule the harvest campaign starts when the cooperative decides and usually it begins at the same time in nearby regions. The crop and climatic conditions during fruit development strongly influence oil accumulation and determine the oil content at harvest. Determining the oil content and composition using analytical methods is time consuming and expensive. For this reason, in 2013, Bellincontro et al. [15] studied the potential use of Near Infrared Acousto-Optical Tunable Filter (NIR-AOTF) for the determination of the oil content of olive fruit during their development. They calibrated the NIR device against a Nuclear Magnetic Resonance Technique (NMR) technique and they validate the models with cross validation and with different olive varieties. In all the tests the coefficient of regression was greater than 0.90 units. Previously, similar works [118] [89] studied the NIR spectrum for the prediction of moisture and oil content in fresh olives.

Also, other features of the olive fruits were correlated with the NIR spectrum and some spectral bands, such as VIS and MIR combined with NIR, have been studied. In the works of Cayuela and Camino [38], 2010 and Cayuela et al. [39], 2009, the authors investigated the usefulness of a portable VIS/NIR spectrometer using Partial Least Square (PLS) for the prediction of oil free acidity among other features from measures over fresh olives. Figure 2.16 shows the experimental setup for these experiments. The external validation output showed a residual predictive error of 0.03 for free acidity and it is a positive result taking into account the free acidity levels to categorise the olive oil. Moreover, in 2010, Dupuy et al. [62] combined NIR and MIR spectral bands associated with chemometric treatments for a direct



Figure 2.16 Sample turn table accessory of the spectrometer used in [38].

and rapid test method used firstly to predict the water and oil contents of olives and secondly to provide French variety recognition of olive fruit with 100% of good classification. Furthermore, Bellincontro et al. [16] in 2012 employed a portable NIR to identify the harvest period that ensures the ripening stage corresponding to the optimal phenolic content. In this work authors compared the proposed methodology with an analysis in laboratory and they achieved correlation coefficients between 0.874 and 0.942 with residual predictive errors between 0.68 and 6.33 mg/g. These results show how NIR can be considered a feasible on-field tool for the measurement of phenols and demonstrate its potential for adaption to the input of the VOOPP.

Technological variables in the olive oil elaboration process are regulated according to the elaboration objective, which are mainly the oil quality or the yield rate. The elaboration objective depend on the physical and chemical characteristics of the olives. Consequently the measurement of these features on process before the milling process could be useful. An in-line NIR system attached to the conveyor belt was tested by Salguero Chaparro et al. [170] in 2013. They utilised the system in order to determine moisture, fat content and acidity before the milling process. In a previous work [169] the parameters of the NIR instrument – focal distance and integration time – were adjusted. The authors achieved prediction errors of 2.98%, 2.15% and 2.53% respectively. Afterwards, in 2014 the same authors compared the performance of two NIR technologies, diode-array and grating monochromator technology [171]. The experimental setup of this works can be seen in Figure 2.17.

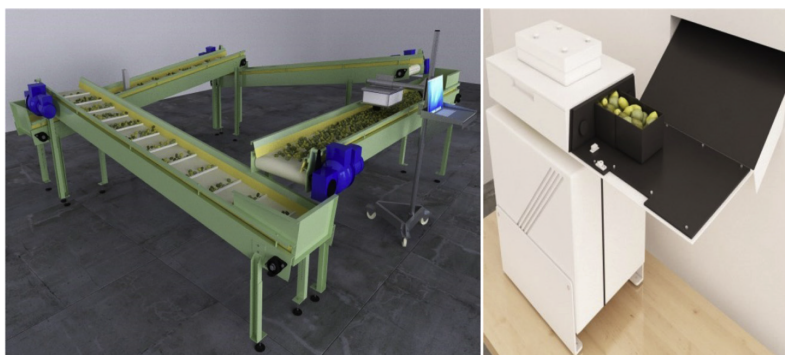


Figure 2.17 Experimental setup used in [171]. CORONA 45 Vis-NIR (left) and Foss NIRSystem 6500 (right) instruments.

The errors estimated indicated the similarity in the accuracy of the on-line analysis in comparison to the traditional off-line approach.

2.4.3 Near infrared spectroscopy applied to olive oil

Armenta et al. [8], 2010 and Casale et al. [35], 2014 reviewed the employment of the NIR technology on olive oil during recent years. They concluded that the use of the olive oil infrared spectrum can be employed for different applications such as detection of adulterates, quantification of parameters related to olive oil quality and classification of olive oils attending to these parameters.

Since the consumption of olive oil has increased, so have fraudulent attempts to adulterate olive oils with cheaper vegetable oils. Different works have demonstrated that the adulterated olive oils can be detected from the infrared spectrum [155] [117] [42] [57] [204]. Furthermore, the infrared spectral band has been studied to classify olive oils according to the varietal and geographical origin of the olive fruits [126] [34] [51] [198] [179] [58] [19] [74] [153] [33]. However the interest of these applications to the industrial plant could be useless because olive oil is directly produced from the olive fruit and usually its variety and geographical origin is well known. On the other hand, the on-line determination of the olive oil quality markers is interesting for monitoring and controlling the process.

The feasibility of the infrared spectroscopy as a rapid and cost-effective technique for quality determination has been tested during recent years. In 2007 Bendini et al. [18] used a NIR instrument for free acidity determination of olive oil samples. Their dataset composed 160 samples for training and 160 samples for validation. They

Table 2.2 Research works about infrared spectroscopy applied to olive fruits.

Main goal	Description of the results	Referenced work
Oil and moisture content	Regression Coefficient (R) value upper than 0.9 and standard errors of prediction lowers than 2.21%	[124]
Oil, moisture, oleic acid and linoleic acid contents	R values of 0.94, 0.93, 0.84, and 0.88 and Root Mean Square Error of Cross Validation (RMSECV) values of 1.33, 1.88, 4.73, and 2.91 respectively	[123]
Moisture, oil, oleic acid and linoleic acid contents	R values of 0.8, 0.83, 0.77 and 0.81 respectively	[122]
Mesocarp oil content in vivo	R values in calibration and validation were 0.921 and 0.903 respectively	[15]
Quality features, including firmness, oil content and colour	Firmness: R = 0.77 and RMSECV = 1.36; Oil content: R = 0.64 and RMSECV = 0.05; Olive colour: R between 0.85 and 0.92	[118]
Determination of oil and moisture contents in intact olive fruits	Oil content: R = 0.89 and RMSECV = 1.99; Moisture content: R = 0.88 and RMSECV = 2.06	[89]
Oil moisture, dry matter, oil content, oil free acidity and fruit maturity index	Oil moisture: RMSECV = 1.74; Dry matter: RMSECV = 1.71; Oil content: RMSECV = 0.90; Fruit maturity index: Root Mean Square Error of Prediction (RMSEP) = 0.51; Oil free acidity: RMSECV = 0.03	[38]
Total phenols, verbascoside, oleuropein, 3,4-DHPEA-EDA	R values in prediction ranged between 0.874 and 0.942	[16]

Table 2.3 Research works about infrared spectroscopy applied to olive fruits (continuation).

Main goal	Description of the results	Referenced work
On-line determination of acidity, moisture and fat content in intact olive fruits	RMSECV values of 2.54-3.26 for moisture, 2.35-2.71 for fat content and 2.50-3.26 for acidity parameter	[170]
Influence of parameters such as focal distance and integration time	A focal distance of 13 mm and an integration time of 5 s have been defined as the optimal operational conditions	[169]
Fat content, moisture and free acidity with on-line and off-line devices	Off-line = [Fat content: RMSEP = 2.28; Free acidity: RMSEP = 2.95; Moisture: RMSEP = 3.30]; On-line = [Fat content: RMSEP = 2.02; Free acidity: RMSEP = 2.41; Moisture: RMSEP = 2.89]	[171]
Free Acidity (FA) in olive oil, Oil Content Referring to Fresh Weight (OCFW), Oil Content Referring to Dry Matter (OCDM) and fruit Moisture Content (MC)	RMSEP values were 0.05, 7.98, 12.3 and 1.52 respectively	[39]
Oil and water contents and classify samples into five principal French cultivar origins	Oil content: R = 0.98 and Square Error of Prediction (SEP) = 0.72; Water content: R = 0.98 and SEP = 1.28; 100% of good classification	[62]

achieved a prediction error lower than 0.04% and a regression coefficient value of 0.987. Also, Muik et al. [150] in 2003 and El-Abassy et al. [64] in 2009 used infrared spectroscopy for free acidity determination in oil samples. The lowest values of error achieved for each work was 0.29% and 0.14% respectively, although the last one was employing the Raman technique. Other works [7] [17] appended in their researches the study of the relationship between the infrared spectra and the index of peroxides. They achieved errors of 1.20 m.e.q. $O_2 \text{ kg}^{-1}$ and 3.96 m.e.q. $O_2 \text{ kg}^{-1}$ respectively. Also, the sensory attributes have been correlated with the NIR spectra. Particularly, in 2009 Sinelli et al. [180] employed the NIR spectral band to classify between low fruity, medium fruity and high fruity olive oils with high accuracy.

As can be seen, the most cited works in the literature are based on the estimation of the major quality components, free fatty acids and peroxide value. However, the fatty acid profiles, polyphenol content, and chlorophyll levels contribute to the sensory and stability qualities of the oil. The estimation of these parameters using the NIR spectrum was demonstrated in 2004 by Mailer [129]. The best results were reached with the estimation of major fatty acids (palmitic, oleic, linoleic and linolenic) and chlorophylls. Also, the correlation between the Infrared (IR) spectrum and other parameters such as water content, total phenol amount and antioxidant activity was study by Cerretani et al. [41] in 2010 with regression coefficients greater than 0.8 for the first and second parameter. Furthermore, the assessment of olive oil minor components from NIR spectrum was recently investigated in 2012 by Inarejos García et al. [101] Acceptable models for several minor components – in particular hydroxytyrosol derivatives and C6 alcohols as well as for positive sensory attributes — fruity and bitter were obtained.

Also, NIR spectrum combined with MIR and Visible spectrum have been employed by Dupuy et al. [61] in 2010 and Cayuela Sánchez et al. [40] in 2013 respectively. The main objective of the first work was to quantify fatty acids and triacylglycerols in olive oil samples. MIR spectra gave better results than NIR spectra and the model built on the concatenated vector (NIR+MIR) did not improve the results. The second work reported good results in the determination of the olive Oil Stability Index (OSI) by multivariate models from the VIS and NIR spectrum.

One of the last parameters included in the European Normative in order to determine olive oil quality are the fatty acid alkyl esters. The content of this compound is increased when the olive fruit is over-ripening or it has suffered mechanical damage during storage and transport. In 2013, Valli et al. [190] released a study about the correlation between the mid-infrared spectrum of the olive oil and the total amount of methyl and ethyl esters of fatty acids. The correlation coefficients were greater than 0.8. Nevertheless, they indicated that the results obtained here needed to be

confirmed through the acquisition of a larger set of olive oils in order to increase the robustness of the model.

The aforementioned works have dealt with off-line setups, that is, the infrared sensor was placed in laboratory and the samples were collected from the process and carried to the sensor. Jimenez Márquez adapted an off-line NIR sensor to the line process in order to carry out on-line experiments. With the help of a pump he created a by-pass at the exit of the centrifugal machine (Figure 2.18). The by-pass pumped the oil sample from the main oil current and gave it back once it had been scanned. In 2003 the same author [111] used this setup to assess chlorophylls and carotenoids in virgin olive oil good correlation coefficients, 0.993 and 0.985 respectively. Later, in 2005, Jimenez et al. [135] studied the feasibility of predicting acidity value, bitter taste and fatty acid composition on process. In all cases the correlation coefficients were greater than 0.93.

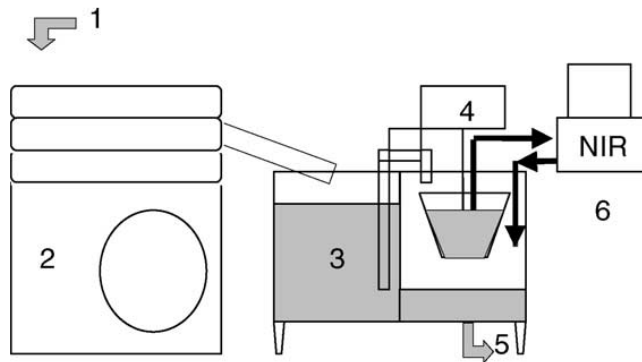


Figure 2.18 Schematic diagram of localization of the sensor NIR in the last phase of the olive oil extraction process used in [135]: (1) oil from horizontal centrifuge decanter; (2) vertical centrifuge for oil clarification; (3) tank for oil sedimentation; (4) continuous oil-weigher; (5) oil to storage; (6) NIR equipment.

The infrared spectrum of the olive oil samples can be disturbed by different variables such as the turbidity, temperature, moisture, optical path length or intensity of the halogen lamp light for example. However, there are no works in the literature that deal with the influence of these variables in the outcomes. Only Manley and Eberle [131] in 2006 and García Martín [76] in 2015 researched different path length in order to select the optimal one according to the achievement in the prediction of olive oil quality parameters.

The aforementioned works show that the NIR and MIR bands are highly correlated with different olive oil parameters and the methodology could work together with the official analytics, because they are necessary in order to calibrate the NIR and MIR systems.

2.4.4 Near infrared spectroscopy related to olive pastes and pomaces

Basically two types of slurries appear into the olive oil production process: olive paste and pomace (also olive cake). Olive paste is produced when the olive fruit are crushed by the mill and normally it is processed into the thermomixer. Pomace is the product obtained at the exit of the separation process and ideally it is free of oil.

Fundamentally the application of infrared spectroscopy in pomace has been oriented to assess the quantity of oil leaked during the olive oil production process and the quantity of water contained in the slurries. In 2004, Muik et al. [151] compared two spectroscopy techniques, Raman and NIR reflectance for determination of the oil and water content of olive pomace. The prediction models were obtained with PLS and the best results were reached when infrared spectra were transformed into their first derivative. The authors got a prediction error of 0.20% and 0.21% for oil content and 2.0% and 1.8% for water content regarding the official methods, using Raman and NIR, respectively. When the information acquired by the above techniques were appended, the models got better performance for oil content (0.17%) but no improvement for water content. Furthermore in 2005, García Sanchez et al. [77] carried out a similar work with good results. In this case the correlation coefficients were greater than 0.95 in all cases. In 2009 Barros et al. [13] studied the same issue but in this case the authors employed samples picked up from different harvest seasons. The error prediction in this work was 0.86% for oil content and 1.87% for water content.

Interesting experiments on process were done in 2008 by Jimenez et al. [108] the authors studied an automatic sensor software based on artificial networks to predict moisture and fat content of olive pomace during the production process. They placed a sensor NIR at the exit of the horizontal centrifugal machine (Figure 2.19) and they achieved prediction errors of 0.20% and 0.55% for fat content and moisture respectively.

The employment of infrared spectrum with olive paste has been focused on the assessment of two parameters: moisture and fat content. Usually the percentage of moisture into the olive paste is an interesting parameter in order to regulate the quantity of water that needs to be added to the slurry previous to the separation

Table 2.4 Research works about infrared spectroscopy applied to olive oil.

Main goal	Description of the results	Referenced work
Determination of adulteration of olive oil by vegetable oils	The correlation coefficients of actual versus predicted concentrations were between 0.90 and 0.99	[155]
Discrimination and determination of adulterants in pure olive oil	The optimum PLS calibration models yielded the prediction results for the adulterants of corn oil, hazelnut oil, soya oil, and sunflower oil with the error of 0.33%, 0.465%, 0.47%, and 1.05% w/w, respectively	[117]
Classification and quantification of the adulteration of pure olive oil by soya oil, sun flower oil, corn oil, walnut oil and hazelnut oil	The models predicted the adulterants, corn oil, sun flower oil, soya oil, walnut oil and hazelnut oil involved in olive oil with error limits ± 0.57 , ± 1.32 , ± 0.96 , ± 0.56 and $\pm 0.57\%$ weight/weight	[42]
Detect and quantify the sunflower oil adulterant	Prediction of adulterant content was possible with a standard error equal to 0.8%	[57]
Quantify the amount of olive pomace oil adulteration in extra virgin olive oil	Fourier Transform-Raman spectroscopy gave the highest correlation with a correlation coefficient of 0.997 and a standard error of prediction of 1.72%	[204]
Identify the geographical origin of olive oils	The recognition ratio was 97%	[126]

Table 2.5 Research works about infrared spectroscopy applied to olive oil (continuation I).

Main goal	Description of the results	Referenced work
Classification of extra virgin olive oils from Liguria according to their geographical origin	Accuracy up to 100%	[34]
Discrimination of the geographical origin of Italian olive oil	Accuracy of 86.3% with NIR dataset and 82.2% with Fourier Transformations Infrared (FT-IR)	[51]
Confirm the geographical origin of European olive oil samples	The best models correctly predicted the origins of samples in the prediction set up to 92.8 and 81.5% for Ligurian and non-Ligurian olive oil samples	[198]
Classify oils on the basis of their geographical origin	Accuracy of 90% with NIR and upper to 95% with MIR spectrum	[179]
Classify extra virgin olive oils from the eastern Mediterranean on the basis of their geographic origin	A correct classification rate of 93.9% was obtained	[58]
Discriminating virgin olive oils of two very similar and geographically close denominations of origin	The results thus obtained revealed 11.6% of the samples in the calibration set to be misclassified	[19]

Table 2.6 Research works about infrared spectroscopy applied to olive oil (continuation II).

Main goal	Description of the results	Referenced work
Quantify fatty acid and triacylglycerol contents in virgin olive oil to discriminate French Registered Denomination of Origin (RDO)	Nine fatty acids and fifteen triacylglycerols were correctly predicted with Relative Error Prediction (REP) <10%. French RDO were well predicted	[74]
Authentication study of the Italian Protected Denomination of Origin (PDO) based on near-infrared and UV-Visible spectroscopy, artificial nose and artificial tongue	From the explorative and modelling results, it arises that the spectroscopic techniques worked better than artificial nose and tongue for characterising the Chianti Classico PDO olive oil	[153]
Classification of samples representative of the whole Chianti Classico production area	Classification rate up to 100% of accuracy	[33]
Fat content, moisture and free acidity with a device located in an industrial mill	Oil content: $R = 0.92$ and $RMSEP = 0.67\%$; Moisture: $R = 0.913$ and $RMSEP = 1.29\%$; Free acidity: $R = 0.987$ and $RMSEP = 0.0364\%$	[18]
Free fatty acid content in olive oils and olives	$RMSEP$ was 0.29% for oil and 0.28% for olives	[150]
Free fatty acid content of extra virgin olive oil	R value was 0.96 and Root Mean Square Error of Validation ($RMSEV$) was 0.034%	[64]
Determination of acidity and peroxide index in edible oils	The validation set provide a $RMSEP$ value of 0.034% (w/w) for acidity in olive oil and $1.87 \text{ m.e.q. } O_2 \text{ kg}^{-1}$ for peroxides	[7]

Table 2.7 Research works about infrared spectroscopy applied to olive oil (continuation III).

Main goal	Description of the results	Referenced work
Free fatty acid content and the peroxide value in virgin olive oils	Free acidity: $R = 0.955$ and $RMSECV = 0.15\%$ (w/w); Peroxides: $R = 0.855$ and $RMSECV = 3.96$ m.e.q. $O_2\ kg^{-1}$	[17]
Classification of olive oil samples on the basis of the fruity attribute	The prediction rate varied between 71.6% and 100% as average value	[180]
Determination of quality parameters in olive oil, including Free Fatty Acids, Polyphenol Value, Polyphenol content, induction time, chlorophyll and the major Fatty Acids	Fatty Acids, oleic and linoleic, Free Fatty Acids and chlorophyll provided high levels of accuracy with R values of 0.99, 1.00, 0.97 and 0.98 respectively	[129]
Determination of Water Content (WC), total phenol Total Phenols (TP) amount and antioxidant activity Antioxidant Activity (ABTS)	The limit of detection values were 9.4 (mg/kg oil) for WC, 12.5 (mg gallic acid/kg oil) for TP and 0.76 (mmol Trolox/kg oil) for ABTS	[41]
Prediction of the minor composition of VOO	Fatty Acids Profile: $R = 0.84-0.96$; Hydroxytyrosol derivatives: $R = 0.86-0.88$; C6 alcohols: $R = 0.69-0.80$; Fruity: $R = 0.87$; Bitter: $R = 0.85$	[101]
Quantitative analysis of fatty acids and triacylglycerols and for identifying the Registered Designation of Origin (RDO) of extra virgin olive oils	For the determination of the geographical origins, as the results obtained in the MIR range were very good	[61]

Table 2.8 Research works about infrared spectroscopy applied to olive oil (continuation IV).

Main goal	Description of the results	Referenced work
Determination of OSI	The SEP values were lower than 7.38	[40]
Total amount of methyl and ethyl esters of fatty acids and the ratio between ethyl and methyl esters	Good agreement between predicted and actual values on calibration data sets was found. R values of 0.98 and 0.83, respectively	[190]
Visible-near infrared transmittance spectroscopy was used to determine total levels of chlorophyll and carotenoid in virgin olive oil	The validation set gave a correlation coefficient and standard error of prediction of 0.985 and 0.66 mg kg ⁻¹ for carotene totals and 0.993 and 0.96 mg kg ⁻¹ for chlorophyll totals	[111]
On-line control quality and characterization of virgin olive oils	The validation set gave correlation coefficients and standard error of prediction of 0.999 and 0.35, 0.936 and 0.058, 0.998 and 0.604, 0.992 and 0.674% for Acidity Value (AV), k225, fatty acid oleic and fatty acid linoleic, respectively	[135]
Determination of several extra virgin olive oil quality parameters	In general, reliable prediction results were obtained for the peroxide value [SEP = 4.15 meq O ₂ kg ⁻¹ , R = 0.87], K232 (SEP = 0.94, R = 0.94), linoleic acid (SEP = 0.83%, R = 0.90) and saturated acid content (SEP = 0.91%, R = 0.88)	[131]
Evaluation of different optical path lengths for olive oil's free acidity determination	With a 10-mm cuvette the SEP achieved was 0.051%	[76]

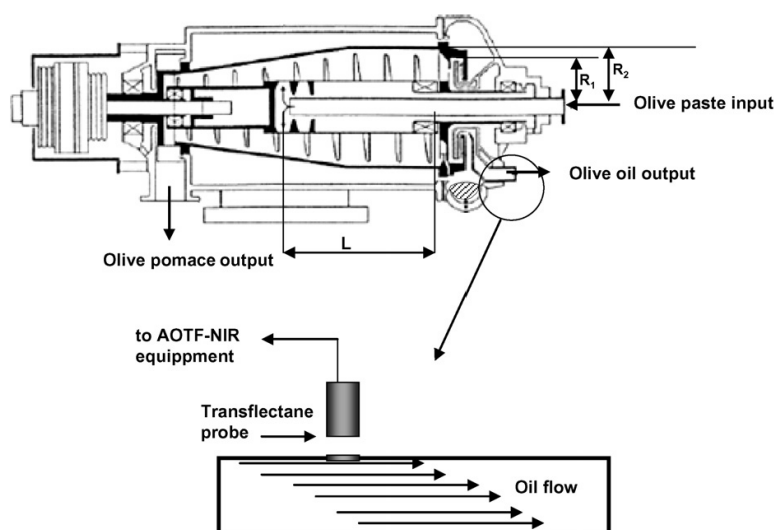


Figure 2.19 Schematic decanter and localization for reflectance probe used in [108]. R_2 is the decanter radius; R_1 is the level output oil; L is the length of horizontal zone of decanter.

process inside the horizontal centrifugal machine. Moreover, the estimation of the quantity of dispersed oil into the olive paste at the beginning of the process is an interesting data because it allows to know the potential olive oil that can be extracted. Using the NIR technique, in 2000 Jimenez et al. [109] achieved a standard error of prediction of 0.92% for moisture and 0.811% for fat content. Furthermore, in 2007 Bendini et al. [18] got an estimation error of 1.29% and 0.67% for moisture and fat content respectively. Later, similar works were carried out in 2009 by Barros et al. [13] with prediction errors of 1.08% for moisture and 2.64% for fat content.

In 2005 Ayora et al. [9] proved the potential of diffuse reflectance near-infrared spectroscopy combined with pattern recognition to discriminate between olive fruit of different qualities. The sample set was formed by sound olive fruit and those showing the most common alterations of olives which lead to decreased oil quality, namely freeze damage, harvest after falling on the ground, fermentation due to prolonged storage time, and olive tree diseases. Discriminant analysis provided prediction abilities of 100% for sound, 79% for frostbite, 96% for ground, and 92% for fermented olives using cross-validation. A similar study was carried out by Guzman et al. [92] in 2012 with a portable Raman sensor to classify slurries coming from sound and ground milled olives. Figure 2.20 shows the experimental setup of this experiment. They compare Soft Independent Modelling of Class Analogies (SIMCA), partial least square discriminant analysis Partial Least Squares Discriminant Analy-

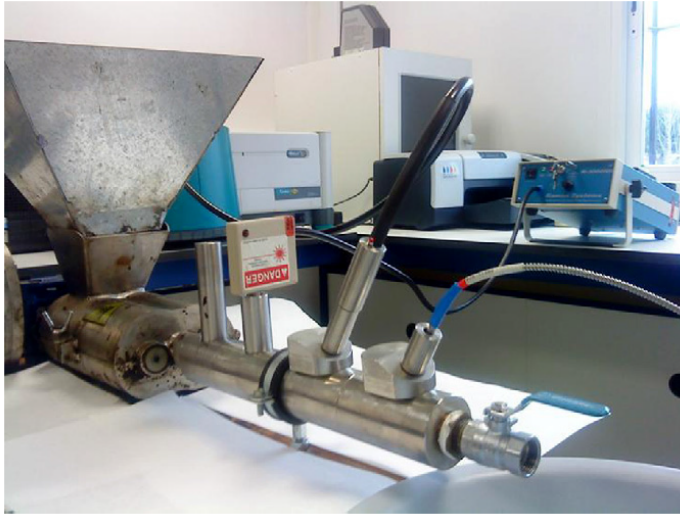


Figure 2.20 System obtained from [92] and designed for this work in order to take Raman measurements.

sis (PLS-DA) and k-Nearest Neighbours (k-NN) algorithms, and the last one was the better with 100% and 97% of success for sound and ground olives respectively.

The prediction of the quality of the dispersed oil into the olive paste is another studied application. In 2003 Muik et al. [150] researched the feasibility of using Raman spectrometry to determine the free fatty acids content in olive oils from olive paste measures with a root mean square of prediction of 0.28%. In the literature there is only one research work that appears that uses an on-line NIR system to estimate the oil quality from olive paste measures. So in 2015 Allouche et al. [2] employed an artificial neural network, operating online, to predict olive fruit parameters (pulp/stone ratio, extractability index, moisture and oil contents) and the potential characteristics of the extracted oil (free acidity, peroxide index, K_{232} and K_{270} , pigments and polyphenols) in olive paste. These predictions were obtained in part of on-line NIR spectral information on freshly crushed olive paste prior to the kneading stage. In this work the NIR device was placed on the first section of the thermomalaxer in the olive oil industrial plant (Figure 2.21). The regression coefficients were greater than 0.80 in all cases.

Table 2.9 Research works about infrared spectroscopy applied to olive slurries.

Main goal	Description of the results	Referenced work
Determination of the oil and water content of olive pomace	The values of RMSEP were 0.19 and 0.20–0.21 for oil content and 2.0 and 1.8 for water content, respectively	[151]
Determination of moisture and fat content of olives and pomace	R value was 0.9198 for moisture and 0.9782 for fat content	[77]
Reagent- free determination fat and moisture content in milled olive and olive pomace	Water content: R = 0.96 and RMSECV = 1.08%; Oil content: R = 0.88 and RMSECV = 2.64%	[13]
Real-time prediction of the moisture and fat content in olive pomace	Values of R=0.961 and RMSEP=0.32% for fat content and R=0.970 and RMSEP=1.01% for moisture were obtained	[108]
Measuring oil content and moisture in olives	Olive oil content: R = 0.9908 and SEP = 0.811; Moisture: R = 0.9939 and SEP = 0.928	[109]
Discrimination between olives of different qualities	Discriminant analysis provided prediction abilities of 100% for sound, 79% for frostbite, 96% for ground, and 92% for fermented olives	[9]

Table 2.10 Research works about infrared spectroscopy applied to olive slurries (continuation).

Main goal	Description of the results	Referenced work
Discrimination of olives before the oil processing stage in order to detect whether they have been collected directly from the tree or not	The best results were obtained using the k-NN method, with prediction abilities of 100% for 'sound' and 97% for 'ground' in an independent validation set	[92]
Real-time characterisation of olive fruit and the potential characteristics of the extracted oil in olive paste	RMSEP values for extractability index: 0.04%; OCFW: 1.11%; MC: 1.46%; Pulp/stone ratio: 0.31; Free acidity: 0.11%; Peroxide value: 0.45 mEq O ₂ kg ⁻¹ ; UV absorption at 232nm: 0.04; UV absorption at 270nm: 0.01; Carotenoid content: 0.58 mgkg ⁻¹ ; Chlorophyll content: 0.96 mgkg ⁻¹ ; Polyphenol content: 61.50 mgkg ⁻¹	[2]

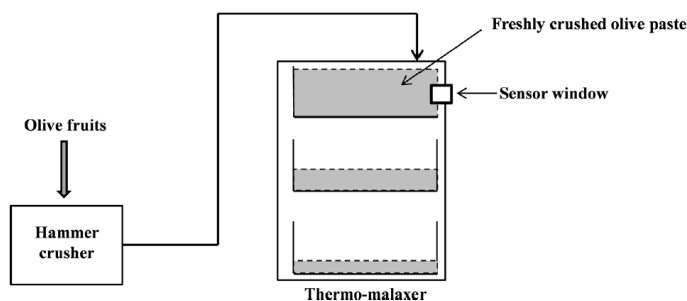


Figure 2.21 Schematic diagram of localisation of the sensor AOTF-NIR used in [2].

2.5 Computer vision

Computer vision technology tries to emulate the vision system of living beings where images are captured by the eye and they are transformed into nerve impulses which serve as inputs to the brain, in order to make inferences and take actions based on these inputs [85]. In this context, computer vision is a technique based on imaging acquisition and image processing. The images are obtained through imaging devices (camera or digital video recorder) and imaging processing [87] is employed to extract the information which describes the scene [86]. Then this information can be considered as inputs of different algorithms to learn and take decisions. Nowadays computer vision is widely implemented in most production processes technologically advanced.

Furthermore, this technology have been increasingly useful in the agrifood industry, especially for tasks related to quality control such as detection of imperfections in meat structures and the onset of food deterioration [128] [3] [59] [23] [24]. However the application int the olive oil industry is not widespread: there are hardly any commercial systems and most of the research works are mainly focused on the automatic classification of olive fruit according to quality, oil yield, maturity index, variety or origin, among others.

2.5.1 Principles of operation

A computer vision system generally is composed of five basic components: illumination system, a camera device, an image capture board (frame grabber or digitiser), computer hardware and software. Figure 2.22 shows an example of computer vision set-up. The electromagnetic radiation emitted by the lighting system is reflected of transmitted by the sample toward the camera device. This device disposes of

arranged light sensitive elements which turn the incident radiation into electronic charges. Finally these charges are transformed into digital data stream and the information is transmitted to the computer memory.

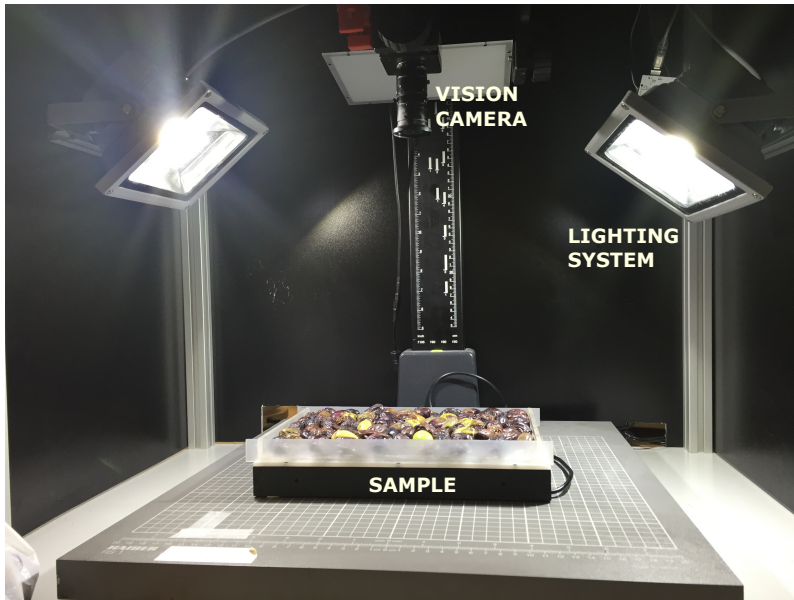


Figure 2.22 Structure of an image analysis system.

Since the images acquired by a computer vision system are affected by the intensity of lighting, the quality of the lighting system and the effects of other external illumination, it needs to be properly designed. Fluorescent and incandescent bulbs are the most frequently used. However, in recent years, the use of LED has been increased because of the low cost, low consumption, high light intensity and high durability. Also it allows to design different arrangements of lamps which can be adapted to the morphology of the sample, such as circular systems for flat samples or scattered systems for ball-shaped samples.

Typically the image sensors used in computer vision are based on Charged Coupled Device (CCD) array or Complementary Metal Oxide Semiconductor Array (CMOS) [23] [182]. CCD array consists of a hundreds of thousand of light sensitive elements directly related to the number of pixels contained in the acquired images. The array is covered with a crystal silicon plate. When the light reaches the plate, it ejects electrons from particular pixels. The number of ejected electrons are proportional to light intensity at a given pixel. Electronic charges are transferred from the pixel to

the output register by means of adjacent pixels, and then they are transformed into voltage levels. CMOS arrays are based on the same principle. The main difference is that each pixel has its own voltage transducer [50].

2.5.2 Computer vision applied to olive fruit

The main goals of the computer vision applied to olive fruit are focused on the best harvesting date estimation and the incoming inspection.

Determining the optimum harvest date of the olives allows to increase the profits of the farmer and the companies that produce olive oil. Correctly choosing this parameter has a significant influence on both the quality and quantity of the oil produced [49]. Different systems have been proposed for determining the optimal time to harvest. Some of them are based on computer vision techniques to determine parameters related to the collection, such as the size of olives, maturity index or oil content. In these works, images from olives or even olive branches recently harvested are studied using these techniques.

For example, Gatica et al. [78] in 2013 presented a method for olive fruit recognition using pictures taken in the tree. The target of this work was to estimate the size and weight of the olives in order to obtain the best time for harvesting. They worked taking images of olive tree branches where there were leaves and fruits. The recognition was performed in two stages: first, it was decided if the segmented object was an olive or not and later if the olives were overlapped. In both stages, they used a feedforward Artificial Neural Network (ANN) with a back propagation learning algorithm [14]. The performance achieved by the first stage was at 97% of correct classifications, while the second stage achieved a performance of 88.8%. The proposed model was based on the measurement of significant parameters of the olive as equatorial and polar diameters, eccentricity, volume, weight and average colour. The diameter of the olives was the most significant parameter for calculating the weight of the fruit and estimating the best harvest date.

On the other hand, the correct classification of olives in the oil mill yard before the extraction step is crucial because it determinates the best attainable olive oil quality [189]. Nowadays, in most of the factories, the evaluation of the incoming fruit batches depends on human visual inspection and the classification between olive fruit conditions are subject to personal criteria.

There are few works studying the automatic selection of olives in the reception before the mill. Some of them were focused on the detection of defective olives and

the determination of the quantity of impurities (leaves, pebbles, ...) mixed with the olive batch; while others are based on the prediction of different fruit parameters such as the maturity index, geographical origin, variety, quality parameters or quantity of oil which can be extracted from the fruit.

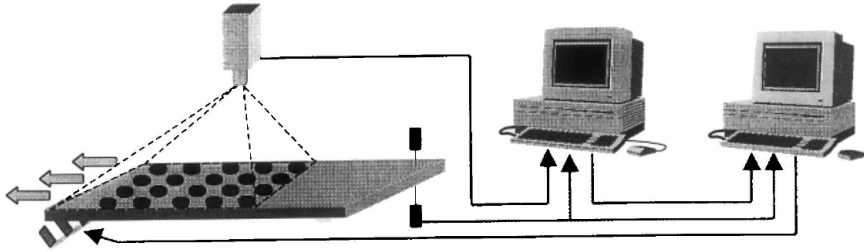


Figure 2.23 The camera takes images which are acquired and processed by the machine vision system. This system sends the parameters of each olive to the control system which decides which class the olive belongs to and when it should be ejected (Figure obtained from [54]).

