SULTAN MOULAY SLIMANE UNIVERSITY
Faculty of Sciences and Techniques of Beni Mellal

Doctoral Center of Sciences and Techniques

PhD Program of Natural Resources, Environment and Health

DOCTORAL THESIS

Presented By
El Orche Aimen

To obtain

PhD degree

Speciality: Chemistry

Option: Analytical Chemistry, Chemometrics and Quality Control

Application and development of spectroscopic fingerprints and chemometric methods to assess the quality of olive and argan oil: Geographical traceability, shelf life and adulteration

Defended on July 10, 2021 in front the committee members:

• Professor. **Ahmed GAMOUH**,	FST Beni Mellal	**President**
• Professor. **Hassan OULFAJRITE,**	FST Beni Mellal	**Reporter**
• Professor. **Jaouad EL HARTI**,	FMP Rabat	**Reporter**
• Professor. **Amine MOUBARIK**,	FP Beni Mellal	**Reporter**
• Professor. **Tarik AINANE**,	EST Khenifra	**Examiner**
• Professor. **Tarik EL OUAFY**,	FP Khouribga	**Examiner**
• Professor. **Mustapha BOUATIA**,	FMP Rabat	**Co-Supervisor**
• Professor. **Mohamed MBARKI**,	FST Beni Mellal	**Supervisor**

Foreword

This thesis is the result of collaboration between the University Sultan Moulay Slimane and the University Mohammed V. The work was conducted in alternation between the Team of Analytical & Computational Chemistry, Nanotechnology and Environment, and the Laboratory of Analytical Chemistry and Bromatology, Faculty of Medicine and Pharmacy.

This thesis work is focused on developing a rapid optical method to evaluate the quality of olive and argan oil. During this thesis work, I have realized several internships and training

- Training of one year in sustainable development at the Faculty of environmental sciences and biochemistry in Toledo.
- Four-month internship in the national laboratory for the drug control in the department of drugs and pharmacy
- Six-month internship in the laboratory of femtoscience and time-resolved microscopy, faculty of environmental sciences and biochemistry and INAMOL, campus tecnologico de toledo. Supervised by Professor Abderrazzak Douhal
- Internship in the bioequivalence studies center of the sheikh zaid foundation

During this Ph.D. period, we have published 9 articles in Scopus indexed journals with an important impact factor, we also participated in national and international congresses with oral and poster presentations.

The realization of this thesis was made possible thanks to the financing provided by CNRST, Morocco.

Finally, i cannot forget all the friendships that were formed during these three years of the thesis. I would like to thank all the researchers and professors of our team, for their cordial reception and the pleasant working environment that they have established within the team

INDEX

List of Tables

Table 4: results of validation of the three spectroscopic methods developed by PLS regression using accuracy profile

List of figures

Figure 1: visualization of FT-MIR (A), UV-Visible (B) and fluorescence (C) spectra of pure argan oil (0%) and adulterated argan oil at different levels (from 0.5% to 32%).

Figure 2: PLS-DA score plot of the models developed by FT-MIR (A), UV-visible (B) and fluorescence (C) for the discrimination of pure argan oil and adulterated argan oil.

Figure 3: Actual vs. predicted adulteration levels of the best PLS model obtained using fluorescence spectroscopy for the quantification of argan adulteration by olive oil. The model was constructed on two latent variables.

Figure 4: Accuracy profiles obtained for the validation of the three spectroscopic approaches FT-MIR (A), UV-visible (B), and Fluorescence (C) to quantify the adulteration rate of argan oil by olive oil considering PLS regression. The red dashed lines are the upper and lower acceptance limits(UAL and LAL) defined at 15%, the solid blue lines are the upper and lower calculated tolerance limits(UTL and LTL) of β expectation (with β= 90%), and the solid black line is the recovery rate (Accuracy).

.

List of abbreviations

ATR: ATTENUATED TOTAL REFLECTANCE

CLS: CLASSICAL LEAST SQUARE

DAD: DIODE ARRAY DETECTOR

FDA: FACTORIAL DISCRIMINANT ANALYSIS

FT-IR: FOURIER TRANSFORM INFRARED

GC: GAS CHROMATOGRAPHY

HPLC: HIGH PERFORMANCE LIQUID CHROMATOGRAPHY

ICA: INDEPENDENT COMPONENT ANALYSIS

ILC: INVERSE LEAST SQUARES

K-NN: K-NEAREST NEIGHBORS

LDA: LINEAR DISCRIMINANT ANALYSIS

MIR: MID-INFRARED

MS: MASS SPECTROMETRY

NIR: NEAR-INFRARED

OPLS: ORTHOGONAL PARTIAL LEAST SQUARE

OPLS-DA: ORTHOGONAL PARTIAL LEAST SQUARE DISCRIMINANT ANALYSIS

PAEAFAC: PARALLEL FACTOR ANALYSIS

PCA: PRINCIPAL COMPONENT ANALYSIS

PCR: PRINCIPAL COMPONENT REGRESSION

PDA: PHOTO DIODE ARRAY

PLS: PARTIAL LEAST SQUARE

PLS-DA: PARTIAL LEAST SQUARE DISCRIMINANT ANALYSIS

SIMCA: SOFT INDEPENDENT MODELING CLASS OF ANALOGY

SVM: SUPPORT VECTOR MACHINE

SVM-DA: SUPPORT VECTOR DISCRIMINANT ANALYSIS

TLC: THIN-LAYER CHROMATOGRAPHY

UPLC: ULTRA-PERFORMANCE LIQUID CHROMATOGRAPHY

UV-VISIBLE: ULTRAVIOLET –VISIBLE

ACKNOWLEDGEMENTS

I would like to thank all those who have provided me with help, advice and support, especially:

My supervisor Mohamed Mbarki, professor at the laboratory of Organic and Analytical Chemistry, Faculty of Sciences and Techniques of Beni Mellal, University of Sultan Moulay Slimane, Morocco, for his help and all the time that he devoted to me during this project.

My co-supervisor Mustapha Bouatia, professor at the Laboratory of Analytical Chemistry and Bromatology and the team of Formulation and Quality Control of Health Products, Faculty of Medicine and Pharmacy, Mohammed V University Rabat, Morocco, for their supervision and having been present every time I needed their advice and support.

I would like also to thank the CNRST (Centre National pour la Recherche Scientifique et Technique) for the financial support of the present thesis.

Finally I would like to thank my parents, without whom I would not be the person that I am today, thank you for your unconditional support in my career and for having confidence in what I do and although the distance separates us, you were always in my heart, I will never have enough words to express my gratitude.

Abstract

The present thesis focuses on the application of chemometrics and spectroscopic methods to evaluate the quality of food (olive and argan oil). The study aims to evaluate spectroscopic sensors combined with chemometric algorithms for the identification of geographical origin, classification according to freshness, and the determination of adulteration.

In this thesis the capability of UV-visible and FT-MIR in coupling with recognition algorithms was evaluated to determinate the geographical traceability of olive oils coming from various provinces of the Beni Mellal Khenifra region. PCA was applied to the spectral data set to represent the data in a small space, then classification methods were applied to the main components synthesized by the PCA. The application of the PCA-LDA approach on the UV-Visible and ATR-FTMIR spectral data demonstrates a strong efficiency in classifying olive oils according to their geographical origin with a correct classification rate (CCR) of 90.24% and 85.87%, respectively. While the use of the PCA-SVM method shows a CCR of 100% and 97.56% respectively. Next, laser-induced fluorescence spectroscopy at 400 nm was evaluated with multivariate analysis methods, supervised and unsupervised, to establish a rapid analytical approach to distinguish freshly produced olive oils from oils that have been stored for a period of time ranging from 12 to 24 months. The spectral fluorescence data were first processed by PCA in order to visualize the samples in a reduced space. This method displays a strong clustering of the three oil groups using the first three components that summarize 96% of overall variability. Subsequently, three discrimination methods were applied on the emission fluorescence data, these approaches indicate a strong ability to classify the three classes of olive oil according to their degree of freshness. Finally, a comparative study of three multivariate approaches to detect the argan oil adulteration by olive oil was performed using fluorescence, UV-Visible, and FT-MIR spectroscopy coupled to chemometric tools such as PLS

regression and PLS-DA. The application of PLS-DA shows a strong ability to discriminate between pure argan oils and falsified oils. The Validation of the approaches developed by PLS-DA indicates a sensitivity, specificity, and CCR of 100% in all spectroscopic methods. For the quantification of the adulteration rate, the application of PLS also shows a high performance expressed by the high value of R-square and low value of RMSE. The validation of the models developed by PLS using the accuracy profile shows that the PLS approaches guarantee reliable and valid results between 0.5% to 32%, 7% to 32%, and 10% to 32% using respectively fluorescence, FT-MIR, and UV-Visible spectroscopies.

Keywords: chemometrics, spectroscopy, quality control, classification, oils.

Résumé

La présente thèse porte sur l'application des méthodes chimiométriques et spectroscopiques pour évaluer la qualité des aliments (huile d'olive et huile d'argan). L'étude se concentre sur l'évaluation de capteurs spectroscopiques combinés à des algorithmes chimiométriques pour l'identification de l'origine géographique, la classification selon la fraîcheur et la détermination de l'adultération.