Although table olive classification applying computer vision is not directly related to the main topic of this chapter, the research works associated with this purpose used similar procedures and techniques. In 2000 Diaz et al. [54] designed a system to sort table olives into four categories automatically using machine vision (Figure 2.23). They studied the effect of the colour and the image resolution per olive in the classification results. Furthermore, a comparison between human and machine results was conducted. In 2004 Diaz et al. [55] presented a new work on the table olives automatic inspection system, based on computer vision, that took into account aspects such as the colour of the skin or the presence of defects on the surface of the olive (Figure 2.24). In this work three algorithms were tested: Bayesian discriminant analysis, PLS-DA and ANN. They obtained the best results with the ANN approach with an accuracy of over 91%. Other authors [164] worked in the same way. They classify defective olives for table application using an artificial vision machine. In 2008 Riquelme et al. [165] proposed a classification system with eight categories of defective olives, using colour features of the fruit with morphological characteristics of external defects and applying discriminant analysis. The results of the machine vision system depended strongly on those categories and they obtained between 38% and 100% of classification success rates. The aforementioned works used fluorescent lamps with filters or indirect illumination as light sources. The images were acquired in the visible spectrum.

Methodologies and results with table olives can be extrapolated to olives for oil



Figure 2.24 Damage types identified in olives by expert technicians at processing companies: (a) undamaged olives; (b) 'serpeta'; (c) 'granizo'; (d) 'rehús'; (f) wrinkled; (g) purple olive; (e) and (h) others. (Figure obtained from [165]).

production. Furferi and Carfagni [71] in 2010 employed a machine vision with the objective of finding defects in olive surfaces and evaluating the ripening index of the olives according to the colour of the fruit. In this case, the computer vision setup with halogen lamps was located in a real oil mill during the blowing and washing processes. K-means clustering and thresholding were used as classification algorithms in order to determine the colour and superficial defects.

An offline method to detect defective areas in olives for oil production was developed by Guzman et al. [91] in 2013. The setup was composed by a digital monochrome camera with band-pass filters on IR and VIS spectra. Healthy and defective olive percentages were obtained with high accuracy. They were also classified into four categories depending on the defect size. The light source was a halogen light which provided good illumination conditions in both infrared and visible spectra. In 2013 Guzman et al. [93] presented another similar work but, in this case, they added colour information to estimate the ripening index and the images were acquire in line.

Olive ripening index is an useful parameter for choosing the best harvesting time. The common way to determine this index is by means of visual inspection (skin and flesh colour change during the ripening of fruit). In this context, Bibiloni and Radic

[20] in 2005 tried to measure the ripening index of olives by means of computer vision in the visible colour space. White colours related to background and sparkles were removed. They reported as conclusion that the green space channel was the most representative to estimate the maturity index in olives.

Furferi et al. [73] in 2010 employed computer vision and chemical parameters (sugar, oil and phenolic content) to estimate and validate an automatic system to estimate the ripening index. They used as inputs the colour of the olives in the Pantone colour space and the chemical indexes. Results show that the maximum gap between the visual ripening index and the predicted was equal to 0.25.

In 2013 Guzman et al. [94] predicted by means of computer vision the maturity index of different samples of olives with good accuracy. They also presented an automatic methodology to calculate the ripening index for each olive individually according to the colour, size and weight of the olives.

On the other hand, computer vision technology has been applied to the automatic prediction of the percentage of impurities mixed with olive batches. With this aim, Guzman et al. [90] in 2013 studied a new methodology to estimate the amount of leaves, branches and stones in olive batches. It was based on visible colour segmentation and classification through k-NN and clustering algorithms.

Image processing has also been used to predict the olive oil content in intact fruits. Ram et al. [162] in 2010 developed two prediction models based on visual image analysis and the implementation of linear regressions and artificial neural networks to estimate the oil accumulation in olives during the harvest process. The best results were reached using artificial neural networks, obtaining linear correlations of 0.87 for Picual variety of olives.

Furthermore, organoleptic and chemical features of olive oil depend on the olive variety. In this context, computer vision has been useful to classify different olive cultivars. In 2014 Vanloot et al. [192] presented a machine learning system based on computer vision in order to classify five main French cultivars (Aglandau, Bouteillan, Lucques, Picholine and Tanche). They reached 100% of positive results in classification. Images of the frontal and the profile of the olive pits were treated to extract significant features related to the shape, colour and texture. Principal Component Analysis (PCA) was used to simplify the information and PLS-DA as algorithm of classification.

Furferi et al. [70] and Carfagni et al. [31] in 2008 developed a new system based on artificial vision to estimate free acidity and peroxide index from the olive images

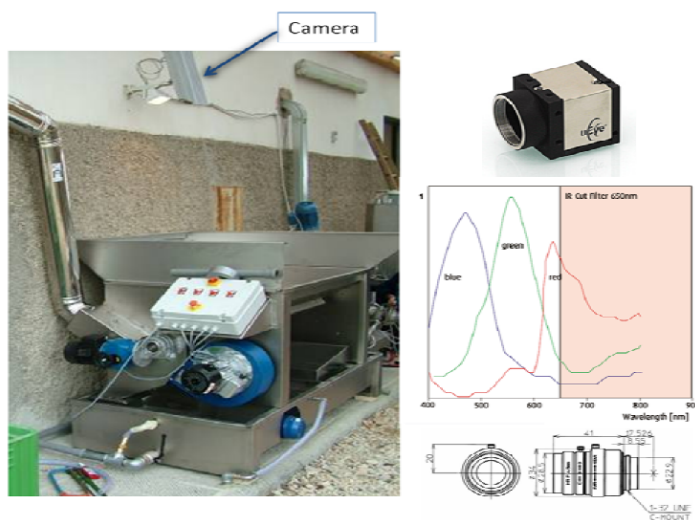


Figure 2.25 Positioning of the CMOS camera, camera mode, spectral response and optical (Figure obtained from [31]).

during the production process. Figure 2.25 presents the experimental setup. In this case, Red Green Blue (RGB) image channels were used to calculate the ripening and healthy degree of the olives through colour segmentation. The olive oil at the end of the process was chemically analyzed to calculate acidity and peroxides. They employed neural networks obtaining relatively good results with mean errors of 3.8% for each parameter. Thereupon, in 2007 the these authors [72] presented the same work however, in this case, they did not take into consideration the sanitary conditions and they reached worse results.

2.5.3 Computer vision applied to olive oil

Moisture and impurities content into the olive oil can affect to the oil deterioration. The determination of these two parameters is an important task before the storage process.

In this context, in 2013 Marchal et al. [133] presented a work in which, using artificial vision (Figure 2.26), the oil samples were classified into three categories according to the impurities content after the drying process. The classification was performed by treating the colour information of the CIELab and Hue Saturation Value (HSV) colour spaces with different algorithms: PCA, Kernel Principal Component Analysis (KPCA), Linear Discriminant Analysis (LDA) and Kernel Linear Discriminant Analysis (KLDA). Once the main features were extracted, the clas-

Table 2.11 Research works about computer vision applied to olive fruit.

Main goal	Description of the results	Referenced work
To estimate olives size and weight in order to obtain the best harvesting moment	97% of samples correctly classified	[78]
To classify table olives in four categories depending on the surface quality	Good classification results from 62.97% to 99.26% of success	[54]
To classify table olive in 4 different categories according to the defects over the surface of the fruits, using different algorithms	Different accuracy: Mahalanobis algorithm (43% to 86%); PLS algorithm (40% to 90%); ANN (91% to 100%)	[55]
Detection of external faults in table olives based on different commercial categories	38% to 100% of classification success rates	[165]
Evaluation of ripening index and the presence of superficial defect in olives for oil production	Good determination of defective areas according to the visual perception	[71]

Table 2.12 Research works about computer vision applied to olive fruit (continuation D).

Main goal	Description of the results	Referenced work
Evaluation of olive quality based on detecting external defects using infrared machine vision system	IR vision system is a useful technology for olive inspection and defect sorting based on surface damages	[91]
On-line prediction of ripening index and sanitary conditions in five categories of olive for oil production using visible and infrared vision system	This methodology could be used for on-line classification of fruit in different categories in olive oil extraction process	[93]
Estimation of olives superficial color to obtain the best harvesting moment and the ripening index	The reflection index of green colour is very appropriate measure to evaluate ripening index of the olive	[20]
Prediction of ripening index based on olive skin color and a set of chemical parameters provided by historical curves, using ANN	The ripening index values provided are very well correlated with the ones provided by the panel of experts. Maximum prediction gap of 0.25 in 8 different values	[73]
Estimation of ripening index, size and the average weight of olives for oil production using visible and infrared machine vision	The predicted ripening index was close to the one visually estimated by experts. The evaluation of weight of fruit also provided good results ($R = 0.91$)	[94]

Table 2.13 Research works about computer vision applied to olive fruit (continuation II).

Main goal	Description of the results	Referenced work
To estimate the percentage of impurities (leaves, branches and pebbles) mixed with olive batches	Results very close to the real values provided by manual method ($R = 0.88$ for leaves and $R = 0.82$ for pebbles)	[90]
To predict the olive oil content based on a computer vision system	Two prediction models were used: linear regressions and artificial neural networks. The results were correlations of 0.81 and 0.87 respectively	[162]
To assess the varietal origin of olive fruits from olive pits	100% of correctly classified samples	[192]
Estimation of acidity level and peroxides by means of computer vision over olive fruit	The mean error of validation was 3.8% for each parameter	[31]
Estimation of acidity level and peroxides by means of computer vision over olive fruit	The maximum error was about 10% for both parameters	[70]
Estimation of acidity level and peroxides by means of computer vision over olive fruit	The maximum error was about 12% for both parameters	[72]

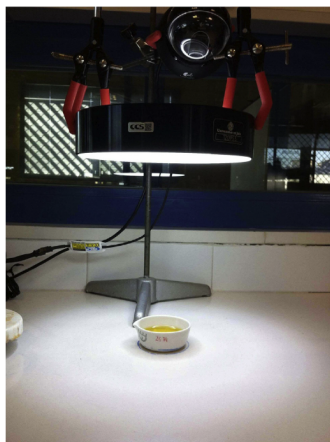


Figure 2.26 Image acquisition set-up, composed by the illumination ring and the camera (top). In the bottom of the image is the porcelain basin containing the olive oil sample. (Figure obtained from [133]).

sification was carried out with Support Vector Machine (SVM) and an ANN. The best results were obtained using KPCA and SVM. They reach a classification rate of 87.66%.

Several research results have demonstrated that plant pigments play important roles in health [83] and they could be estimated in line using non invasive technology such as computer vision. With this aim, in 2014 Cayuela et al. [37] compared the NIR spectrum with the VIS spectrum in order to determine total chlorophylls and carotenoids from 258 samples of olive oil. The optimization of the calibration for total chlorophylls was set by using Multiple Linear Regression (MLR) from the wavelengths exclusively visible. Its satisfactory performance was proven from the model coefficients: Standard Error of Calibration (SEC) of 2.63 and R equal to 0.97, and the Residual Predictive Error (RPD) of 5.76 from the external validation. However, the model built using the entire spectrum (VIS-NIRS) was slightly better for total carotenoids determination, showing an RPD of 3.86.

While the previous section has presented different works dealing with olive fruit classification according to variety and geographical origin, the same topic could be applied to olive oil. In 2015, Milanez and Pontes [188] proposed a new methodology based on digital images and supervised pattern recognition methods for the classification of extra virgin olive oil samples with respect to brand and verification of adulteration with soybean oil. Figure 2.27 presents the experimental

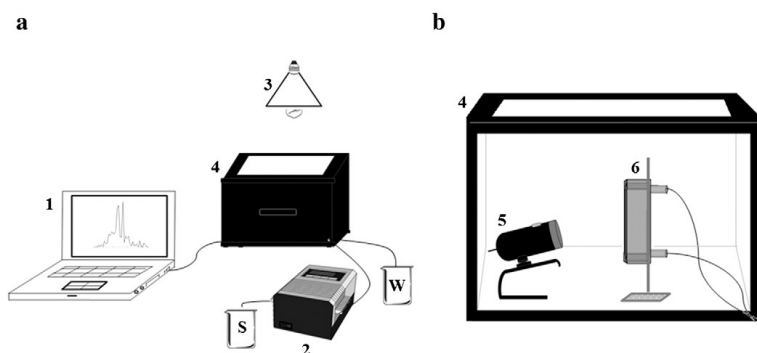


Figure 2.27 (a) System used in the acquisition of images of the edible vegetable oil samples where (1) is the notebook; (2) is a peristaltic pump; (3) is the fluorescent lamp; (4) is the closed compartment. (b) Detailed compartment where (5) is a webcam and (6) is the flow cell. (Figure obtained from [188]).

setup. They evaluate two classification models, LDA with Successive Projections Algorithm (SPA), and PLS-DA. With respect to brand, a correct classification rate of 100% was achieved for both models. Classification results for the second task showed an accuracy upper than 88%. Previously, the same authors in 2014 [145] presented a simple and non-expensive method based on digital image and pattern recognition techniques for the classification of edible vegetable oils with respect to the type and the conservation state. In both cases authors reached high rates of success, however they did not consider olive oil samples.

Furthermore, in 2015 Kadiroglu and Korel [113] studied the classification of Turkish commercial extra virgin olive oil samples according to geographical origins by using surface acoustic wave sensing electronic nose and machine vision system analyses in combination with chemometric approaches. The data analyses were performed with PCA class models, PLS-DA and Hierarchical Cluster Analysis (HCA). The results showed that PLS-DA and HCA methods provided clear differentiation among the Extra Virgin Olive Oil (EVOO) samples in terms of electronic nose and color measurements.

Table 2.14 Research works about computer vision applied to olive oil.

Main goal	Description of the results	Referenced work
Classification of olive oil according to the impurities content after the drying process	Classification rate of 87.66%	[133]
To use the visible spectrum in order to determine total chlorophylls and carotenoids	Residual predictive deviations of 5.76 and 3.86 respectively	[37]
Olive oil classification according to brand and adulteration with soybean oil	Classification rates of 100% and 88% respectively	[188]
Classification of edible vegetable oils with respect to the type and the conservation state	In both cases authors reached high rates of success	[145]
Classification of Turkish commercial extra virgin olive oil samples according to geographical origins	PLS-DA and HCA methods provided clear differentiation among the EVOO samples	[113]

3 System based on hyperspectral information to estimate acidity, moisture and peroxides in olive oil samples

Contents

3.1. Introduction	61
3.2. Materials and methods	63
3.3. Results and discussion	68
3.4. Conclusions and future works	71

3.1 Introduction

In general, the quality of a product is determined by the set of characteristics that allow it to be classified as equal, better or worse than those of similar nature. For olive oil, the parameters that determine its quality are influenced by the olive fruit (origin, variety, maturity and agronomic conditions, mainly) and the elaboration process variables (thermomixer temperature, centrifugal decanter differential velocity and others) [25].

Currently, the official analysis of the phycochemical and organoleptic quality parameters of olive oil are performed according to the methods established in the European Normative CE 1348/2013 [66]. The procedures are manual, work-intensive

methods and not eligible for their on-line implementation in order to adjust the elaboration parameters automatically.

In recent years, there have been several advances related to non-invasive techniques for automatic food quality analysis [10] [100] [106] [186] including olive oil [168] [178]. They have been focused in the infrared spectral range and have achieved good results. For instance, in [64] a Raman spectroscopy in the spectral window $945 - 1600\text{ cm}^{-1}$, which includes carotenoid bands, was found to be a useful fingerprint region, being statistically significant for the prediction of the free fatty acids. Also, in [7], one FT-NIR (Fourier Transform Near Infrared) spectrophotometer device was employed for the determination of acidity and peroxides index in edible olive, sunflower seed and maize oils [101] [129] [111]. Moreover, in [108] one AOTF-NIR (Acousto-Optic Tunable Filter Near Infrared) spectrophotometer working in the spectral range between 1100 and 2300 nm was used for real-time prediction of the moisture and fat content in olive pomace using two-phase olive oil processing. To our knowledge, there are no contributions dealing with the selection of the best wavelengths for the evaluation of olive oil quality parameters from images captured by an imaging spectrograph.

In this context, and using the images acquired by an hyperspectral device, the goal of this work was to select the optimal wavelengths which are better correlated with the olive oil parameters: free acidity, peroxide index and moisture. With this information, the computation time of the regression algorithms and the hardware costs for building new hyperspectral sensors could be significantly reduced. To this end, three methods were studied in order to select the optimal wavelengths: Genetic Algorithm (GA) [63], Least Absolute Shrinkage and Selection Operator (LASSO) [209] and SPA [6]. To our knowledge, these methods were selected for our experimentation because they are widely used in the literature and they have achieved good results with spectral data inputs [4] [6]. However, in the last years other variable selection algorithms and variations [148] [121] have arisen that could be used in future work. The correlation between the selected wavelengths and the values of the parameters obtained by means of analysis performed in a laboratory were evaluated with MLR [99].

This chapter has been structured as follows: Section 3.2 describes the olive oil samples used for the experiments, the analytical methods employed in order to reach the reference parameters, the experimental set-up built for acquiring images and the mathematical algorithms used for selecting wavelengths. Then, Section 3.3 shows the results and different comments. Finally, Section 3.4 presents the conclusions.

3.2 Materials and methods

The procedure followed in this work has been the selection and preparation of the olive oil samples, the measurement of these samples by a hyperspectral sensor, their analysis in a certified laboratory, and the study of the resulting data.

3.2.1 Olive oil samples

A total of 133 olive oil samples of around 30cl per sample were provided by the olive oil laboratory CM Europa S.L. (www.cmeuropa.com). This number included different types of olive oil -virgin, extra virgin, lampante- from the campaigns of 2012-2013 and 2013-2014. A database with acidity, peroxides levels, moisture and other chemical analysis applied to the samples were given too.

Of these samples, 56 of them included peroxides index values, 69 of them with moisture values and 133 samples with free-acidity analysis. Maximum, minimum and average values are shown in Table 3.1.

Table 3.1 Reference parameters from laboratory.

	Number of samples	Maximum	Minimum	Average $\pm$ S.D.
Acidity %	133	1.99	0.12	0.48 ± 0.45
Peroxides index meq. O_2	56	6.70	3.90	5.31 ± 0.83
Moisture %	69	0.43	0.07	0.15 ± 0.07

3.2.2 Analytical methods

The analytical methods were carried out by CM Europa. Thus, the acidity index is determined by acid-base titration of the free fatty acids with potassium hydroxide of the olive oil sample dissolved in ethanol. In turn, the peroxides index is determined dissolving the sample in a mixture of acetic acid and chloroform, later in a potassium iodide solution and titrating the freed iodine. The moisture content is determined relating the weight of the sample before and after a drying process held in a drying oven.

3.2.3 Experimental setup

The hyperspectral camera device used was composed of a Xeva-1.7-320 digital camera with a thermo-electrically cooled InGaAs detector head, an ImSpector N17E spectrograph and a 8mm lens. The integration time was set in 1ms. It behaves like a lineal camera capturing one line of the sample composed of 320 pixels over 256

different wavelengths. Its spectral range captured is between 900nm and 1700 nm, with a resolution of 4 nm. The result is an image with 320 columns and 256 rows, equal to 81920 pixels codified by 14 bits of resolution. The device captures up to 120 frames per second which can be transmitted to a PC through its CameraLink interface.

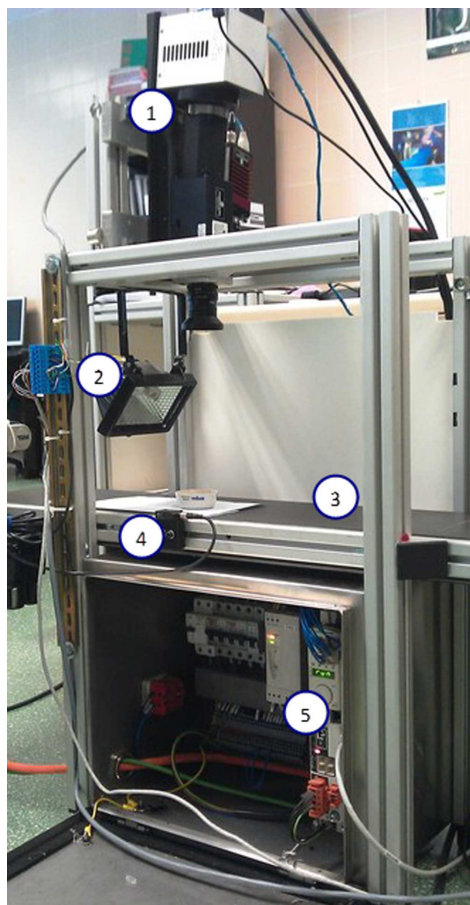


Figure 3.1 Vision system configured. In the figure, label 1 indicates the InGaAs hyperspectral camera, label 2 the illumination system used, label 3 the conveyor belt, label 4 the infrared sensor and label 5 indicates the variable speed drive.

An automatic sampler was built with the hyperspectral camera device, a 100W halogen lamp, a conveyor controlled by a LXM32M speed shifter and an infrared sensor used to detect the presence of objects. The setup is shown in Figure 3.1. The images were captured using the external trigger of the camera connected to the

infrared sensor. In measuring time the belt was stopped.

The software employed to control the LXM32M speed shifter was SoMove Lite V.1.4.4.0. On the other hand, Matlab 7.11.0 (R2013a) was used for the development of the software for the capture and analysis of the hyperspectral images.

The hyperspectral images were first preprocessed to obtain the spectra of the samples and also to reduce the noise introduced by the lighting system and other sources. Then, component selection algorithms were employed based on the following prefilters.

3.2.4 Hyperspectral images preprocessing

Before starting the capture of the images, the samples were prepared by pouring a small quantity of oil in a basin and placing them on the conveyor of the experimental setup over a white paper. The line of the sample which crosses the center of the basin was captured by the camera device (Figure 3.2a). An example of captured image is shown in the Figure 3.2b, where the X axis of the image belongs to a line in the sample, and the Y axis contains the wavelengths.

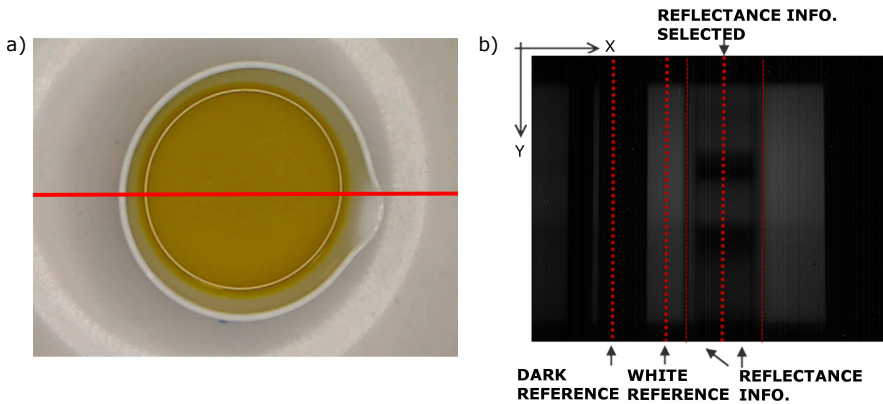


Figure 3.2 a) Image of an olive oil sample contained in a basin. The red line denotes the piece of sample to analyze. The hyperspectral image for this line is shown in the Figure b. b) Image (320x256) captured by hyperspectral camera. Discontinuous lines emphasize the white reference (white paper), dark reference (conveyor belt) and olive oil reflectance information.

The spectra of the olive oil samples are coded in each column of the monochromatic image. The first step consists of eliminating the blind frequency bands that appear due to the model of hyperspectral camera employed and do not provide any

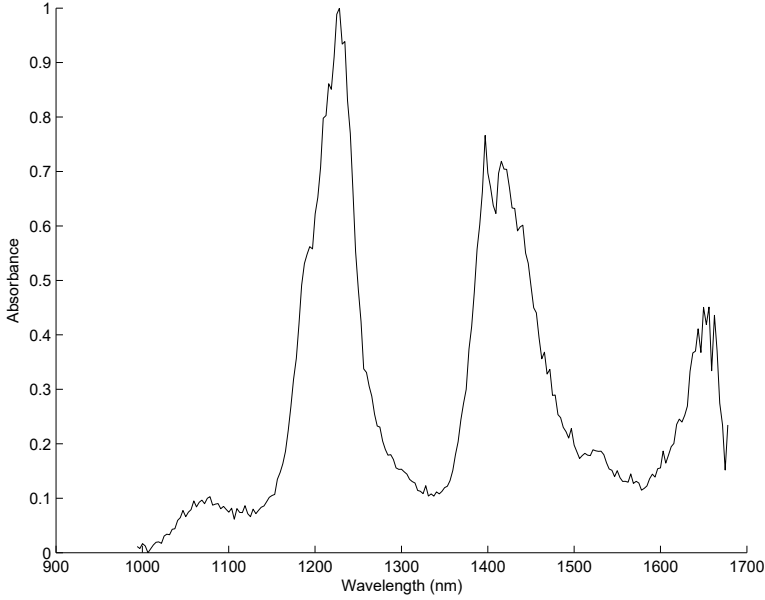


Figure 3.3 Spectrogram without any preprocessing of an olive oil hyperspectral image.

information. Specifically, these are the wavelengths included between 900-994 nm and 1678-1700 nm. After this, an average spectrum is calculated according to Eq. 3.1.

$$\bar{E}(y) = \frac{\sum_{x=x_0-\Delta x}^{x_0+\Delta x} E(x,y)}{2\Delta x} \quad (3.1)$$

where x_0 represents the selected pixel of the sample, Δx is half the number of pixels to average and $E(x,y)$ is the hyperspectral image. The variable x is the pixel of the line in the captured sample, and y is the hyperspectral information for the different wavelengths.

Afterwards, the image is calibrated with respect to two surfaces with non-variable spectrum according to the Eq. 3.2. These were the white paper under the basin and the black conveyor belt.

$$I = \frac{I_0 - D}{W - D} \times 100 \quad (3.2)$$

where I_0 is the spectrum to calibrate, D is the black surface spectrum and W is the white surface spectrum.

Then, the spectra were turned into energy absorbed by the oil samples according to Eq. 3.3.

$$A = \log \frac{1}{I} \quad (3.3)$$

Figure 3.3 shows the resultant spectrum. After that, the near infrared spectra were standardized by using Standard Normal Variate (SNV) and filtered by Savitzky-Golay (SG) method [174]. The SG input parameters were tuned with the objective of achieving good correlation results. Also, the aforementioned methods were applied in order to remove and minimize any unwanted spectral contributions such as light scatter [205].

3.2.5 Wavelength selection methods

Three known component selection algorithms were evaluated with the objective of searching for the best wavelengths which are correlated with the aforementioned parameters. They were Genetic Algorithm [63], Least Absolute Shrinkage and Selection Operator [209] and Successive Projection Algorithm [6].

The genetic algorithm was coded by using the GA Toolbox of Matlab in order to select the spectral components of the samples that optimize the multiple linear regression method [205]. Specifically, for the problem treated in this work, the steady-state GA model was employed, where the population of individuals is formed by binary vectors and every component of the spectrum (220 components) is represented by one bit. If the component is selected, the value of the bit is "1", being "0" otherwise. The genetic algorithm uses a cost function in order to evaluate the selected spectral components. Individuals showing the lowest values of the cost function are more likely to be propagated to the next generation. In this case, the cost function was the mean squared error obtained with the MLR method. The parameters of the genetic algorithm empirically selected to control the convergence speed were:

- Number of individuals per generation: 100.
- Number of generations: 70.
- The genetic algorithm stops when the cost function is not improved during 10 iterations.

On the other hand, a LASSO regression algorithm was applied to the spectra data with the goal of selecting wavelengths. LASSO uses an l_1 -penalty and continuously shrinks the smallest estimated regression coefficients towards zero. The number of zero-valued regression coefficients increases as a function of the regularization

parameter λ . Eq. 3.4 explains the problem.

$$\min_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (3.4)$$

where n is the number of olive oil samples, p is the number of wavelengths (220 in this case), β_j are the regression coefficients and λ is the regularization parameter. Matlab Statistic Toolbox was used for this method.

Successive projections algorithm was also evaluated as selection method. This algorithm starts with one wavelength, then incorporates a new one at each iteration, until a specified number N_{max} of wavelengths is reached. Its purpose is to select a number of wavelengths whose information content is minimally redundant and it solves collinearity problems. From $N = 1$ to N_{max} selected wavelengths, a MLR calibration model was employed in order to find the minimal mean square error of cross validation.

In summary, three models were used: GA-MLR, LASSO and SPA-MLR. The performance of calibration models was evaluated in terms of Mean Square Error of Validation (MSEV) and RPD [197].

$$MSEV = \frac{\sum_{i=1}^{N_v} (\hat{y}_{iv} - y_i)^2}{N_v} \quad (3.5)$$

being y_i the value measured in laboratory for the sample i , $\hat{y}_{iv}$ is the predicted value by validation model for the sample i , and N_v is the number of validation samples. To validate the model the leave-one-out approach was used.

$$RPD = \frac{SD}{RMSEV} \quad (3.6)$$

where Standard Deviation (SD) is the standard deviation of the reference data which were measured in laboratory and RMSEV is the root of MSEV.

3.3 Results and discussion

Figure 3.3 shows the raw NIR spectra of one olive oil sample. The spectra did not evidence any obvious differences from visual inspection on the basis of their compositional features. However, in the NIR region, bands around 1200nm arise from 2nd overtones of C-H stretching vibrations while those at 1400nm and 1410nm are due to the combination band of C-H [42] [47].

Table 3.2 shows the results for acidity determination by employing different

component selection algorithms. The best results were obtained with the dataset turned into Absorbance (ABS) levels and standard normal variate transformation without the Savitzky-Golay filter. The last method could be unnecessary because the spectrum has been averaged by the spectral resolution of the hyperspectral device (3 nm). On the other hand, the best RMSEV value was reached with Genetic Algorithm with Multiple Linear Regression (GA-MLR) method. In this case, an error in the estimation of free acidity of 0.0949% was obtained and the predicted values were fitted with a regression coefficient of 0.98 (Figure 3.4b). This result is appreciably good in order to classify on-line olive oil quality where the free acidity value between virgin and extra virgin class is in 0.8%. The components selected by GA algorithm are shown in Figure 3.4a. Most components are located between 1000-1160 nm (23% of the total), 1280-1350 nm (13%), 1480-1500 nm (5%) and 1570-1640 nm (14%). Also, the validation carried out by the orthogonal components selected by Successive Projection Algorithm with Multiple Linear Regression (SPA-MLR) method obtained good results with an error of 0.1628% (lower than the GA one). Finally, the best error obtained with LASSO method was reached with 22 components (0.3114%).

Table 3.2 Results for free acidity estimation.

Algorithm applied	Pre-processing method	Number of components	RMSEV	R_v	RPD
GA-MLR	ABS, SNV	106	0.0949	0.98	4.77
SPA-MLR	ABS, SNV	106	0.1628	0.93	2.78
LASSO	ABS, SNV, SG	22	0.3114	0.84	1.45

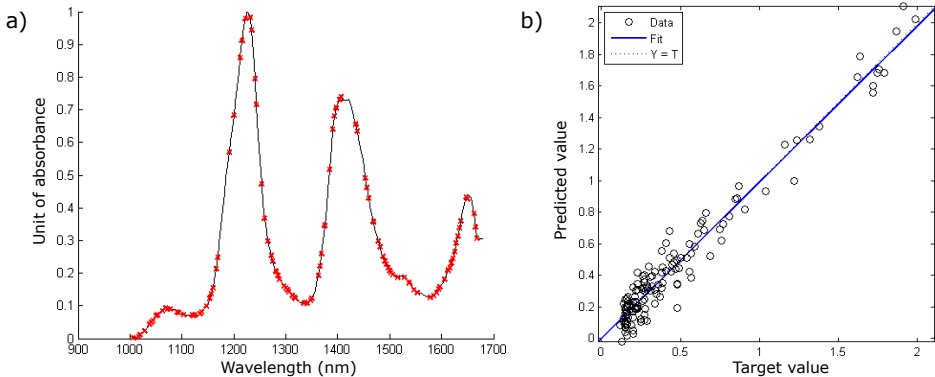


Figure 3.4 a) Components selected by GA method, marked with x marks. b) Regression fitted by cross validation..

On the other hand, Table 3.3 presents the minimum error values obtained for each method for peroxide index determination. In this case, the best result was reached with the SPA selection method and the spectral responses transformed

into ABS and SNV. The SPA with 52 components adjusted the response variables with a regression coefficient of 0.99 and 0.03 of MLR residual error (Figure 3.5b). Figure 3.5a shows the selected wavelengths. These were grouped around the three main peaks: 11 components or (21% of the total) are located between 1210-1240 nm, the same quantity between 1390-1430nm and 10 components (20%) between 1630-1670 nm.

Table 3.3 Results for peroxide index estimation.

Algorithm applied	Pre-processing method	Number of components	RMSEV	R_v	RPD
GA-MLR	ABS, SNV, SG(3,0,0)	105	0.1830	0.97	4.55
SPA-MLR	ABS, SNV	52	0.0316	0.99	26.38
LASSO	ABS, SNV, SG(7,0,0)	7	0.5280	0.84	1.58

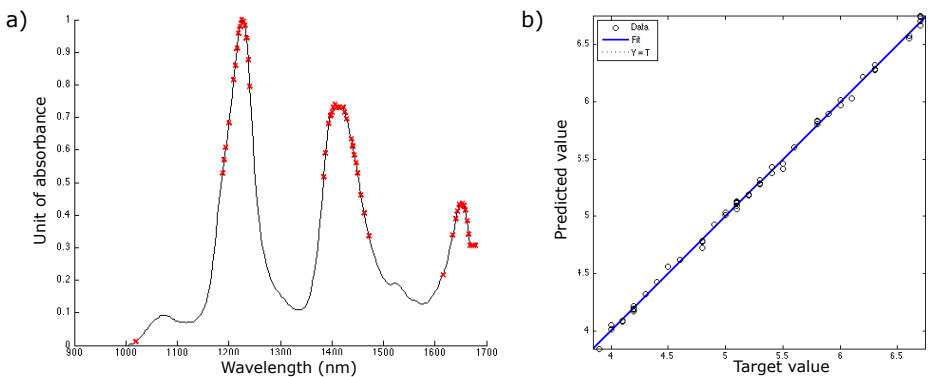


Figure 3.5 a) Components selected by SPA method, pointed with x marks. b) Regression fitted by cross validation..

Finally, the last table (Table 3.4) shows the results for moisture determination. Also, for this parameter the best result was obtained with SPA selection method and the three preprocessing methods, ABS , SNV and SG. The last one with (3,3,1) parameters, that is, mobile average with three points, 3rd degree polynomial and first derivative. These parameters reached an MLR error value of 0.0006 and a regression coefficient of 0.99 (Figure 3.6b). The components selected by SPA algorithm are shown in Figure 3.6a. The components were distributed around the three main peaks (11 (16%) in 1206-1241 nm, 16 (24%) in 1390-1440 nm, 8 (12%) in 1640-1660nm) and 15 wavelengths (22%) in the valley located between 1447-1594 nm.

Table 3.4 Results for moisture estimation.

Algorithm applied	Pre-processing method	Number of components	RMSEV	R_y	RPD
GA-MLR	ABS, SNV, SG(3,0,0)	103	0.0242	0.94	2.89
SPA-MLR	ABS, SNV, SG(3,3,1)	67	0.0006	0.99	116.66
LASSO		18	0.0600	0.80	1.16

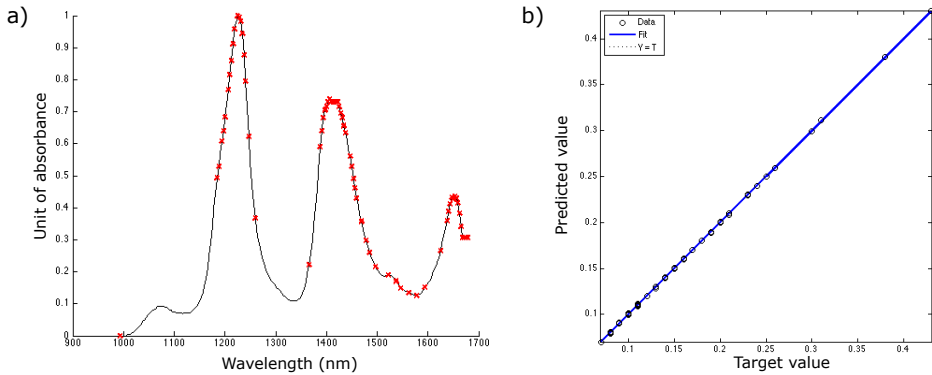


Figure 3.6 a) Components selected by SPA method, pointed with x marks. b) Regression fitted by cross validation..

3.4 Conclusions and future works

The work documented in this chapter has tested the feasibility of using a non-invasive hyperspectral device in order to determinate three interesting parameters in the olive oil extraction process: free acidity, peroxide index and moisture. The optimal wavelengths for the parameter estimation have been identified by using three component selection algorithms: genetic algorithm, least absolute shrinkage and selection operator and successive projection algorithm. The best results have been achieved according to the minimum cross validation error after applying multiple linear regression. In particular, the error values obtained were 0.09% for acidity percentage, $0.03meq.O_2$ for peroxide index and 0.0006% for moisture percentage.

Considering these reduced error values and the well-fitted regression lines, the suitability of using the device in an industrial plant could be concluded. Also, the identification of the optimal wavelengths could allow the reduction of the array sensor size and thus the sensor cost. The chemical reference parameters have been taken from the analysis of the real olive oil samples carried out by the accredited laboratory of CM Europa S.L company.

4 Non-invasive methodology to estimate polyphenol content in EVOO based on NIR and wavelength selection

Contents

4.1. Introduction	73
4.2. Problem description and earlier works	75
4.3. Materials and methods	76
4.4. Experimental results and discussions	86
4.5. Conclusions and future works	90

4.1 Introduction

The commercial olive oil quality is determined by different chemical parameters related to olive oil deterioration. These parameters are, among others, the free acidity and the oxidation state, which are measured by the index of peroxides, K_{232} , K_{270} and ΔK absorption indexes, according to the European Normative 1348/2013 regulation and the International Olive Council [66]. Other parameters such as phenolic compounds, tocopherols or carotenoids, are related to the quality and sensory attributes of olive oil playing a healthy role in the humans [127]. Furthermore, the concentration of these healthy compounds could be interesting for the consumer choice and the competitive strategy of producers.

A modern Virgin Olive Oil factory usually has three different zones which are a reception yard, an extraction facility and the oil cellar. Briefly, the Virgin Olive Oil Production Process is as follows. It begins with the reception of the olives in the factory. These olives are washed in order to remove the dust, small pebbles and leaves that arrive with them. After this preliminary stage, the olives are stored in hoppers and fed into the mill, where they are crushed to form the olive paste. This paste in this state would allow poor separation of the oil, so it is pumped into a thermomixer to be slowly stirred and heated in order to better its conditions for the oil separation. This separation is carried out in a solid-bowl horizontal centrifuge, and yields olive oil and pomace as by-product. The humidity and impurity content of this oil is still undesirably high, so further separation in a vertical centrifuge or a settling tank is performed. After this operation, the oil might be filtered or directly pumped to its final tank to be stored. A more detailed explanation of the process can be found in [25].

Basically the olive oil compounds depend on the fruit physicochemical features, although they are modulated according to the process variables. Through the years, different works have examined the influence of technological operations of olive processing on oil yields and quality [53] [84]; being an initial step for the Process Analytical Technology concept into olive oil sector. Although this concept was introduced in 2004 [69] focused on the pharmaceutical industry, it has been introduced in the food industry recently [187]. The approach involves multivariate data collection through in-line and on-line analytical techniques (such as near infrared spectroscopy, biosensors, Raman spectroscopy by fiber optics) coupled to multivariate data analysis (chemometrics) for continuous learning and information build-up [181]. Thus it has four main components [191]: studying the relationship between product features and process factors; exploring how the process dynamic can affects the sampling procedure; choosing the analytical instruments for in-line or on-line process stream measurements; employing algorithms to multivariate data acquisition and analysis of massive processes data flows.

Also the wavelength or the set of wavelengths that are correlated with the interested compound need to be identified in order to reduce the system cost and the acquisition and computational time. Mainly when the hardware price is the limiting condition of the NIR system deployments.

The goals of this chapter are focused on the second and fourth points of the PAT concept. While the first point develop the correlation model between the olive oil NIR spectra and high concentrations of polyphenols by the employment of algorithms based on feature extraction and feature selection approaches; the second determinates the influence in the prediction models of several acquisition parameters such as number of acquisitions per sample and light source power. All

in the context of the application of an on-line NIR sensor within the VOOPP.

This chapter has been organized as follows: Section 4.2 presents the problem description and describes previous works; Section 4.3 shows the proposed methodology, the experimental setup and the mathematical resources employed to get the results; Section 4.4 explains the outcomes; Finally, Section 4.5 presents the conclusions.

4.2 Problem description and earlier works

Nowadays the master miller is the person in the factory who adjusts the process variables based on his own expertise because the olive oil elaborated is analysed in laboratories far from the factory. It produces lack of objectivity and sometimes the decisions can not be based on the results because of the delay. Furthermore, the methods employed to assess the olive oil quality are invasive, time consuming, expensive and require qualified staff.

Different works have evaluated the use of non-invasive technologies, such as near infrared spectrometers, computer vision or electronic noses and tongues, to monitor the product qualities at all stages of the production process [154] [138] [133] [160] [82] [149]. Particularly, the application of the Near Infrared (NIR) technology on the olive oil industry has been reported in [8]. In that case, authors reviewed the literature related to the use of NIR spectrometry in the olive oil industry over olive fruits, slurries and olive oils. Furthermore an interesting review was done in [35] where authors divided the main applications of olive oil NIR spectroscopy into adulteration detection, prediction of geographical origin, quality parameters determination, online process monitoring and oxidative stability studies.

Particularly the use of NIR in the minor components determination has been reported in several studies. One of the first works that employed NIR to predict the phenolic content in olive oil is detailed in [129]. In that work the authors got a regression coefficient value of 0.89 and root mean square error in cross validation of 58.669 ppm. Similarly, in [131] different PLS regression models for the determination of several extra virgin olive oil minor components were developed from NIR spectra collected on two different spectrophotometers. However they reached poor correlations for total polyphenol content prediction. The R^2 value was 0.34 at best. Most recently the total phenol content prediction from near infrared spectra was studied in [101]. They reached a R^2 value of 0.85 with a PLS regression model and RMSECV of 45.10 ppm. To our knowledge, the last work that deals with NIR spectra to predict phenol content was reported in [2]. In that work a system based on an artificial neuronal network and NIR was designed for real-time

characterization of olive oil during the kneading process. The system was trained to predict polyphenol content among other parameters with a R^2 of 0.89 and root mean square error of prediction of 61.50 ppm. All referred works used olive oil samples with polyphenol concentrations below 1000 ppm.

On the other hand sources of error and how they affect to the calibration need to be considered, analysed and interpreted in the context of the specific application [48]. In fact there are different variables related to the NIR spectra acquisition set-up which need to be studied and adjusted in order to reach a reproducible and repeatable system, such as light source power and optical distance from the lens to the sample among others. For example, when the olive oil contains microdroplets of vegetation water and solid particles from the olive fruit its turbidity gets increased [88] and the power received by the NIR sensor is reduced. Specifically, in [169], the authors evaluated the effect of some parameters such as focal distance and integration time on the spectral repeatability for the analysis of intact olive fruits on a conveyor belt. They employed an InGaAs diode array spectrometer concluding that a focal distance of 13 millimeters and an integration time of 5 seconds gave a higher repeatability than other values. Also, [76] studied a Partial Least Square model to predict free acidity in olive oils. Their aim was to determine the optimal path length and spectral wavelengths. The work showed that the increase of the optical path length in data acquisition results in more reliable PLS models. Other variables such as number of acquisitions and halogen light power have been evaluated in this research.

4.3 Materials and methods

The proposed methodology that has been carried out to correlate and identify the EVOO near infrared spectra with the polyphenols content has been structured in five blocks. The first stage was to select the production process point to get the data source, that is the samples of EVOO, and it depended on the olive oil turbidity. Secondly the spectra acquisition was carried out using the experimental setup built ad-hoc. Also, as will be shown, different acquisition parameters were modified at this stage. Consecutively, in the third stage, the raw frequency values were processed with different pretreatment algorithms to remove non-desired effects into the NIR spectra. After that, in the fourth stage, the spectra were filtered in order to select certain wavelengths related to the interested compound. Finally, the last phase consisted of building the regression model which correlates the absorbance values with the concentration of total polyphenols. At that point two different regression models were proved: Partial Least Squares Regression and Stepwise Multilinear

Regression. Figure 4.1 shows the aforementioned work flow while the details are explained in the following sections.

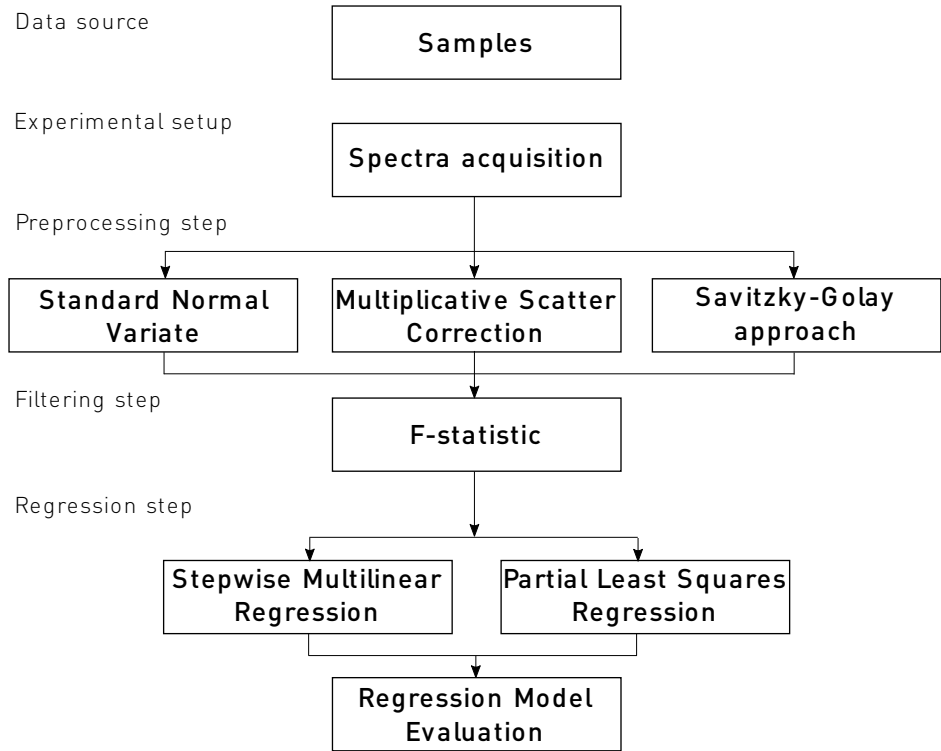


Figure 4.1 Work flow followed for the proposed approach.

4.3.1 Extra virgin olive oil samples

The olive oil samples were obtained from the oil mill directly, at the exit of the vertical centrifugal machine and before the filtering step. At this point the micro-droplets of water and different solid particles have been removed from the olive oil. The samples were elaborated with different process variables such as malaxing temperature, malaxing time and water addition before the oil extraction. Then, olive oils with different content of minor components were reached. Figure 4.2 presents the sampling process. The initial set was composed of 11 samples with polyphenol concentrations from 1017 ppm to 2169 ppm. Finally, these samples were blended to increase the dataset to 21 olive oil samples. The olive oil was monovarietal and the olive fruit variety was mainly Picual. On the other hand, the samples were stored at 6°C in darkness until their chemical and spectral analysis.



Figure 4.2 Olive oil sampling process into the factory.

The analysis of the total content of polyphenols was performed by the Food Biotechnology Department (www.ig.csic.es). Briefly, the method is as follow; 0.6 g of olive oil was extracted using 3*0.6 mL of DMF; the extract was then washed with hexane, and N_2 was bubbled into the DMF extract to eliminate residual hexane. Finally, the extract was filtered through a 0.22 μm pore size and injected into the chromatograph. This method is more detailed in [167].

4.3.2 Experimental setup

The experimental setup is shown in Figure 4.3 where the main components are labelled with numbers. It was composed by the ARCoptix Fourier Transform Near Infrared spectrometer (label number 1) with an spectral range from 900nm to 2600nm and a spectral resolution of 8cm^{-1} . The halogen light source was the model HL-2000-HP-232R from OceanOptix with 20W of nominal bulb power and output power regulation (label number 2). That lamp gives off radiation within the NIR band. The probe (label number 3) was machined to carry out measurements in transreflectance mode. The devices were connected by fiber optic cable. Although the ARCoptix software was used to control the FT-NIR device, the statistical computations were carried out with Matlab R2013b.

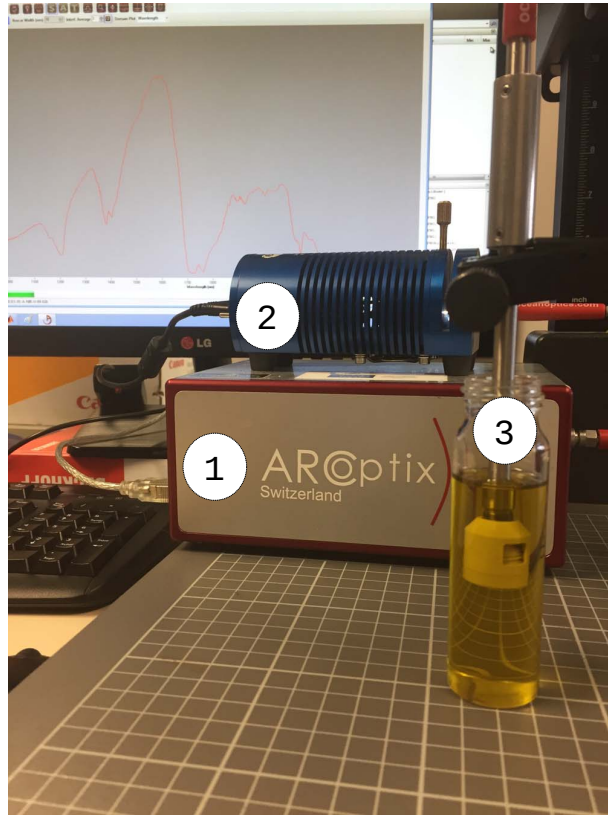


Figure 4.3 Experimental setup where label 1 is the NIR sensor, label 2 is the light source and label 3 is the transmittance probe.

The FT-NIR employed in this work is an interferometer with a Michelson configuration [185]. A beam-splitter is used for dividing the light to be spectrally analyzed into two beams. After they have been reflected on two distinct mirrors, the two beams were recombined by the same beam-splitter and sent to a detector. One of the mirrors is fixed, while the other is a movable mirror. When the fixed and the movable mirrors are equidistant from the beam splitter both beams travel the same distance before reaching the detector. However, if the movable mirror moved away, one of the light beams has to travel an additional distance called retardation. This retardation is used to build the near infrared spectrum.

4.3.3 Spectra acquisition procedure

Firstly, the transmittance spectra of the whole set of olive oil samples were acquired at room temperature ($20 \pm 0.5^\circ\text{C}$) with a constant optical pathlength equal to 5mm.

The transreflectance method is a measurement method employed in spectroscopy and it consists in acquiring the spectrum of the light when previously the light has gone through a liquid sample, posteriorly it has been reflected by a specular surface and finally it has gone through the sample for the second time.

The number of captures were 10 for each sample; however, each acquisition was saved independently. Then the transreflectance spectra were coded in matrix form (Eq. 4.1) where rows (I) are the sample number and columns (Υ) are the light power transmitted by the sample for each wavelength. In our case I was 21 and Υ was equal to 921.

$$X = (x_{iv}, i = 1, 2, \dots, I; v = 1, 2, \dots, \Upsilon) \quad (4.1)$$

The power received by the NIR device follows the classical energy balance and it depends on the transmitted power by the halogen lamp (P_t), the power absorbed by the olive oil samples (P_{abs}) and different losses caused by the connectors, joints, optical fiber paths and other devices as multiplexors in a multipoint system L . The energy balance equation for each wavelength (v) is treated in Eq. 4.2.

$$P_r(v) = P_t(v) - P_{abs}(v) - L(v) \quad (4.2)$$

One of the aims of this work is to present the effect of the halogen light power on prediction results. Therefore, three $P_r(v)$ targets were fixed for each acquisition, 200 μt , 300 μt and 400 μt , where μt are the units of transreflectance and show the raw value obtained by the analogue to digital converter of the NIR device. Also the referenced wavelength v was set in 1600nm because it is the maximum value of olive oil transreflectance spectrum (Figure 4.4). Then, three groups of spectra (S_1 , S_2 and S_3) were built and the output light power was adjusted for each group in order to reach the targeted $P_r(1600nm)$ value (Table 4.1).

Table 4.1 Different groups of spectra considered.

Set number	Number of samples	Target value	Maximum value	Standard Deviation
S_1	21	200	202.39	2.53
S_2	21	300	301.74	2.37
S_3	21	400	403.75	3.38

Target value is the expected power at 1600nm) in units of absorbance in transreflectance mode (μt); Reach value is the average value of the received power for the whole dataset (21 samples) in μt when the power P_t was adjusted.

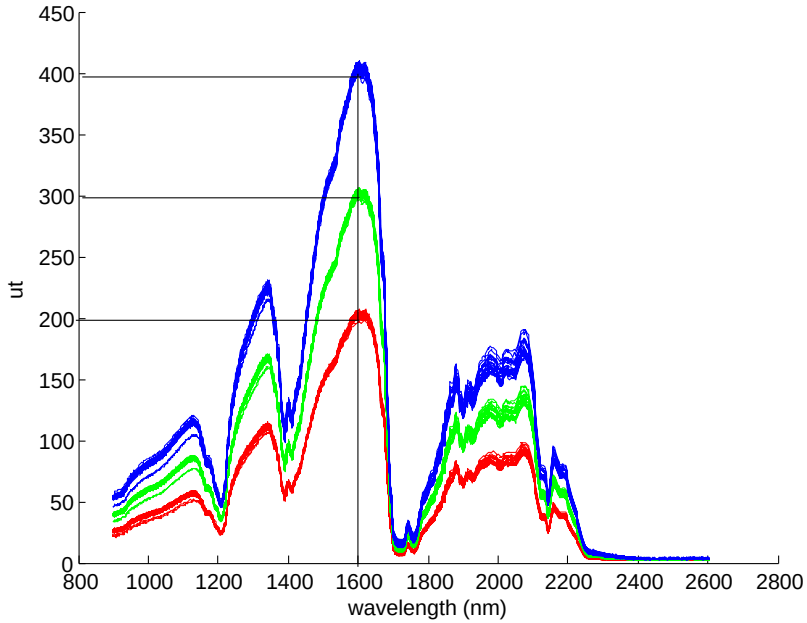


Figure 4.4 Olive oil sample spectra for the three groups: S_1 in red, S_2 in green and S_3 in blue..

4.3.4 Spectra preprocessing

Firstly the raw spectra were transformed from transreflectance values to absorbance values. The spectra of absorbance show the quantity of light which is absorbed at different wavelengths. It is useful when the purpose is to build a calibration model [203]. Therefore, Eq. 4.3 was applied to the whole dataset.

$$X_{abs}(v) = \log_{10} \left(\frac{1}{X_{trans}(v)} \right) \quad (4.3)$$

where X_{abs} are the absorbance spectra and X_{trans} are the transreflectance spectra.

The NIR sensor sensitivity is not the same among the different wavelengths and there are spectral band where the signal-to-noise ratio turns bad owing to low gain (see Figure 4.4, from 2300 nm to 2600 nm). Thus, the Coefficient of Variation (CV) (Eq. 4.4) was applied for each sample in order to detect these spectral bands.

$$CV(v) = \frac{s(v)}{\bar{x}(v)} \times 100 \quad (4.4)$$

where s is the sample standard deviation and $\bar{x}$ is the sample mean for each wavelength v .

Furthermore, spectra preprocessing algorithms were used for correcting low disturbances on the signal. These ones could be produced by scattering effects in the olive oil due to micro-particles or inhomogeneities of the surface and also by differences in the spectrum baseline due to differences in the optical path length or interferences from external light and random noise. The choice of preprocessing methods is difficult to assess before model evaluation. For this reason, in the literature, numerous authors have tested different pretreatments in order to determine the most suitable one. In this work were studied as pre-treatment methods the standard normal variate [11], Multiplicative Scatter Correction (MSC) [79] and Savitzky-Golay [174] derivative method.

The SNV is a pretreatment method used quite often in NIR to remove the scatter added to the spectrum. Its approach is to calculate the average and standard deviation of all the data points for that spectra and each data point of the spectra is subtracted from the mean and divided by the standard deviation. It is applied to all the spectrum individually and it allows to compare different spectra on the basis of the same reference (Eq. 4.5).

$$x_{iv,SNV} = \frac{(x_{iv} - \bar{x}_i)}{\sqrt{\frac{\sum_{v=1}^Y (x_{iv} - \bar{x}_i)^2}{Y-1}}} \quad (4.5)$$

where $x_{iv,SNV}$ is the transformed element for original element x_{iv} , and $\bar{x}_i$ is the mean of spectrum i and Y is the number of variables in the spectrum.

On the other hand, the MSC is a transformation method used to compensate additive and/or multiplicative effects in spectral data. It removes not only physical effects like particle size changes, but also surface blaze from the spectra, which do not carry any chemical or physical information. Firstly, each spectrum is fitted to the average spectrum by least squares (Eq. 4.6).

$$x_i = a_i + b_i \bar{x}_v + e_i \quad (4.6)$$

where x_i is the spectrum of an individual sample i , $\bar{x}_v$ is the mean spectrum of the group and the error e_{ik} corresponds to all other effects in the spectrum that cannot be modelled by an additive and multiplicative constant, in other words, it represents the chemical differences among the olive oil samples.

The corrected spectrum $x_{i,MSC}$ is calculated using the fitted constants a_i (intercept) and b_i (slope) with the Eq. 4.7.

$$x_{i,MSC} = \frac{x_i - a_i}{b_i} \quad (4.7)$$

Alternatively, the derivative approach named Savitzky-Golay method [174] was evaluated in order to remove possible overlapping peaks and correct the spectra baseline. It is a spectral low-pass filtering method where a convolution is used instead of the mean of the spectrum. Its mathematical expression appears in Eq. 4.8.

$$x_{iv,SG} = \frac{\sum_{n=-N/2}^{N/2} C_n x_{i,v+n}}{N} \quad (4.8)$$

where C_n is the convolution coefficient for the spectral value $v + n$, N is the size of the window. The convolution coefficients are obtained using minimum squares adjustment of the points of the spectrum to a polynomial of determined grade.

The result is a spectrum similar to the input one, but smoother. The main advantage of this approach is that it tends to preserve characteristics of the original distribution such as the relative maxima and minima and the width of the peaks, which tends to disappear with other averaging techniques.

4.3.5 Spectra filtering

The amplitude of the absorption at any particular wavelength is determined by its absorptivity and the number of molecules encountered within the beam path of the measuring instrument. This relationship is described by Beer's law [202]. Then the absorptivity is not the same throughout the spectrum and there are wavelengths that provide more information than others related to the studied compounds.

Since the objective was to search the most discriminant wavelengths, certain wavelengths were removed from the dataset by employing the one-way Analysis of Variance (ANOVA). In order to perform this analysis the samples were divided in two groups, the first one with the highest polyphenol concentrations samples and five with the lowest ones.

Then ANOVA compares the relation between inter-class and intra-class variabilities. If the ratio of within-group variation to between-group variation is significantly high, we can conclude that the group means are significantly different from each other. We can measure this using a test statistic that has an F - *distribution* with $(k - 1, N - k)$ degrees of freedom. This test was applied for each wavelength. If the p - *value* for the F - *statistic* is smaller than the significance level, then the test rejects the null hypothesis that all group means are equal and concludes that at least one of the group means is different from the others. The most common significance levels are 0.05 and 0.01. In our case features with p - *value* lower than 0.05 were candidates to be removed.

4.3.6 Dimensionality reduction with regression models

There are two general approaches for performing dimensionality reduction: feature selection models and feature extraction models [207]. The first one identifies a subset of features without applying any transformations and the second one employs mathematical operations in order to transform the original features into a lower dimensional space. In this study has been evaluated one model for each type. Step-wise Multilinear Regression (SMLR) [99] as feature selection model and Partial Least Squares Regression (PLSR) [21] as feature extraction model, using the polyphenol content as dependent variable (Y) and wavelengths as independent variables (X).

SMLR uses forward and backward stepwise regression to build the final model. At each step, the algorithm searches for wavelengths to add or remove from the model according to a specific criterion. In our case the criterion was to use the statistical p -value and F -value to test models with and without a potential wavelength at each step. Then if a wavelength is not currently in the model, the null hypothesis (H_0) is that the coefficient attached to the wavelength would have a zero value if it is added to the model (forward approach). If the null hypothesis is rejected the wavelength is added to the model. Furthermore, if a wavelength is currently in the model, the H_0 is that the coefficient of the wavelength is equal to zero. If there is insufficient evidence to reject the null hypothesis, the wavelength is removed from the model (backward approach). The algorithm was applied throughout the next steps:

- *Step 1:* Fit a linear model with β_0 and the set of selected (B) wavelengths:

$$Y = \beta_0 + X * B \quad (4.9)$$

- *Step 2:* For β_k (k from 1 to number of wavelengths) not in B iterate from 3 to 6.
- *Step 3:* Adjust the lineal model:

$$y_n = \beta_0 + \sum_B x_{n,k} \beta_k \quad (4.10)$$

- *Step 4:* Null hypothesis: $H_0 : \beta_k = 0$
- *Step 5:* F -value and p -value are obtained.
- *Step 6:* If p -value < 0.05 : β_k is added to the set B and go to Step 1.

On the other hand, PLSR is a bilinear regression method that extracts a small number of estimated factors t_a , $a = 1, 2, \dots, A$ that are linear combinations of the X variables, and uses these factors as regressors for y . In addition, they are also used as regressors for X itself, in order to obtain residual spectra, which are useful as

interpretation and diagnostic tools [21]. Then, the X variables themselves are also modelled by these regression factors. The model for regressions with estimated latent variables can be summarized by Eq. 4.11-4.14.

$$T = w(X) \quad (4.11)$$

$$X = p(T) + E \quad (4.12)$$

$$y = q(T) + f \quad (4.13)$$

$$y = q(w(X)) + f = b(X) + f \quad (4.14)$$

where w , p , q and b are lineal functions and their values are estimated statistically based on empirical data. It employs the principle of least squares in this estimation. On the other hand E and f are residuals and they represent the noise and the model errors.

Finally, the regression model was evaluated with leave one out cross validation [183]. It is an iterative validation wherein one sample x_i is removed from the dataset in each iteration. The number of iterations is equal to the number of samples (N). Then the model is built with the rest of samples in the dataset and it predicts the y_i related to the previously removed sample x_i . The predicted value is represented as $\hat{y}_i$. At the end the Root Mean Square Error (RMSE) is obtained according to the Eq. 4.15.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}; \quad (4.15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (4.16)$$

Furthermore, the determination coefficient R^2 is calculated in order to evaluate how the variability of the response variable Y is explained by the regressor variable X . It is explained with the Eq. 4.16.

4.4 Experimental results and discussions

The noisy spectral bands of the NIR sensor was detected by applying the CV coefficient to the whole wavelengths. The criterion was to remove the components whose CV value was greater than its mean value. Then, Figure 4.5 shows that the noisy spectral bands are located between 2240 and 2600 nm and it is due to the lose of power caused by the experimental setup.

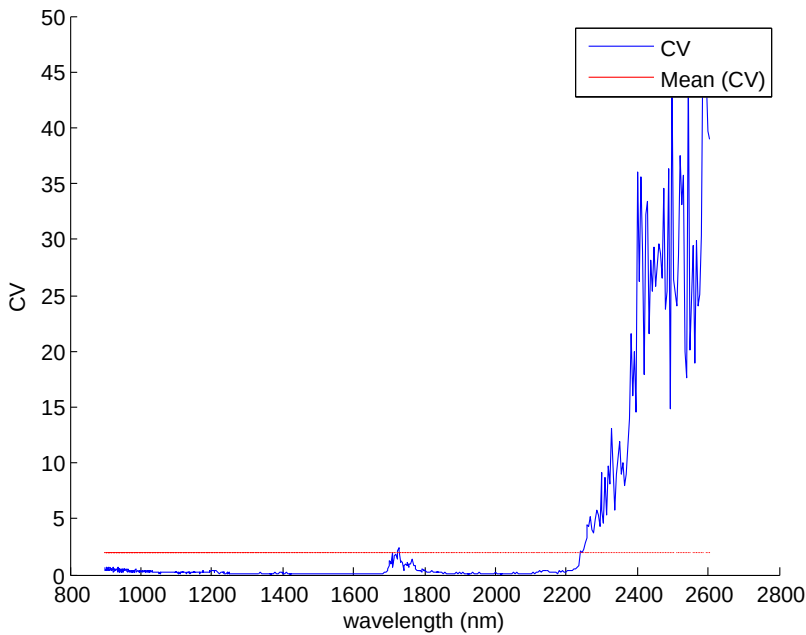


Figure 4.5 Blue line represents the CV coefficients for each wavelength; Red line is the CV mean value.

Spectra were preprocessed with the aforementioned algorithms and the F – statistic and the p – value were obtained for each wavelengths. Significant wavelengths (with p – value lower than 0.05) were selected (up to a limit of 100 components) while the rest were removed from the feature vector. After that, the wavelengths were arranged in descending order according to their F – statistic values. Figure 4.6 shows the preprocessing spectra for the whole dataset and the chosen components. The number of these significant components was 57 for spectra without preprocessing, 100 for SNV, 100 for MSC and 98 for the SG approach. The majority of these components were located at the spectral bands of 1185 nm - 1245 nm and 1673 nm - 1965 nm. Other works concur in the relation between this spectral band and the

OH stretch first overtone and OH deformation combination of hydroxyl groups [47] [101].

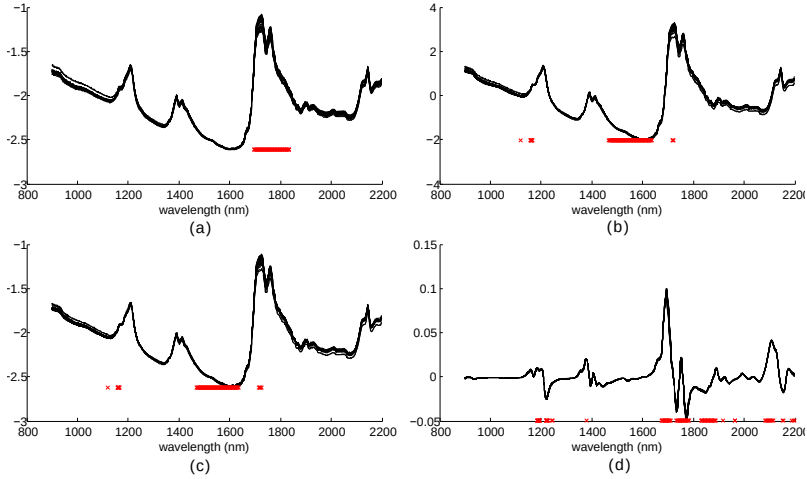


Figure 4.6 Figure shows the selected wavelengths after the filtering step (red marks) and the spectra of the whole dataset for each preprocessing algorithms: (a) without preprocessing, (b) Standard Normal Variate, (c) Multiplicative Scatter Correction and (d) Savitzky-Golay approach.

Regression models were applied with the selected wavelengths and the experimental results were evaluated on the basis of the prediction error. Table 4.2 shows the calibration and validation performance in terms of *RMSE* and R^2 for each preprocessing and regression algorithm. The best result was reached by SMLR method with SG approach, obtaining a R_v^2 coefficient of 0.80 and a *RMSEV* value of 156.14 ppm, that is a 13.55% over the reference range of polyphenol content. Figure 4.7 presents the correlation between the target and the predicted values for the calibration and validation phases. The studies developed in [101] reached similar error values but they used different reference range of polyphenol concentrations (below 1000 ppm) using PLS regression models with nine components, and the standard deviation was smaller than the dataset employed in this work. In our case, the best result was reached when the SMLR algorithm fitted a linear model with three components. That model appears in Eq. 4.17.

$$\hat{y} = \beta_0 + \beta_1 \cdot x_{11} + \beta_2 \cdot x_{38} + \beta_3 \cdot x_{72} \quad (4.17)$$

where $\hat{y}$ is the predicted value, x_{11} , x_{38} and x_{72} are the absorbance values at 1830 nm, 1917 nm and 2189 nm respectively and β_0 , β_1 , β_2 and β_3 are the estimated regression coefficients and their values are -5453 , $-4.16 \cdot 10^5$, $1.56 \cdot 10^5$ and $1.92 \cdot 10^5$ accordingly.

Table 4.2 Experimental results.

<i>RA</i>	<i>PA</i>	<i>NCF</i>	<i>NCR</i>	RMSEC (ppm)	R_c^2	RMSEV (ppm)	R_v^2
<i>SMLR</i>	ABS	57	3	155.05	0.80	188.29	0.70
<i>SMLR</i>	ABS+SNV	100	1	191.56	0.69	211.25	0.63
<i>SMLR</i>	ABS+MSC	100	1	189.62	0.70	208.23	0.64
<i>SMLR</i>	ABS+SG	98	3	124.78	0.87	156.14	0.80
<i>PLSR</i>	ABS	57	5	51.27	0.97	222.69	0.59
<i>PLSR</i>	ABS+SNV	100	5	22.91	0.99	211.71	0.63
<i>PLSR</i>	ABS+MSC	100	5	22.55	0.99	210.00	0.63
<i>PLSR</i>	ABS+SG	98	8	2.49	0.99	178.63	0.73

where *RA* is regression algorithm; *PA* is preprocessing algorithm; *NCF* is number of components after filtering; *NCR* is number of components after regression.

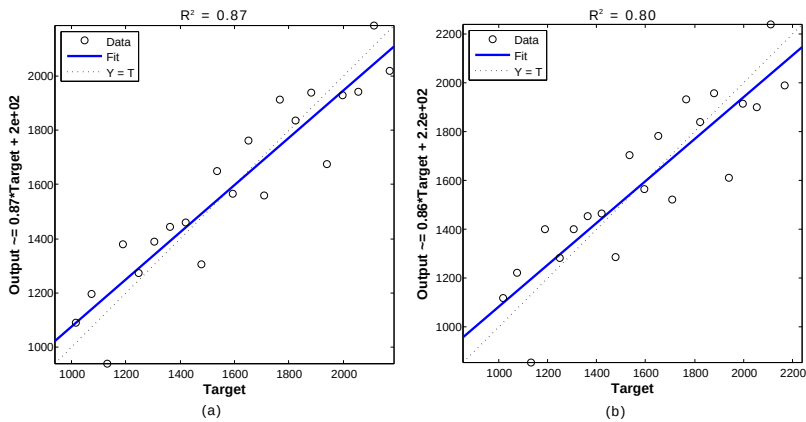


Figure 4.7 Predicted values versus target values when the built linear model is applied. Figure (a) represents the results for calibration and (b) for validation.

On the other hand the preprocessing SG algorithm was employed by applying the first derivative to the spectra, with a fitting polynomial of second degree and a windows size of 11 points. The last parameter was selected experimentally and Figure 4.8 presents the *RMSE* results for each windows size. It suggests that the windows size of 11 components reaches the best Root Mean Square Error of Calibration (RMSEC) and RMSEV results.

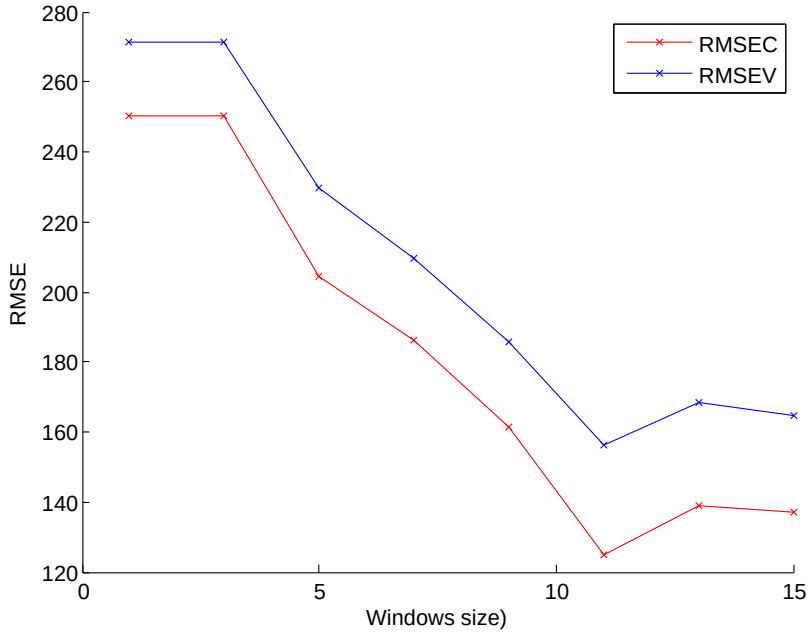


Figure 4.8 Blue and red lines represent the RMSEV and RMSEC values respectively for each windows size when Savitzky-Golay preprocessing algorithm is employed.