Dans cette thèse, la capacité des UV-visible et FT-MIR en couplage à des algorithmes de reconnaissance a été évaluée pour la détermination de la traçabilité géographique des huiles d'olive provenant de différentes provinces de la région de beni mellal khenifra. L'ACP a été appliquée à l'ensemble de données spectrales pour représenter les données dans un petit espace, puis des méthodes de classification ont été appliquées aux principales composantes synthétisées par l'ACP. L'application de l'approche PCA-LDA sur les données spectrales UV-Visible et ATR-FTMIR montre une forte capacité à classer les huiles d'olive en fonction de leur origine géographique avec un taux de classification correcte de 90,24% et 85,87%, respectivement. Alors que l'utilisation de la méthode PCA-SVM montre un taux de classification correcte de 100 % et 97,56, respectivement. Ensuite, la spectroscopie de fluorescence induite par laser à 400 nm a été évaluée avec des méthodes d'analyse multivariée, supervisée et non supervisée, afin d'établir une approche analytique rapide pour distinguer les huiles d'olive fraîchement produites des huiles qui ont été stockées pendant une période de temps allant de 12 à 24 mois. Les données de fluorescence spectrale ont d'abord été traitées par PCA afin de visualiser les échantillons dans un espace réduit. Cette méthode montre un fort regroupement des trois groupes d'huiles d'olive en utilisant les trois premières composantes qui résument 96% de la variabilité globale. Par la suite, trois méthodes de discrimination ont été appliquées sur les données de fluorescence d'émission, ces approches indiquent une capacité forte à classer les trois classes d'huile d'olive selon leur degré de fraîcheur.

Enfin, une étude comparative pour la détection de l'adultération de l'huile d'argan par l'huile d'olive a été réalisée en utilisant la spectroscopie de fluorescence, UV-Visible et FT-MIR combinée avec des outils chimiométriques tels que la régression PLS et la discrimination PLS-DA. L'application de la PLS-DA montre une forte capacité à discriminer entre l'huile d'argan pure et les huiles falsifiées. La validation de l'approche développée par le PLS-DA indique un taux de sensibilité, de spécificité et de classification correcte de 100 % pour toutes les méthodes spectroscopiques. En quantifiant le taux d'adultération, l'application du PLS montre également une performance élevée exprimée par la valeur élevée de R-carré et la valeur faible de RMSE. La validation des modèles développés par PLS à l'aide du profil de précision montre que les approches PLS garantissent des résultats fiables et valides entre 0,5 % à 32 %, 7 % à 32 % et 10 % à 32 % en utilisant respectivement les spectroscopies de fluorescence, FT-MIR et UV-Visible.

Mots clés : chimiométrie, spectroscopie, contrôle qualité, classification, huiles

ملخص

تركز هذه الاطروحة على تطبيق القياسات الكيميائية والطرق الطيفية لتقييم جودة الطعام (زيت الزيتون وزيت الأرغان). بحيث تركز الدراسة على تقييم أجهزة الاستشعار الطيفية مع خوارزميات القياس الكيميائي لتحديد المنشأ الجغرافي وتصنيف الزيوت حسب طراوتها وكشف زيوت الأركان المغشوشة. الجزء الأول من الأطروحة هو عبارة عن مقدمة عامة، أما الجزء الثاني فيعرض نتائج الأطروحة ومناقشتها، أما الجزء الثالث فيقدم خاتمة شاملة.

في الجزء الأول لهذه الأطروحة تم تقييم قدرة الأشعة فوق البنفسجية المرئية ومتوسط الأشعة تحت الحمراء بالإقتران مع خوارزميات التعرف لتحديد التتبع الجغرافي لزيوت الزيتون القادمة من مختلف أقاليم جهة بني ملال خنيفرة. تم تطبيق التحليل الإحصائي للمكونات الرئيسية على مجموعة البيانات الطيفية لتمثيل البيانات في مساحة صغيرة، ثم بعد دلك تم تطبيق طرق التصنيف الإحصائية على المكونات الرئيسية التي تم الحصول عليها بواسطة التحليل الإحصائي للمكونات الرئيسية. يبين تطبيق نهج التحليل الإحصائي للمكونات الرئيسية وخوارزميات التصنيف (PCA-LDA) على البيانات الطيفية للأشعة فوق البنفسجية المرئية ومتوسط الأشعة تحت الحمراء نتائج جيدة في تصنيف زيوت الزيتون وفقًا لأصلها الجغرافي بمعدل تصنيف صحيح (م.ت.ص) يبلغ 90.24٪ و85.87٪ على التوالي. بينما يظهر استخدام طريقة PCA-SVM معدل (م.ت.ص) بنسبة 100٪ و 97.56% على التوالي.

بعد ذلك، تم تقييم التحليل الطيفي الإشعاعي عند 400 نانومتر باستخدام طرق التحليل الإحصائي، لإنشاء نهج تحليلي سريع يمكن من التمييز بين زيوت الزيتون الطازجة والزيوت التي تم تخزينها لفترة زمنية تتراوح من 12 إلى 24 شهرًا. تمت معالجة البيانات الطيفية أولاً بواسطة التحليل الإحصائي للمكونات الرئيسية من أجل تحليل البيانات بشكل جيد. تعرض هذه الطريقة تجميعًا قويًا لمجموعات الزيت الثلاث حسب مدة تخزينها باستخدام المكونات الثلاثة الأولى التي تلخص 96٪ من التباين الكلي. بعد ذلك، تم تطبيق ثلاث طرق إحصائية على بيانات الانبعاث الإشعاعي من أجل تمييز الزيوت القديمة والزيوت الجديدة، ويشير تطبيق هذه الطرق الإحصائية لتحليل البيانات إلى قدرة قوية على تصنيف الأصناف الثلاثة لزيت الزيتون وفقًا لدرجة نضارتها.

أخيرًا، تم إجراء دراسة مقارنة للكشف عن غش زيت الأرغان بزيت الزيتون باستخدام طرق التحليل الطيفي (فوق بنفسجي مرئي ومتوسط بالأشعة تحت الحمراء) جنبًا إلى جنب مع أدوات القياس الكيميائي PLS وPLS-DA. يُظهر تطبيق PLS-DA قدرة قوية على التمييز بين زيت الأرغان الخالص والزيوت المغشوشة. يشير التحقق من صحة النهج الذي طوره PLS-DA إلى

حساسية وخصوصية ومعدل تصنيف صحيح بنسبة 100٪ لجميع طرق التحليل الطيفي. فيما يتعلق بالتقدير الكمي لمعدل الغش، يُظهر تطبيق PLS أيضًا أداءً عاليًا يعبر عنه بالقيمة العالية لـمعامل الانحدار وقيمة الخطأ المنخفضة. يُظهر التحقق من صحة النماذج التي طورتها PLS باستخدام ملف تعريف الدقة, أن مناهج PLS تضمن نتائج موثوقة وصحيحة بين 0.5٪ إلى 32٪، 7٪ إلى 32٪ و10٪ إلى32٪ على التوالي باستخدام طريقة الطيف الإشعاعي، متوسط بالأشعة تحت الحمراء والأشعة فوق البنفسجية المرئية.

الكلمات الأساسية: القياس الكيميائي، التحليل الطيفي، مراقبة الجودة ٍالتصنيف والزيوت

Introduction

The science of analytical chemistry can be defined as a field of chemistry that aims at the identification, characterization, and quantification of chemical substances and the development of methods necessary for this analysis. Furthermore, it is also interested in understanding the phenomena involved in analytical practices and techniques in order to continuously improve them. The sample under investigation may be solid, liquid, or gaseous compounds, and the results of the analysis are expressed as data related to the initial question raised on the sample. Based on the data collected during the analysis, information about the sample can be extracted. This information may be either qualitative or quantitative (or both).

Olive oil and argan oil are considered among the most important oils from a nutritional point of view. Olive oil is considered as an important element of the Mediterranean diet, since it is the main source of fat of the Mediterranean diet. It provides essential fatty acids and its high proportion of unsaturated fatty acids gives it beneficial properties for health. Argan oil is considered as a very valuable cosmetic or edible oil, which has been used for centuries as the main ingredient of the Moroccan Amazigh diet. Nowadays, the origin, freshness, and purity of these oils constitute an important criterion for the import and export market in order to guarantee their traceability and their quality, because many consumers require pure, fresh products and from a specific origin. The information concerning the origin of food is needed to ensure its quality since products of different origins can have different qualities.

The traceability and the control of oil falsification (olive oil and argan oil) have been the subject of the Moroccan authority to protect consumers and increase its economic value. To answer this

question, accurate and rapid analytical approaches are required to establish and evaluate the geographical origin, authenticity, and freshness of oils.

Spectroscopic techniques are generally fast and the analysis requires only a few seconds to a few minutes. Moreover, spectroscopy produces a considerable amount of data for each sample scanned. The use of chemometric tools with many variables provides many advantages in qualitative and quantitative spectroscopic analysis. Generally, methods become more reliable, more accurate, and less sensitive to spectral artifacts. Therefore, the multivariate analysis could be considered as the optimal choice for the evaluation of spectroscopic data in order to develop analytical tools able to assess food quality.

The general objective of the present PhD thesis is to develop rapid multivariate analysis methods and their validation for quality authentication. This includes the investigation and application of various spectroscopic methods that are generally used with spectral pre-treatment and new approaches of classification and quantification in the area of food.

The experimental developments that determine this overall purpose lead to three different sub-objectives:

- **Evaluate the UV-Visible and FT-MIR signal for its use in multivariate approaches to determine the geographical traceability of olive oil.**

Due to the increasing interest of many scientists towards the development of rapid analytical methods for the determination of food traceability using spectroscopic methods. This present work evaluates the ability of two spectroscopic techniques in combination with chemometrics classification tools to answer the question of geographical traceability.

- **Develop multivariate approaches based on non-targeted modeling for studying the freshness of olive oils.**

Most studies based on multivariate qualitative analysis have been focusing on the multi-class approach to address the analytical problems of classifying foods according to their freshness. This PhD thesis exploits the important potential of the single-class approach to deal with food safety issues. For this purpose, well-established class modeling techniques such as PLS-DA, PCA-LDA, and SVM have been used to classify olive oils according to their storage time based on fluorescence spectroscopy.

- **Establish and validate rapid chemometric approaches for the detection and quantification of argan oil adulteration based on spectroscopic sensors.**

The use of multivariate methods in combination with spectroscopic sensors is continuously increasing. This thesis applies three spectroscopic approaches for the detection of adulteration through the development of qualitative and quantitative models. However, there are no established criteria for their validation at a global level. For this purpose, a classical validation approach has been established for the validation of the qualitative model while a statistical approach named "accuracy profile" has been applied for the validation of the quantitative models.