Finally, the lineal model was evaluated with different acquisition parameters (maximum absorbance value and number of acquisitions per sample). As indicated in the previous section different set of samples were built (S_1 , S_2 and S_3). The same *SMLR* algorithm was applied for each set and the results are supplied on Figure 4.9a. It shows a decreasing trend in *RMSE* value according to the increase of the maximum absorbance value. However, it depends on the maximum output power of the halogen lamp as restriction. Furthermore, the number of acquisitions per sample was modified (from 1 to 10) in order to evaluate how it affects to the performance. The S_3 set was employed for this experiment and Figure 4.9b shows the results. Since the spectra is averaged the noise is reduced. It occurs up to 6 acquisitions where the minimum *RMSEV* value is reached. These results are useful to adjust the optimal number of acquisitions and the halogen lamp features in order to adjust and import the setup in the olive oil elaboration process.

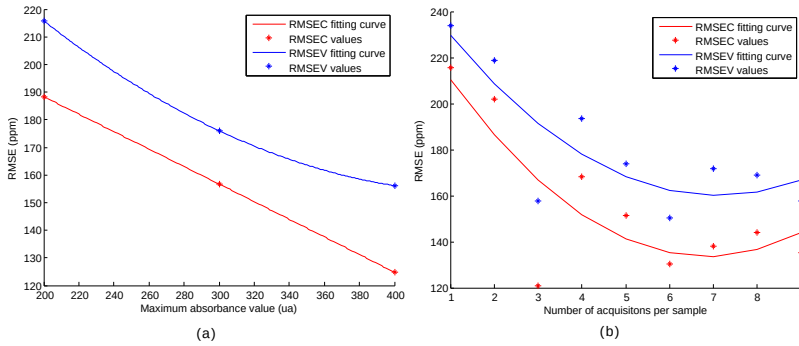


Figure 4.9 Figures (a) and (b) show the *RMSEC* and *RMSEV* values when different light intensities were employed and different number of acquisitions per sample were averaged respectively.

4.5 Conclusions and future works

This work describes a methodology to obtain and to improve the correlations between olive oil NIR spectra and polyphenol content. The best correlated wavelengths were identified and two acquisition parameters were modified in order to reduce the cross validation error. The proposed approach was based on five stages: selection of samples, spectra acquisition with the experimental setup, preprocessing step, filtering step and regression model implementation and validation. Different preprocessing algorithms were evaluated such as Standard Normal Variate, Multi-variate Scatter Correction and Savitzky-Golay method. The preprocessed spectra were filtered by selecting the significant components and two specific regression algorithms were employed (PLSR and SMLR) based on feature extraction and feature selection approaches, respectively.

The results showed the existence of correlations between olive oil NIR spectra and polyphenol content by using the proposed methodology. The best results were reached with the lineal model obtained by the *SMLR* regression algorithm with Savitzky-Golay preprocessing algorithm. It got a *RMSE* value of 156.14 ppm with a R^2 value of 0.80 when leave one out cross validation was applied. Furthermore, the significant wavelengths included on the linear model were 1830 nm, 1971 nm and 2189 nm.

Finally, the linear regression model was applied with different acquisition parameters such as halogen lamp power and number of acquisition per sample. These parameters need to be taken into consideration in order to design an industrial system. The model performance improved lineally with the increase of the halogen lamp output power. Also the optimal number of acquisition to average were 6 for

our purpose, where $RMSEV$ decreased to 150.23 ppm. As future work other factors could be taken into consideration such as olive oil temperature or moisture content.

5 Expert system for monitoring the malaxing state of the olive paste based on computer vision

Contents

5.1. Introduction	93
5.2. Problem description	95
5.3. Materials and methods	97
5.4. Experimental results and discussions	105
5.5. Conclusions and future works	108

5.1 Introduction

Within the elaboration process of virgin olive oil, the malaxing of the olive paste is one of the most important subprocesses. Mainly, the initial state of the olive paste that enters to the mixing process depends on the physical and chemical conditions of the olive milled. In turn, these conditions depend on the harvest season of the fruit and its variety. An incomplete or defective kneading affects the subsequent extraction process adversely, and may reduce both the quality and the quantity of the elaborated oil.

There are published studies that have examined the influence of the variables involved in the malaxing subprocess on the quality and extraction yield for the

olive oil elaboration process [45]. Specifically, the increase of the malaxing time increases the extraction yield up to a certain time threshold. If this limit is exceeded, the nutritional characteristics of the processed oil will be decremented. Likewise, the temperature increase has positive effects in the extraction yield, but it causes negative results in the content of polyphenols and in the presence of volatiles compounds.

On the other hand, fuzzy techniques are employed in the food industry in order to supervise processes and reach determined quality objectives. These techniques allow to model to some extent the expert knowledge of the operators of the process and to use it as a sensor for variables where there are no hardware sensors available [158].



Figure 5.1 Malaxing of olive paste inside the thermomixer.

Currently, the automation degree of the olive oil elaboration process is in a preliminary stage [25]. Particularly, the malaxing state control of the olive paste is performed by the master miller, who observes periodically the paste inside the thermomixer and acts on variables manually. This operating mode depends on the frequency that the master miller looks at the paste and his subjective judgment. It can cause absence of uniformity in making the decisions and lack of robustness.

The automatic control of the malaxing state of the olive paste in real time allows adjustment of the malaxing time needed to prepare the olive paste to the minimum value, which results in an increase in the productive capacity of the plant. Also it lets the rapid detection of difficult pastes, by means of the adaptation of the beat process to its features. Thus, it avoids olive oil leak in the pomace due to olive paste unprepared. Furthermore, the aforementioned supposes a significant

improvement in the energy consumption of the plant.

The first contribution of this paper is based on the determination of the olive paste parameters which determine the malaxing state through image processing and the second one is the implementation of an automatic control system of the malaxing state of the olive paste by using the previous parameters. To this purpose, we have included computer vision sensors in the thermomixer. From this hardware setup, we propose the use of the following variables to control the process [44]:

- Floating Oil (FO): This is the most obvious indicator. The presence of floating oil on the olive paste, makes the subsequent separation process easier.
- Viscosity (V): The emulsion of oil droplets with the other olive paste components increases the paste viscosity. This state is indicative of unprepared or difficult pastes.
- Granularity (G) of the paste surface: The oil droplets coalescence causes the clustering of the particles of the olive paste solid phase. Therefore the size of the lumps of paste formed on its surface is an indicator of the paste state.

The resulting system has a dual purpose: assisting the master miller to characterize the paste continuously and to close the control loop in order to guarantee the state of the olive paste. To the best knowledge of the author, is the first time that the aforementioned methodology is employed to achieve the referred objectives.

The structure of this chapter is as follows: the next Section 5.2 exposes the description of the problem; then, the Section 5.3 explains the design of the proposed solution. It describes the methodology used for the implementation of the malaxing state sensor of the paste and it shows the controller design; the Section 5.4 presents the results and discusses the experimental results obtained after applying the proposed methodology; finally, the Section 5.5 summarizes the conclusions.

5.2 Problem description

Within the elaboration process of virgin olive oil, the malaxing of the olive paste is the subprocess after the grinding stage and the previous stage to the centrifugal subprocess (see Section 2.3 for a complete description of the process). Its purpose is to reach the first separation of the different phases of the olive paste, so that it is ready for the subsequent extraction process. This sub-phase completes the grinding of the olive paste parts treated inadequately in the mill, while its temperature is increased to cause the aggregation of the oil droplets dispersed in the ground paste.

The malaxing process is carried out in the thermomixer machines. In the olive oil

elaboration process are used basically two types of mixers that are distinguished by the arrangement of the blade axis, such as vertical and horizontal thermomixers (Figure 5.1). The horizontal thermomixers are usually distributed in different sections and stacked vertically. The number of sections mainly depends on the milling capacity of the extraction line. Inside these sections is provided a system of blades that beats the paste and gradually leads it through the section, where the paste exits by an overflow. The walls of the mixer are hollow so as to permit the hot water circulation inside them. This fact produces the heating of the olive paste. To increase the heat exchange surface, each section of the mixer is usually divided into several separated compartments by heat exchangers. This distribution makes the heating more progressive and avoids drastic changes of temperature that can affect to the quality of the elaborated product adversely.

Nowadays the master miller assesses the degree of difficulty of the olive paste by means of the direct observation of the mixing process inside the thermomixer periodically. Mainly, the aspects on which he bases his assessment are the degree of soiling of the blades, the granularity of the olive paste and the presence of floating oil on the paste surface. Thus, if the olive paste has low granularity, the blades of the thermomixer are completely dirty and there is no floating oil, the paste will be considered as unprepared. In this situation, with the objective of the extraction yield increase, the master miller should act on the hot water inlet valve to rise its flow, and therefore the temperature of the paste too. Otherwise, if the olive paste is very grainy, shows the oil phase and the blades are completely clean, the paste will be considered as prepared.

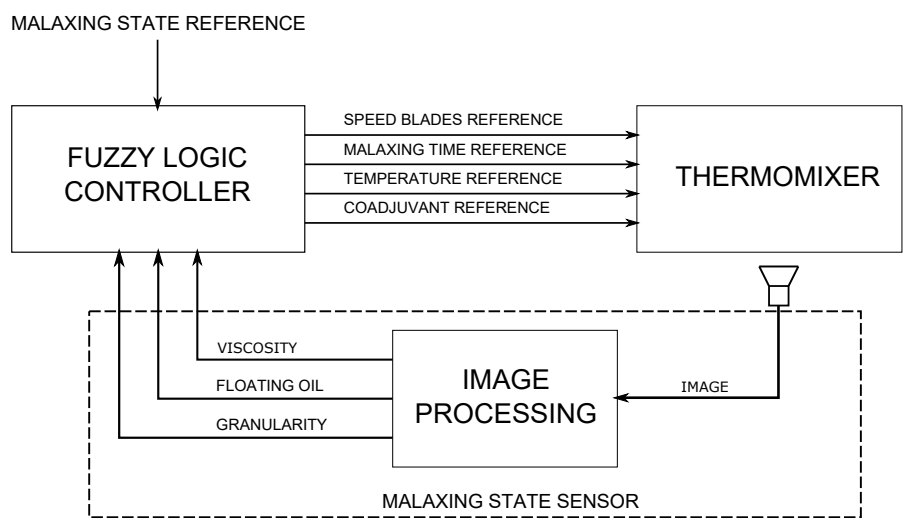


Figure 5.2 Block diagram of the control system for the malaxing paste state.

The works carried out in this chapter tries to solve the problems inherent to manual control of the malaxing state, daily carried out by the master miller.

5.3 Materials and methods

The proposed solution is based on designing a control system as is shown in the Figure 5.2. This system consists in two distinct parts. The first one is the software implementation of a paste state sensor that obtains the olive paste characteristics from the acquired images inside the thermomixer. And the second one is the design of a fuzzy logic based controller in order to reach the paste state decided by the master miller, with the information received by the sensor.

In the following sections, the experimental setups employed to configure the malaxing state sensor will be explained together with the main features extracted from the images. The controller design will be presented subsequently.

5.3.1 Experimental setup

The experimentations were carried out with the hardware deployed in a real oil mill. Two experimental setups were installed. The first computer vision system was installed in the thermomixer in order to acquire on-line images. Furthermore, the second setup was located next to the thermomixer and it was employed to capture images at-line by means of sampling.

Figure 5.3 shows the three sections thermomixer used for this study. The vision system was installed in the lower section. The absence of a vision system in the upper sections was due to inadequacy of its location. Moreover one computer PC was installed to receive and store the images, and all the associated electronics.

The vision system was composed of one 100W halogen lamp and one multispectral vision camera which model is JAI AD-080 GE with F1.8/6mm lens. This camera was used to acquire visual information from the olive paste in the infrared and visible spectrum. The camera was protected by a stainless steel encapsulation, with forced ventilation built to avoid the cameras overheating due to the high temperature reached by the thermomixer.

On the other hand, Figure 5.4 presents the second experimental setup installed next to the thermomixer in order to acquire images at-line. It was composed by one JAI AD-080 GE camera and two 100W halogen lamps. Also a conveyor belt was employed to locate the sample within the field of view of the camera.

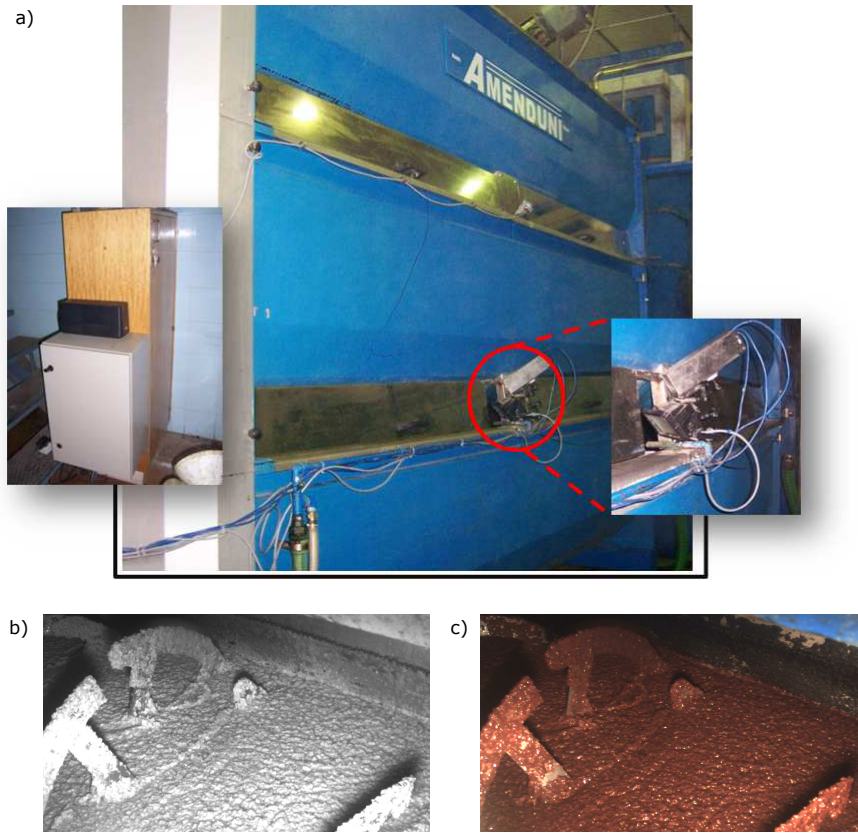


Figure 5.3 Image acquisition set-up. Subfigure (a) is composed by the vision camera and the illumination lamp which are installed in the thermomixer. At the left of this subfigure is the cabinet that houses the computer PC and the electronic associated. Subfigure (b) presents an example of image acquired by the infrared sensor. Subfigure (c) displays the same scene in the visible colour channel.

The first experimental setup was used to the viscosity and floating oil determination, whereas the second one was employed to obtain the granularity parameter.

5.3.2 Viscosity determination

The viscosity of the olive paste during the malaxing subprocess was estimated by the monitoring of the amount of adhered paste to the blades of the thermomixer in the lower section. This thermomixer section was chosen because it is the only one in which the blades are always visible, and they are kept above the level of the paste. So that the paste is more viscous when blades area are dirtier and vice versa.

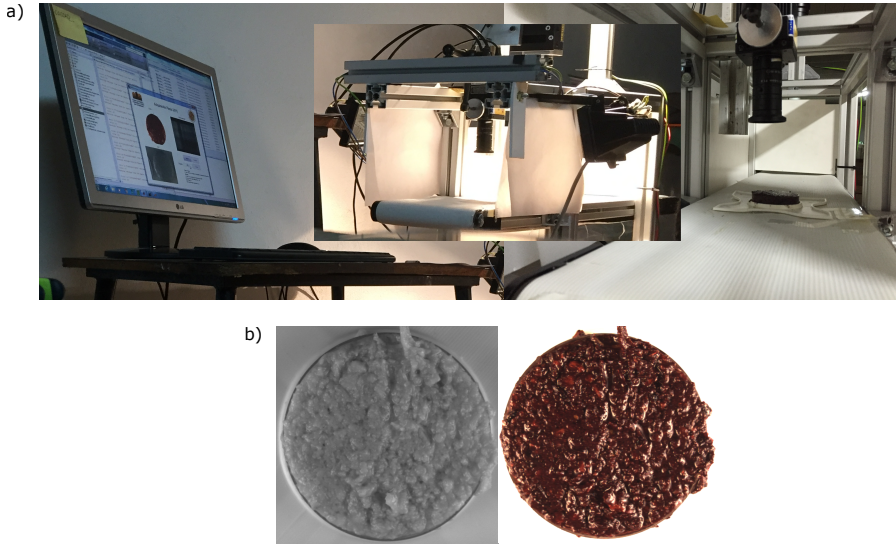


Figure 5.4 Image setup for at-line acquisitions. Subfigure (a) is composed by the vision camera, two halogen lamps and the conveyor belt. At the left of this subfigure appears the PC which handles the experimental setup. To the left of this subfigure is the Petri dish with a sample of paste below the vision device. Subfigure (b) displays the acquired images in the infrared (left) and colour (right) channels.

The quantity of paste adhered to the blades was determined by the processing of the images captured with the JAI AD-080 GE vision device.

A sequence of images were monitored from dirty to clean shovels. Figure 5.5 displays the first image of the sequence for each camera vision channel. It shows the difference of tones between the olive paste and the stainless steel blade. The difference is higher in the infrared channel. Then a binary mask was applied in order to consider only the shovel area as the region of interest. After that, the quantity of paste attached to the shovel was computed as the average intensity according to the Eq. 5.1. This operation was performed for each image of the sequence.

$$\bar{I}_c = \frac{\sum_{n=1}^N \sum_{m=1}^M I_c(n,m) \times I_{mask}(n,m)}{N \cdot M} \quad (5.1)$$

where $I_c(n,m)$ denotes the gray value for each pixel in the c image channel (red, blue, green and infrared); N and M are the number of rows and columns respectively; $I_{mask}(n,m)$ is the designed mask where the pixels out of the shovel have a value of 0 and the pixels which match with the the shovel have a value of 1.

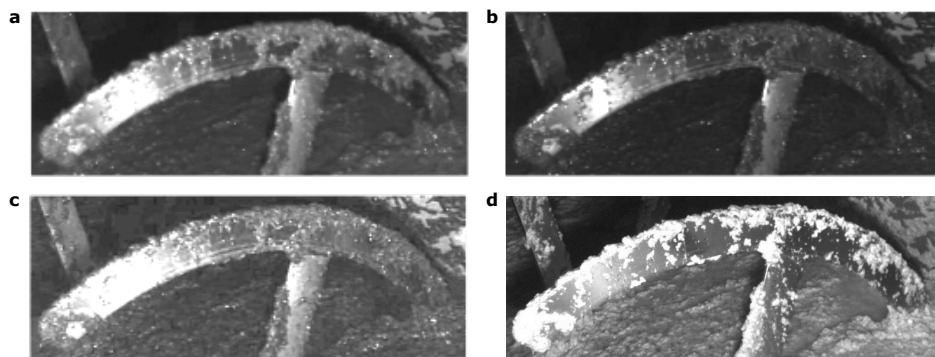


Figure 5.5 These figures show the olive paste adhered to one of the blades of the thermomixer in the red channel (a), green channel (b), blue channel (c) and infrared channel (d). In the last one the olive paste is more contrasted.

5.3.3 Floating oil determination

The images acquired by the JAI AD-080 GE vision device were analysed in the infrared and colour spectrum, to estimate the amount of oil floating on the olive paste. Usually, the floating oil is accumulated close to the wall for this thermomixer type. Then the region of interest was selected to obtain a sequence of olive paste surface images most likely to have floating oil and without the presence of shovels.

As in the previous parameter, a sequence of images were acquired from absence to presence of floating oil. The Figure 5.6 shows two images, the first and the last images of the sequence. Also in this case the Eq. 5.1 was applied to obtain the average intensity for each colour channel.

5.3.4 Olive paste granularity determination

The texture of the olive paste inside the thermomixer can be influenced by the movement of the shovels and by the presence of floating oil too. It could be fixed with a structural change in the thermomixer (e.g. by removing one shovel). Therefore, an experimental setup was built at-line in order to evaluate the granularity of the olive paste surface.

A batch with 4000 Kg of olives was milled and then malaxed for 120 minutes. The purpose of this experiment was to obtain an evolution of olive pastes from unprepared to prepared in excess. Three samples of olive paste were manually taken from the thermomixer every 10 minutes. A total of 36 samples were collected and prepared in Petri dishes.

The setup was composed by the JAI AD-080 GE vision device and images in the

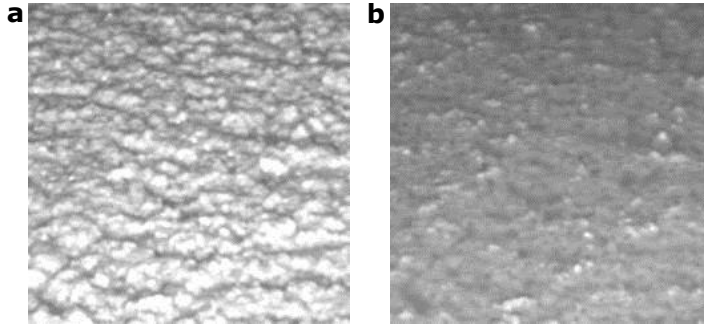


Figure 5.6 Infrared images of the surface of the olive paste inside the thermomixer. The image (a) is an example of paste without floating oil, and the image (b) shows the olive paste flooded by oil. Note their different tones.

infrared and colour space were also acquired. The Petri dish shape was the region of interest for each image. For this Region of Interest (ROI), the gray level variation between neighboring pixels produces the needed information to apply the statistical analysis method based on the Gray Level Co-occurrence Matrix (GLCM) [98] [97].

Grey level co-occurrence matrix is one of the mostly used statistical texture analysis method for the evaluation of food qualities [208], in which texture feature is extracted by some statistical approaches from the co-occurrence matrix, $p(i,j)$. Before building the matrix, two parameters θ (direction of the pixel pairs) and d (distance between the pixel pairs) need to be chosen. The directions θ can be selected from four different values of 0, 45, 90, and 135 while distance depends on the resolution of texture. In most cases in the food industry, d is usually selected according to the properties of the foodstuff being studied. In order to obtain the olive paste GLCM matrix, the matrix was averaged firstly for all possible directions θ and d equal to 1. Secondly, the matrix was normalized with the Eq. 5.2. In this case, one GLCM matrix was obtained for each colour channel, sample and number of repeated sample.

$$P_{c,n,m}(i,j) = \frac{p_c(i,j)}{I \cdot J} \quad (5.2)$$

where $(I \cdot J)$ is the sum of all matrix elements; c denotes the colour channel (red, green, blue or infrared); n is the number of sample during the malaxing process (from 1 to 12); m is the number of the repeated sample (from 1 to 3).

Different features such as contrast, correlation and homogeneity were obtained from this co-occurrence matrix. Then, in order to quantify the olive paste granularity,

Eqs. 5.3 5.4 and 5.5 were used.

$$C_{c,n,m} = \sum_{i,j} |i-j|^2 P_{c,n,m}(i,j) \quad (5.3)$$

$$E_{c,n,m} = \sum_{i,j} P_{c,n,m}(i,j)^2 \quad (5.4)$$

$$H_{c,n,m} = \sum_{i,j} \frac{P_{c,n,m}(i,j)}{1 + |i-j|} \quad (5.5)$$

where $C_{c,n,m}$ is the contrast parameter and it returns a measure of the intensity contrast between a pixel and its neighbour over the whole image and it is 0 for a constant image; $E_{c,n,m}$ is the energy parameter and it returns the sum of squared elements in the GLCM; $H_{c,n,m}$ is the homogeneity parameter and it returns a value that measures the closeness of the distribution of elements in the GLCM matrix to the matrix diagonal. It is 1 for a diagonal GLCM. The final goal of this experimentation was to evaluate what parameter is the most sensitive to tiny changes in lump sizes over the olive paste surface.

5.3.5 Fuzzy controller design

The control system was designed according to the diagram of blocks that appears in the Figure 5.7. It is composed of the following units: the fuzzy unit, in which the physical values obtained from the malaxing state sensor are transformed into fuzzy variables; the inference system was based on the Mamdani's method [130] and it implements the fuzzy set outputs as a result of the group of defined rules that are related to the expert knowledge of the master miller; and, finally, the defuzzyfier unit transforms the fuzzy sets obtained from the fuzzy unit into updated set-points for different output parameters. The membership functions have been adjusted manually with the Fuzzy Logic Toolbox of Matlab.

Firstly, the membership functions of the input variables to the controller were defined (Figure 5.8). The variation ranges of the input parameters of these membership functions were set according to the features obtained from image processing which have been discussed in the previous subsections. The input parameters to the controller were: paste granularity, amount of floating oil and dirt of the blades of the thermomixer.

Thus, the universe of discourse for each variable obtained from the image processing step was scaled from 0 to 1 relating to the minimum and maximum value for each parameter. In addition, in order to relate the human visual perception with

the variable values, different fuzzy sets were defined: 'Very Low' (VL), 'Low' (L), 'Normal' (N), 'High' (H) and 'Very High' (VH).

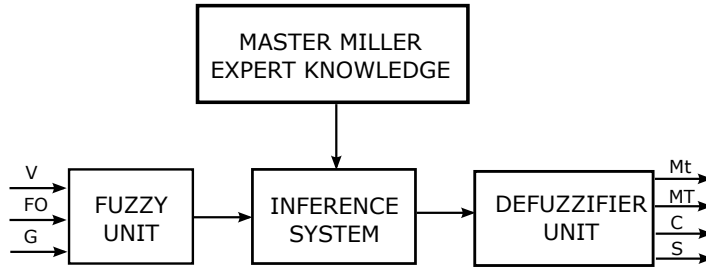


Figure 5.7 Fuzzy Rule-Based System. Viscosity, floating oil and granularity (V, FO and G respectively) are the inputs. Malaxing time, malaxing temperature, coadjuvant addition and speed of the shovels (Mt, MT, C and S respectively) represents the outputs.

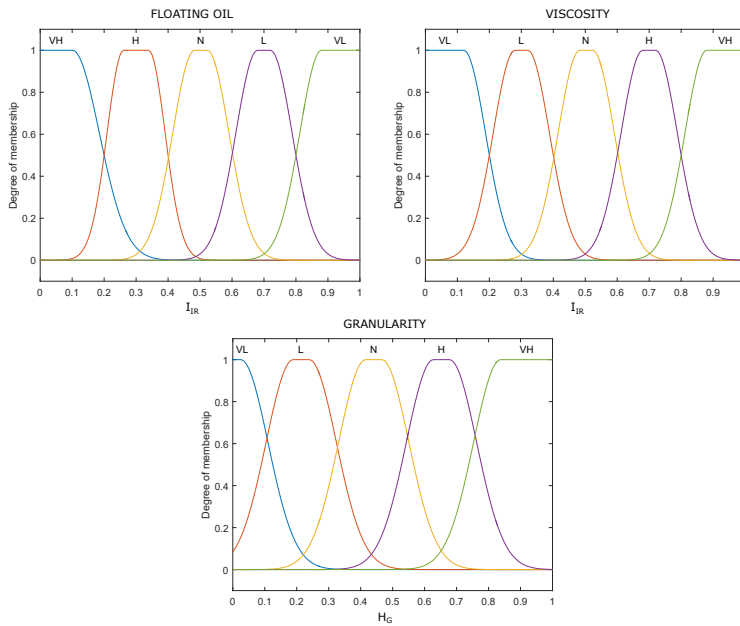


Figure 5.8 Input fuzzy membership functions.

On the other hand, the membership functions of the output variables are shown in Figure 5.9. They were configured as follow:

- Malaxing time (Mt): The malaxing time variable denotes the amount of time the olive paste is stirring inside the thermo-mixer. The universe of discourse of

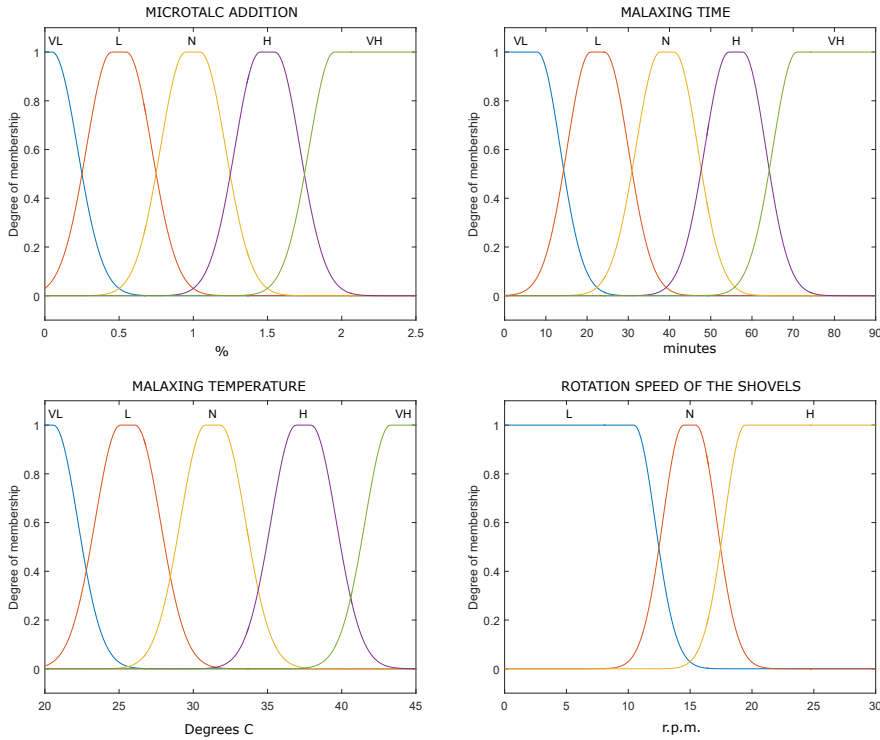


Figure 5.9 Input fuzzy membership functions.

this variable is between the minimum value of time needed to prepare the olive paste (around 20 min) and the maximum value of time to avoid formations of emulsions (close to 90 min) [44]. The malaxing time in the thermomixer can be modified by acting on the olive paste outflow.

- **Malaxing Temperature (MT):** The reference temperature ranges between 20°C and 40°C [44]. Usually the temperature of the paste can be controlled by acting in the servo-valve that regulates the inflow of hot water.
- **Coadjuvant (C) addition:** Coadjuvant addition is done by pumping from the container. In this study has been considered microtalc whose percentage varies from 0% to 4% [44], depending on the type of pasta.
- **Rotating speed of the shovels (S):** Normally, its value ranges between 10 and 20 rpm. This value needs to be adjusted to the minimum with difficult pastes for not favouring the appearance of emulsions [44].

Finally, the combinational logic rules were defined according to the common operation mode where the elaboration objective is to maximize the quantity of olive

oil produced without loss of quality. They were formulated as follows:

$$\text{If : } V \text{ is VL \& FO is VH \& G is VH} \quad (5.6)$$

$$\text{Then : } S \text{ is H \& MT is VL \& Mt is VL \& C is VL}$$

$$\text{If : } V \text{ is L \& FO is H \& G is H} \quad (5.7)$$

$$\text{Then : } S \text{ is H \& MT is L \& Mt is L \& C is L}$$

$$\text{If : } V \text{ is N \& FO is N \& G is N} \quad (5.8)$$

$$\text{Then : } S \text{ is N \& MT is N \& Mt is N \& C is N}$$

$$\text{If : } V \text{ is H \& FO is L \& G is L} \quad (5.9)$$

$$\text{Then : } S \text{ is L \& MT is H \& Mt is H \& C is H}$$

$$\text{If : } V \text{ is VH \& FO is VL \& G is VL} \quad (5.10)$$

$$\text{Then : } S \text{ is L \& MT is VH \& Mt is VH \& C is VH}$$

5.4 Experimental results and discussions

The viscosity index was evaluated from the colour intensity index $\bar{I}_c$ computed for each image channel by monitoring the quantity of olive paste attached to the shovels. Figure 5.10 represents the $\bar{I}_R$, $\bar{I}_G$, $\bar{I}_B$ and $\bar{I}_{IR}$ parameters versus the malaxing time for the red, green, blue and infrared channels respectively. As the quantity of paste attached to the shovel is decreased, the $\bar{I}_c$ index is decremented. The results prove that the infrared channel was more sensitive than the rest of the channels. The slope was negative to the point number six when there are no visual changes.

Furthermore the floating oil index was achieved by monitoring the olive paste surface inside the thermomixer. Also the average intensity for each channel was computed. As in the viscosity index case, the results show that the infrared channel (I_{IR}) was more sensitive than other channels to the presence of floating oil over the paste. The differences between channels can be clearly observed in the Figure 5.11.

The textural parameters $C_{c,n,m}$, $E_{c,n,m}$ and $H_{c,n,m}$ were individually computed for each image. Then the m dimension was removed by averaging the values which are related to the same sample. After that, these parameters were scaled from 0 to 1 in order to compare the resultant values. Finally, the lineal relation between the textural parameters and the malaxing time was evaluated by fitting the different sequences (n from 1 to 12) with a one-degree polynomial. This procedure was repeated for each colour channel c .

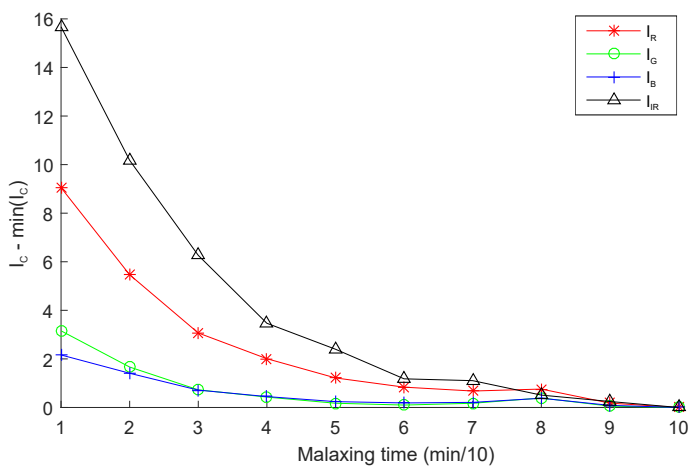


Figure 5.10 Viscosity index for different channels versus malaxing time.

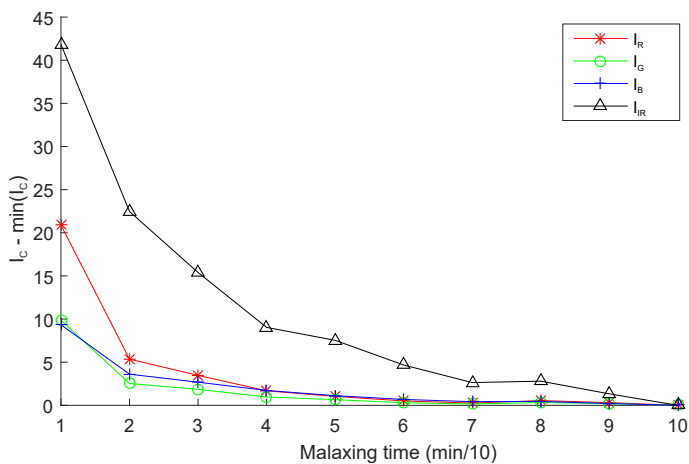


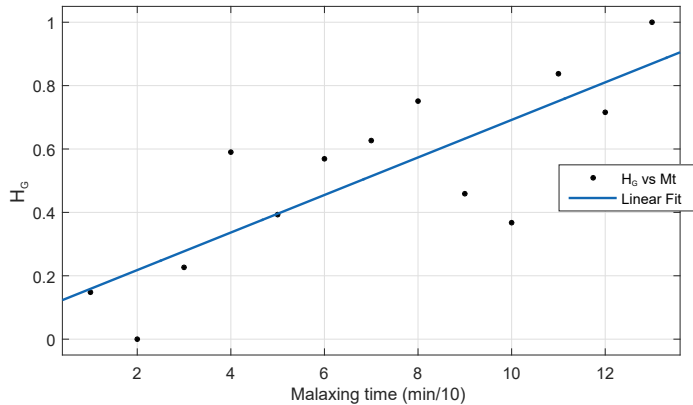
Figure 5.11 Floating oil index for different channels versus malaxing time.

Table 5.1 presents the results. It shows that the best correlation was reached with the homogeneity parameter in the green channel (H_G) with the least value of root mean square error. On the other hand, the Figure 5.12 presents this linear relationship. It means that the GLCM matrix is diagonalized according to the malaxing time increment. It could be due to the lumps over the olive paste surface tend to be larger at the end of the malaxing time when olive paste can be considered as prepared.

Finally, the parameters H_G , I_{IR} (floating oil) and I_{IR} (viscosity) were considered as controller inputs (Figure 5.13) and the outputs were evaluated according to the

Table 5.1 Textural parameters.

Parameter	Mean	SD	R ²	RMSE
C_R	0.44	0.29	0.48	0.21
C_G	0.46	0.27	0.54	0.18
C_B	0.51	0.28	0.35	0.22
C_{IR}	0.61	0.26	0.49	0.19
E_R	0.41	0.27	0.37	0.21
E_G	0.28	0.30	0.51	0.20
E_B	0.32	0.28	0.13	0.26
E_{IR}	0.41	0.27	0.42	0.20
H_R	0.50	0.30	0.56	0.19
H_G	0.51	0.28	0.62	0.17
H_B	0.46	0.27	0.38	0.21
H_{IR}	0.38	0.26	0.54	0.17

**Figure 5.12** Homogeneity parameter in the green channel versus malaxing time.

sequence of images acquired during the malaxing process. For each malaxing state the system recommended four set points (C , Mt , S and MT) (Figure 5.14). For example, in order to prepare difficult pastes the system recommended high temperature, high addition of microtalc, long malaxing time and low rotating speed of the shovels. On the other hand, the variables temperature, microtalc addition and malaxing time were decreased and the speed of the shovels was increased as the paste trends to be prepared.

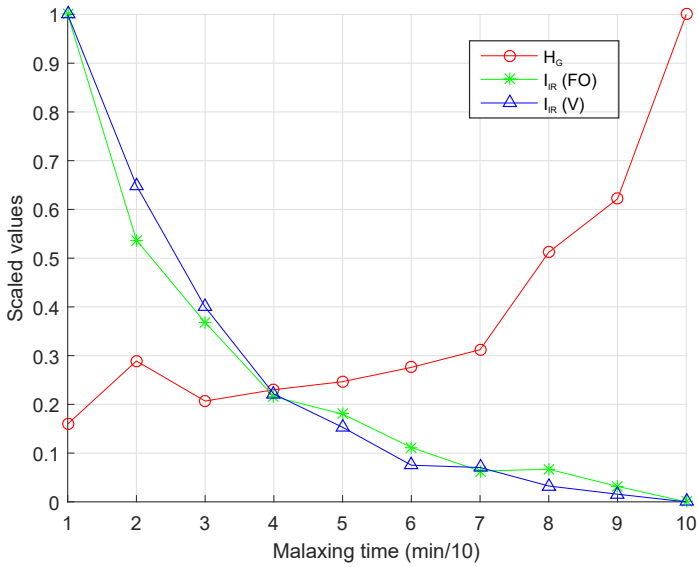


Figure 5.13 Inputs values for the controller where H_G is the textural parameter homogeneity in the green channel, $I_{IR}(FO)$ is the intensity value of the infrared channel for floating oil and $I_{IR}(V)$ is the same value for viscosity.

5.5 Conclusions and future works

In this work a new control system of the olive oil malaxing process has been designed to evaluate the olive paste and to act accordingly. In order to determinate the olive paste conditions a malaxing state sensor has been developed. It was composed by a vision camera device and different processing algorithms to extract features from the acquired images. These features were floating oil, viscosity and granularity. The infrared channel (I_{IR}) revealed more sensitivity for the determination of floating oil and viscosity. On the other hand, the textural parameter homogeneity in the green channel (H_G) showed the best correlation with the evolution of the olive paste granularity. From these parameters and the control miller expertise a fuzzy logic controller was implemented. Different technological variables were considered as controller outputs, such as malaxing time, malaxing temperature, addition of coadjuvant and rotating speed of the shovels. Furthermore the controller were evaluated from a sequence of images acquire from a real plant and it obtained the expected reference values of the involved variables in the thermomixer. The configured system lets the detection of difficult pastes, and it permits the adaptation of the malaxing process to its features. Thus, it avoids olive oil leak in the pomace due to olive paste unprepared and it could suppose a significant improvement in

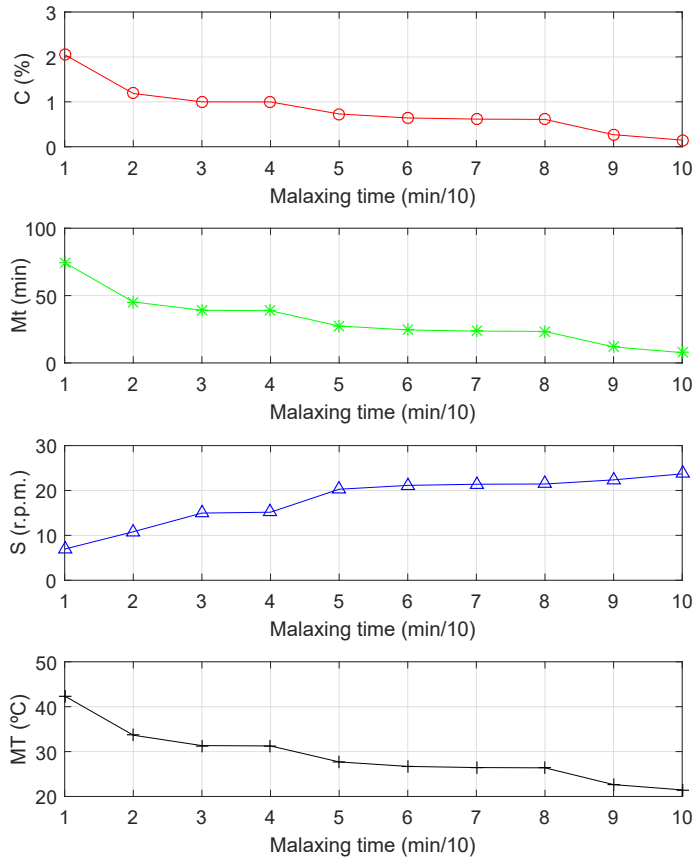


Figure 5.14 Outputs for different olive paste conditions.

the energy consumption of the plant.

For future works we intend to develop a software system based on interactive forms, to acquire the expert knowledge of the master miller, so that a system that allows an on-line validation of the results.

6 Sorting olive batches for the milling process using image processing

Contents

6.1. Introduction	111
6.2. Materials and methods	113
6.3. Experimental results and discussions	122
6.4. Conclusions and future works	127

6.1 Introduction

Currently, the main use of olives worldwide is to produce olive oil (about 88% of the total olive production). The main factor which affects the olive oil quality is the raw material: the olive fruit. As a result, olive quality is related to agronomic conditions and harvesting conditions. In [12] it is stated the influence of the cultivar, kind of soil, level of fertilization, harvest day and other significant parameters.

Besides, the olive oil production process could affect the oil quality. Many research works have been carried out to determine the influence of the manufacturing process of olive oil quality. In [53] the authors compare different processes and extraction conditions and their influence in main olive oil quality parameters. This process consists of several phases. It depends on a great number of physical variables and presents different control setups, which are described in [26] and [25]. It is therefore a complex process.



Figure 6.1 An example of commercial classification machine for table olives. Figure obtained from www.multiscan.eu.

One of the key factors of the olive oil production is the correct reception and classification of olives before the milling, in order to produce the maximum quality level with the maximum process performance. In [189] the fruit reception is classified as one of the most important steps in the oil extraction process. There are an increased use of automated classification machines in other agro-food industries such as the tomato or table olives sector. Figure 6.1 presents an example of commercial classification machine for table olives.

A first classification that could be done of fruits in the incoming of the mill is the differentiation between ground and tree olive, that is, the fruit that has fallen to the soil and later has been picked up and olives which have been picked up directly from the tree. This is because the fruit on the ground undergoes a series of changes that deteriorate the quality of the oil obtained. For example the acidity increases when the time of staying on the soil is longer. As well as this, the organoleptic quality is affected; the tree oil is usually extra virgin, with a great fragrance and positive attributes; conversely, soil oil deteriorates to lampante virgin olive oil with major defects and unfit for direct consumption [110]. Olives coming from the ground also contain waste such as soil, pieces of branches, and some pebbles.

Some discriminatory features can be observed by looking the olives harvested from tree and soil. These features are mainly related to the colour and the morpho-

logical appearance of their skin. The olives picked up from the tree are normally green, purple or black (according to the ripening index) [67] and smooth skinned. What is more, when the olives have remained several days on the ground, they turn into brownish due to the oxidation process and their skin changes into wrinkled olives because they start to dry up. In addition, olives batches coming from the soil are less homogeneous than olives from the tree due to the fact that they have fallen in different moments and they have other elements such as ground grains, pebbles or leaves.

Nowadays, in the olive oil production, the quality evaluation of the incoming fruit batches depends on manual inspection and an operator will decide if the batch is from the tree, the soil, mixed or even if it has other kind of defect. Obviously, this selection is not often carried out properly since it is subjected to personal criteria and it is a complex decision to be taken in a quick inspection. Later, depending on this previous classification, the oil extracted in the milling goes to different tanks classified by its quality. Olives should have been previously classified by the farmer, as it is reported in [43], but sometimes the fruit has been mixed in order to reduce the harvest cost. In [196] the authors expose the costs and profits in different kind of olive fields. It is absolutely necessary to collect, transport and process fruits from soil and tree separately. Similarly, in [110] it is stated that a small amounts of olives coming from soil can significantly modify the organoleptic characteristics of olive oil from the tree.

As soon as the olives lots have been separated, independent production lines should be used to manufacture the olive oil. In this way, it would be easy to keep the best possible olive oil quality until the end of the production process.

In this work we propose an automatic classification methodology of olives based on computer vision. The aim of the classification is the differentiation between tree and soil olives lots for oil production, trying to minimize possible grading mistakes.

The structure of this chapter is as follows: Section 6.2 presents the materials and the methods used for our experimentation are exposed. Secondly, Section 6.3 shows the experimental results. Finally, in Section 6.4 the conclusions and future work are presented.

6.2 Materials and methods

The hardware setup, the image processing and the algorithms used for the classification are explained in the following parts.

6.2.1 Experimental set-up

A basic vision system was build in laboratory for the first experiments. The type of computer used was a personal computer (PC) with the Matlab acquisition toolbox installed. The vision device was a Logitech QuickCam Sphere AF camera with a resolution of 2 Megapixels; it acquires images with 24 bits of depth and 1600x1200 pixels of size. The lighting system was a low-angle led ring with a white diffuser and 16W of power (INFAIMON), which provides appropriate measurement conditions.

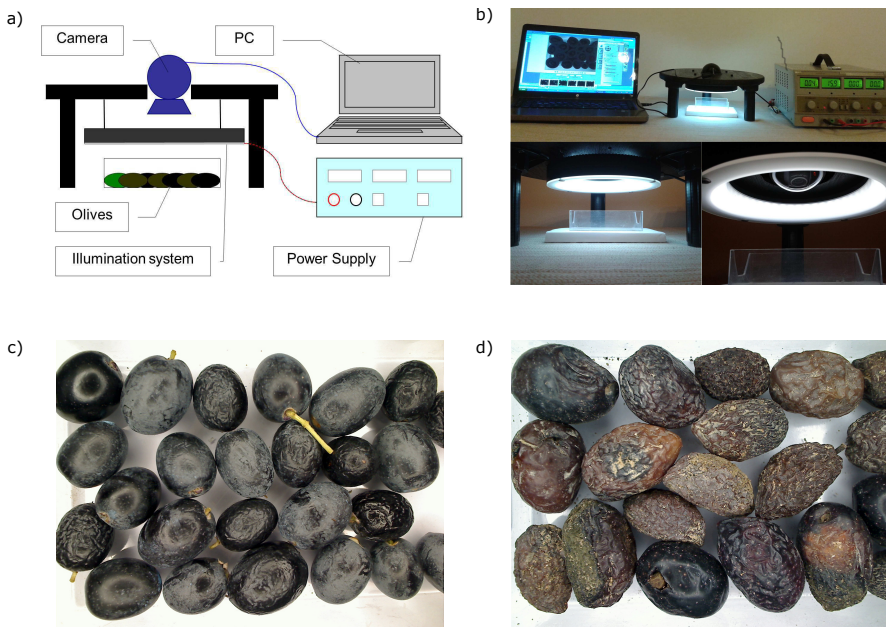


Figure 6.2 At-line acquisition system setup. Subfigure (a) shows the setup as sketch and Subfigure (b) is the real setup. On the other hand, Subfigure (c) presents an example of tree olives and Subfigure d) is an example of soil olives.

The camera was rigidly attached and also the LED ring. A white plastic tray, where the olives are placed before for the acquisition, was placed under the camera on the illumination system. The image processing and the statistical analysis software were programmed in Matlab over Windows SO. Figure 6.2 shows the prototype used to acquire the images of olives.

The selection and disposition of the illumination system was crucial. It was found problems related to the brightness of the light in the surface of some kinds of fruit. There was too much light and the intensity of the shine complicated the process to

see the defects or wrinkles in the olive skin. The same issue has been reported in other works [55].

The problem was solved in different ways. Firstly, different lighting systems were tested (two LED rings and a multipoint square light source) and the led ring was the best solution for this application. An small LED ring of 17 cm diameter was chosen. After this, the light intensity was tested. As a result, the direct current power supply was connected to the illumination system in order to change the voltage and the current with the goal of regulating the light intensity. Then the supplied voltage to the illumination system was adjusted in order to avoid the saturation of the pixels over the green olives. After that, it was supplied a constant voltage of 15,9 V and 0,04 A to the LED ring. Figure 6.3 shows 4 pictures of the same olives with different light intensity.



Figure 6.3 Difference in brightness with different light intensity.

Furthermore, different distances between the light source, the camera and the tray of olives were evaluated. The distance was fixed in order to see the complete tray of olives and to have the best possible lighting conditions. Therefore, the distance from the LED ring to the tray of olives was 7 cm. In addition, the camera was attached to the support with a distance of 9 cm from the olives tray. The plastic tray that contains the olives during the image acquisition had a size of 6,5 cm by 10,5 cm.

After that, the experimental setup was replicated with some modifications in a real olive oil mill. The goal was to extrapolate the proposed methodology from laboratory to factory conditions. In this case, two computer vision were installed. While the first vision system was placed over the line which is dedicated to convey olives coming from the soil, the second one was installed over the line which conveys olives coming from the tree.

The vision cameras and the lighting system were attached to a plastic box made of PVC plastic sheet and aluminium profiles. Each box was composed of one industrial vision camera model Mako G-223C (PoE power supply, 2/3" sensor size, 2048x1088 of resolution, frame rate of 50i/s, Color CMOS sensor, GigE interface) and six distributed white LED lamps (in particular two LED lamps of 30 W and four LED Lamps of 20 W). A semitransparent paper was attached for each lamp in order to diffuse the LED light and reach good lighting conditions.

These boxes were joined to the metallic structure of the conveyor belts which normally transport the olives between different production phases. They were placed after the blowing and washing machine to acquire images without pebbles, mud and leaves. The image processing and the data analysis were programmed in Matlab over Windows SO. Figure 6.4 shows a sketch of the prototype used to acquire the images of olives.

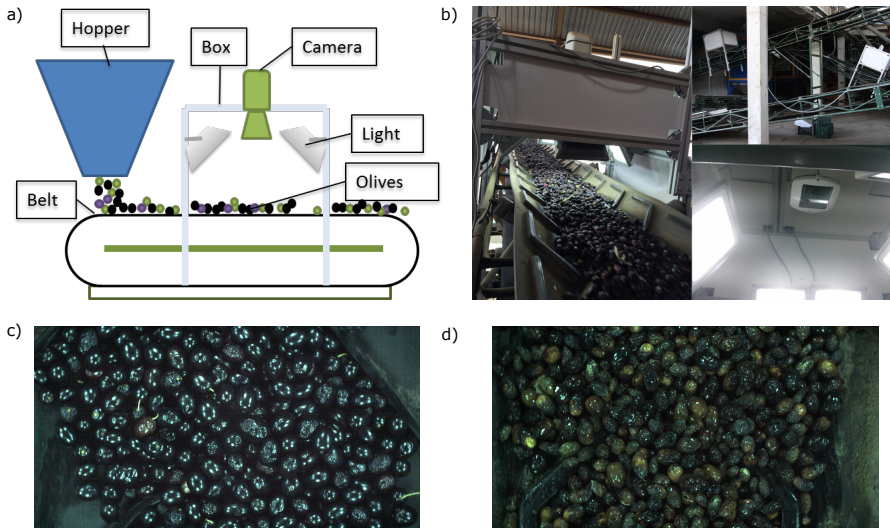


Figure 6.4 Figure a) shows the different components of the computer vision system; Figure b) is a real image of the vision system installed over the conveyor belt; Figure c) presents an example of tree olives; Figure d) is an example of soil olives.

6.2.2 Olive fruit samples

The experiments in laboratory were carried out with samples of olives harvested during January and February in 2014, in the region of Priego de Cordoba, Spain. A sample set with 176 olive batches among different varieties (*Picudo*, *Picual* and *Hojiblanco*) was randomly recollected by hand. Of these, 77 samples were picked from the soil and 99 samples were directly picked from the tree. For the objective classification of this work, it was taken into account olives with mid and advanced maturity indexes. This maturity point is an interesting case of study because normally early olives (green ones) do not fall at these stages and, therefore, the two classes are not available.

Olive images were obtained under static conditions the same day of the harvest. Each batch contains between 15 to 25 olives depending on the size of olives and the size of the olive is directly linked with the variety and the agronomic conditions of the tree.

On the other hand, the vision systems employed in the production plant were installed during December in 2014. In this case 4287 images were acquired (1432 corresponding to soil olives and 2855 with tree olives).

6.2.3 Image analysis

Different authors have worked with images where the olives have been segmented to find some defects on the fruit surface [165]. The proposed methodology of this work is based on the use of entire images for the classification task instead of segmenting each olive fruit. This proposal is interesting because undesirable components in the images could provide extra information about the class. Nevertheless, this methodology is not able to identify the olive pieces belonging to each class in the same batch and then it is not possible to classify each olive fruit individually. However, this ability could be less efficient in terms of time consumption on an industrial scale.

Two image processing methods were applied: gradient image and differences in the RGB channels [157]. The first one was focused on the detection of sudden changes into the intensity profile of images. This fact is related to detect olives with wrinkled skin. We can see the results of applying this filter in the second row as shown in Figure 6.5, where abrupt changes in the olive skin have been contrasted. The darker they are, the more wrinkled they are. By employing this filter the differences between tree and soil samples are more obvious.

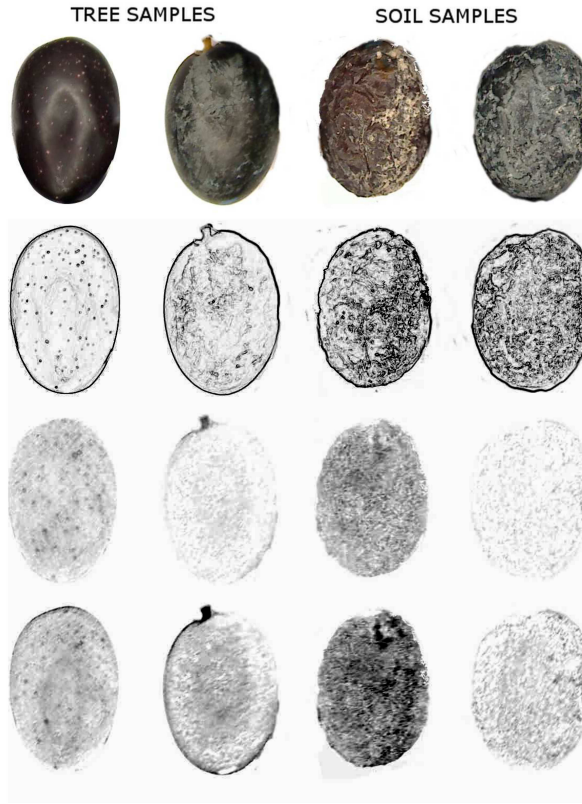


Figure 6.5 Olives coming from different locations and different image analysis. The first row shows the original image of the samples (RGB); the second row presents the result after applying the textural filter (I) and the third and fourth rows show the same samples where the brownish colour has been detected (C_{rg} and C_{rb} respectively).

The gradient image was obtained by applying the textural filter according to the Eq. 6.1.

$$\nabla I(x,y) = G(x,y) = \sqrt{G_x^2 + G_y^2} \quad (6.1)$$

where $I(x,y)$ is the original gray-scale image of each sample and $G(x,y)$ is its gradient image. This gradient image was obtained by using the Roberts operator [166] and by convolving the $I(x,y)$ image with two convolution masks, which are shown in Eqs.

6.2 and 6.3.

$$G_x = I(x,y) * \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (6.2)$$

$$G_y = I(x,y) * \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (6.3)$$

The gradient will be zero in areas with constant intensity levels and it will be proportional to the intensity level variations in the other ones. In this case, the image processing toolbox of Matlab was employed.

On the other hand, the brownish colour appears in the RGB images when the red channel intensity levels are greater than the green and blue ones. Therefore, the presence of brownish tones in the images was identified with the difference between red and green channels, including the difference between red and blue channels. Thus, two new images $C_{rg}(x,y)$ and $C_{rb}(x,y)$ are obtained (Eqs. 6.4 and 6.5).

$$C_{rg}(x,y) = RGB_R(x,y) - RGB_G(x,y) \quad (6.4)$$

$$C_{rb}(x,y) = RGB_R(x,y) - RGB_B(x,y) \quad (6.5)$$

where RGB is the original image.

The third and fourth rows in Figure 6.5 show an example of the image analysis aforementioned for four olives. In this case, dark tones indicate brownish colours and the differences between soil and tress classes can be observed.

6.2.4 Feature vector

All samples were translated into feature vectors with the goal to be mathematically usable in the next steps. Before using the classification algorithms the gray-level histogram of each image (I , C_{rg} and C_{rb}) was obtained. In order to build the feature vector of each sample, the histograms were concatenated to obtain a vector of 768 components (256 components per image). Each component was modified to have mean zero and scaled to have a standard deviation unity.

There are essentially two approaches for feature selection: filter-based feature selection and wrapper-based feature selection. Their differences can be seen in [207]. Stated briefly, the filters evaluate feature subsets by their information content and the wrappers set the objective function as a pattern classifier. A filter algorithm

has been selected due to its fast execution and generality for future samples. The one-way analysis of variance is like a filter algorithm where the distance between classes was used to evaluate the difference of 768 components for each olive class. For each component, if $F - value$ is greater than 2 the distance inter-classes is the double of the distance intra-classes and we can consider the component as discriminant feature. Moreover, if $p - value$ is greater than 0.05 the $F - value$ is significant statistically by agreement. The optimal components were selected to discriminate among these classes.

Finally, the principal component analysis [112] was carried out on the resultant histograms to investigate possible clustering of samples on the basis of sample origin. PCA is a linear and unsupervised method helpful for reducing the dimensionality of numerical datasets in a multivariate problem. This method can extract useful information from the data and identify the relation among different classes. The statistic toolbox of Matlab was used for above tasks and the algorithm to implement PCA was based on eigenvalue decomposition.

After that, linear and non-linear algorithms were compared for classifying among classes. The lineal method was Fisher discriminant analysis (FDA) [68] and the non-linear one was an artificial neural network [56]. The accuracy of each method was proved with cross validation for fold size of 2, that is, 50% of samples for the training phase and the other 50% for the validating phase. The 2-fold cross validation has been chosen because our objective was to avoid the over-training and maximize the number of validation samples at the same time. This fact will allow to generalize the results for future samples. The results can be seen in the experimental results and discussion section.

6.2.5 Fisher Discriminant Analysis

FDA is a simplified case of LDA [120] in which only two classes appear. Both of them are supervised methods and seek to find a new reduced space which maximizes the separation among classes. For two classes, the solution proposed by Fisher was to maximize a function that represents the difference between the means, normalized by a measure of the within-class variability. Thus, the Fisher linear discriminant is defined as the linear function (Eq. 6.6) which maximizes the objective function presented in the Eq. 6.7.

$$y = w^T x \quad (6.6)$$

$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|}{s_1^2 + s_2^2} \quad (6.7)$$

where $\bar{\mu}_1 - \bar{\mu}_2$ is the difference between the different class means and $\bar{s}_1^2 + \bar{s}_2^2$ measures the variability within the two classes after projection, hence it is called within-class scatter of the projected samples.

The next step for this classification algorithm is to fit a multivariate normal density of each group in the new space. Finally, for grading samples, the probability of belonging to each group (G_S and G_T) is obtained for each sample (Figure 6.6). The statistic toolbox of Matlab was employed as well.

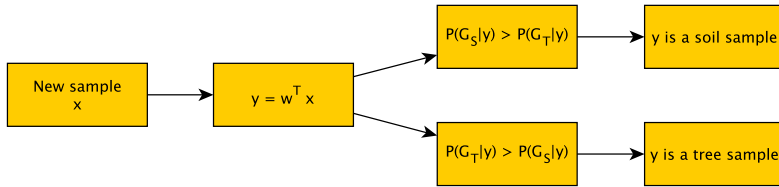


Figure 6.6 Proposed FDA classifier diagram.

6.2.6 Artificial Neuronal Network

The selected artificial neural network was Multilayer Perceptron (MLP) as non-linear classification method.

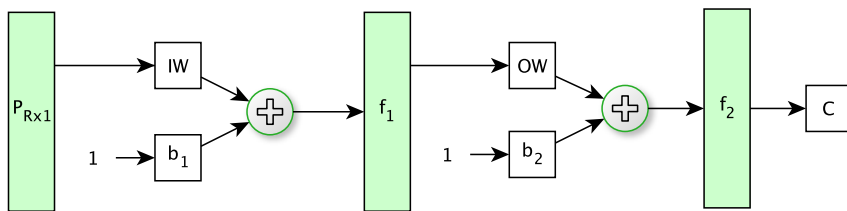


Figure 6.7 Neuronal network diagram where P is the input vector; IW is the input weights; OW is the output weights; b_1 and b_2 are the bias; f_1 and f_2 are the sigmoid functions and C is the classification result.

Only one hidden layer was considered since this configuration is capable to learn any pattern and less prone to getting caught in local minima those networks with a higher number of hidden layers [60]. The number of input nodes was the same as vector components selected in the previous step. Furthermore, the neural network

was tested with different number of neurons into the hidden layer. The classification results were similar, therefore, one neuron was established. Finally the output layer was configured with one neuron because it is sufficient to classify between soil and tree classes. Figure 6.7 shows the net topology which has been employed.

The ANN was trained with the backpropagation algorithm named Scaled Conjugate Gradient [206] because it has been proven to be efficient in pattern recognition problems [96]. In contrast, the activation function for hidden layer was sigmoid in order to learn nonlinear relationships between input and output vectors. Additionally, the activation function of the output layers were sigmoid to constrain the network output. In this case, Neural Network toolbox of Matlab was used.

6.3 Experimental results and discussions

The histograms of the processed images for the whole dataset can be seen in Figure 6.8. The average values for each component and their standard deviation appear in this figures. The difference between two olive classes is more evident in the textural image histograms (Figure 6.8a) than in the other ones. This fact proves that the wrinkling of the olive surface is an important factor to arrange classes. However, in the other histograms (Figure 6.8b and Figure 6.8c) several components appear that could be interesting in order to build the feature vector.

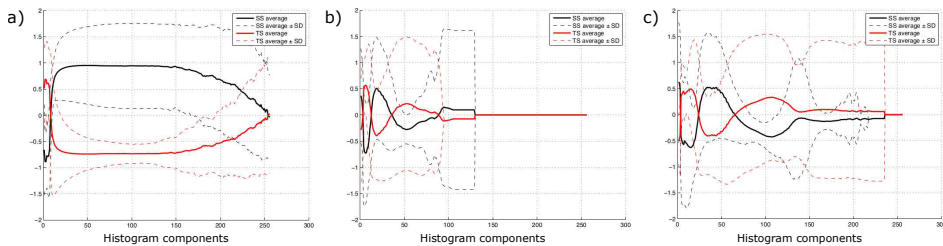


Figure 6.8 This figure shows the normalized image histograms after applying the textural filter (a), channel R minus channel G operation (b) and channel R minus channel B operation (c).

The variance analysis was carried out in order to search the histogram components more discriminants between classes. The results of this test for each component are shown in Figure 6.9. The F – value indicates the distance between two olive classes and p – value is the significant parameter. In these figures we can see that F values in texture histogram are in general greater than colour histogram ones. As a result, we can confirm the high weight of the textural information in the classification task.

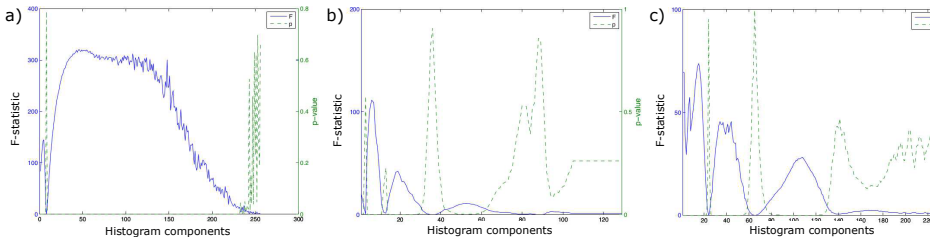


Figure 6.9 This figure shows the statistical F – values and p – values between the aforementioned two classes of olives for the histograms in Figure 6.8.

Then, for building the feature vector, the component was excluded if its p – value is upper than 0.05 and its F – value is lower than 2, indicating that the gray intensity level of the histogram has little or no influence on the discrimination of classes. In contrast, if the p – value is lower than 0.05 and its F – value is upper than 2, the histogram component is included in the feature vector. Finally, the number of 768 components was reduced to 405 significant components. The feature vector for the whole dataset is shown in Figure 6.10.

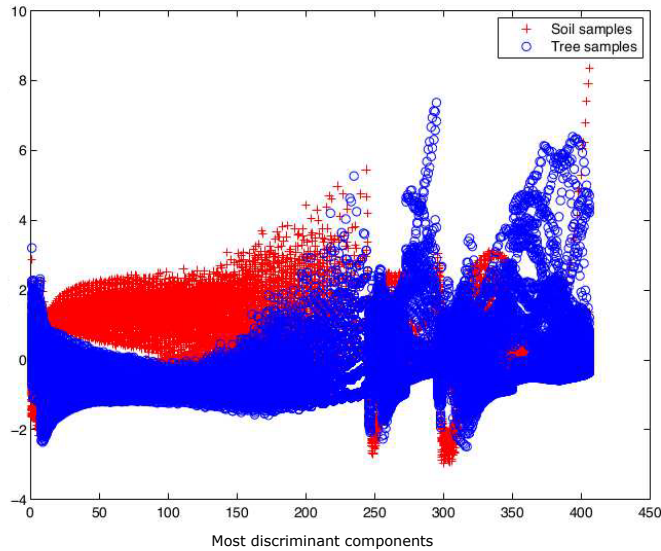


Figure 6.10 Feature vectors of the whole dataset after filtering by ANOVA test.

The next step was to eliminate co-linearity and reduce the amount of components in the feature vector. The clustering among classes is evidenced from the results obtained by PCA applied to the matrix of the entire set of scaled histogram data. Figure 6.11 shows the samples projected in the three firsts components. In this case,

the explained variance was about 75%.

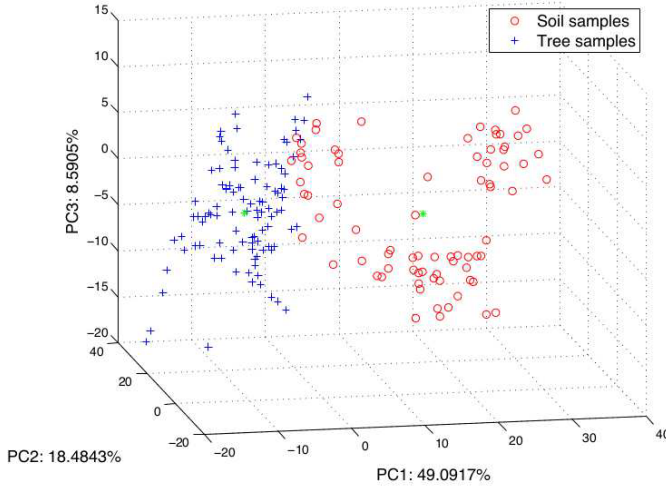


Figure 6.11 All the dataset is projected in the new 3D space which is created by three principal components PC1, PC2 y PC3.

Table 6.1 shows the results for the two classifiers considered. The second column in the table presents the algorithms followed to process the feature vectors of each sample, where NORM means normalization; ANOVA is the feature selection algorithm and PCA is the feature extraction algorithm. These algorithms have been explained in the Materials and Method section. The third column shows the image information used to assemble the feature vector, where TEXT is textural information, RG is red and green channel and RB is red and blue channel. Finally, it shows the classification results for a number of PCA components.

The dataset was randomly divided in two groups with the same size for training and validating purposes. The correct classification percentage was the average of two iterations according to 2-kfold cross validation. For each iteration, this percentage was obtained from dividing the number of samples well classified in each class (tree and soil) by the amount of samples in the dataset (Eq. 6.8). Since the dataset was randomly divided in two groups, the aforementioned validation process was repeated three times and the average value was selected.

$$CCP = \frac{\text{sum}(\text{diag}(\text{CM}))}{TS} \times 100 \quad (6.8)$$

where CCP is the correct classification percentage, CM is the confusion matrix and TS is the total number of samples.

Table 6.1 Classification results in laboratory.

Classifier	Feature vector processing	Image information used	Number of PCA components	Correct classification
FDA	NORM + PCA	TEXT	12	93.75%
		TEXT + RG	19	94.32%
		TEXT + RG + RB	27	99.43%
	NORM + ANOVA + PCA	TEXT	12	92.61%
		TEXT + RG	12	95.45%
		TEXT + RG + RB	15	100%
ANN	NORM + PCA	TEXT	11	98.86%
		TEXT + RG	6	96.02%
		TEXT + RG + RB	7	98.86%
	NORM + ANOVA + PCA	TEXT	6	96.60%
		TEXT + RG	5	98.30%
		TEXT + RG + RB	13	100%

On the other hand, the classification results obtained with the linear classification algorithm (FDA) denotes that it is a good approach to classify olive batches. These results were expected because the PCA-scores with the three principal components are almost linearly separable (see Fig. Figure 6.11). In this case, the best result was reached with 15 PCA-components, the three channels and the feature selection algorithm.

The correct classification result was also 100% with the neuronal network. Nevertheless, ANN obtains this percentage with a lower value of PCA components. It means that the data are linearly separable but the neural network could be more successful with future samples.

On the other hand, we can see that ANOVA filtering is useful because it reduces the number of PCA-components for the same success ratio. In addition to this, this ratio is slightly decremented in both classifiers with only textural information of the samples which is used. Nonetheless, these results are improved when the colour information of the sample images is added. This fact justifies the use of both histograms.

Moreover, the optimal number of PCA-components was selected with an iterative process: firstly the PCA-components were ordered according to the explained variance; for each iteration a new PCA-component was added and the sorting accuracy

was obtained. Figure 6.12 show the results of this iterative process for the nine cases in Table 6.1. In this figure we can observe that when the number of component increases the accuracy of the classifiers decreases because of the over-fitting. On the contrary, the performance of the ANN is better than FDA with the firsts PCA-components are employed. Then we can reduce the number of features without high losses and could be improved the computation time. This margin could be positive in order to use the classification system into an industrial classification machine of olive batches.

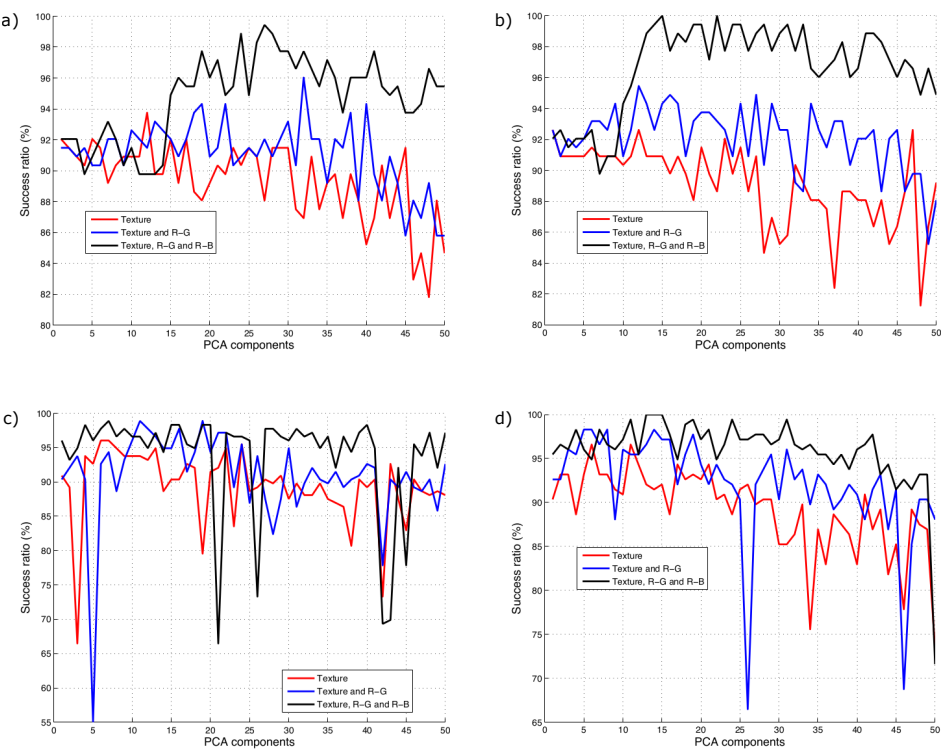


Figure 6.12 a) Lineal classification results without ANOVA component filter; b) Lineal classification results with ANOVA component filter; c) ANN classification results without ANOVA component filter; d) ANN classification results with ANOVA component filter.