Structure of the thesis

This thesis is organized into three parts: Part I: Introduction, Part II encompasses the experimental part, results, and discussion and the third part contains conclusions and perspectives.

The first part: It includes the section "State of the art", containing a bibliographical overview of recent literature. In addition, conceptual backgrounds that the authors consider insufficiently

exposed or established in the documents, which are relevant for a proper understanding of the current work, are presented in the sub-section on the state of the art.

The second part: includes the results obtained during this thesis work on the spectral characterization of olive oil and argan oil, geographical classification, authentication, freshness and quality control. This part also evaluates practical aspects of implementation and the use of chemometric tools in the context of oil quality control.

This part is generally divided into three chapters

Chapter 1 describes a study that focuses on the evaluation of the capability of horizontal ATR-FTMIR and UV-Visible spectroscopy in the discrimination of virgin olive oils from the Moroccan region of Beni Mellal-Khenifra

Chapter 2 investigates the feasibility to use fluorescence emission spectroscopy for rapid determination of olive oil freshness.

Chapter 3 compares three multivariate approaches based on three spectroscopic methods for the detection and quantification of argan oil falsification.

The third part is a concluding section, which summarizes the results achieved, future trends and perspectives regarding the application of multivariate analysis and spectroscopy for food control.

PART 1: State -of-the art of the thesis topic.

I. Plants materials

1. Olive tree

The olive tree is a sub-tropical evergreen tree and its edible fruit. The fruit of the olive tree and its oil are considered as key elements of the Mediterranean diet and are appreciated outside the region [1], [2]. The cultivation of the olive tree has several aspects: wood (fuel), livestock food, as fertilizer, utilization in the diet (use of oil), as cosmetics, and the economic importance of the tree and its fruits [3]. All these issues have been the subject of vigorous discussion in recent decades. For this reason, olive tree cultivation has spread to all countries bordering the Mediterranean, and the tree is also planted as an ornamental plant in appropriate climates.

The olive tree measures between 3 and 12 meters in height and has many branches. Its leaves, are dark green on top and silvery on the underside and are paired opposite each other on the branch, its wood is resistant to rot. If the top dies, a new trunk will often be born from the roots [4].

Olive trees bloom in late spring; small whitish flowers are born in loose clusters in the axils of the leaves. The flowers are divided into two types: perfect, containing both male and female parts, which can grow to give the fruits of the olive tree; and male, containing only the pollen-producing parts, the olive is pollinated by the wind. The fruit formation of the olive is often irregular. In some regions, particularly where irrigation and fertilization are not carried out, alternate fruiting is the rule. Trees may produce a plentiful crop one year and not bloom the following year [5].

In Morocco, olive cultivation (olive growing) is an important source of employment providing more than 51 million working days per year, the equivalent of 380,000 permanent jobs. With good reason, the olive tree is the main cultivated fruit sector and represents 65% of the national arboricultural sole. Olives are cultivated mainly for the production of olive oil.

2. Argan tree

The argan tree Known as the "tree of life" by the Berbers for its numerous health benefits, Argania spinosa L. Skeels is a tropical plant and is the single endemic species of the Argania genus in Morocco. This tree is considered an important socio-economic pillar of southern Morocco. Since antiquity, it has played an important role in the culture and symbolic customs, economy, healthcare and gastronomy of the Moroccan Amazigh people [6]. As the only home of this rare and endangered species, the argan grove was classified as an International Biosphere Reserve by UNESCO in 1998 [7].

At present, the Moroccan argan tree forest occupies an overall area of 840,000 ha, including the fertile Souss Valley area, the Anti-Atlas foothills, and the coastal area between Essaouira and Agadir [8]. Since antiquity, argan oil has been a basic material for cosmetics, hair care, skin treatments, and body beauty.

The argan tree is extremely resistant to unfavorable environmental factors and grows in arid and semi-arid regions, although agricultural production is affected consequently [9]. Therefore, suitable care and monitoring of the crop (irrigation or rainfall, appropriate harvesting and ripening practices) are crucial to achieving high fruit and oil yield. Argan fruit weight increases and attains its maximum maturity between June and August, but it fluctuates depending on the latitude of the cultivation area, soil quality, temperature, and water availability [10]. Argan oil is extracted from the argan fruit (stones), while the extraction process differs between edible and cosmetic oil.

The cultivation of argan fruit represents economic importance in Morocco, enabling the subsistence of part of the Moroccan population during centuries. Its socio-economic pattern has stayed traditional, structured first around a particular socio-legal structure that operates on the basis of various productions. Today, it is a source of additional income for local rural populations, the major

part of the population is involved in agriculture, in general, the different sectors of the argan tree's activity became an opportunity for sustainable development.

The argan forest formerly extended over 1.5 million hectares. Nowadays it covers an area of more than 800,000 hectares, while less than half of the forest has disappeared and around 600 hectares of forest are removed each year. However, the Moroccan agriculture and forestry authorities are initiating the reforestation of those areas with argan trees [11]. The argan tree provides a livelihood for 3 million people, 2.2 million of whom live in rural areas. The diverse productions of the argan tree ensure more than 20 million days of work, of which 7.5 million days are mainly for females to extract argan oil alone [12].

II. Oil extraction

1. Olive oil extraction

The process of producing olive oil is relatively simple but requires careful adherence to the following steps. The extraction process is outlined briefly in Figure 1.

- **Washing, crushing and mixing.**

The olives destined for the production of oil, they are sorted to remove twigs, leaves and then washed with cold water. Afterward, the olives are immediately crushed to avoid oxidation, with the pits containing antioxidants, considered as natural preserving agents. Crushing constitutes the first phase of the actual extraction process, which causes the rupture of cell walls and membranes resulting in the release of cell juices and oil. This operation can be carried out with stone grinders or with a metal grinder. Grinding is not enough to break all the vacuoles containing the oil, a mixing is then applied to the dough until a smooth dough is obtained to facilitate oil extraction.

The grinding and mixing allow obtaining a paste that contains solid matter (fragments of kernels, epidermis, and cell walls ...) and fluid (oil and plant fluids of the cell content).

- **Extraction**.

The extraction of the oil is always carried out cold (27° C according to the international standard) either by centrifugation or by pressure. The centrifugation is done in a decanter, a metal cylinder rotating at high speed (4000 rpm), in which the different components are separated based on the difference in density between water, oil and residue, centrifugation separates the oil (about 20%) from the water and the pomace. pressure or solid-liquid extraction is the oldest process which is often used to obtain olive oil. Generally, the pressure is obtained in an open hydraulic press by arranging the oil paste in thin layers alternating with fiber discs, called ***scourtins***, in a mobile tower. The ***scourtins*** consist of a synthetic fiber disc pierced in the center so that it can be threaded onto the needle. On the first ***scourtin***, placed on the bottom of the tray, a 3 cm thick layer of paste is placed, a second ***scourtin*** and a second layer of paste are superimposed, and so on. For every three layers of dough, we superimpose a ***scourtin*** without dough and a steel disc to distribute the pressure evenly. Finally, the liquid part composed of vegetation water (margine) and oil flows out while the solid part (cores and pulp) remains between the scourtins, this is what is known as the pomace, it is during this process that the bitter-tasting oleuropein is eliminated in the vegetation water.

- **Separation**

It is done by difference in density between fluids (olive oil being lighter than water).

To isolate the oil from the margins, natural decantation or centrifugation separation can be used. Once the separation process is complete, the olive oil is immediately stored in stainless steel tanks to avoid oxidation. The oil can then be filtered to make it clear and shiny, and then bottled. Once bottled, the olive oil should be stored in the dark and protected from heat and light.