The proposed methodology was validated with the images acquired by the experimental setup placed in the oil mill over the conveyor belts. Table 6.2 shows the results for the different configurations. In this case, the best results were obtained with the linear classifier. It reaches an accuracy of 96.85% with the information of the textural, red, green and blue channels. While the neural network obtained similar result, it was reached with fewer components than FDA. On the other hand,

the percentages are worse than those which were achieved with the laboratory setup. This fact is mainly due to the batches which contains a lot of sound olives and they were classified as defective class by the person who decides the input line to the process.

Table 6.2 Classification results in oil mill.

Classifier	Feature vector processing	Image information used	Number of PCA components	Correct classification
FDA	NORM + PCA	TEXT	27	87.26%
		TEXT + RG	302	92.25%
		TEXT + RG + RB	320	96.05%
	NORM + ANOVA + PCA	TEXT	16	86.98%
		TEXT + RG	314	92.74%
		TEXT + RG + RB	341	96.85%
ANN	NORM + PCA	TEXT	42	87.56%
		TEXT + RG	163	91.06%
		TEXT + RG + RB	93	95.24%
	NORM + ANOVA + PCA	TEXT	26	87.38%
		TEXT + RG	108	91.39%
		TEXT + RG + RB	118	95.73%

6.4 Conclusions and future works

This paper presents the intelligent block of an automatic classifying system of olives based on computer vision before being processed for the oil extraction. Different problems have been solved such as the brightness produced by the olive skin or the hardware setup. This system was able to distinguish between soil and tree olives batches for olive oil production, and it uses information such as the wrinkle of their skin and their colours. The first experiments were carried out in laboratory with success ratio of 100%. Furthermore, the same methodology was applied to images acquire in line with the experimental setup installed in the oil mill over the conveyor belts. In this case the system also reaches good results (greater than 95%). The grading algorithms proposed have been linear discriminant analysis and neuronal networks. The goodness of the results has allowed us to start different

tests in a real plant with the building of an automatic sorting machine. This fact has awoken the interest of different manufacturing companies of olive mill machinery.

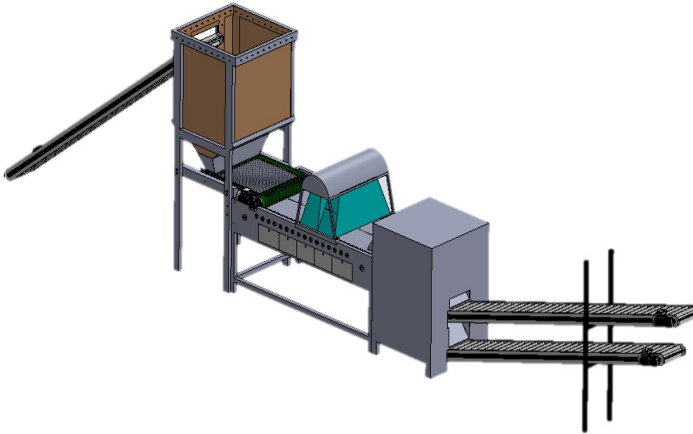


Figure 6.13 Concept of machine developed by GRAV.

As future work, it could be interesting to test the same classifier using olives with different ripening index and prove the robustness of the algorithm with batches of olives previously cleaned (which reduces the dirt over the skin of the olives). To improve the classification, an option could be having in consideration other features such as the difference in olives size. Furthermore, the detection of batches of olives with diseases and other problems such as frozen or mixed olives could also be an interesting work to be done.

There is another category of the olives batches in mills called mixed olives being a mixture of olives coming from the soil and from the tree in the same lot. It could be very interesting for future work to study the classifier testing lots of mixed olives in different percentage of olives coming from the soil or the tree.

Last but not least, the classification of the olives individually could be also interesting as future work. Actually, the research group of Robotics, Automation and Computer Vision is developing this concept of machine and the design of the model can be seen in Figure 6.13.

7 Results and discussion

The works set out in this doctoral thesis intend to deepen the knowledge about the potential use of two non-invasive technologies (near infrared spectroscopy and computer vision) in order to determinate different parameters which have direct effects on the olive oil quality and on the industrial production performance.

While there are a considerable amount of research works which have proved the feasibility of using the technologies evaluated in this doctoral thesis (e.g. detecting defects over the olive skin using computer vision or predicting quality parameters by means of NIR devices), most the techniques have been evaluated at laboratory level and there are scarce works dealing with them at industrial level or reproducing them in line analysis conditions. In this context three VOOPP points have been studied: the final phase where the olive oil is finally produced, the preparation step of the olive paste before the extraction process and the input step to the process where olives are received.

Most of the works that use NIR spectroscopy to predict different parameters related to the olive oil quality such as FA or Peroxide Index (PI), employ measurement probes which need to be in contact with the sample to acquire its spectral absorption curve. However, this methodology requires periodical maintenance due, for example, to the accumulation of impurities in the optical path. This fact, together with the in line monitoring, are the main reasons to study, in the first work carried out in this doctoral thesis, the feasibility of determining in line three of the main control parameters in the VOOPP (MC, FA and PI) by means of employing a spectrograph device based on hyperspectral technology (it was composed by a InGaAs diode-array detector model Xeva-1.7-320 plus the spectrograph model ImSpector N17E). In addition to being non-invasive it allows to acquire spectral information without contact

with the sample [138]. This study was basically focused on the determination of two quality parameters (FA and PI) and one control parameter (MC) using 133 aliquots of olive oil as experimental samples and an in line hyperspectral acquisition system located in laboratory and based on conveyor belt as experimental setup. Different regression and wavelength selection algorithms were evaluated taking into account the reference values obtained by an olive oil chemical laboratory.

The results that have been achieved with this research work revealed the capability of this technology in order to assess the evaluated parameters without contact. The determination of FA in olive oil samples was evaluated and the lower RMSEV value was reached with the algorithm GA-MLR and it was 0.09%. The R value was 0.98 and the RPD value was 4.77 units. In this case 133 samples of VOO were employed with FA average of 0.48% and SD of 0.45%. Furthermore, the PI was cross-validated by using 56 samples of VOO with an average of 5.31 *meq.O₂/Kg* and SD of 0.83 *meq.O₂/Kg*. For this parameter the lower RMSEV value was 0.03 *meq.O₂/Kg*, R value was 0.99 and RPD value was 26.38 units. They were reached with the SPA-MLR algorithm. But undoubtedly, the MC was the best estimated parameter according to NIR spectrum. The experiment was performed with a total of 69 VOO samples and the best result was achieved also applying the SPA-MLR algorithm. It showed a RMSEV value of 0.0006%, with a R value of 0.99 and RPD value of 116.66 units, when the VOO considered samples presented an average of 0.15% and SD of 0.07%. This result could be attributed to the strong absorption of the water molecule on the acquired bands, located in 1200 and 1450 nanometers [199]. The first two results of this work can be compared with the study published by Inarejos et al. in 2013 [101] when authors got R value of 0.76 and 0.92 for FA and PI respectively. In this work the measurements were carried out with the FT-NIR technology in transmittance mode. As conclusion, the hyperspectral technology can be considered as useful attending to the thresholds of quality established by European Norm (0.8% and 2.0% for FA and 20 *meq.O₂/Kg* for PI).

In addition to produce olive oil with quality, there is also other research line related to enhancing the production of functional oils with high content of antioxidant components which have beneficial effects on health. In this context, it could be interesting to know the content of these chemical compounds in line. Several researchers have studied the linkage between the NIR spectrum of the olive oil and its polyphenol content. Most of them are directly based on the application of PLS correlations between the whole spectrum and the analytical results obtained from chemical laboratories. However, not many of them have dealt with the identification of the most correlated bands with the functional group of this type of molecules. Identifying these bands could be useful to increase the spectral resolution within

the spectral band of interest and to develop low-cost devices by means of reduce the number of photocells inside the diode array sensor. For these reasons, the aim of the second study presented in this doctoral thesis is to identify the absorption bands relating to polyphenol content of AOVE samples obtained in line and elaborated with different process variables [80]. Furthermore the effects of the lighting intensity and the number of acquisitions per sample on the prediction results were evaluated. The NIR technology employed in this work was FT-NIR (model ARCoptix Rocket 0.9-2.6 nm) and the measurements were carried out in transmittance mode.

In this work, the results achieved by the algorithms PLS and SMLR were compared. These approaches were based on feature extraction and feature selection respectively. The best results were reached by SMLR with RMSEV value of 156.14 ppm, R value of 0.80 and RPD equal to 2.3 units. On the other hand, the sampling conditions were maximum light intensity and averaging six measurements per sample. The resulting linear model showed that the absorption bands located in 1830, 2189 y 1917 nm were the most correlated bands with the polyphenol content. The two most referenced studies in the literature related to this topic, [101] and [129], achieved similar RPD coefficients. However they employed more than three features as predictor variables and they employed VOO samples with concentrations in different ranges.

One of the main step within the VOOPP is the olive paste preparation before the extraction process. Different technological variables such as time and temperature can affect both the quantity and the quality of the olive oil produced. An olive paste well prepared allows to avoid leaks of olive oil at the exit of pomaces (decreasing the industrial performance) and leaks of volatile compounds (decreasing the organoleptic features in the final olive oil). Normally, is the master miller the person who adjust the process variables according to the periodical inspection of the olive paste and expertise. In order to remove this repetitive task and to bring objectivity to this phase of VOOPP, it was decided to implement an expert system based on computer vision [82]. The goal of the system was to determine in line the olive paste state and to recommend the values of the process variables according to different rules gathered from the master miller knowledge. The images were acquired by a double CCD camera (model JAI AD-080 GE) working at visible and infrared spectra.

Three features of the olive paste were analysed: viscosity, granularity and presence of floating olive oil. The viscosity was related to the quantity of paste attached to the blades of the thermomixer. On the other hand, the presence of floating oil was evaluated by means of the images acquired in the infrared spectrum. Finally the granularity was determined by analysing the texture of the olive paste surface. Four process variables were selected: malaxing time, malaxing temperature, rotating

speed of the blades and coadjuvant addition. The recommendation system was based on fuzzy logic and it was validated with the sequence of images acquired from the experimental setup during the malaxing process. The results were as expected. Firstly the system recommended coadjuvant addition, high temperature, long malaxing time and low rotation speed. Lately the value of these variable were progressively reversing up to reach the preparation state.

The olive oil quality significantly depends on the olive fruit at the input of the process. The physico-chemical features of the olives can be affected by extrinsic factors such as the olive variety and intrinsic factors such as the agri-environmental conditions, the harvesting method and the transport conditions from the field to the industrial plant. Nowadays is the yard manager the person who decides the reception hopper where the farmer discharges their olives. This decision is based on visual inspection together with the good faith of the farmer. Normally, the number of lines at the input and the quantity of olives for each batch is such that would be impractical to supervise and classify the total amount of olives by the yard manager. In this context, the fourth work contained in this doctoral thesis develops a new system to automatically characterize olives according to the harvesting method (soil or tree) [160]. The system was based on computer vision in the visible spectrum and firstly it was evaluated at laboratory level. Posteriorly it was installed in a real oil mill.

The features extracted for each image were related to the texture and the color of the olive skin such as the morphology (smoothed or wrinkled) and the presence of brownish tones. The feature vector was composed after filtering the most discriminant features by means of applying the ANOVA analysis. Finally, two classifiers were evaluated: FDA and ANN. While the laboratory experimental setup reaches 100% of accuracy, the experimental setup installed in the factory achieved 95% of success. In both cases, color and texture were significant features.

8 Conclusions and future works

This doctoral thesis has been mainly related to the application of non-invasive sensor technologies and the processing of the acquired data in order to characterize the product at different stages of the process and also to improve the process from the points of view of quality, quantity and nutritional value. The experimental phase has been one of the great strength of this thesis and it has been possible thanks to the collaboration between the research group (GRAV) and different oil mills within and outside the province of Jaén, (Spain). The experimental works performed in the production plants have allowed to dispose of the product with different conditions which frequently occur during the olive oil campaign and to reach the aforementioned results.

The characterization of the olive fruit at the input of the process is surely one of the main tasks to perform in order to achieve any elaboration objective. The olives delivered to the oil mills present different features and they affect to the subsequent products in the process (olive paste and olive oil). The research work developed in this thesis and related to this phase of the process has intended to characterise olive batches according to the harvesting method (soil or tree) by means of using computer vision and automatically classify them. To this end, the most discriminant features for each class were filtered and two classification algorithms were assessed, one of them based on linear discriminant and the other one based on neural network.

Later, the olives are crushed in the milling process and they turn into olive paste. The malaxing process inside the thermomixer is like a biological reactor where the transference of heat and movement to the olive paste produces new molecular

structures, while at the same time it modifies the visual aspect of the olive paste. The characterisation of the olive paste inside the thermomixer has been a central objective of this thesis and it has been performed by using computer vision in the visible and infrared spectrum. One of the main task of this work was to identify the color and textural features which are closely correlated with the malaxing time, with the objective of implementing the so-called malaxing state sensor. Furthermore, it was developed a recommendation system in order to suggest the value of different process variables according to the information delivered by the malaxing state sensor and the expert knowledge of the master miller.

Finally, it could be useful to assess in line the features of the elaborated olive oil in order to take appropriate decisions to reach the proposed elaboration objective. One of the research lines developed in this thesis has been focused on this topic. The final goal of this line was to infer olive oil features related to the quality and nutritional value by means of processing the near infrared absorption spectra of different olive oil samples. Furthermore, the spectral bands highly correlated with the parameters of interest were identified and the effects of different measurement conditions over the prediction results were evaluated.

Even though each chapter of this doctoral thesis includes their own conclusions, the next section deals with the main conclusions for each research line: Virgin olive oil characterization (Chapters 3 and 4), malaxing state assessment of the olive paste (Chapter 5) and olive fruit classification at the input of the VOOPP (Chapter 6). Subsequently, the future works for each subject will be exposed.

8.1 Summary of conclusions

Below is a summary of the main conclusions which have been obtained from the contributions of this doctoral thesis.

- **Olive fruit classification at the input of the VOOPP**

A system based on computer vision has been developed in order to classify olives batches at the oil mill yard according to the harvesting method (soil or tree). It has been carried out by means of processing the images acquired by the vision sensor and extracting the features related to color and textural information for each olive batch. The referred system has been validated at laboratory and industrial level.

The input dataset was preprocessed by applying the analysis of variance algorithm for the purpose of identifying the most discriminant features between the two aforementioned classes and removing the undesired features. Furthermore the analysis of principal components was performed in order to reduce the number of features. Two classification algorithms were evaluated: one of them based on Fisher discriminant and the other based on neural network. Both classifiers achieved similar results, however the results obtained by the neural network were achieved with fewer characteristics.

- **Malaxing state assessment of the olive paste**

A malaxing state sensor based on computer vision has been developed for the purpose of determining the malaxing state of the olive paste inside the thermomixer. Different features were extracted from the images acquired by the vision sensor such as viscosity, granularity and floating oil over the olive paste surface. The best descriptors related to the previous features were the average intensity in the infrared channel and the homogeneity textural parameter in the green channel. The proposed methodology was evaluated in plant during the olive oil campaign.

Furthermore it has been implemented a recommendation system based on fuzzy logic. It recommends the value of different process variables, such as malaxing time, temperature, addition of coadjuvant and rotating speed of the blades, from the information delivered by the malaxing state sensor and the knowledge of the master miller.

- **Virgin olive oil characterization**

A system based on hyperspectral image technology operating in the near infrared spectrum was developed in order to assess olive oil quality (free acidity and peroxides index) and moisture content in a non-invasive, contactless and online form. The acquired images were processed and the spectral bands which are correlated with the aforementioned parameters of interest were identified. For this purpose different feature selection algorithms were coded, such as GA-MLR, SPA-MLR and LASSO. The best results were achieved with the moisture content parameter.

Within this research line, a second system based on FT-NIR was developed with the objective of determining the polyphenols content in olive oil samples, also in a non-invasive form. For this purpose, two multivariate calibration algorithms, PLS and SMLR, were evaluated. SMLR with a previous SG filtering

step was slightly more accurate than PLS and it only used three bands.

8.2 Future researches

The future research lines for each thematic area that has been studied in this thesis are outlined below.

- **Olive fruit classification at the input of the VOOPP**

While the classification of the olive fruit according to the harvesting method at the input of the VOOPP has been studied in the present thesis, it could be interesting to consider other classification criteria such as presence of contaminant substances in the batch, pest infestations or different maturity indexes since they affect directly to the olive oil quality. These parameters, together with the work performed in this research line, will be taken into account to develop the software of the industrial olive classification machine. This machine is currently being constructed and it will allow the oil mill to characterize and classify olives individually.

- **Malaxing state assessment of the olive paste**

In addition to evaluate the physical parameters related to the visual aspect of the olive paste during the malaxing process, it could be interesting to assess the chemical parameters during the same process by means of using non-invasive technology such as the well-known FT-NIR or hyperspectral imaging technologies. In the last campaign, as part of the works carried out in CARDIOLIVE project, olive paste was sampled from the thermomixer at different times and temperatures. The analysis of the olive oil extracted from this olive paste will be useful to provided added value to the developed malaxing sensor.

- **Virgin olive oil characterization**

There are other factors which can affect to the results of the implemented system, such as the temperature or the flow rate of the elaborated olive oil. It could be interesting to evaluate this influence in order to add ruggedness to the developed system. Currently GRAV is designing and implementing a NIR sensor of low cost in order to be installed in a distributed form in all the lines of the production plant. The results and the expertise acquired with this research line will be used to implement this distributed system.

Appendix A

Documentación en castellano

El acuerdo del 29 de julio de 2013 de la Comisión de Doctorado de la Universidad de Jaén sobre los idiomas en que deben redactarse las tesis doctorales establece que todas las tesis doctorales redactadas en un idioma diferente al castellano deberán incluir además en la edición del volumen que se deposite para exposición pública los siguientes apartados en castellano: título, índice, introducción, resumen y conclusiones. En base a este acuerdo se ha redactado el presente anexo.

A.1 Título

APLICACIÓN DE TECNOLOGÍAS SENSORIALES NO CONVENCIONALES PARA LA
MONITORIZACIÓN Y CONTROL AUTOMÁTICO DEL PROCESO DE ELABORACIÓN DEL
ACEITE DE OLIVA VIRGEN

A.2 Índice

1. Introducción	1
1.1. Contexto científico	5
1.2. Hipótesis y objetivos	7

1.3.	Contribuciones	8
1.4.	Estructura del documento	10
2.	Revisión bibliográfica	13
2.1.	Introducción	13
2.2.	Calidad del aceite de oliva	16
2.3.	Etapas del proceso de producción	18
2.4.	Espectroscopía infrarroja	23
2.4.1.	Conceptos básicos	23
2.4.2.	Estado del arte relacionado con la aplicación de la espectroscopía infrarroja sobre aceituna entera	30
2.4.3.	Espectroscopía infrarroja aplicada a muestras de aceite de oliva	33
2.4.4.	Trabajos de investigación relacionados con la aplicación de la espectroscopía infrarroja sobre pasta de aceituna y orujos	38
2.5.	Visión por computador	48
2.5.1.	Principios de operación	48
2.5.2.	Visión por computador aplicada a aceitunas enteras	50
2.5.3.	Visión por computador aplicada sobre muestras de aceite de oliva	54
3.	Sistema basado en información hiperespectral para estimar acidez, contenido en humedad e índice de peróxidos en muestras de aceite de oliva	61
3.1.	Introducción	61
3.2.	Materiales y métodos	63
3.2.1.	Muestras de aceite de oliva	63
3.2.2.	Métodos analíticos	63
3.2.3.	Setup experimental	63
3.2.4.	Preprocesamiento de imágenes hiperespectrales	65
3.2.5.	Métodos de selección de longitudes de onda	67
3.3.	Resultados y discusión	68
3.4.	Conclusiones y líneas futuras	71
4.	Metodología no invasiva para estimar el contenido total en polifenoles en muestras de AOVE basada en FT-NIR y selección de bandas de interés	73
4.1.	Introducción	73
4.2.	Descripción del problema y trabajos previos	75

4.3.	Materiales y métodos	76
4.3.1.	Muestras de aceite de oliva virgen	77
4.3.2.	Setup experimental	78
4.3.3.	Proceso de adquisición de datos	79
4.3.4.	Preprocesado del espectro	81
4.3.5.	Filtrado del espectro	83
4.3.6.	Reducción de dimensionalidad con modelos de regresión	84
4.4.	Resultados experimentales y discusión	86
4.5.	Conclusiones y líneas futuras	90
5.	Sistema experto basado en visión por computador para la monitorización del estado de batido de la pasta de aceituna	93
5.1.	Introducción	93
5.2.	Descripción del problema	95
5.3.	Materiales y métodos	97
5.3.1.	Setup experimental	97
5.3.2.	Determinación de viscosidad	98
5.3.3.	Determinación de aceite flotante	100
5.3.4.	Determinación de la granularidad de la pasta	100
5.3.5.	Diseño del controlador difuso	102
5.4.	Resultados experimentales y discusión	105
5.5.	Conclusiones y líneas futuras	108
6.	Clasificación de lotes de aceitunas para el proceso de producción de AOVE empleando procesado de imágenes	111
6.1.	Introducción	111
6.2.	Materiales y métodos	113
6.2.1.	Setup experimental	114
6.2.2.	Muestras de aceitunas	117
6.2.3.	Análisis de imágenes	117
6.2.4.	Vector de características	119
6.2.5.	Análisis del Discriminante de Fisher	120
6.2.6.	Red Neuronal Artificial	121
6.3.	Resultados experimentales y discusión	122

6.4. Conclusiones y líneas futuras	127
7. Resultados y discusión	129
8. Conclusiones y líneas futuras	133
8.1. Resumen de conclusiones	134
8.2. Líneas futuras de investigación	136
A. Documentación en castellano	137
A.1. Título	137
A.2. Índice	137
A.3. Introducción	140
A.3.1. Contexto científico	144
A.3.2. Hipótesis y objetivos	145
A.3.3. Contribuciones	147
A.3.4. Estructura del documento	150
A.4. Resumen	151
A.5. Conclusiones generales y líneas futuras	151
A.5.1. Resumen de conclusiones	152
A.5.2. Líneas futuras de investigación	154

A.3 Introducción

La producción de aceite de oliva virgen es una actividad económica importante llevada a cabo en más de treinta países y ha experimentado una evolución significativa durante los últimos años desde el punto de vista de la calidad y de la cantidad de aceite producido [105]. La instalación generalizada del sistema de extracción de dos fases ha permitido incrementar la capacidad de producción y, al mismo tiempo, reducir el personal necesario para operar la planta. Como dato de interés la cantidad de aceite de oliva producido en la última campaña fue superior a los 2.4 millones de toneladas [105]. Los trabajos de investigación, la transferencia al sector de buenas prácticas y la necesidad de producir aceite con el mejor índice de calidad posible para poner en valor el producto, han incrementado el nivel medio de la calidad del aceite de oliva producido [189]. Por otro lado, en un mercado cada vez más creciente y competitivo, la introducción en el proceso de avances tecnológicos puede ser útil como ventaja competitiva frente al resto de competidores. Los sistemas de automatización y control aplicados a las almazaras pueden considerarse como uno

de estos avances tecnológicos que proporcionan una ventaja competitiva para las fábricas de aceite de oliva [43].

Automatizar un proceso consiste en controlar de forma automática los parámetros fundamentales que determinan el funcionamiento de dicho proceso [175]. Para ello es necesario instalar una serie de sensores y actuadores en la planta de producción, así como también incorporar elementos o dispositivos de control como son los reguladores y los sistemas de supervisión y adquisición de datos (SCADA). Los sensores son capaces de medir en cada momento las variables fundamentales del proceso y enviar electrónicamente dicha información a los dispositivos de control. En el caso de la automatización del proceso de extracción del aceite de oliva estos sensores son principalmente sondas de temperatura y caudalímetros, que miden la temperatura y el caudal de los distintos fluidos (masa de aceituna, agua, aceite y alpeorujo), así como amperímetros y frecuenciómetros que miden el consumo y la velocidad de los diferentes motores de la almazara. El proceso de elaboración de aceite de oliva virgen (PEAOV) es un proceso complejo compuesto por diferentes etapas interconectadas y ofrece un amplio rango de posibilidades para controlarlo de forma automática. La Figura A.1 y la Figura A.2 resumen las distintas opciones de automatización para cada etapa del proceso.



Figure A.1 Posibilidades de automatización de las fases del proceso.



Figure A.2 Posibilidades de automatización de las fases del proceso (continuación).

Por otro lado diferentes grupos de investigación han evaluado la posibilidad de usar tecnologías no invasivas para estimar propiedades del aceite de oliva y de la aceituna a nivel científico. Las tecnologías no invasivas más estudiadas durante los últimos años han sido espectroscopía en el infrarrojo cercano (NIRS), visión por computador, sensores olfativos electrónicos (e-nose) y sensores gustativos electrónicos (e-tongue). Mientras que la mayoría de las aplicaciones basadas en visión por computador han estado enfocadas fundamentalmente en la caracterización de aceitunas detectando defectos externos sobre su piel y en la estimación del índice de madurez de lotes de aceituna, el resto de aplicaciones han sido desarrolladas para detectar aceites adulterados, predecir su origen geográfico y estimar parámetros químicos y organolépticos relacionados con la calidad del aceite.

Además de la sensorística, es necesario instalar actuadores que puedan ser controlados electrónicamente y que son los que modifican el régimen de funcionamiento del proceso. Para el caso de la automatización de almazaras estos actuadores son normalmente electro-válvulas, con las que se regulan los diferentes caudales y se controlan las temperaturas de los distintos fluidos. También existen variadores de velocidad con los cuales se regula la velocidad de giro de los motores de las distintas máquinas y bombas de impulsión.

Tanto los sensores como los actuadores suelen estar gobernados por un programa de control instalado en un autómata programable (PLC), o un computador personal (PC), que permite fijar los parámetros de funcionamiento del sistema, estableciendo

unas consignas, bien de forma manual o por el contrario de manera automática. A partir de los datos recibidos de los sensores acerca del estado del proceso y con las consignas marcadas, el programa calcula la velocidad de los motores y la apertura de las electro-válvulas necesarias para alcanzar los parámetros de funcionamiento establecidos por el usuario.

El control automático aplicado al (PEAOV) puede ser interesante para alcanzar el objetivo de elaboración propuesto, el cual será a su vez función de otros objetivos tales como agotamiento, calidad y eficiencia energética. La Figura A.3 muestra la superficie que relaciona los parámetros mencionados. Normalmente no es posible alcanzar de forma simultánea el valor máximo de todos los parámetros. Por ejemplo, en términos generales de un lote de aceitunas no es posible elaborar aceite con la máxima calidad posible y al mismo tiempo extraer la máxima cantidad posible de aceite. Los trabajos que se desarrollan en esta tesis doctoral están enfocados en determinar en línea el punto real de operación de la planta.

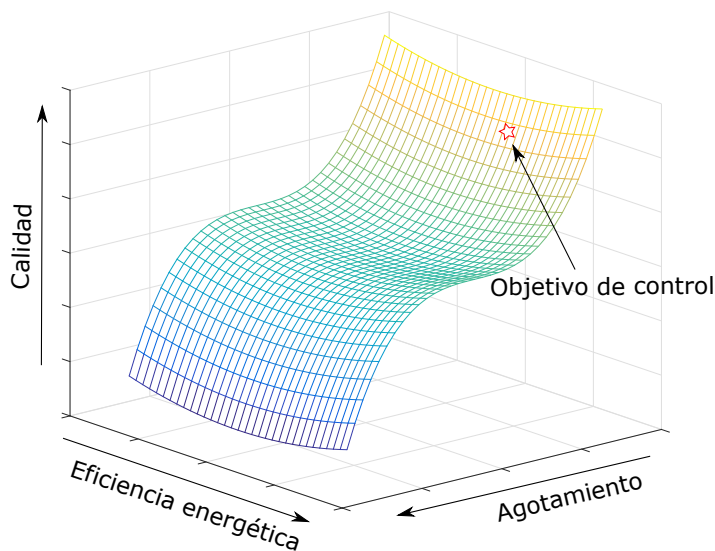


Figure A.3 Relación entre calidad, agotamiento y eficiencia energética.

Actualmente el nivel de automatización de las plantas de producción de aceite de oliva virgen es inferior al de otras industrias agro-alimentarias [25]. En la mayoría de los casos el maestro de almazara es la persona encargada de tomar decisiones en el patio de almazara (por ejemplo supervisa la calidad de la aceituna transportada por los agricultores y selecciona la tolva de recepción) y dentro de fábrica para regular manualmente las variables que intervienen en el proceso de extracción en base a su experiencia (por ejemplo regula la cantidad de agua añadida a la pasta

de aceituna antes de su entrada al decánter en base al aspecto visual de la pasta). Por otro lado, tanto la aceituna a la entrada del proceso como el aceite elaborado son muestreados y analizados por laboratorios externos de acuerdo a diferentes normativas europeas [66]. Normalmente las analíticas que se les practica a las muestras ocupan bastante tiempo y no sería posible tomar decisiones para reajustar el proceso a tiempo en caso de detectar desviaciones (por ejemplo la adición de agua modifica de forma cuantitativa el perfil fenólico del aceite producido). Además de consumir bastante tiempo, las técnicas analíticas son destructivas y requieren de personal cualificado para llevarlas a cabo. En este contexto las tecnologías analíticas de proceso (PAT) no invasivas pueden ser una herramienta útil para adquirir información del producto procesado y para proveer de información a los lazos de control de nivel superior.

El objetivo final de la presente tesis es proporcionar valor añadido al PEAOV a través de tecnologías sensoriales no invasivas que permitan inferir propiedades del producto en sus diferentes fases de producción. Estas propiedades son útiles para caracterizar el producto elaborado y clasificarlo automáticamente atendiendo a su calidad o para tomar decisiones destinadas a alcanzar el objetivo de elaboración propuesto.

A.3.1 Contexto científico

Los estudios que se recogen en la presente tesis doctoral están enmarcados en la línea de investigación relacionada con la aplicación del control automático a la olivicultura y elaiotecnía, línea que está siendo desarrollada actualmente por el Grupo de Robótica, Automática y Visión por Computador (GRAV) de la Universidad de Jaén, TEP-237 (grav.ujaen.es). Por otro lado los trabajos de investigación que se han desarrollado han sido financiados por diferentes proyectos. En orden cronológico estos son:

- **AGR-6616:** "Modelado y optimización del proceso de elaboración de aceite de oliva virgen desde el punto de vista de la calidad, productividad y eficiencia energética desarrollando metodologías analíticas y de control avanzadas"
- **NUTRAOLEUM:** "Estudio de nuevos procesos en las industrias almazarera andaluza para la elaboración de nuevos productos de alto valor biológico con aplicación en salud humana"
- **CARDIOLIVE:** "Desarrollo de suplementos alimenticios a partir de productos del olivar dirigidos a la protección cardiovascular humana"

El primer proyecto, AGR-6616, comenzó en 2011 dentro del programa de Proyectos de Excelencia de la Junta de Andalucía. Dentro del proyecto se plantearon tres

objetivos: Caracterizar el PEAOV con el objetivo de regular la calidad del aceite de oliva elaborado; Determinar la referencia de control de acuerdo a las condiciones de producción; Desarrollar e implementar estrategias de control multivariable y multi-objetivo dentro del PEAOV. Desde entonces los recursos del GRAV relacionados con esta línea de investigación han incrementado de forma significativa y actualmente dispone de una almazara piloto experimental en el parque tecnológico de GEOLIT, de sensores no invasivos (por ejemplo de un equipo NIR y de una nariz electrónica), cámaras de visión por computador y sistemas de iluminación en el espectro visible e infrarrojo.

El segundo proyecto, NUTRAOLEUM, fue financiado por el programa FEDER Innterconecta en 2013. Este proyecto cubrió diferentes áreas de conocimiento, desde experimentación a nivel de campo hasta pruebas clínicas con personas. El objetivo final del proyecto fue la elaboración de aceites de oliva funcionales con alto contenido en componentes nutricionales. Los experimentos se llevaron a cabo con diferentes variedades de aceituna y con diferentes variables de proceso en sucesivas jornadas de experimentación en planta de producción. Los aceites de oliva obtenidos se introdujeron en la dieta de una selección de personas para evaluar la influencia de los componentes estudiados sobre el síndrome metabólico.

El tercer proyecto, CARDIOLIVE, comenzó en 2015 financiado también por el programa FEDER Innterconecta y actualmente se encuentra en progreso. Los objetivos del proyecto son dos: El primero es la producción de productos nutracéuticos a partir de los principales compuestos nutricionales extraídos del aceite de oliva; El segundo es la investigación de nuevos sistemas de extracción de aceite que fomenten la presencia de estos compuestos nutricionales.

Por otro lado, los trabajos de experimentación llevados a cabo para desarrollar los estudios presentes en esta tesis doctoral han sido parcialmente llevados a cabo gracias a acuerdos de colaboración entre el GRAV y diferentes almazaras de la provincia de Jaén (p. ej. la almazara San Bartolomé del grupo Oleocampo, Torredelcampo, Jaén), y también con laboratorios de análisis químicos de aceite (p. ej. el laboratorio CM Europa de Martos, Jaén). Además, en el contexto de los proyectos que se han mencionado, se firmaron diferentes contratos de transferencia tecnológica entre el GRAV y la compañía GEA Westfalia Separator Ibérica S.A. Estas colaboraciones han permitido validar las propuestas tecnológicas presentadas en esta tesis doctoral en la cooperativa San Francisco de Asís S.C.A. (Montefrío, Granada).

A.3.2 Hipótesis y objetivos

En el contexto del sector agro-alimentario relacionado con el proceso de elaboración de aceite de oliva virgen, la hipótesis central de esta tesis doctoral se basa en la



Figure A.4 Logotipo de los proyectos.

introducción de tecnologías sensoriales no invasivas en el PEAOV para mejorarlo desde el punto de vista de la calidad, la cantidad y el valor nutricional del producto elaborado, y al mismo tiempo conocer el punto de operación del proceso. Ésto, junto con el empleo de tecnologías de control automático, supondrá un avance científico en la producción de aceite de oliva virgen.

Bajo esta premisa, los estudios de investigación llevados a cabo en esta tesis han estado enfocados sobre los cuatro objetivos principales que se enumeran a continuación:

1. Evaluar el empleo de la tecnología de imagen hiperespectral como sistema automático, no invasivo y sin contacto para la caracterización y el control de calidad del aceite de oliva producido.
 - a) Desarrollar metodologías predictivas para evaluar la calidad del aceite producido de forma rápida y no destructiva determinando el contenido en humedad, el índice de peróxidos y la acidez libre a partir de información hiperespectral.
 - b) Identificar las bandas de absorción que están mejor correlacionadas con los parámetros del aceite de oliva considerados.
2. Desarrollar modelos predictivos a partir de la información capturada por el equipo FT-NIR para determinar de forma automática la concentración de polifenoles totales en muestras de aceite de oliva.
 - a) Aplicar métodos de selección de características con el objetivo de minimizar el error de predicción, identificando las bandas mejor correlacionadas con el contenido en polifenoles.
 - b) Determinar la influencia del número de capturas por muestra y del sistema de iluminación sobre los resultados de predicción.

3. Asistir al maestro de almazara para regular las variables tecnológicas que están involucradas en el proceso de batido.
 - a) Desarrollar un sensor software para determinar el estado de batido de la pasta de aceituna de acuerdo a diferentes parámetros tales como aceite sobrenadante, granularidad y viscosidad.
 - b) Caracterizar la pasta de forma continua y recomendar consignas de acuerdo al estado de batido de la pasta.
4. Caracterizar y clasificar de forma automática aceitunas a la entrada del PEAOV empleando para ello visión por computador.
 - a) Identificar y extraer de las imágenes capturadas las características más discriminantes entre aceitunas sanas y aceitunas con defectos.
 - b) Clasificar lotes de aceitunas de forma automática y de acuerdo al método de recolección empleado (suelo o vuelo).

A.3.3 Contribuciones

Esta sección muestra los trabajos desarrollados de acuerdo a los objetivos planteados en la tesis. Los trabajos que se enumeran han sido publicados en revistas científicas y en congresos y jornadas nacionales e internacionales. La lista de contribuciones se ha organizado en dos bloques principales. El primer bloque contiene las contribuciones en las que la tecnología basada en espectroscopía en el infrarrojo cercano ha sido evaluada para determinar el estado de operación del proceso. A continuación, el segundo bloque describe las contribuciones relacionadas con el empleo de visión por computador en el proceso.