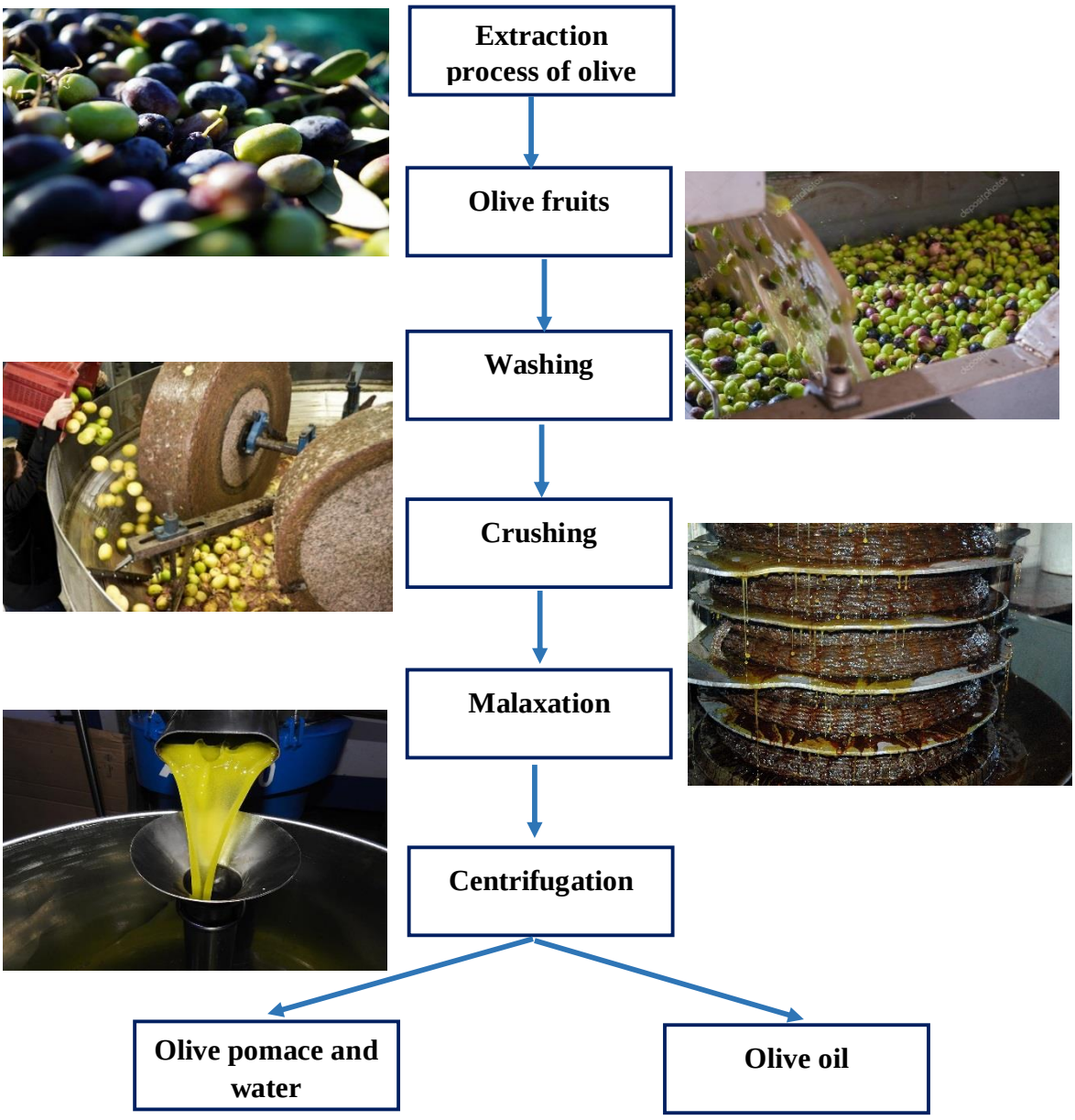

Figure 1: General process for the extraction of olive oil.

2. Argan oil extraction

Basically, the argan oil extraction is carried out using the unroasted argan kernels or the roasted kernels to produce respectively cosmetic oil or edible oil. Argan oil is produced through two techniques; traditional extraction or mechanical extraction (figure 2). The roasting operation is used

only for the production of edible quality oil. The roasting process allows eliminating the bitterness of the oil by reducing the moisture content of the kernels and gives the oil a hazelnut taste and a dark brown color. The artisanal process of extraction is only used to produce edible argan oil. The kernels obtained from the crushing are put in a container usually made of terracotta (Taflounte), then they are heated over a low fire. From time to time, the kernels are stirred so that they take on a brown tint. According to the women, the purpose of this operation is to develop the color, smell, and taste of the oil to be extracted. If the fire is increased, the color of the oil becomes browner. Then the roasted kernels are crushed using a stone grinder, usually similar to the one used in the artisanal milling of cereals (Azerg). The extracted paste is accumulated in a pottery container for mixing. The mixing is carried out manually with the addition of a small amount of warm water to obtain a creamy paste (Tazguemmoute). However, this operation conditioning the quality of the oil. Therefore, a significant increase in water will reduce the quality of the oil. The resulting paste is pressed by hand, releasing oil in the form of droplets.

The mechanical process is performed on both roasted and unroasted kernels to produce an edible or cosmetic grade oil, respectively. During this process the kernels are crushed in a press that separates the argan oil from the residue (cake).

The oil is then decanted and filtered to remove the suspended particles generated by the crushing.

A summary of the various stages of argan oil extraction is shown in Figure 2.

Fruit pulping by hand

There are two ways of pulping, either it is done by women, in this case it consists of a light crushing of the whole fruit against a stone that serves as a support, then the separation of the pulp from the nut is done manually.

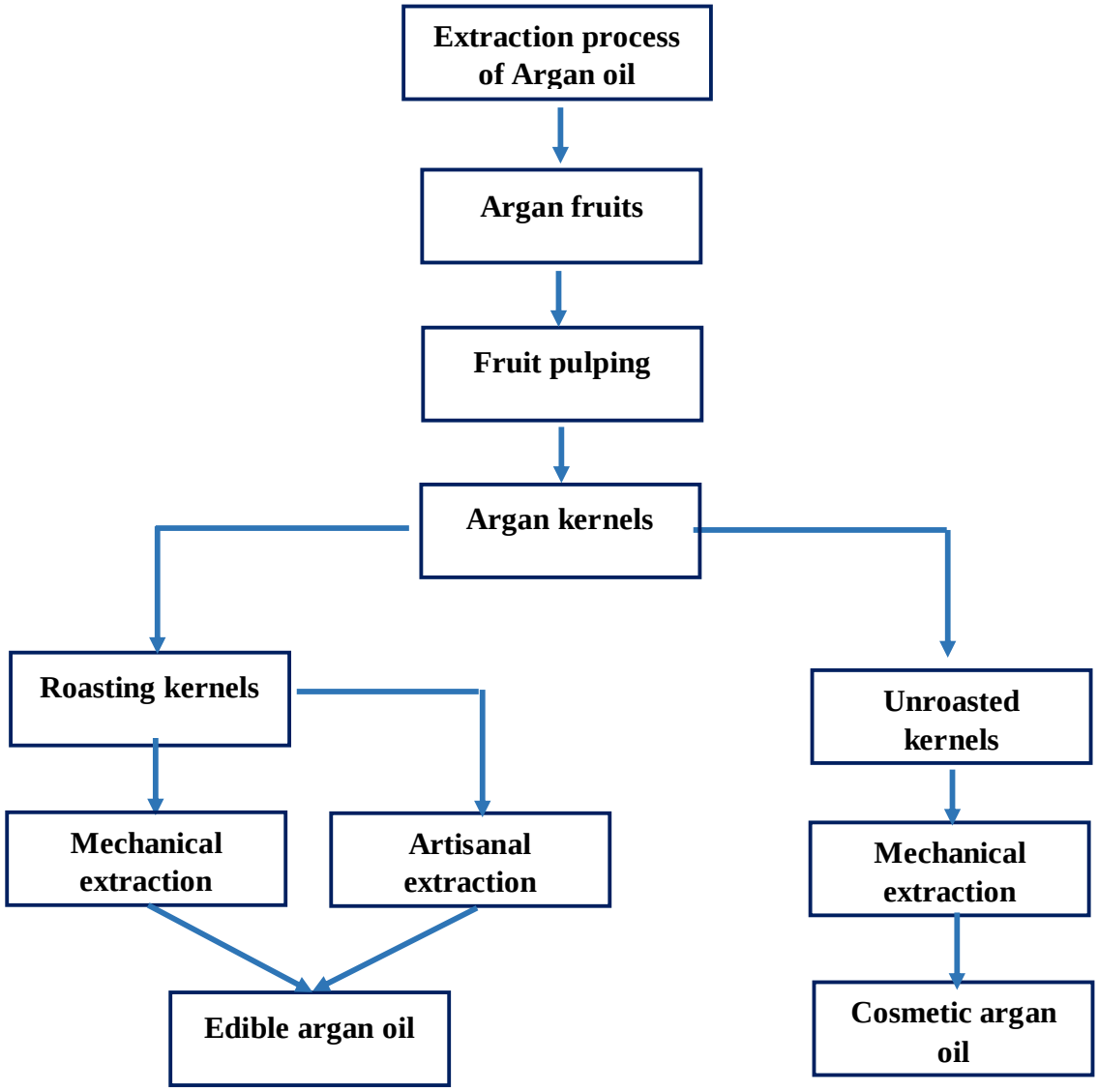

Figure 2: production process of argan oil.

III. Plant oil composition and sensory properties

1. Olive oil

Olive oil is considered one of the most important elements from a nutritional and cosmetic aspect since it contains chemical compounds highly beneficial for human health. Olive oil is also rich in phenolic compounds known for their antioxidant properties that could delay the development of

certain cardiovascular or degenerative diseases [13]. Its composition in fatty acids and phenolic compounds is strongly correlated to the varietal origin, geographical origin, and ripeness degree. In addition, the quality of olive oil depends on several factors such as environmental and agronomic factors, the manufacturing process, and also the packaging. Generally, its quality can be evaluated by Physico-chemical parameters and organoleptic characteristics regulated by international standards. Limit values are set to define different quality categories, the best of which is "extra virgin oil". Although it is of very good quality, during its use or storage, an oil deteriorates and loses its properties [14]. Table 1 provides information on the composition of 100 g of olive oil based on the United States Department of Agriculture (USDA) National Nutrient Database for standard reference.

Table 1: The nutritional composition contained in 100 of olive oil.

Nutrients name	Units	Amount
Energy	Kcal	884
Calcium, Ca	mg	1
Iron, Fe	mg	0.56
Sodium, Na	mg	2
Potassium, K	mg	1
Choline, total	mg	0.3
Vitamin E (alpha-tocopherol)	mg	14.35
Vitamin K (phylloquinone)	µg	60.2
Fatty acids, total saturated	g	13.808
Fatty acids, total monounsaturated	g	72.961
Fatty acids, total polyunsaturated	g	10.523

Olive oil is approximately composed of 99% fat. The remaining 1% represents minor compounds; they are mainly made up of: squalene, triterpene alcohols, sterols, phenols and tocopherol derivatives (Respectively in order of importance).

Olive oil fat is composed mainly of triglycerides. These are made up of different kinds of fatty acids, whose distribution is characteristic of olive oil and with a more detailed level regarding the varieties or the geographical origin. The content in each of these fatty acids (table 2), especially oleic acid, should not be confused with the acidity of an oil, which is expressed in grams of free oleic acid per 100 grams of oil.

Sensory analysis is an essential technique for characterizing olive oils and studying consumer preferences. International cooperative studies, supported by the International Olive Council (IOC), have provided a codified sensory methodology for virgin olive oils, known as the "COI Panel test". Assessing exactly the way in which various factors affect the sensory quality of the final product is crucial in distinguishing between different types of oil. The following factors all play an important role in the production of high quality olive oil: cultivar, cultivation techniques and maturation of the olive, harvest and storage of the olive, de-leafing and washing of the olive, pressing, centrifugation. Kneading, extraction and conservation conditions of the oil [15].

Table 2: Fatty acid composition of olive oils (%)

Fatty acids	Denomination	Average centered	Minimum value	Maximum value
C16:0	Palmitic acid	11.8	8.53	14.49
C16:1ω9	Hypogeic acid	0.12	0.09	0.20
C16:1ω7	Palmitoleic acid	0.81	0.26	1.76
C17:0	Margaric acid	0.08	0.03	0.20
C17:1ω8	Margaroleic acid4	0.15	0.06	0.36
C18:0	Stearic acid	2.2	1.3	3.3
C18:1ω9	Oleic acid	72.6	64.5	80.3
C18:1ω7	Vaccenic acid	2.3	1.2	3.9
C18:2ω6	Linoleic acid	7.9	3.6	16.8
C18:3ω3	Linolenic acid	0.65	0.39	0.98
C20:0	Arachidic acid	0.37	0.23	0.49
C20:1ω9	Gondoic acid	0.28	0.21	0.40
C22:0	Behenic acid	0.11	0.07	0.16
C24:0	Lignoceric acid	0.05	0.03	0.08
Saturated fatty acids		14.8	11.75	17.77
Monounsaturated fatty acids		76.6	68.5	83.4
Polyunsaturated fatty acids		8.6	4.23	17.46

2. Argan oil

argan oil is considered among the most well-used oils for food consumption due to its balanced chemical composition. Argan oil has numerous pharmacological activities and health benefits due to its chemical composition profile. This oil is therefore composed of a glyceric fraction (99%) and an unsaponifiable fraction. It is an oil of the oleic-linoleic group, composed of 80% unsaturated fatty acids. Compared to olive oil, argan oil is more unsaturated and contains less oleic acid (45%); its low content of unstable linoleic acid, and its high content (35%) of linoleic acid, polyunsaturated, make it an oil of high food value, with hypocholesterolemic and anti-atherogenic properties.

The unsaponifiable fraction (about 1%) contains carotenes (37.5%), tocopherols (7.5%), sterols and methyl sterols (20%), triterpene alcohols (20%) and xanthophylls (6.5%). the color of argan oil is mainly due to carotenoids and particularly xanthophylls. Argan oil is particularly rich in phenolic compounds and tocopherols (620 mg/kg vs. olive oil: 320 mg/kg). These tocopherols make it an excellent source of vitamin E, which is well known for its eutrophic properties that contribute to organ development. The richness of argan oil in vitamin E leads to the stimulation of cellular enzymatic activities related to detoxification and antioxidant defenses.

Quality aspects being more and more crucial for customers purchasing decisions, it is necessary for manufacturers and marketers of argan oil to have in-depth knowledge of the factors that can influence the quality of the oil during the treatment, storage, and transport. They include the oxidation state and shelf life of the oils, but especially the sensory quality because only products that fulfill the expectations of the consumer can be successful commercially. This is especially true for a high priced product such as argan oil. The study of the sensory properties of argan oil showed that the extraction process has an important effect on the sensory quality of argan oil, because the

roasting step of the kernels and the argan oil extraction technique play a determining role on the quality and quantity of the aroma fraction.

IV. Fingerprinting concept for food control.

Fraudulent food scammers employ a variety of techniques to make financial gains. Such methods include mislabeling, falsifying or not documenting, performing unauthorized processes, and substituting, diluting, or incorporating ingredients in a product without declaring it. As ingredients from national suppliers or specific regions are often expensive, fraudsters can use ingredients from lower-priced regions or alternatively substitute synthetic substances. Food products that are commonly adulterated or misrepresented include olive oil, milk, honey, orange juice, fish, coffee, and herbs and spices.

Food fingerprints could be understood as analytical molecular markers that represent a characteristic food quality state or condition, enabling better identification of products. Basically, it is a label or set of labels that help us to address many questions about the authenticity of food, such as "Are these oils really organic? Is this oil really from Morocco? Can we distinguish between adulterated and non-adulterated olive, and argan oil? Can we discriminate between fresh and expired oil? Unfortunately, looking for these fingerprints is not only concerned with the quality of the products, but also with their innocuity for human health. The ability to authenticate and verify food products is currently a major focus for the food industry, but also for governments. This is why there is extensive legislation concerning food safety and authenticity around the world, of which the United States and the European Union have the most comprehensive and strictest directives. In the United States, the Food Safety Modernization Act (FSMA) [16] defined by the Food and Drug Administration (FDA) establishes several rules to prevent the adulteration of food.

On the other hand, the European Food Safety Authority (EFSA) [17] includes a set of laws and guidelines for food quality and safety assurance, designed as general food laws.

1. Foods and their untargeted/targeted fingerprints

In the field of food authentication, we can find different analysis techniques that can be based on targeted and non-targeted approaches. The first approach mainly focuses on the analysis of a particular metabolite or group of metabolites, while in the non-targeted approach, the main objective is to discriminate between foods which may change regarding many factors such as environmental, variety or human modifications which is often the case of adulteration.

The right choice of a fingerprinting technique depends on the characteristics of the constituents of the food. Several targeted techniques, as UPLC, HPLC, GC, CE, and TLC coupled to many detectors, eg. UV-DAD, UV-PDA, UV-FID, MS / MS are applied to develop fingerprints based on the study of the chemical composition of food. While spectroscopic techniques, eg. FTIR, NIR, NMR and UV can also be used to construct food-based fingerprints.

The implementation of chemometric tools on fingerprint data shows potential performance to assess the complex composition of foods such as oils. Chemometric tools including experimental designs, exploratory multivariate data analysis, mathematical and statistical data preprocessing tools, variable selection, regression tools, discrimination, and pattern recognition techniques form the basis for managing fingerprints of food and other products.

The general procedure of analysis operations and the chemometric tools used for fingerprint analysis are summarized in the figure 4. It gives a graphical overview of the analysis of foods (oils) using non-targeted and targeted fingerprint techniques coupled with multivariate data analysis methods or chemometric methods to assess geographical origin, genetic background, and falsification.

Vegetable oils (olive oil and argan oil) have been widely used for hundreds of years in food, prevention, and treatment of human diseases. A specific chemical profile or fingerprint of oil can be obtained using spectroscopic, chromatographic, or electronic techniques. This food fingerprint can be crucial for quality assessment, identification of adulteration, and classification of oils according to varietal and geographical origin. The exploitation of analytical data (spectroscopic, chromatographic ...) by multivariate data analysis tools such as exploratory analysis, modeling, and pattern recognition is useful for obtaining typical chemical information.

The fingerprinting approach has become an extremely powerful investigative tool for these purposes. Although the use of fingerprints in non-targeted/targeted approaches remains limited for the analysis of oils, it provides possibilities of discrimination according to geographical origin, taxonomic identification, control of adulteration, optimization, and quality control of oils by the determination of physicochemical parameters.

Figure 3 shows a brief summary of the general analytical procedure and the chemometric tools employed in fingerprint analysis. It gives a graphical insight of plant analysis using non-targeted and targeted fingerprinting techniques in combination with multivariate data analysis.

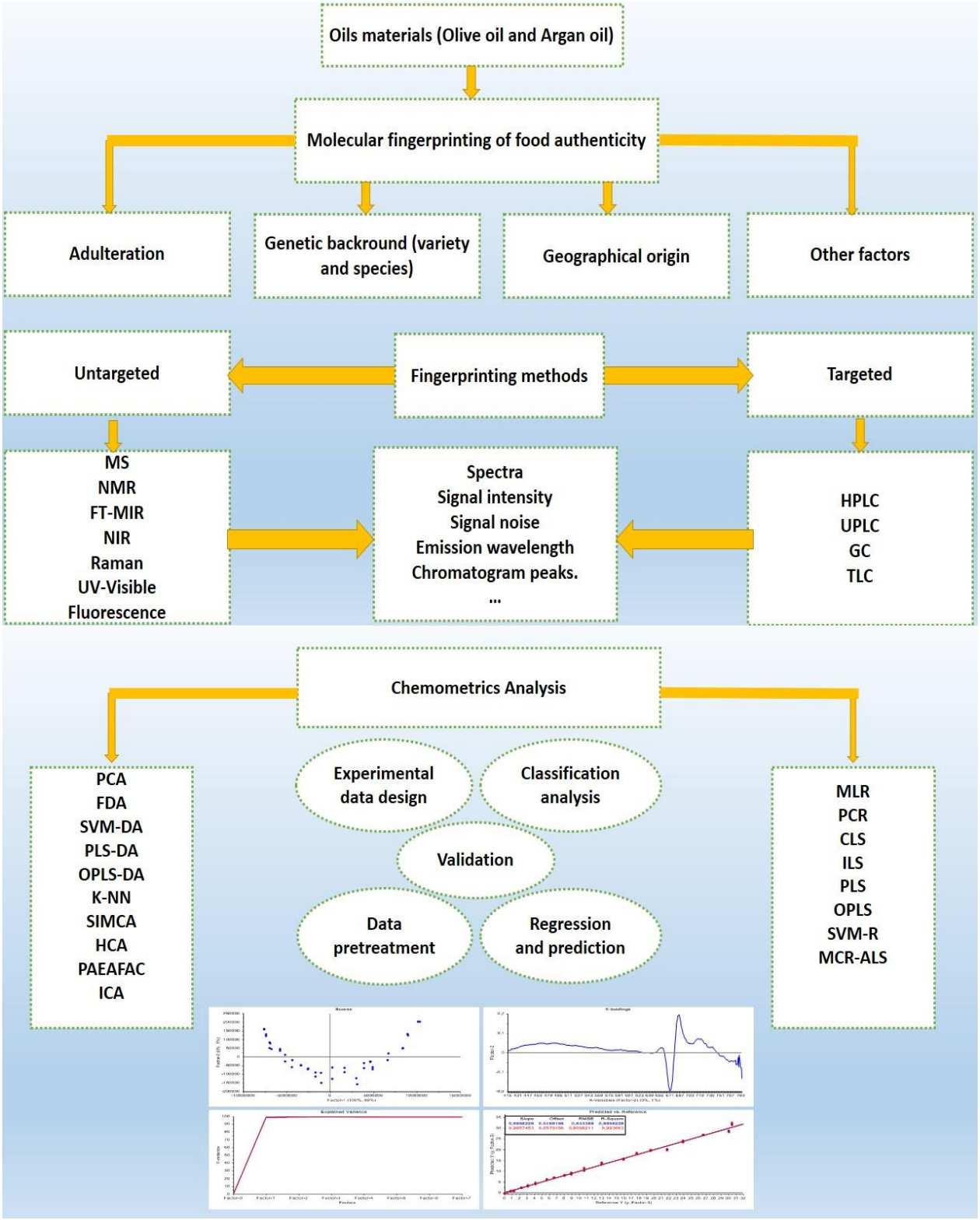

Figure 3: general procedure of analytical operation and chemometric tools used in the analysis of fingerprints of foodstuffs (oils).