1. Publicaciones relacionadas con la aplicación de tecnologías no invasivas para determinar parámetros del aceite de oliva virgen relacionados con la calidad, humedad, valor nutricional y presencia de impurezas:
 - [138] **Martínez Gila, D.M.**, Cano Marchal, P., Gámez García, J., Gómez Ortega, J., *On-line system based on hyperspectral information to estimate acidity, moisture and peroxides in olive oil samples*. In Journal of Computers and Electronics in Agriculture, 2015, 116:1–7.
DOI: 10.1016/j.compag.2015.06.002
 - [133] Cano Marchal, P., **Martínez Gila, D.M.**, Gámez García, J., Gómez Ortega, J., *Expert system based on computer vision to estimate the content of impurities in olive oil samples*. In Journal of Food Engineering, Elsevier

Ltd, 2013, 119(2):220–228.

DOI: 10.1016/j.jfoodeng.2013.05.032

- [154] Beltran Ortega, J., **Martínez Gila, D.M.**, Aguilera Puerto, D., Gámez García, J., Gómez Ortega, J., *Novel Technologies for Monitoring the In-Line Quality of Virgin Olive Oil during Manufacturing and Storage*. Journal of Science in Food and Agriculture, 2016. Unedited article.
DOI: 10.1002/jsfa.7733
 - [140] **Martínez Gila, D.M.**, Cano Marchal, P.C., Gámez García, J., Gómez Ortega, J., *Non-invasive methodology to estimate polyphenol content in EVOO based on Stepwise Multilinear Regression*. In Journal of Computers and Electronics in Agriculture, 2016, Under review.
 - [81] **Martínez Gila, D.M.**, Cano Marchal, P., Gámez García, J., Gómez Ortega, J., *Hyperspectral imaging for determination of some quality parameters for olive oil*. In Proceedings of IEEE 18 International Conference on Automation and Computing (ICAC), 2012.
 - [80] **Martínez Gila, D.M.**, Beltran Ortega, J., Gámez García, J., Gómez Ortega, J., *Rapid quantification of total polyphenol content in EVOO using NIR sensor with wavelength selection and FS-MLR*. In Proceedings of IEEE International Conference on Imaging Systems and Techniques (IST), 2015.
DOI: 10.1109/IST.2015.7294542
 - [134] Cano Marchal, P., **Martínez Gila, D.M.**, Gámez García, J., Gómez Ortega, J., *Application of computer vision and Support Vector Machines to estimate the content of impurities in olive oil samples*. In Proceedings of IEEE 18 International Conference on Automation and Computing (ICAC), 2012.
 - [139] **Martínez Gila, D.M.**, Beltran Ortega, J., Gámez García, J., Gómez Ortega, J., *Tecnología no invasiva para la determinación automática del contenido total de polifenoles en el aceite de oliva virgen*. In Proceedings of XXXVI Jornadas de Automática, 2015.
2. Publicaciones relacionadas con el empleo de visión por computador para caracterizar y clasificar aceitunas a la entrada del PEAOV, y para el control automático del estado de batido de la pasta de aceituna:
- [160] Aguilera Puerto, D., **Martínez Gila, D.M.**, Gámez García, J., Gómez Ortega, J., *Sorting Olive Batches for the Milling Process Using Image Processing*. In Journal of Sensors, 2015, 15(7):15738–15754.
DOI: 10.3390/s150715738

- [82] **Martínez Gila, D.M.**, Cano Marchal, P., Gámez García, J., Gómez Ortega, J., *Control System of the Malaxing State for the Olive Paste Based on Computer Vision and Fuzzy Logic*. In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2013, pp. 1951–1956. DOI: 10.1109/SMC.2013.335
- [137] **Martínez Gila, D.M.**, Cano Marchal, P., Gámez García, J., Gómez Ortega, J., *Extractability control for the virgin olive oil elaboration process*. In Proceedings of IEEE International Conference on Industrial Technology (ICIT), 2015, pp. 240–245. DOI: 10.1109/ICIT.2015.7125105
- [136] **Martínez Gila, D.M.**, Aguilera Puerto, D., Gámez García, J., Gómez Ortega, J., *Automatic classification of olives for oil production using computer vision*. In Proceedings of IEEE International Conference on Industrial Technology (ICIT), 2015, pp. 1651–1656. DOI: 10.1109/ICIT.2015.7125334
- [149] Cáceres Moreno, O., **Martínez Gila, D.M.**, Aguilera Puerto, D., Gámez García, J., Gómez Ortega, J., *Automatic determination of peroxides and acidity of olive oil using machine vision in olive fruits before milling process*. In Proceedings of IEEE International Conference on Imaging Systems and Techniques (IST), 2015. DOI: 10.1109/IST.2015.7294543
- [141] **Martínez Gila, D.M.**, Cano Marchal, P., Gámez García, J., Gómez Ortega, J., *Determinación del estado de batido de la pasta de aceituna empleando visión por computador*. In Proceedings of XXXIV Jornadas de Automática, 2013, pp. 826–833.
- [159] Aguilera Puerto, D., Cáceres Moreno, O., **Martínez Gila, D.M.**, Gámez García, J., Gómez Ortega, J., *Olive batches automatic classification in mill reception using computer vision*. In Proceedings of IEEE International Conference on Imaging Systems and Techniques (IST), 2015. DOI: 10.1109/IST.2015.7294544
- [1] Aguilera Puerto, D., Cáceres Moreno, O., **Martínez Gila, D.M.**, Gámez García, J., Gómez Ortega, J., *Clasificación automática de lotes de aceitunas en almazara mediante visión por computador*. In Proceedings of XXXVI Jornadas de Automática, 2015, pp. 59–65.

A.3.4 Estructura del documento

El documento de tesis se ha estructurado siguiendo el conjunto de capítulos que se describen a continuación:

- El **Capítulo 1** contiene la introducción de la tesis. Además este capítulo presenta el contexto en el que la tesis ha sido desarrollada, los objetivos a ser alcanzados con el desarrollo de la línea de investigación, las contribuciones que se han realizado y la estructura del documento.
- El **Capítulo 2** presenta el estado del arte a partir del cual se ha desarrollado el trabajo descrito. En él se describen los principales parámetros que determinan la calidad del aceite de oliva y presenta las principales fases del proceso productivo. Además contiene una revisión completa a fecha de mayo de 2016 de los principales trabajos de investigación que han utilizado tecnologías no invasivas dentro del PEAOV.
- El **Capítulo 3** describe el trabajo que se ha realizado para estimar el contenido en humedad, el grado de acidez y el índice de peróxidos sobre muestras de aceite de oliva a partir de imágenes hiperespectrales.
- El **Capítulo 4** muestra la metodología empleada para estimar el contenido de polifenoles totales en muestras de aceite de oliva empleando la tecnología FT-NIR.
- El **Capítulo 5** presenta el procedimiento desarrollado para describir el estado de batido de la pasta de aceituna en el interior de la termobatidora y recomendar el valor de las variables de proceso de acuerdo al aspecto visual de la pasta.
- El **Capítulo 6** describe el sistema desarrollado para clasificar automáticamente lotes de aceitunas antes del proceso de molturación y en base al método de recolección empleado.
- El **Capítulo 7** expone la discusión de los resultados obtenidos a partir del trabajo realizado.
- Por último, el **Capítulo 8** describe las conclusiones que han sido obtenidas para cada una de las líneas de investigación desarrolladas y enumera los trabajos de investigación futuros.

A.4 Resumen

En los últimos años la producción de aceite de oliva ha experimentado un crecimiento desde el punto de vista cualitativo y cuantitativo, en parte gracias a la transferencia de conocimiento desde los centros de investigación hacia la planta de producción. Sin embargo, el grado de automatización del sistema productivo se encuentra de momento en una etapa inicial del recorrido que supone ponerse al nivel de otras industrias agroalimentarias. En este contexto, el trabajo que se desarrolla en la presente Tesis Doctoral tiene como objetivo profundizar en la aplicación de dos tecnologías no invasivas, espectroscopía en el infrarojo cercano y visión por computador, para adquirir información a nivel de proceso, y en sus distintas etapas, y fundamentalmente con dos fines: (1) ayudar a determinar de forma cualitativa qué está ocurriendo en el proceso desde el punto de vista físico-químico y organoléptico, y (2) poder utilizar la información adquirida en estructuras de control automático de nivel local (subprocesos), global (proceso) o incluso en la toma de decisiones sobre el modo de operación de la planta como es el caso de la planificación de la producción.

A.5 Conclusiones generales y líneas futuras

Los trabajos que se han llevado a cabo en esta tesis doctoral han estado relacionados fundamentalmente con la aplicación de tecnologías sensoriales no invasivas y con el procesamiento de la información capturada para caracterizar el producto en las distintas etapas del proceso de elaboración y para mejorar el proceso desde el punto de vista de la calidad, la cantidad y el valor nutricional del producto elaborado. La fase de experimentación ha sido uno de los puntos fuertes de la investigación que se ha desarrollado y ésta ha sido posible gracias a la colaboración del grupo de investigación (GRAV) con diferentes almazaras de dentro y fuera de la provincia de Jaén. Los trabajos de experimentación en planta han permitido disponer del producto objeto de estudio en las diferentes condiciones que naturalmente se dan durante la campaña de producción de aceite y obtener los resultados ya descritos.

La caracterización del fruto de entrada al proceso es sin duda una de las principales tareas a realizar para cualquier objetivo de elaboración que se plantee conseguir. Existe una gran variabilidad de condiciones en las cuales el fruto llega al patio de la almazara y estas tienen un gran impacto en las características de los productos que subsiguen en el proceso (pasta y aceite). El trabajo de investigación desarrollado en esta tesis y relacionado con esta etapa del proceso ha tenido la intención de caracterizar el fruto en base a su método de recolección (suelo o vuelo)

empleando visión por computador, para posteriormente poder clasificar los lotes de aceituna de forma automática. Para ello se han extraído y evaluado las características más discriminantes entre las dos clases consideradas y se han implementado y evaluado dos algoritmos de clasificación, uno basado en discriminante lineal y otro en redes neuronales.

A continuación, la aceituna que entra al proceso de elaboración se tritura en la fase de molienda dando lugar a la pasta de aceituna. El batido de la pasta de aceituna en el interior de la termobatidora es lo más parecido a un reactor biológico en el que la aportación de calor y movimiento durante el proceso da lugar a la conversión de unas formas moleculares en otras a la vez que modifica el aspecto visual de la pasta. Caracterizar la evolución de la pasta de aceituna durante el proceso de batido ha sido otro de los objetivos de esta tesis doctoral y para ello se ha empleado visión por computador en el espectro visible e infrarrojo. Una de las principales tareas de este trabajo ha sido identificar y extraer las características de color y textura que mejor correlacionan con el tiempo de batido para, a partir de ellas, implementar un sensor de estado de batido de la pasta. Además se ha implementado un sistema de recomendación para regular las variables de proceso tomando como base la información del sensor y el conocimiento experto del maestro de almazara.

Finalmente, conocer en tiempo de proceso las características del producto final es interesante para poder regular el proceso con el fin de alcanzar el objetivo de elaboración propuesto. A este tema se le ha dedicado una de las líneas de investigación desarrolladas en esta tesis doctoral. Mediante la implementación de metodologías de preprocesado y análisis del espectro de absorción en el infrarrojo cercano se ha tratado de inferir características del aceite relacionadas con su calidad y valor nutricional. Además se han identificado las bandas mejor correlacionadas con el parámetro de interés implementando metodologías de extracción de características y se ha evaluado la influencia de las condiciones de medida sobre los resultados de predicción.

Si bien, cada uno de los capítulos de esta tesis incluye las conclusiones de las investigaciones realizadas, se recogen, a continuación, las conclusiones más relevantes de cada una de las líneas de trabajo que se han tratado: caracterización del aceite elaborado (Capítulos 3 y 4), caracterización de la pasta de aceituna (Capítulo 5) y caracterización del fruto de entrada al proceso (Capítulo 6). Además, se detallan las líneas futuras de investigación en cada una de estas temáticas.

A.5.1 Resumen de conclusiones

A continuación se resumen las principales conclusiones obtenidas a partir de las aportaciones introducidas por la presente tesis.

- **Caracterización del fruto de entrada al proceso.**

Se ha desarrollado un sistema basado en visión por computador para la clasificación de aceitunas suelo-vuelo en el patio de la almazara, a partir de las imágenes adquiridas por el sensor de visión y de la extracción de características relacionadas con el color y la textura de cada uno de los lotes analizados. El sistema desarrollado ha sido validado a nivel de laboratorio y a nivel industrial.

Con el fin de identificar las características más discriminantes entre las dos clases consideradas y eliminar las no deseadas se aplicó sobre el conjunto de datos de entrada el algoritmo de análisis de varianza. Para reducir el número de variables de entrada al clasificador se realizó un análisis de componentes principales y posteriormente se evaluaron dos algoritmos de clasificación: clasificador basado en el discriminante de Fisher y clasificador basado en red neuronal. Ambos clasificadores obtuvieron resultados similares aunque para la red neuronal se necesitaron menos características descriptoras de la clase.

- **Determinación del estado de batido de la pasta de aceituna.**

Se ha desarrollado un sensor de batido basado en visión por computador para la determinación del estado de batido de la pasta de aceituna. A partir de las imágenes capturadas por el sensor de visión se han extraído características relacionadas con la viscosidad, la granularidad y la presencia de aceite sobrenadante sobre la superficie de la pasta. Los parámetros descriptivos mejor relacionados con las características anteriores fueron la intensidad media del canal infrarrojo y el parámetro homogeneidad en el canal verde, este último relacionado con la textura de la pasta. La metodología desarrollada se evaluó en planta durante el proceso de batido de un lote de aceitunas.

Además se ha desarrollado un sistema de recomendación basado en lógica difusa, en el que a partir de la información generada por el sensor de batido y la experiencia del maestro de almazara se recomienda el valor de diferentes consignas tales como tiempo de batido, temperatura, adición de coadyuvante y velocidad de rotación de las palas de la termobatidora.

- **Caracterización del aceite de oliva elaborado.**

Se ha desarrollado un sistema basado en imagen hiperespectral en el infrarrojo cercano para la determinación no invasiva, sin contacto y en línea de dos de los principales parámetros que determinan la calidad del aceite de oliva (grado de acidez e índice de peróxidos) y de su contenido en humedad. Por otro lado se han identificado las bandas del espectro que mejor correlacionan con los parámetros de interés. Para ello se han evaluado diferentes algoritmos que permiten la selección de características tales como GA-MLR, SPA-MLR y

LASSO. Los mejores resultados se obtuvieron para el parámetro contenido en humedad.

Dentro de esta línea de investigación también se ha desarrollado un segundo sistema basado en FT-NIR para la determinación no invasiva del contenido en polifenoles totales sobre muestras de aceite de oliva. Para la identificación de las bandas espectrales mejor relacionadas con el compuesto de interés se evaluaron dos algoritmos de calibración multivariable, PLS y SMLR. El mejor error de predicción se obtuvo con SMLR empleando únicamente tres bandas de absorción cuando previamente los datos espectrales fueron preprocesados con el filtro SG.

A.5.2 Líneas futuras de investigación

A continuación se exponen las líneas de investigación futuras relacionadas con cada una de las temáticas estudiadas en la presente tesis.

- **Caracterización del fruto de entrada al proceso.**

En esta tesis se ha estudiado la caracterización del fruto de entrada al proceso en base al método de recolección suelo o vuelo. Además del método de recolección existen otras características tales como la presencia de sustancias contaminantes, la infestación por plagas o el índice de madurez de la aceituna que pueden influir sobre los parámetros de calidad del aceite elaborado. Estos parámetros unidos al trabajo desarrollado en esta línea se tendrán en cuenta para desarrollar el software de la máquina industrial clasificadora de aceituna que actualmente se está construyendo y que permitirá la caracterización y clasificación individual aceituna a aceituna.

- **Determinación del estado de batido de la pasta de aceituna.**

Además de evaluar los parámetros físicos relacionados con el aspecto visual de la pasta de aceituna durante el proceso de batido sería interesante determinar la evolución de los parámetros químicos durante el mismo proceso mediante el empleo de las tecnologías FT-NIR e hiperespectral ya conocidas. Durante la última campaña y como parte de los trabajos realizados dentro del proyecto CARDIOLIVE se recogieron muestras de pasta de aceituna durante el batido con diferentes tiempos y temperaturas. El análisis del aceite obtenido de estas muestras de pasta será interesante para añadir valor al sensor de batido que se ha desarrollado.

- **Caracterización del aceite de oliva elaborado.**

Además de los factores evaluados existen otras variables de proceso que

pueden influir en los resultados del sistema implementado, tales como temperatura o caudal del aceite a la salida del proceso de elaboración. Sería interesante evaluar esta influencia para aportarle robustez al sistema desarrollado. Por otro lado, actualmente el GRAV está diseñando e implementando un sensor NIR de bajo coste para ser instalado de forma distribuida en todas las líneas de la planta de producción. Parte de los resultados obtenidos y relacionados con la identificación de las banda de absorción de interés serán utilizados para la selección de este sensor en un futuro próximo.

Bibliography

- [1] Daniel Aguilera Puerto, Oscar Cáceres Moreno, Diego M. Martínez Gila, Javier Gámez García, and Juan Gómez Ortega, *Clasificación automática de lotes de aceitunas en almazara mediante visión por computador*, XXXVI Jornadas de Automática, 2015, pp. 59–65.
- [2] Yosra Allouche, Estrella López, Gabriel Beltrán Maza, and Antonio Jiménez Márquez, *Near infrared spectroscopy and artificial neural network to characterise olive fruit and oil online for process optimisation*, Journal of Near Infrared Spectroscopy **23** (2015), no. 2, 111.
- [3] Ayman H. Amer Eissa and Ayman A. Abdel Khalik, *Understanding Color Image Processing by Machine Vision for Biological Materials*, Structure and Function of Food Engineering, InTech, aug 2012, pp. 227–274.
- [4] C. M. Andersen and R. Bro, *Variable selection in regression-a tutorial*, Journal of Chemometrics **24** (2010), no. 11-12, 728–737.
- [5] Franca Angerosa, Roberta Mostallino, Carla Basti, and Raffaella Vito, *Influence of malaxation temperature and time on the quality of virgin olive oils*, Food Chemistry **72** (2001), no. 1, 19–28.
- [6] Mário César Ugulino Araújo, Teresa Cristina Bezerra Saldanha, Roberto Kawakami Harrop Galvão, Takashi Yoneyama, Henrique Caldas Chame, and Valeria Visani, *The successive projections algorithm for variable selection in spectroscopic multicomponent analysis*, Chemometrics and Intelligent Laboratory Systems **57** (2001), no. 2, 65–73.

- [7] S. Armenta, S. Garrigues, and M. de la Guardia, *Determination of edible oil parameters by near infrared spectrometry*, *Analytica Chimica Acta* **596** (2007), no. 2, 330–337.
- [8] S Armenta, J Moros, S Garrigues, and M de la Guardia, *The use of near-infrared spectrometry in the olive oil industry*, *Critical reviews in food science and nutrition* **50** (2010), no. 6, 567–82.
- [9] María José Ayora-Cañada, Barbara Muik, Jose Antonio García-Mesa, Domingo Ortega-Calderón, and Antonio Molina-Díaz, *Fourier-Transform Near-Infrared Spectroscopy as a Tool for Olive Fruit Classification and Quantitative Analysis*, *Spectroscopy Letters* **38** (2005), no. 6, 769–785.
- [10] Antonietta Baiano, Carmela Terracone, Giorgio Peri, and Roberto Romaniello, *Application of hyperspectral imaging for prediction of physico-chemical and sensory characteristics of table grapes*, *Computers and Electronics in Agriculture* **87** (2012), 142–151.
- [11] R. J. Barnes, M. S. Dhanoa, and Susan J. Lister, *Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra*, *Applied Spectroscopy* **43** (1989), no. 5, 772–777 (EN).
- [12] Diego Barranco Navero, Ricardo Fernandez Escobar, and Luis Rallo Romero, *El cultivo del olivo*, 6 ed., Madrid, Spain, 2008.
- [13] António S. Barros, Alexandra Nunes, Joana Martins, and Iyonne Delgadillo, *Determination of oil and water in olive and olive pomace by NIR and multivariate analysis*, *Sensing and Instrumentation for Food Quality and Safety* **3** (2009), no. 3, 180–186.
- [14] IA Basheer and M. Hajmeer, *Artificial neural networks: fundamentals, computing, design, and application*, *Journal of Microbiological Methods* **43** (2000), no. 1, 3–31.
- [15] Andrea Bellincontro, Giovanni Caruso, Fabio Mencarelli, and Riccardo Gucci, *Oil accumulation in intact olive fruits measured by near infrared spectroscopy-acousto-optically tunable filter*, *Journal of the Science of Food and Agriculture* **93** (2013), no. 6, 1259–1265.
- [16] Andrea Bellincontro, Agnese Taticchi, Maurizio Servili, Sonia Esposto, Daniela Farinelli, and Fabio Mencarelli, *Feasible Application of a Portable NIR-AOTF Tool for On-Field Prediction of Phenolic Compounds during the Ripening of*

- Olives for Oil Production*, Journal of Agricultural and Food Chemistry **60** (2012), no. 10, 2665–2673.
- [17] Alessandra Bendini, Lorenzo Cerretani, Fabio Di Virgilio, Paolo Belloni, Matteo Bonoli-Carbognin, and Giovanni Lercker, *Preliminary evaluation of the application of the FTIR spectroscopy to control the geographic origin and quality of virgin olive oils*, Journal of Food Quality **30** (2007), no. 4, 424–437.
- [18] Alessandra Bendini, Lorenzo Cerretani, Fabio Di Virgilio, Paolo Belloni, Giovanni Lercker, and Tullia Gallina Toschi, *In-process monitoring in industrial olive mill by means of FT-NIR*, European Journal of Lipid Science and Technology **109** (2007), no. 5, 498–504.
- [19] E. Bertran, M. Blanco, J. Coello, H. Iturriaga, S. MasPOCH, and I. Montoliu, *Near infrared spectrometry and pattern recognition as screening methods for the authentication of virgin olive oils of very close geographical origins*, Journal of Near Infrared Spectroscopy **8** (2000), no. 1, 45.
- [20] C.V.L. Bibiloni and J.M.A. Radic, *Digital measurement of color evolution in olives according to fruit maturity degree*, (2005), 33–40.
- [21] Hans-René Bjørsvik and Harald Martens, *Data Analysis: Calibration of NIR Instruments by PLS Regression*, Handbook of Near-Infrared Analysis, 2008, pp. 189–206.
- [22] K. Brkić Bubola, O. Koprivnjak, B. Sladonja, and I. Belobrajić, *Influence of storage temperature on quality parameters, phenols and volatile compounds of Croatian virgin olive oils*, Grasas y Aceites **65** (2014), no. 3, e034.
- [23] Brosnan and Sun, *Improving quality inspection of food products by computer vision—a review*, Journal of Food Engineering **61** (2004), no. 1, 3–16.
- [24] Tadhg Brosnan and Da-Wen Sun, *Inspection and grading of agricultural and food products by computer vision systems—a review*, Computers and Electronics in Agriculture **36** (2002), no. 2-3, 193–213.
- [25] Pablo Cano Marchal, Juan Gómez Ortega, Daniel Aguilera Puerto, and Javier Gámez García, *Situación actual y perspectivas futuras del control del proceso de elaboración del aceite de oliva virgen*, Revista Iberoamericana de Automática e Informática Industrial RIAI **8** (2011), no. 3, 258–269.
- [26] Pablo Cano Marchal, Diego Manuel Martínez Gila, Javier Gámez García, and

- Juan Gómez Ortega, *Optimal Production Planning for the Virgin Olive Oil Elaboration Process*, Proceedings of the 19th IFAC World Congress, 2014 (Boje Edward, ed.), 2014, pp. 8921–8926.
- [27] Francesco Caponio, Tommaso Gomes, Carmine Summo, and Antonella Pasqualone, *Influence of the type of olive-crusher used on the quality of extra virgin olive oils*, European Journal of Lipid Science and Technology **105** (2003), no. 5, 201–206.
- [28] Francesco Caponio, Carmine Summo, Vito M. Paradiso, and Antonella Pasqualone, *Influence of decanter working parameters on the extra virgin olive oil quality*, European Journal of Lipid Science and Technology **116** (2014), no. 12, 1626–1633.
- [29] Rafael Cárdenas and Javier Marcos, *Proceso de extracción de aceite de oliva*, El Manual del Maestro de Almazara (GEA Westfalia Separator Iberica, ed.), Jaén, España, 2 ed., 2013, pp. 87–143.
- [30] Rafael Cárdenas García and Javier Marcos Braojos, *Normativa y análisis*, El Manual del Maestro de Almazara (GEA Westfalia Separator Ibérica, ed.), Jaén, España, 2 ed., 2013, pp. 157–200.
- [31] Monica Carfagni, Marco Daou, Rocco Furferi, and Santa Marta, *Real-time estimation of olive oil quality parameters : a combined approach based on ANNs and Machine Vision*, International Conference on Computers (2008), 387–392.
- [32] Ana I. Carrapiso, Amparo García, M. Jesús Petrón, and Lourdes Martín, *Effect of talc and water addition on olive oil quality and antioxidants*, European Journal of Lipid Science and Technology **115** (2013), no. 5, 583–588.
- [33] M Casale, P Oliveri, C Casolino, N Sinelli, P Zunin, C Armanino, M Forina, and S Lanteri, *Characterisation of PDO olive oil Chianti Classico by non-selective (UV-visible, NIR and MIR spectroscopy) and selective (fatty acid composition) analytical techniques*, Analytica Chimica Acta **712** (2012), 56–63.
- [34] Monica Casale, Chiara Casolino, Giuseppe Ferrari, and Michele Forina, *Near infrared spectroscopy and class modelling techniques for the geographical authentication of Ligurian extra virgin olive oil*, Journal of Near Infrared Spectroscopy **16** (2008), no. 1, 39.

- [35] Monica Casale and Remo Simonetti, *Review: Near infrared spectroscopy for analysing olive oils*, Journal of Near Infrared Spectroscopy **22** (2014), no. 2, 59.
- [36] Pietro Catania, Mariangela Vallone, Felice Pipitone, Paolo Inglese, Giuseppe Aiello, and Giada La Scalia, *An oxygen monitoring and control system inside a malaxation machine to improve extra virgin olive oil quality*, Biosystems Engineering **114** (2013), no. 1, 1–8.
- [37] José A. Cayuela, Khaled Yousfi, M. Carmen Martínez, and José M. García, *Rapid Determination of Olive Oil Chlorophylls and Carotenoids by Using Visible Spectroscopy*, Journal of the American Oil Chemists' Society **91** (2014), no. 10, 1677–1684.
- [38] José Antonio Cayuela and María Del Carmen Pérez Camino, *Prediction of quality of intact olives by near infrared spectroscopy*, European Journal of Lipid Science and Technology **112** (2010), no. 11, 1209–1217.
- [39] José Antonio Cayuela, José María García, José María García, Nuria Caliani, and Nuria Caliani, *NIR prediction of fruit moisture, free acidity and oil content in intact olives*, Grasas y Aceites **60** (2009), no. 2, 194–202.
- [40] José a. Cayuela Sánchez, Wenceslao Moreda, and José M. García, *Rapid Determination of Olive Oil Oxidative Stability and Its Major Quality Parameters Using Vis/NIR Transmittance Spectroscopy*, Journal of Agricultural and Food Chemistry **61** (2013), no. 34, 8056–8062.
- [41] Lorenzo Cerretani, Angela Giuliani, Rubén M. Maggio, Alessandra Bendini, Tullia Gallina Toschi, and Angelo Cichelli, *Rapid FTIR determination of water, phenolics and antioxidant activity of olive oil*, European Journal of Lipid Science and Technology **112** (2010), no. 10, 1150–1157.
- [42] Alfred A. Christy, Sumaporn Kasemsumran, Yiping Du, and Yukihiro Ozaki, *The Detection and Quantification of Adulteration in Olive Oil by Near-Infrared Spectroscopy and Chemometrics*, Analytical Sciences **20** (2004), no. 6, 935–940.
- [43] Luis Civantos, *Obtención del aceite de oliva virgen*, Editorial Agrícola Española, Madrid, Spain, 1999.
- [44] Luis Civantos Lopez-Villalta, *Preparación de la pasta: molienda y batido*, Obtención del aceite de oliva virgen, Editorial Agrícola Española, S.A., Madrid, Spain, 1998, pp. 95–121.

- [45] Maria Lisa Clodoveo, *Malaxation: Influence on virgin olive oil quality. Past, present and future – An overview*, Trends in Food Science & Technology **25** (2012), no. 1, 13–23.
- [46] Maria Lisa Clodoveo and Rim Hachicha Hbaieb, *Beyond the traditional virgin olive oil extraction systems: Searching innovative and sustainable plant engineering solutions*, Food Research International **54** (2013), no. 2, 1926–1933.
- [47] D. Cozzolino, I. Murray, A. Chree, and J.R. Scaife, *Multivariate determination of free fatty acids and moisture in fish oils by partial least-squares regression and near-infrared spectroscopy*, LWT - Food Science and Technology **38** (2005), no. 8, 821–828.
- [48] Daniel Cozzolino, *Sample presentation, sources of error and future perspectives on the application of vibrational spectroscopy in the wine industry.*, Journal of the science of food and agriculture **95** (2015), no. 5, 861–8.
- [49] Arnon Dag, Zohar Kerem, Nir Yogev, Issac Zipori, Shimon Lavee, and Eric Ben-David, *Influence of time of harvest and maturity index on olive oil yield and quality*, Scientia Horticulturae **127** (2011), no. 3, 358–366.
- [50] E.R. Davies, *Machine vision: Theory, algorithms, practicalities*, Academic Press, London, UK, 1997.
- [51] Olivier Devos, Gerard Downey, and Ludovic Duponchel, *Simultaneous data pre-processing and SVM classification model selection based on a parallel genetic algorithm applied to spectroscopic data of olive oils*, Food Chemistry **148** (2014), 124–130.
- [52] Luciano Di Giovacchino, Nello Costantini, Arnaldo Serraiocco, Giulio Surricchio, and Carla Basti, *Natural antioxidants and volatile compounds of virgin olive oils obtained by two or three-phases centrifugal decanters*, European Journal of Lipid Science and Technology **103** (2001), no. 5, 279–285.
- [53] Luciano Di Giovacchino, Simona Sestili, and Daria Di Vincenzo, *Influence of olive processing on virgin olive oil quality*, European Journal of Lipid Science and Technology **104** (2002), no. 9-10, 587–601.
- [54] R. Diaz, G. Faus, M Blasco, J. Blasco, and E. Moltó, *The application of a fast algorithm for the classification of olives by machine vision*, Food Research International **33** (2000), no. 3-4, 305–309.

- [55] R. Diaz, L. Gil, C. Serrano, M. Blasco, E. Moltó, and J. Blasco, *Comparison of three algorithms in the classification of table olives by means of computer vision*, Journal of Food Engineering **61** (2004), no. 1, 101–107.
- [56] S. Domenico and W. Gary, *Machine vision and neuronal nets in food processing and packaging-natural way combinations*, Proceedings of the Food Processing Automation III Conference (Orlando, FL, USA), 1994, pp. 9–12.
- [57] Gerard Downey, Peter McIntyre, and Antony N. Davies, *Detecting and Quantifying Sunflower Oil Adulteration in Extra Virgin Olive Oils from the Eastern Mediterranean by Visible and Near-Infrared Spectroscopy*, Journal of Agricultural and Food Chemistry **50** (2002), no. 20, 5520–5525.
- [58] Gerard Downey, Peter McIntyre, and Antony N Davies, *Geographic Classification of Extra Virgin Olive Oils From the Eastern Mediterranean by Chemometric Analysis of Visible and Near-Infrared Spectroscopic Data*, Applied Spectroscopy **57** (2003), no. 2, 158–163.
- [59] Cheng-Jin Du and Da-Wen Sun, *Learning techniques used in computer vision for food quality evaluation: a review*, Journal of Food Engineering **72** (2006), no. 1, 39–55.
- [60] R.O. Duda, P.E. Hart, and D.G. Stock, *Pattern classification*, John Wiley and Sons, Chichester, England, 2012.
- [61] N. Dupuy, O. Galtier, D. Ollivier, P. Vanloot, and J. Artaud, *Comparison between NIR, MIR, concatenated NIR and MIR analysis and hierarchical PLS model. Application to virgin olive oil analysis*, Analytica Chimica Acta **666** (2010), no. 1-2, 23–31.
- [62] Nathalie Dupuy, Oswin Galtier, Yveline Le Dréau, Christian Pinatel, Jacky Kister, and Jacques Artaud, *Chemometric analysis of combined NIR and MIR spectra to characterize French olives*, European Journal of Lipid Science and Technology **112** (2010), no. 4, 463–475.
- [63] A.E. Eiben and James E Smith, *Genetic Algorithms*, Introduction to Evolutionary Computing, Springer-Verlag, Berlin Heidelberg, 1 ed., 2003, pp. 37–69.
- [64] Rasha M. El-Abassy, Patrice Donfack, and Arnulf Materny, *Rapid Determination of Free Fatty Acid in Extra Virgin Olive Oil by Raman Spectroscopy and Multivariate Analysis*, Journal of the American Oil Chemists' Society **86** (2009), no. 6, 507–511.

- [65] Gamal ElMasry and Da-Wen Sun, *Principles of Hyperspectral Imaging Technology*, Hyperspectral Imaging for Food Quality Analysis and Control (Da-Wen Sun, ed.), Elsevier, London, UK, 1 ed., 2010, pp. 3–43.
- [66] European Union, *Commission Implementing Regulation (EU) No. 1348/2013 amending Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis*, Off. J. Eur. Union **31** (2013).
- [67] J. Ferreira, *Explotaciones olivareras colaboradoras*, Ministerio de Agricultura, Madrid, Spain, 1979.
- [68] RA Fisher, *The use of multiple measurements in taxonomic problems*, ANNALS OF EUGENICS **7** (1936), no. 2, 179–188.
- [69] Food and Drug Administration, *Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*, 2004.
- [70] Furferi, Monica Carfagni, and Marco Daou, *Real-time estimation of peroxides and acidity level of extra-virgin olive oil : an integrated approach*, International Journal of Mathematics and Computers in Simulation **2** (2008), 206–214.
- [71] R. Furferi and M. Carfagni, *A machine vision system for real-time and automatic assessment of olives colour and surface defects*, Int. J. Comput. Res. (2010).
- [72] Rocco Furferi, Monica Carfagni, and Marco Daou, *Artificial neural network software for real-time estimation of olive oil qualitative parameters during continuous extraction*, Computers and Electronics in Agriculture **55** (2007), no. 2, 115–131.
- [73] Rocco Furferi, Lapo Governi, and Yary Volpe, *ANN-based method for olive Ripening Index automatic prediction*, Journal of Food Engineering **101** (2010), no. 3, 318–328.
- [74] O Galtier, N Dupuy, Y Le Dréau, D Ollivier, C Pinatel, J Kister, and J Artaud, *Geographic origins and compositions of virgin olive oils determinated by chemometric analysis of NIR spectra*, Analytica Chimica Acta **595** (2007), no. 1-2, 136–144.
- [75] José M. García, Francisca Gutiérrez, María J. Barrera, and Miguel A. Albi, *Storage of Mill Olives on an Industrial Scale*, Journal of Agricultural and Food

- Chemistry **44** (1996), no. 2, 590–593.
- [76] Juan Francisco García Martín, *Optical path length and wavelength selection using Vis/NIR spectroscopy for olive oil's free acidity determination*, International Journal of Food Science & Technology **50** (2015), no. 6, 1461–1467.
- [77] A García Sánchez, N Martos Ramos, and E Ballesteros, *Estudio comparativo de distintas técnicas analíticas (espectroscopía NIR y RMN y extracción mediante Soxhlet) para la determinación del contenido graso y de humedad de aceitunas y orujo de Jaén*, Grasas y Aceites **56** (2005), no. 3, 220–227.
- [78] Gabriel Gatica, Stanley Best, José Ceroni, and Gastón Lefranc, *Olive Fruits Recognition Using Neural Networks*, Procedia Computer Science **17** (2013), 412–419.
- [79] P. Geladi, D. MacDougall, and H. Martens, *Correction for Near-Infrared Reflectance Spectra of Meat*, Applied Spectroscopy **39** (1985), no. 3, 491–500 (EN).
- [80] D. M. Martínez Gila, J. Beltran Ortega, J. Gamez Garcia, and J. Gomez Ortega, *Rapid quantification of total polyphenol content in EVOO using NIR sensor with wavelength selection and FS-MLR*, 2015 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, sep 2015, pp. 1–6.
- [81] D Martínez Gila, P Cano Marchal, J Gámez García, and J Gómez Ortega, *Hyper-spectral imaging for determination of some quality parameters for olive oil*, the 18 International Conference on Automation & Computing, sep 2012, pp. 1–6.
- [82] D. Martinez Gila, P. Cano Marchal, J. Gamez Garcia, and J. Gomez Ortega, *Control System of the Malaxing State for the Olive Paste Based on Computer Vision and Fuzzy Logic*, 2013 IEEE International Conference on Systems, Man, and Cybernetics, IEEE, oct 2013, pp. 1951–1956.
- [83] Daniele Giuffrida, Francesco Salvo, Andrea Salvo, Lina Cossignani, and Giacomo Dugo, *Pigments profile in monovarietal virgin olive oils from various Italian olive varieties*, Food Chemistry **124** (2011), no. 3, 1119–1123.
- [84] Aurora Gómez-Rico, Antonio M Inarejos-García, M Desamparados Salvador, and Giuseppe Fregapane, *Effect of malaxation conditions on phenol and volatile profiles in olive paste and the corresponding virgin olive oils (Olea europaea L. Cv. Cornicabra).*, Journal of agricultural and food chemistry **57** (2009), no. 9, 3587–95.