The choice of an adequate approach (non-targeted or targeted) must be established to build the required fingerprint from a specific foodstuff material (oil, plant extract, food), and the result should comply with several performance criteria (reproducibility, efficiency, robustness, time and cost...). The literature provides clear proof of the applicability of the fingerprint concept for the development of analytical methods to achieve multiple purposes at the laboratory and industry levels [18], [19]. The development of fingerprint methods, combined with chemometric tools and their validation, can contribute to providing robust analytical results that answer many questions related to food safety certification. Besides qualitative and quantitative applications, the developed fingerprint can be implemented as a key element in food quality assurance for international marketing. In addition, further instrumental enhancements could improve the sensitivity and specificity of fingerprint tests.

2. Spectroscopic fingerprinting

Spectroscopy is the study of how light (electromagnetic waves) interacts with matter. Spectroscopic analysis methods allow us to investigate the matter by various methods (FT-IR, UV-Visible, NMR…) to obtain information on the structure of the molecules that make up this matter. The energy of a molecule results from four contributions: Electron energy (**Ee**), translation energy (**Et**), vibration energy (**Ev**), and rotation energy (**Er**). These energies are quantified. Each energy corresponds to a type of spectroscopy, and each type of spectroscopy will give different information on the nature of the chemical compound studied, i.e. the mode of establishment, the functions, the environment of the atoms, and the number of atoms through UV-visible spectroscopy, fluorescence, infrared, nuclear magnetic resonance.

The various types of molecular energies (electronic, vibrational, or nuclear spin) are quantified which means that only certain energies are permitted. The molecule can be excited from its lower

energy state (ground state) to a higher energy state (excited state) by a photon (energy quantum) of electromagnetic radiation of the appropriate wavelength.

According to quantum theory, each body (atom, ion or molecule) can only exist in certain discrete energy states. Moreover, the transition of a body from its ground state to its excited state can only be done by absorption of the energy quantum which represents the energy difference between the ground state and the excited state. Thus, a body crossed by electromagnetic radiation will only absorb the photons having the energy allowing this body to reach an excited state. It will then remain in this state for a brief moment (10^{-13} s) before relaxing to its ground state by re-emitting the absorbed energy in the form of heat and/or an electromagnetic wave.

Food analysis using spectroscopic fingerprinting techniques has become more and more common and widespread. These approaches are very practical and easy to apply at the laboratory and industrial levels since they are typically fast, inexpensive, and non-destructive. For many purposes, the application of spectroscopic methods requires no sample pre-treatment and the acquisition of the spectrum can be performed in about one minute without the use of solvents and reagents. Consequently, the investment in this spectroscopic approach for the development of rapid food authentication instruments has attracted many industries and laboratories.

Various quick and non-destructive instrumental approaches have been proposed to overcome the challenges faced by food control units in the implementation of the process analytical technology (PAT). These include UV-Visible, Mid-infrared, Near-infrared, and fluorescence spectroscopy which have proven to be successful analytical approaches for food analysis due to a number of important advantages [20].

The chemical information contained in the spectral data located in the positions, intensities, and shapes of the bands, provide information about the molecular structure and properties of the chemical compounds, the intensities of the bands are correlated with the concentration of these

compounds, as demonstrated by the Lambert-Beer law. However, this is possible for a purely component system, but in the case of foods containing numerous components leading to complex spectra with overlapping peaks.

Indeed, in order to take advantage of these spectroscopic fingerprinting techniques, it is necessary to overcome the sensitivity and selectivity limitations that result from the bands being relatively weak and strongly overlapping at the spectral level.

Therefore, in order to carry out a successful analytical procedure, and to obtain the maximum amount of information, it is fundamental to use chemometric methods which are based on mathematical and statistical algorithms to extract as much information as possible from the spectra of the sample.

3. UV-Visible spectroscopy

UV-Vis spectroscopy is an analytical method for monitoring and measuring the interactions of UV-visible light with different chemical compounds in the wavelength range between 200 and 800 nm. The technique exploits different physical responses of light and analytes in the sample, such as absorption, scattering, diffraction, refraction, and reflection [21]. The aim of this method is to study the transitions between the electronic levels of the molecule. These transitions take place when the molecule absorbs a photon of the same energy as the difference in energy between the electronic levels. Therefore, characteristic absorption spectra can be obtained for individual molecules as the electrons of these chromophores are excited [21]. Quantitative analysis based on UV-Vis spectroscopy has been described by Beer-Lambert's law. The method can be used to determine and quantify the concentration of the target molecule in the food matrix.

4. Fluorescence spectroscopy

Fluorescence is the phenomenon that results from the emission of photons by a molecule excited by light radiation in the ultraviolet and visible [22]. When a photon encounters a molecule, an electron absorbs this energy and passes from the fundamental singlet level S_0 to a singlet level S_1 Figure 4. The excited state persists for a finite characteristic time of the molecule. Following collisions with other molecules, a deactivation takes place and the electron returns to the first vibrational level S_1. If the molecule has enough energy, after a short time that is characteristic of each molecule (10^{-9}, 10^{-7} seconds), the electron returns to the S_0 level by emitting a photon. This photon emission is called fluorescence and the signal recorded at different wavelengths constitutes the fluorescence emission spectrum. The emitted photon will thus have lower energy than the excitation photon: for a given molecule, the emission wavelengths will thus be higher than the excitation wavelengths and the energy of the fluorescence photons will be lower than that of the excitation photons. The whole of the radiation emitted during fluorescence deactivation constitutes the emission spectrum.

Fluorescence spectroscopy is a sensitive, rapid and non-invasive method of analysis. It provides information on the presence of fluorescent molecules and their environment in analyzed samples. Thus, the fluorescence properties of aromatic amino acids in proteins [22]–[24] and of the fluorescent products resulting from the oxidation of lipids. It is also used for the evaluation of the quality of olive oils, geographical origin, detection of adulteration and monitoring the degradation of oils according to their shelf-life [25]–[30].

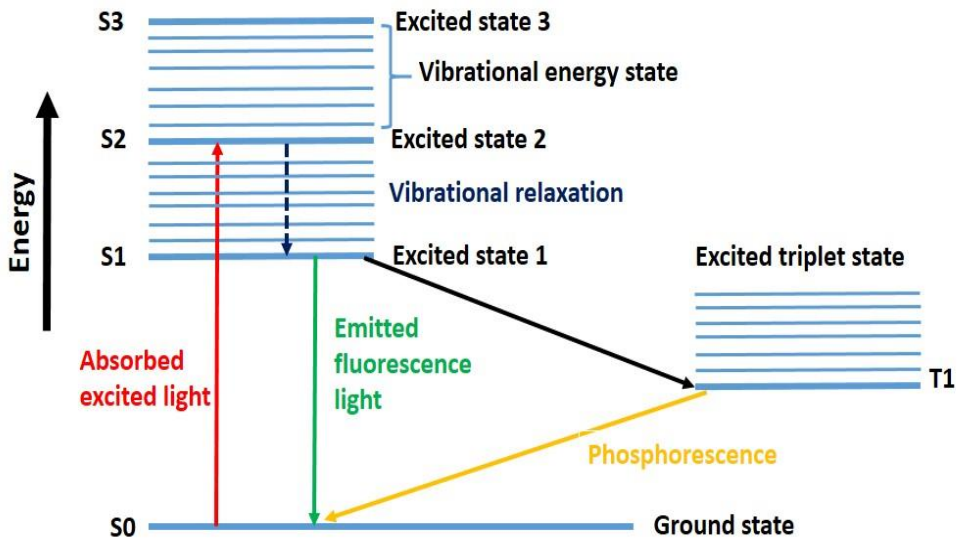

Figure 4: Deactivation of molecules and emission of fluorescence or phosphorescence.

5. FT-MIR spectroscopy

The infrared range is subdivided into three categories according to frequency: near-infrared (NIR) which ranges from 750 nm to 2500 nm, mid-infrared (MIR) ranging from 2500 nm to 25000 nm, and far-infrared at wavelengths above 25000 nm.

MIR is considered to be the most informative part of the infrared spectrum. The absorption bands observed in MIR are mainly associated to fundamental valence bond vibrations (ν) of functional groups of molecules. The MIR spectra of many molecules are already known. The allocations of the spectral bands in MIR are described by *luna et al* and *Gouvinhas et al* [31], [32].

In the field of food processing, MIR spectroscopy has been used less for on-line measurements than PIR. This is explained by the fact that water is a major constituent of these products and contributes strongly to the MIR spectrum.

However, the development of Fourier Transform Infrared Spectroscopy (FTIR) in recent years has given the possibility of obtaining interesting information on lipid structures, fatty acid composition, and proteins [33]. The development of the Attenuated Total Reflection (ATR) technique has proved to be very useful for the acquisition of MIR spectra of solid, liquid, and pasty food products simply spread on a crystalline slide of zinc selenide (ZnSe), silicon (Si), or germanium (Ge).

V. Chemometric tools applied in fingerprint analysis

Chemometrics or multivariate data analysis is defined as the science of data acquisition, validation, and processing in the field of analytical chemistry. It includes mathematical signal processing and statistical methods for extracting information from spectral data. The development of analytical methods (particularly those based on spectroscopic techniques) is closely linked to chemometric progress. Indeed, the analysis of spectral collections with small differences requires the use of chemometric methods to evaluate the data, an approach that has been successfully used in the field of infrared spectroscopy for many years. Applying these statistical tools to data collections containing a large number of variables measured for a large number of samples allows the extraction of relevant information and provides conclusions on the level of statistical significance of the small spectral differences observed.

Chemometrics covers several objectives such as the application of pretreatments to experimental data to improve signal quality, the construction of models for pattern recognition, and quantitative determination [20]. There are many chemometric techniques that can be used to carry out these objectives; this section describes those that have been used in this report.

- Exploratory Analysis
- Quantitative Predictive Modelling
- Classification

To perform chemometric studies, the first step of the analysis involves performing exploratory analysis of the multivariate data, this chemometric processing known as qualitative analysis or unsupervised analysis, which is performed without prior knowledge about the nature and group membership of the samples. Before processing exploratory analysis, the data must be pre-processed or "cleaned". This operation is often performed using various algorithms.

1. Data Preprocessing

Chemometric pretreatment of data is commonly used before performing modeling in order to reduce noise and undesirable signal interference. Several pretreatment methods have been developed in spectroscopy (MIR, PIR ...), due to its sensitivity to the external environment (temperature, humidity, etc.) [20]. Consequently, pretreatment of the fingerprint data is often required to obtain accurate results for the desired purpose. Among the main pretreatments applied to multivariate data are mean centering and scaling.

When performing a PCA, PLSR, or PLS-DA analysis it is usually necessary to center the data on the mean. This pre-treatment calculates the average value of each column of the data matrix and subtracts this value from the column, moving the axes of the coordinate system to the center of the data and makes each sample display only the differences it has with respect to the average of the original data sample.

- **Standard normal variate (SNV).**

SNV is a transformation that is usually applied to spectroscopic data to minimize the effects of light scattering. It uses centering and scaling of each individual spectrum (i.e., it standardizes each spectrum by manipulating only the data of that spectrum). The practical result of the SNV is that it minimizes the multiplicative scattering interferences in the spectral data produced by the different

particle sizes in the sample; one effect of the SNV is that, on the vertical scale, each spectrum is centered at zero.

- **Normalization.**

In many analytical methods, the variables measured for a given sample are increased or decreased from their authentic value by a multiplication factor. Normalization methods attempt to correct these types of effects by identifying some aspect of each sample that should be substantially constant from one sample to another and to correct the scale of all variables using that characteristic. When building discriminant analysis models such as PLS-DA (partial least squares - discriminant analysis) or SIMCA (soft independent modeling of class analogy), normalization is performed if the relationship between the variables and not the absolute magnitude of the response that is the most critical aspect of the data for the identification of a species; for example, the concentration of a compound is not so relevant, just the fact that it is in a detectable quantity. The use of normalization under these conditions should be considered after evaluating how the response of the variables for the different classes of interest changes.

- **Scaling**

The scaling of the data of a matrix between a minimum and a maximum value is a particular case of normalization that can be applied before the construction of the mathematical models. This pre-treatment can be useful to avoid the presence of extreme values in the scaling of data in some samples of natural origin and is preferred when it comes to quantitative applications.

- **Baseline correction.**

There are different ways to make the baseline correction; in the present memory, the Weighted Least Squares (WLS) method has been used. This method is commonly used in spectroscopic (or chromatographic) applications where the signal of some variables is due only to the background signal. These variables serve as a reference to determine how much background signal should be removed from the nearby variables. The WLS algorithm uses an automatic approach to determine which points are most likely to be baseline only; it does this by iteratively adjusting a baseline to each spectrum (or chromatogram) and determining which variables are clearly above the baseline (i.e., the signal) and which are below it. It is assumed that the points below the baseline are more significant in adjusting the signal from the bottom of the spectrum.

- **Smoothing (Savitzky-Golay)**

In spectrometry, the most common noise reduction algorithm is the Savitzky-Golay (SG) algorithm [34]. It is a polynomial smoothing applied on a moving window. At each point i of the spectrum, the raw value Xi is replaced by the value Zi of a polynomial fitted on a window around point i, as shown in figure 5. The spectrum is not replaced by a piece of polynomial. It is the central ordinate of the window that is replaced by the central ordinate of the polynomial.

Two parameters must be set for this algorithm: the window width w, and the degree of the polynomial d. The higher w is larger more the resulting spectrum will be smoothed. The higher d is larger less the resulting spectrum will be smoothed. Note that for computational purposes, we must have $w > d$.

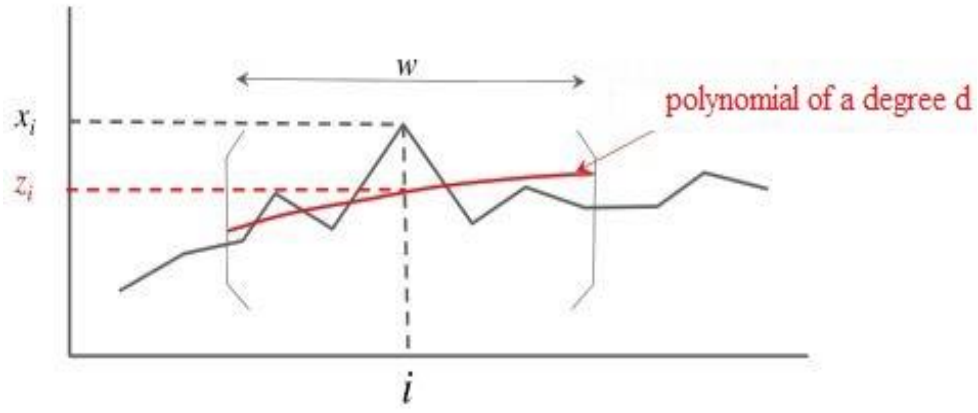

Figure 5: Examples of Smoothing by the Savitsky-Golay algorithm.

- **Detrend correction**

Detrend correction is applied to spectra to eliminate curvilinearity and baseline shifts. It consists of removing from the spectrum its global trend modeled by a polynomial. This polynomial can be of different degrees:

Detrend of order 0: the polynomial removed in this case is a constant which is equal to the average of the spectrum. In the case of the maize spectra figure 6, pre-processing is clearly not sufficient.

Detrend of order 1: by linear regression, the line which fits best to all points of the spectrum are identified and subtracted. In the case of figure 6, the residuals on this line are the absorption peaks corresponding to the chemical compounds in the sample. Applying it to the whole dataset allows observing that the transformation has highlighted the differences in absorption.

Detrends of order 2 or higher: a parabola, cube, etc. are identified and then removed from the spectrum. These polynomial degrees are rarely used.

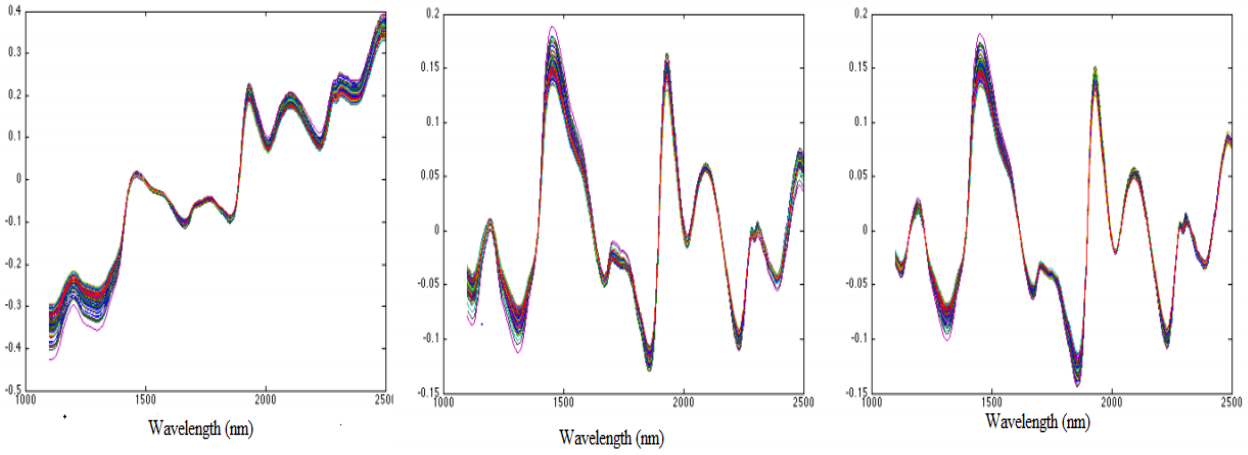

Figure 6: From left to right; shows the detrend effect for orders 0, 1 and 2 on the maize spectra. We can clearly see that order 0 reduces pure vertical translations, then order 1 reduces affine baselines (of the form aλ + b). Finally, we see that order 2 does not bring much[35]**.**

2. Exploratory analysis

- **Principal component analysis (PCA)**

The Principal Component Analysis (PCA) is one of the most widely used methods of multivariate data analysis. It allows the exploration of multidimensional data sets constituted by quantitative variables. It is widely used in biostatistics, chemistry, social sciences and other fields. PCA also named factorial analysis, in the sense that it produces the factors (or main axes) that are derived from linear combinations of initial variables, hierarchical and independent of the others. In some cases, these factors are called variables latent. Not all main components contain the same information; the first ones are those that describe the major variability of data, which is associated with the most relevant information, while the last ones describe variations in the data that can be

due to noise or experimental error, or to an over-fit of the model and can be discarded, thus achieving a significant reduction in the number of variables. The following equation applies:

$$X = \sum_{i=0} t_i * P_i^T + E$$

where X is the data matrix to be treated (in this thesis it will be spectral data after its pre-treatment), a is the number of main components that contain the desired information, t_i is the scores for each main component and P_i is the loadings and E is an error matrix, meaning the residual variation of X that is not explained by the model with the main component. The superscript T indicates the transposed matrix. The equation can be expressed as follows:

$$\mathrm{X} = \mathrm{T} \times P^T + \mathrm{E}$$

The X matrix, is written as the product between the T matrix of scores (or factorial coordinates), and the transposed P matrix of loadings. To this is added the matrix E corresponding to the residual variance.