- [85] Rafael Gonzalez, Richard E. Woods, and Steven L. Eddins, *Introduction*, Digital Image Processing Using MATLAB, Gatesmark Publishing, 2 ed., 2009, pp. 1–12.
- [86] Rafael C. Gonzalez, Richard Woods, and Steven L. Eddins, *Object Recognition*, Digital Image Processing Using MATLAB, Gatesmark Publishing, 2 ed., 2009, pp. 674–706.
- [87] Rafael C. Gonzalez, Richard E. Woods, and Steven Eddins, *Representation and description*, Digital Image Processing Using MATLAB, Gatesmark Publishing, 2 ed., 2009, pp. 597–672.
- [88] Belen Gordillo, Leonardo Ciaccheri, Anna G. Mignani, M. Lourdes Gonzalez-Miret, and Francisco J. Heredia, *Influence of Turbidity Grade on Color and Appearance of Virgin Olive Oil*, Journal of the American Oil Chemists' Society **88** (2011), no. 9, 1317–1327.
- [89] a. Gracia and L. León, *Non-destructive assessment of olive fruit ripening by portable near infrared spectroscopy*, Grasas y Aceites **62** (2011), no. 3, 268–274.
- [90] E. Guzman, *Automatic and on-line method for detection of dirt on olive oil production process*, Servicio de Publicaciones de la Universidad de Córdoba, Córdoba, 2013.
- [91] Elena Guzmán, Vincent Baeten, Juan Antonio Fernández, and José a. García-Mesa, *Infrared machine vision system for the automatic detection of olive fruit quality*, Talanta **116** (2013), 894–898.
- [92] Elena Guzmán, Vincent Baeten, Juan Antonio Fernández Pierna, and José a. García-Mesa, *A portable Raman sensor for the rapid discrimination of olives according to fruit quality*, Talanta **93** (2012), 94–98.
- [93] Elena Guzmán, Vincent Baeten, Juan Antonio Fernández Pierna, and José A. García-Mesa, *Using a Visible Vision System for On-Line Determination of Quality Parameters of Olive Fruits*, Food and Nutrition Sciences **04** (2013), no. 07, 90–98.
- [94] Elena Guzmán, Vincent Baeten, Juan Antonio Fernández Pierna, and José A. García-Mesa, *Determination of the olive maturity index of intact fruits using image analysis*, Journal of Food Science and Technology **52** (2015), no. 3, 1462–1470.
- [95] Rim Hachicha Hbaieb, Faten Kotti, Rosa García-Rodríguez, Mohamed Gargouri, Carlos Sanz, and Ana G Pérez, *Monitoring endogenous enzymes during olive*

- fruit ripening and storage: correlation with virgin olive oil phenolic profiles.*, Food chemistry **174** (2015), 240–7.
- [96] M.T. Hagan, H.B. Demuth, and M.H. Beale, *Neuronal Network design*, Boston, MA, USA, 1996.
- [97] R M Haralick and L G Shapiro, *Computer and robot vision*, Computer and Robot Vision, no. 1, Addison-Wesley Pub. Co., 1992.
- [98] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein, *Textural Features for Image Classification*, IEEE Transactions on Systems, Man, and Cybernetics **3** (1973), no. 6, 610–621.
- [99] Mark Howard, *Data Analysis: Multilinear Regression and Principal Component Analysis*, Handbook of Near-Infrared Analysis, 2008, pp. 151–188.
- [100] Min Huang, Qibing Zhu, Bojin Wang, and Renfu Lu, *Analysis of hyperspectral scattering images using locally linear embedding algorithm for apple mealiness classification*, Computers and Electronics in Agriculture **89** (2012), 175–181.
- [101] Inarejos-García, A.M., S. Gómez-Alonso, G. Fregapane, and M.D. Salvador, *Evaluation of minor components, sensory characteristics and quality of virgin olive oil by near infrared (NIR) spectroscopy*, Food Research International **50** (2013), no. 1, 250–258.
- [102] Antonio M. Inarejos-García, G. Fregapane, and M. Desamparados Salvador, *Effect of crushing on olive paste and virgin olive oil minor components*, European Food Research and Technology **232** (2010), no. 3, 441–451.
- [103] Antonio M Inarejos-García, Aurora Gómez-Rico, M Desamparados Salvador, and Giuseppe Fregapane, *Effect of preprocessing olive storage conditions on virgin olive oil quality and composition.*, Journal of agricultural and food chemistry **58** (2010), no. 8, 4858–65.
- [104] Antonio M. Inarejos-García, Aurora Gómez-Rico, M. Desamparados Salvador, and Giuseppe Fregapane, *Influence of malaxation conditions on virgin olive oil yield, overall quality and composition*, European Food Research and Technology **228** (2008), no. 4, 671–677.
- [105] International Olive Council, *103nd session of the IOOC*, Tech. report, Madrid, Spain, 2015.
- [106] Bahareh Jamshidi, Saeid Minaei, Ezzedin Mohajerani, and Hassan Ghassemian,

- Reflectance Vis/NIR spectroscopy for nondestructive taste characterization of Valencia oranges*, Computers and Electronics in Agriculture **85** (2012), 64–69.
- [107] Tina Jerman Klen and Branka Mozetič Vodopivec, *The fate of olive fruit phenols during commercial olive oil processing: Traditional press versus continuous two- and three-phase centrifuge*, LWT - Food Science and Technology **49** (2012), no. 2, 267–274.
- [108] A. Jiménez, G. Beltrán, M.P. Aguilera, and M. Uceda, *A sensor-software based on artificial neural network for the optimization of olive oil elaboration process*, Sensors and Actuators B: Chemical **129** (2008), no. 2, 985–990.
- [109] A. Jiménez, E. Izquierdo, F. Rodríguez, J. I. Dueñas, and C. Tortosa, *Determination of fat and moisture in olives by nearinfrared reflectance spectroscopy*, Grasas y Aceites **51** (2000), no. 5, 311–315.
- [110] B. Jiménez Herrera and A. Crapio Dueñas, *La cata de aceites: Aceite de oliva virgen. Características organolépticas y análisis sensorial*, Tech. report, IFAPA, Junta de Andalucía, Seville, Spain, 2008.
- [111] a. Jiménez Márquez, *Monitoring carotenoid and chlorophyll pigments in virgin olive oil by visible-near infrared transmittance spectroscopy. On-line application*, Journal of Near Infrared Spectroscopy **11** (2003), no. 1, 219.
- [112] I. Jolliffe, *Principal Component Analysis*, John Willey and Sons, Chichester, England, 2005.
- [113] Pinar Kadiroğlu and Figen Korel, *Chemometric Studies on zNose™ and Machine Vision Technologies for Discrimination of Commercial Extra Virgin Olive Oils*, Journal of the American Oil Chemists' Society **92** (2015), no. 9, 1235–1242.
- [114] Nick Kalogeropoulos, Andriana C. Kaliora, Anna Artemiou, and Ioannis Giogios, *Composition, volatile profiles and functional properties of virgin olive oils produced by two-phase vs three-phase centrifugal decanters*, LWT - Food Science and Technology **58** (2014), no. 1, 272–279.
- [115] C.M. Kalua, M.S. Allen, D.R. Bedgood, A.G. Bishop, P.D. Prenzler, and K. Roberts, *Olive oil volatile compounds, flavour development and quality: A critical review*, Food Chemistry **100** (2007), no. 1, 273–286.
- [116] Curtis M Kalua, Danny R Bedgood, Andrea G Bishop, and Paul D Prenzler, *Changes in virgin olive oil quality during low-temperature fruit storage.*, Journal

- of agricultural and food chemistry **56** (2008), no. 7, 2415–22.
- [117] Sumaporn Kasemsumran, Naroo Kang, Alfred Christy, and Yukihiro Ozaki, *Partial Least Squares Processing of Near Infrared Spectra for Discrimination and Quantification of Adulterated Olive Oils*, *Spectroscopy Letters* **38** (2005), no. 6, 839–851 (en).
- [118] Ismail Kavdir, M. Burak Buyukcan, Renfu Lu, Habib Kocabiyik, and Murat Seker, *Prediction of olive quality using FT-NIR spectroscopy in reflectance and transmittance modes*, *Biosystems Engineering* **103** (2009), no. 3, 304–312.
- [119] S. Kawano, *Sampling and sample presentation*, *Near-Infrared Spectroscopy: Principles, Instruments and Applications* (H.W. Siesler, Y. Ozaki, S. Kawata, and H.M. Heise, eds.), Wiley-VCH, Weinheim, Germany, 2002, pp. 115–124.
- [120] W.J. Krzanowski, *Principles of Multivariate Analysis*, Oxford University Press, New York, USA, 2000.
- [121] Pedro Latorre Carmona, José Martínez Sotoca, and Filiberto Pla, *Filter-Type Variable Selection Based on Information Measures for Regression Tasks*, *Entropy* **14** (2012), no. 12, 323–343.
- [122] L León, L Rallo, and A Garrido, *Análisis de aceituna intacta mediante espectroscopia en el infrarrojo cercano (NIRS): una herramienta de utilidad en programas de mejora de olivo*, *Grasas Y Aceites* **54** (2003), 41–47.
- [123] Lorenzo León, Ana Garrido-Varo, and Gerard Downey, *Parent and Harvest Year Effects on Near-Infrared Reflectance Spectroscopic Analysis of Olive (Olea europaea L.) Fruit Traits*, *Journal of Agricultural and Food Chemistry* **52** (2004), no. 16, 4957–4962.
- [124] L. León-Moreno, *Usefulness of portable near infrared spectroscopy in olive breeding programs*, *Spanish Journal of Agricultural Research* **10** (2012), no. 1, 141.
- [125] Alessandro Leone, Roberto Romaniello, Riccardo Zagaria, Erika Sabella, Luigi De Bellis, and Antonia Tamborrino, *Machining effects of different mechanical crushers on pit particle size and oil drop distribution in olive paste*, *European Journal of Lipid Science and Technology* **117** (2015), no. 8, 1271–1279.
- [126] Ping Lin, Yongming Chen, and Yong He, *Identification of Geographical Origin of Olive Oil Using Visible and Near-Infrared Spectroscopy Technique Combined*

- with *Chemometrics*, Food and Bioprocess Technology **5** (2012), no. 1, 235–242.
- [127] J López-Miranda, F Pérez-Jiménez, E Ros, R De Caterina, L Badimón, M I Covas, E Escrich, J M Ordovás, F Soriguer, R Abiá, C Alarcón de la Lastra, M Battino, D Corella, J Chamorro-Quirós, J Delgado-Lista, D Giugliano, K Esposito, R Estruch, J M Fernandez-Real, J J Gaforio, C La Vecchia, D Lairon, F López-Segura, P Mata, J A Menéndez, F J Muriana, J Osada, D B Panagiotakos, J A Paniagua, P Pérez-Martínez, J Perona, M A Peinado, M Pineda-Priego, H E Poulsen, J L Quiles, M C Ramírez-Tortosa, J Ruano, L Serra-Majem, R Solá, M Solanas, V Solfrizzi, R de la Torre-Fornell, A Trichopoulou, M Uceda, J M Villalba-Montoro, J R Villar-Ortiz, F Visioli, and N Yiannakouris, *Olive oil and health: summary of the II international conference on olive oil and health consensus report, Jaén and Córdoba (Spain) 2008.*, Nutrition, metabolism, and cardiovascular diseases : NMCD **20** (2010), no. 4, 284–94.
- [128] Ji Ma, Da-Wen Sun, Jia-Huan Qu, Dan Liu, Hongbin Pu, Wen-Hong Gao, and Xin-An Zeng, *Applications of Computer Vision for Assessing Quality of Agri-food Products: A Review of Recent Research Advances.*, Critical reviews in food science and nutrition **56** (2016), no. 1, 113–27.
- [129] Rodney J. Mailer, *Rapid evaluation of olive oil quality by NIR reflectance spectroscopy*, Journal of the American Oil Chemists' Society **81** (2004), no. 9, 823–827.
- [130] E.H. Mamdani and S. Assilian, *An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller*, International Journal of Human-Computer Studies **51** (1999), no. 2, 135–147.
- [131] Marena Manley and Kerstin Eberle, *Comparison of Fourier transform near infrared spectroscopy partial least square regression models for South African extra virgin olive oil using spectra collected on two spectrophotometers at different resolutions and path lengths*, Journal of Near Infrared Spectroscopy **14** (2006), no. 1, 111.
- [132] Ángela Manzano Coca, Rafael Cárdenas García, and Javier Marcos Braojos, *Almacenamiento y conservación*, El Manual del Maestro de Almazara (GEA Westfalia Separator Ibérica, ed.), Jaén, España, 2 ed., 2013, pp. 146–156.
- [133] P. Cano Marchal, D. M. Martínez Gila, J. Gamez García, and J. Gómez Ortega, *Expert system based on computer vision to estimate the content of impurities in olive oil samples*, Journal of Food Engineering **119** (2013), no. 2, 220–228.

- [134] P Cano Marchal, D Martínez Gila, J Gámez García, and J Gómez Ortega, *Application of computer vision and Support Vector Machines to estimate the content of impurities in olive oil samples*, the 18 International Conference on Automation & Computing, sep 2012, pp. 1–6.
- [135] a. Jiménez Marquez, a. Molina Díaz, and M.I. Pascual Reguera, *Using optical NIR sensor for on-line virgin olive oils characterization*, Sensors and Actuators B: Chemical **107** (2005), no. 1, 64–68.
- [136] D. Martinez Gila, D. Aguilera Puerto, J. Gamez Garcia, and J. Gomez Ortega, *Automatic classification of olives for oil production using computer vision*, 2015 IEEE International Conference on Industrial Technology (ICIT), IEEE, mar 2015, pp. 1651–1656.
- [137] D. Martinez Gila, P. Cano Marchal, J. Gamez Garcia, and J. Gomez Ortega, *Extractability control for the virgin olive oil elaboration process*, 2015 IEEE International Conference on Industrial Technology (ICIT), IEEE, mar 2015, pp. 240–245.
- [138] D. Martínez Gila, P. Cano Marchal, J. Gámez García, and J. Gómez Ortega, *On-line system based on hyperspectral information to estimate acidity, moisture and peroxides in olive oil samples*, Computers and Electronics in Agriculture **116** (2015), 1–7.
- [139] Diego M. Martínez Gila, Julio Beltrán Ortega, Javier Gámez García, and Juan Gómez Ortega, *Tecnología no invasiva para la determinación automática del contenido total de polifenoles en el aceite de oliva virgen*, XXXVI Jornadas de Automática, 2015, pp. 1–8.
- [140] Diego Manuel Martínez Gila, Pablo Cano Marchal, Javier Gámez García, and Juan Gómez Ortega, *Non-invasive methodology to estimate polyphenol content in EVOO based on Stepwise Multilinear Regression*, Computers and Electronics in Agriculture (Under Review) (2016).
- [141] D.M. Martínez Gila, P. Cano Marchal, J. Gámez García, and J. Gómez Ortega, *Determinación del estado de batido de la pasta de aceituna empleando visión por computador*, XXXIV Jornadas de Automática, 2013, pp. 826–833.
- [142] Piernicola Masella, Alessandro Parenti, Paolo Spugnoli, and Luca Calamai, *Influence of Vertical Centrifugation on Extra Virgin Olive Oil Quality*, Journal of the American Oil Chemists' Society **86** (2009), no. 11, 1137–1140.

- [143] William J. McCarthy and Gabor J. Kemeny, *Fourier Transform Spectrophotometers in the Near-Infrared*, Handbook of Near-Infrared Analysis (D.A. Burns and E.W. Ciurczac, eds.), CRC Press, Taylor & Francis Group, Boca Raton, FL, EE.UU, 3 ed., 2007, pp. 79–91.
- [144] W. McClure, *Review: 204 years of near infrared technology: 1800-2003*, Journal of Near Infrared Spectroscopy **11** (2003), no. 6, 487–518.
- [145] Karla Danielle Tavares Melo Milanez and Márcio José Coelho Pontes, *Classification of edible vegetable oil using digital image and pattern recognition techniques*, Microchemical Journal **113** (2014), 10–16.
- [146] C.E. Miller, *Chemical principles of near infrared spectroscopy*, Near-Infrared Technology in the Agricultural and Food Industries (P.K. Williams and K. Norris, eds.), American Association of Cereal Chemists, Inc. St. Paul, Minnesota, EE.UU, 2 ed., 2001, pp. 19–37.
- [147] José Luis Molina de la Rosa, Brígida Jiménez Herrera, Félix Ruiz Coletto, Francisco García Zamorano, Juan Cano Rodríguez, and Julián Pérez García, *Influencia de los factores agronómicos en la calidad del aceite de oliva*, Técnicas de cultivo: Recolección mecanizada del olivar, Instituto Andaluz de Investigación y Formación Agraria, Pesquera, Alimentaria y de la Producción Ecológica (I.F.A.P.A.), Córdoba, España, 2 ed., 2004, pp. 15–17.
- [148] S. T. Monteiro and Y. Kosugi, *Particle Swarms for Feature Extraction of Hyperspectral Data*, IEICE Transactions on Information and Systems **E90-D** (2007), no. 7, 1038–1046.
- [149] O. Caceres Moreno, D. M. Martinez Gila, D Aguilera Puerto, J. Gamez Garcia, and J. Gomez Ortega, *Automatic determination of peroxides and acidity of olive oil using machine vision in olive fruits before milling process*, 2015 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, sep 2015, pp. 1–6.
- [150] Barbara Muik, Bernhard Lendl, Antonio Molina-Díaz, and María José Ayora-Cañada, *Direct, reagent-free determination of free fatty acid content in olive oil and olives by Fourier transform Raman spectrometry*, Analytica Chimica Acta **487** (2003), no. 2, 211–220.
- [151] Barbara Muik, Bernhard Lendl, Antonio Molina-Díaz, Luis Pérez-Villarejo, and María José Ayora-Cañada, *Determination of oil and water content in olive pomace using near infrared and Raman spectrometry. A comparative study*,

- Analytical and Bioanalytical Chemistry **379** (2004), no. 1, 35–41.
- [152] I. Murray and D. Williams, *Chemical principals of Near-Infrared Technology*, Near Infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists, St. Paul, Minnesota. USA., 1987.
- [153] Paolo Oliveri, Monica Casale, M Chiara Casolino, M Antonietta Baldo, Fiammetta Nizzi Grifi, and Michele Forina, *Comparison between classical and innovative class-modelling techniques for the characterisation of a PDO olive oil*, Analytical and Bioanalytical Chemistry **399** (2011), no. 6, 2105–2113.
- [154] Julio Beltrán Ortega, Diego M. Martínez Gila, Daniel Aguilera Puerto, Javier Gámez García, and Juan Gómez Ortega, *Novel Technologies for Monitoring the In-Line Quality of Virgin Olive Oil during Manufacturing and Storage*, Journal of the Science of Food and Agriculture (2016), n/a–n/a.
- [155] Betül Öztürk, Ayşegül Yalçın, and Durmuş Özdemir, *Determination of olive oil adulteration with vegetable oils by near infrared spectroscopy coupled with multivariate calibration*, Journal of Near Infrared Spectroscopy **18** (2010), no. 1, 191.
- [156] Christopher Palmer, *The physics of diffraction gratings*, Diffraction grating handbook, Newport Corporation, Rochester, New York, 6 ed., 2005, pp. 20–43.
- [157] Henryk Palus, *Representations of colour images in different colour spaces*, The Colour Image Processing Handbook (Stephen J. Sangwine and Robin E.N. Horne, eds.), Springer US, New York, 1998, pp. 67–90.
- [158] N. Perrot, I. Ioannou, I. Allais, C. Curt, J. Hossenlopp, and G. Trystram, *Fuzzy concepts applied to food product quality control: A review*, Fuzzy Sets and Systems **157** (2006), no. 9, 1145–1154.
- [159] D Aguilera Puerto, O. Caceres Moreno, D. M. Martinez Gila, J. Gamez Garcia, and J. Gomez Ortega, *Olive batches automatic classification in mill reception using computer vision*, 2015 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, sep 2015, pp. 1–6.
- [160] Daniel Puerto, Diego Gila, Javier García, and Juan Ortega, *Sorting Olive Batches for the Milling Process Using Image Processing*, Sensors **15** (2015), no. 7, 15738–15754 (en).
- [161] Jianwei Qin, *Hyperspectral Imaging Instruments*, Hyperspectral Imaging for

- Food Quality Analysis and Control (Da-Wen Sun, ed.), Elsevier, London, UK, 1 ed., 2010, pp. 129–172.
- [162] Tomer Ram, Zeev Wiesman, Israel Parment, and Yael Edan, *Olive oil content prediction models based on image processing*, Biosystems Engineering **105** (2010), no. 2, 221–232.
- [163] Alfonso Ranalli, Luciano Pollastri, Stefania Contento, Emilia Iannucci, and Lucia Lucera, *Effect of olive paste kneading process time on the overall quality of virgin olive oil*, European Journal of Lipid Science and Technology **105** (2003), no. 2, 57–67.
- [164] M.T. Riquelme, M.R. Altisent, C.V. Ubierna, and P.B. Elorza, *Clasificación de defectos externos en aceitunas mediante análisis de imágenes*, Horticultura **199** (2007), 72–73.
- [165] M.T. Riquelme, P. Barreiro, M. Ruiz-Altisent, and C. Valero, *Olive classification according to external damage using image analysis*, Journal of Food Engineering **87** (2008), no. 3, 371–379.
- [166] L.G. Roberts, *Opt. Electro-Opt. Inf. Process*, Machine perception of 3-D solids, MIT Press, Cambridge, MA, USA, 1965, pp. 159–197.
- [167] Concepción Romero and Manuel Brenes, *Analysis of total contents of hydroxytyrosol and tyrosol in olive oils.*, Journal of agricultural and food chemistry **60** (2012), no. 36, 9017–22.
- [168] M. Ruiz-Altisent, L. Ruiz-García, G.P. Moreda, Renfu Lu, N. Hernández-Sánchez, E.C. Correa, B. Diezma, B. Nicolăi, and J. García-Ramos, *Sensors for product characterization and quality of specialty crops—A review*, Computers and Electronics in Agriculture **74** (2010), no. 2, 176–194.
- [169] Lourdes Salguero-Chaparro, Vincent Baeten, Ouissam Abbas, and Francisco Peña-Rodríguez, *On-line analysis of intact olive fruits by vis-NIR spectroscopy: Optimisation of the acquisition parameters*, Journal of Food Engineering **112** (2012), no. 3, 152–157.
- [170] Lourdes Salguero-Chaparro, Vincent Baeten, Juan A. Fernández-Pierna, and Francisco Peña-Rodríguez, *Near infrared spectroscopy (NIRS) for on-line determination of quality parameters in intact olives*, Food Chemistry **139** (2013), no. 1-4, 1121–1126.

- [171] Lourdes Salguero-Chaparro and Francisco Peña-Rodríguez, *On-line versus off-line NIRS analysis of intact olives*, LWT - Food Science and Technology **56** (2014), no. 2, 363–369.
- [172] M.D. Salvador, F. Aranda, and G. Fregapane, *Influence of fruit ripening on 'Cornicabra' virgin olive oil quality A study of four successive crop seasons*, Food Chemistry **73** (2001), no. 1, 45–53.
- [173] Araceli Sánchez-Ortiz, Carmen Romero, Ana G Pérez, and Carlos Sanz, *Oxygen concentration affects volatile compound biosynthesis during virgin olive oil production.*, Journal of agricultural and food chemistry **56** (2008), no. 12, 4681–5.
- [174] Abraham. Savitzky and M. J. E. Golay, *Smoothing and Differentiation of Data by Simplified Least Squares Procedures.*, Analytical Chemistry **36** (1964), no. 8, 1627–1639.
- [175] Shimon Y. Nof, *Springer Handbook of Automation*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [176] H. Siesler, *Basic Principles of Near-Infrared Spectroscopy*, Handbook of Near-Infrared Analysis (D.A. Burns and E.W. Ciurczac, eds.), CRC Press, Taylor & Francis Group, Boca Raton, FL, EE.UU, 3 ed., 2007, pp. 7–19.
- [177] H.W. Siesler, *Introduction*, Near-Infrared Spectroscopy: Principles, Instruments and Applications (H.W. Siesler, Y. Ozaki, S. Kawata, and H.M. Heise, eds.), Wiley-VCH, Weinheim, Germany, 2002, pp. 1–10.
- [178] Nicoletta Sinelli, Monica Casale, Valentina Di Egidio, Paolo Oliveri, Daniele Bassi, Debora Tura, and Ernestina Casiraghi, *Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy*, Food Research International **43** (2010), no. 8, 2126–2131.
- [179] Nicoletta Sinelli, Ernestina Casiraghi, Debora Tura, and Gerard Downey, *Characterisation and classification of Italian virgin olive oils by near- and mid-infrared spectroscopy*, Journal of Near Infrared Spectroscopy **16** (2008), no. 1, 335.
- [180] Nicoletta Sinelli, Lorenzo Cerretani, Valentina Di Egidio, Alessandra Bendini, and Ernestina Casiraghi, *Application of near (NIR) infrared and mid (MIR) infrared spectroscopy as a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity*, Food Research International **43** (2010), no. 1,

369–375.

- [181] Erik Skibsted and Søren Balling Engelsen, *Encyclopedia of Spectroscopy and Spectrometry*, Elsevier, 2010.
- [182] Magdalena Śliwińska, Paulina Wiśniewska, Tomasz Dymerski, Jacek Namieśnik, and Waldemar Wardencki, *Food analysis using artificial senses.*, *Journal of agricultural and food chemistry* **62** (2014), no. 7, 1423–48.
- [183] M Stone, *Cross-Validatory Choice and Assessment of Statistical Predictions*, *Journal of the Royal Statistical Society. Series B (Methodological)* **36** (1974), no. 2, 111–147.
- [184] Barbara Stuart, *Introduction*, *Infrared Spectroscopy: Fundamentals and Applications*, John Wiley & Sons, 1 ed., 2004, pp. 1–13.
- [185] Barbara H. Stuart, *Experimental Methods*, *Infrared Spectroscopy: Fundamentals and Applications*, John Wiley & Sons, 2004, pp. 15–44.
- [186] Warawut Suphamitmongkol, Guangli Nie, Rong Liu, Sumaporn Kasemsumran, and Yong Shi, *An alternative approach for the classification of orange varieties based on near infrared spectroscopy*, *Computers and Electronics in Agriculture* **91** (2013), 87–93.
- [187] M. Tajammal Munir, W. Yu, B.R. Young, and David I. Wilson, *The current status of process analytical technologies in the dairy industry*, *Trends in Food Science & Technology* **43** (2015), no. 2, 205–218.
- [188] Karla Danielle Tavares de Melo Milanez and Márcio José Coelho Pontes, *Classification of extra virgin olive oil and verification of adulteration using digital images and discriminant analysis*, *Anal. Methods* **7** (2015), no. 20, 8839–8846.
- [189] Marino Uceda, *Trends in olive oil production Olive oil extraction and quality*, *Grasas y Aceites* **57** (2006), no. 1, 25–31.
- [190] Enrico Valli, Alessandra Bendini, Rubén M. Maggio, Lorenzo Cerretani, Tullia Gallina Toschi, Ernestina Casiraghi, and Giovanni Lercker, *Detection of low-quality extra virgin olive oils by fatty acid alkyl esters evaluation: a preliminary and fast mid-infrared spectroscopy discrimination by a chemometric approach*, *International Journal of Food Science & Technology* **48** (2013), no. 3, 548–555.
- [191] Frans van den Berg, Christian B. Lyndgaard, Klavs M. Sørensen, and Søren B.

- Engelsen, *Process Analytical Technology in the food industry*, Trends in Food Science & Technology **31** (2013), no. 1, 27–35.
- [192] Pierre Vanloot, Dominique Bertrand, Christian Pinatel, Jacques Artaud, and Nathalie Dupuy, *Artificial vision and chemometrics analyses of olive stones for varietal identification of five French cultivars*, Computers and Electronics in Agriculture **102** (2014), 98–105.
- [193] Sébastien Veillet, Valérie Tomao, Isabelle Bornard, Karine Ruiz, and Farid Chemat, *Chemical changes in virgin olive oils as a function of crushing systems: Stone mill and hammer crusher*, Comptes Rendus Chimie **12** (2009), no. 8, 895–904.
- [194] Stefania Vichi, Pinar Boynuegri, Josep Caixach, and Agustí Romero, *Quality losses in virgin olive oil due to washing and short-term storage before olive milling*, European Journal of Lipid Science and Technology **117** (2015), no. 12, 2015–2022.
- [195] Stefania Vichi, Agustí Romero, Joan Gallardo-Chacón, Joan Tous, Elvira López-Tamames, and Susana Buxaderas, *Volatile phenols in virgin olive oils: Influence of olive variety on their formation during fruits storage*, Food Chemistry **116** (2009), no. 3, 651–656.
- [196] Juan Vilar Hernández, María Del Mar Velasco Gámez, and Raquel Puentes Poyatos, *Incidencia del modo de explotación del olivo sobre la renta neta del olivicultor. Estrategias para el cultivo extensivo en el contexto de la posible ausencia de subvenciones*, Grasas y Aceites **61** (2010), no. 4, 430–440.
- [197] P. C. Williams, *Implementation of Near-Infrared Technology*, Near-Infrared Technology in the Agricultural and Food Industry (Phil Williams and Karl Norris, eds.), AACCI Press, St. Paul, Minnesota, USA., 2 ed., 2001, p. 223.
- [198] Tony Woodcock, Gerard Downey, and Colm P. O'Donnell, *Confirmation of Declared Provenance of European Extra Virgin Olive Oil Samples by NIR Spectroscopy*, Journal of Agricultural and Food Chemistry **56** (2008), no. 23, 11520–11525.
- [199] Workman and Weyer, *Water*, Practical guide and spectral atlas for interpretive near-infrared spectroscopy, CRC Press, Taylor & Francis Group, Boca Raton, FL, EE.UU, 2 ed., 2012, pp. 56–61.
- [200] Jerome J. Workman, *The NIR Method: Review of the Concepts*, Handbook of

- Near-Infrared Analysis (D.A. Burns and E.W. Ciurczac, eds.), CRC Press, Taylor & Francis Group, Boca Raton, FL, EE.UU, 3 ed., 2007, pp. 124–150.
- [201] Jerome J. Workman and Donald A. Burns, *Commercial NIR Instrumentation*, Handbook of Near-Infrared Analysis (D.A. Burns and E.W. Ciurczac, eds.), CRC Press, Taylor & Francis Group, Boca Raton, FL, EE.UU, 3 ed., 2007, pp. 67–78.
- [202] Jerry Workman and Lois Weyer, *Introduction to Near-Infrared Spectra*, Practical guide and spectral atlas for interpretive near-infrared spectroscopy, CRC Press, Taylor & Francis Group, Boca Raton, FL, EE.UU, 2 ed., 2012, pp. 1–17.
- [203] Jerry Workman Jr., *NIR Spectroscopy Calibration Basics*, Handbook of Near-Infrared Analysis, 2008, pp. 123–149.
- [204] Hong Yang and Joseph Irudayaraj, *Comparison of near-infrared, fourier transform-infrared, and fourier transform-raman methods for determining olive pomace oil adulteration in extra virgin olive oil*, Journal of the American Oil Chemists' Society **78** (2001), no. 9, 889–895.
- [205] Haibo Yao and David Lewis, *Spectral Preprocessing and Calibration Techniques*, Hyperspectral Imaging for Food Quality Analysis and Control (Da-Wen Sun, ed.), London, UK, 1 ed., 2010, pp. 45–78.
- [206] Yudong Zhang, Zhengchao Dong, Lenan Wu, and Shuihua Wang, *A hybrid method for MRI brain image classification*, Expert Systems with Applications **38** (2011), no. 8, 10049–10053.
- [207] Yudong Zhang, Shuihua Wang, Preetha Phillips, and Genlin Ji, *Binary PSO with mutation operator for feature selection using decision tree applied to spam detection*, Knowledge-Based Systems **64** (2014), 22–31.
- [208] Chaoxin Zheng, Da-Wen Sun, and Liyun Zheng, *Recent applications of image texture for evaluation of food qualities—a review*, Trends in Food Science & Technology **17** (2006), no. 3, 113–128.
- [209] Hui Zou and Trevor Hastie, *Regularization and variable selection via the elastic net*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67** (2005), no. 2, 301–320.

Glossary

ABS Absorbance.

ABTS Antioxidant Activity.

ANN Artificial Neural Network.

ANOVA Analysis of Variance.

AOTF Acousto-Optic Tunable Filters.

AV Acidity Value.

C Coadjuvant.

CCD Charged Coupled Device.

CMOS Complementary Metal Oxide Semiconductor Array.

CV Coefficient of Variation.

EU European Union.

EVOO Extra Virgin Olive Oil.

FA Free Acidity.

FDA Fisher discriminant analysis.

FO Floating Oil.

FT-IR Fourier Transformations Infrared.

FT-NIR Fourier Transform Near Infrared.

G Granularity.

GA Genetic Algorithm.

GA-MLR Genetic Algorithm with Multiple Linear Regression.

GLCM Gray Level Co-occurrence Matriz.

GRAV Robotics, Automation and Computer Vision Research Group.

HCA Hierarchical Cluster Analysis.

HSV Hue Saturation Value.

InGaAs Indium Gallium Arsenide.

IR Infrared.

k-NN k-Nearest Neighbours.

KLDA Kernel Linear Discriminant Analysis.

KPCA Kernel Principal Component Analysis.

LASSO Least Absolute Shrinkage and Selection Operator.

LCTF Liquid Crystal Tunable Filters.

LDA Linear Discriminant Analysis.

LED Light Emitting Diodes.

m.e.q. milliequivalents.

MC Moisture Content.

MIR Mid Infrared.

MIRS Mid Infrared Spectroscopy.

MLP Multilayer Perceptron.

MLR Multiple Linear Regression.

MSC Multiplicative Scatter Correction.

MSEV Mean Square Error of Validation.

MT Malaxing Temperature.

Mt Malaxing time.

NIR Near Infrared.

NIR-AOTF Near Infrared Acousto-Optical Tunable Filter.

NIRS Near Infrared Spectroscopy.

NMR Nuclear Magnetic Resonance Technique.

OCDM Oil Content Referring to Dry Matter.

OCFW Oil Content Referring to Fresh Weight.

OSI Oil Stability Index.

PAT Process Analytical Technologies.

PCA Principal Component Analysis.

PDO Protected Denomination of Origin.

PI Peroxide Index.

PLS Partial Least Square.

PLS-DA Partial Least Squares Discriminant Analysis.

PLSR Partial Least Squares Regression.

R Regression Coefficient.

RDO Registered Denomination of Origin.

REP Relative Error Prediction.

RGB Red Green Blue.

RMSE Root Mean Square Error.

RMSEC Root Mean Square Error of Calibration.

RMSECV Root Mean Square Error of Cross Validation.

RMSEP Root Mean Square Error of Prediction.

RMSEV Root Mean Square Error of Validation.

ROI Region of Interest.

RPD Residual Predictive Error.

S Rotating speed of the shovels.

SD Standard Deviation.

SEC Standard Error of Calibration.

SEP Square Error of Prediction.

SG Savitzky-Golay.

SIMCA Soft Independent Modelling of Class Analogies.

SMLR Stepwise Multilinear Regression.

SNV Standard Normal Variate.

SPA Successive Projections Algorithm.

SPA-MLR Successive Projection Algorithm with Multiple Linear Regression.

SVM Support Vector Machine.

TP Total Phenols.

UV Ultraviolet.

V Viscosity.

VIS Visible.

VOO Virgin Olive Oil.

VOOPP Virgin Olive Oil Production Process.

WC Water Content.