PCA decomposition is also considered in the specific case of spectral data. A spectrum measured and belonging to the spectral database is thus decomposed component by component as the product of a score (or factorial coordinate) and loading, which looks like a spectrum. This represents useful information. The residual variance matrix recovers the unexplained part of the signal, i.e. the noise.

3. Discriminant analysis

The discriminant analysis can be a predictive method (linear discriminant analysis - LDA) and a descriptive method (discriminant factor analysis - DFA). It aims at explaining and predicting the membership of individuals to groups (classes), represented by a categorical target variable, from a collection of explanatory/descriptive variables, mainly quantitative, but which can be qualitative

through an adjustment. When the differences between two or more given classes are greater, the Mahalanobis distance between them increases. The discrimination of the samples is achieved by calculating the Mahalanobis distance from each one of them to the centers of the groups considered [36]. An unknown sample is classified as belonging to the group with which it has a closer distance to the center [37]. The DA can be considered similar to the PCA only in the sense that both determine a hyperplane with a smaller number of variables in which the sample data are projected from a plane with more variables. However, the PCA selects the direction that retains the maximum structure between the data, while the DA selects the direction in which a maximum separation between the defined groups is achieved [38]. In the construction of this model, it is important to take into account the fact that it requires a larger number of samples than variables.

Among the methods used to conduct this thesis work we found the Linear Discriminant Analysis (LDA) and the support vector machine (SVM).

The LDA method consists of finding linear combinations of the p variables (X matrix), called discriminant variables, allowing to realize representations of the K groups as compact as possible but also as far from each other as possible (the most separable). It should be noted that the separation of groups in the LDA is done using hyperplanes. For this purpose, the total variability of the data (the X matrix) is decomposed into inter-group and intra-group variability. Good discrimination is obtained when there is high inter-group variability and low intra-group variability. One of the main problems in the application of LDA on spectral data is the collinearity of the variables used in the models. For this reason, spectral selection or spectral reduction methods such as PCA have been used to perform discriminative studies by LDA [39].

- **Support vector machine SVM**

SVM is a supervised machine learning method that can be applied to classification issues. SVM uses a tool called "kernel trick" for transforming input data, and based on these transformations, it finds an optimal limit between the possible outputs [40].

To clarify, the data are transformed into a new space, called the kernel, which allows to the model of non-linearity. In calibration, this matrix is of dimension N×N. Practically there are different kernels (linear, polynomial, radial basis function, and sigmoid) the most common one is the radial basis function which requires a parameter to optimize the width of the gaussian (sigma) for linearity adjustment. To avoid overlearning an optimization of the regularization parameter must be done (C or cost). The adjustment of these two parameters is crucial to obtain efficient and robust models.

4. Regression tools

- **Partial least square regression (PLS-R)**

PLS regression (Partial Least Squares regression or Projection to Latent Structures), i.e. regression in the sense of partial least squares, also allows, like multiple linear regression, to link a set of dependent variables Y, to a set of independent variables X, when the number of variables (independent and dependent) is important. PLS regression tool is considered one of the most widespread regression methods in chemometrics. Instead of decomposing initially the X matrix into a set of loadings vectors and scores, and regressing the scores on the Y in a separate step, PLS uses the information from Y during the decomposition process [41], [42].

PLS is therefore based on the simultaneous modeling of the variability of the predictive X matrix and the dependent Y matrix by calculating Latent Variables (LVs) that maximize the variance extracted from the two matrices as well as their correlation. This procedure consists of performing a decomposition (often based on the NIPALS algorithm - Nonlinear Iterative Partial Least Squares)

of the two matrices X and Y under the constraint that the factorial coordinates T extracted from X are as much correlated as possible with the factorial coordinates U extracted from Y [43].

In order to fully understand how PLS regression works, Figure 7 shows that the matrix X is decomposed into a matrix T that represents the score matrix and a loading matrix P' plus an error matrix E. The matrix Y is also decomposed into a score matrix U and a loading matrix R' and the error term F. The objective of the PLS algorithm is to minimize the error while maintaining the strong correlation between X and Y by the internal relation U = B*T [44].

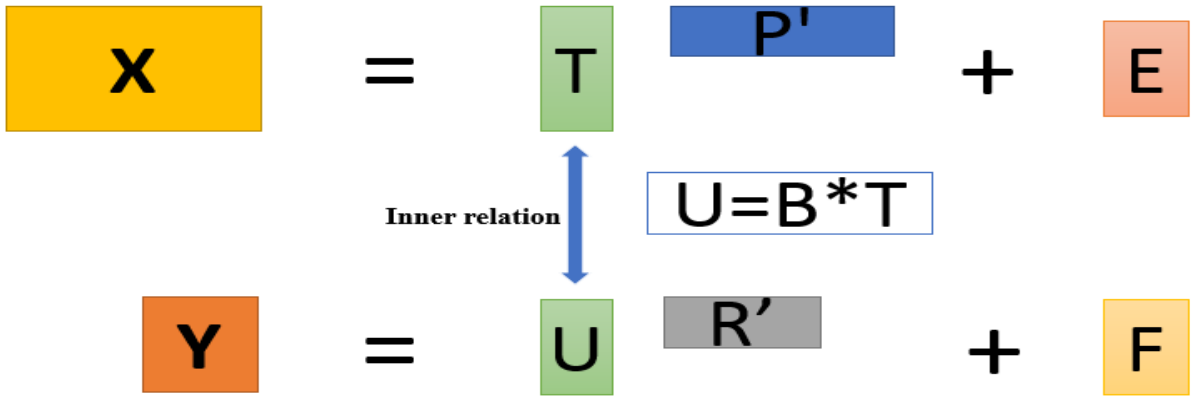

Figure 7: Decomposition of the X and Y matrix for the calculation of PLS factorial coordinates.

For the development of a PLS model, it is important to decide the optimal number of latent variables involved in the PLS model. This optimal number of latent variables can be determined based on the cross-validation using an increasing number of components. The model with the lowest **Pr**edictive **E**rror **S**um of **S**quares (PRESS) or **R**oot **M**ean **S**quared **E**rror of **P**rediction (RMSEP) value and the highest R-square value can be considered as the "best" model [44], [45].

- **Partial least square discriminant analysis PLS-DA**

The partial least squares method was not originally developed for classification and discrimination problems, but it has often been employed for this task [46], [47]. This method, called PLS-DA, is a use of the PLS2 method where the Y matrix is a qualitative variable, this matrix is recoded internally as a block matrix that presents the belonging of each observation, in other words, each response category is coded using an indicator variable. PLS regression (now PLS-DA) is then performed as if Y were a continuous matrix and performs well in practice for large datasets as spectroscopy data where linear discriminant analysis faces problems of collinearity [42].

The first step of PLS-DA consists of constructing the coding of the response variable. This consists of generating a matrix Y of dimension n× K, (where K is the number of groups we want to discriminate) associated with the variable y, and n represents the number of observations.

To clarify how to construct this matrix, an artificial example is given below in figure 8, where n = 10 observations divided into K = 3 classes.

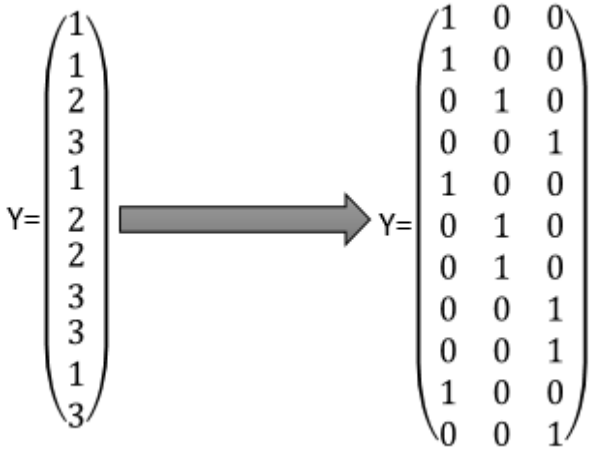

Figure 8: Example of the Y-response coding procedure.

Obviously, it is necessary to carry out cross-validation procedures, to reduce the dimension of the model, the graphical representations are the usual PLS representations. Finally the assignment to the groups is carried out by assigning each observation to the group corresponding to the column of the maximum value of $\hat{Y}$ (predicted value of Y).

VI. References

[1] J. López-Miranda et al., “Olive oil and health: Summary of the II international conference on olive oil and health consensus report, Jaén and Córdoba (Spain) 2008,” Nutr. Metab. Cardiovasc. Dis., vol. 20, no. 4, pp. 284–294, May 2010, doi: 10.1016/j.numecd.2009.12.007.

[2] “Olive Biophenols and Conventional Biotechnology from Mediterranean Aliment Culture,” Olives Olive Oil Health Dis. Prev., pp. 333–340, Jan. 2010, doi: 10.1016/B978-0-12-374420-3.00037-1.

[3] R. Ghanbari, F. Anwar, K. M. Alkharfy, A.-H. Gilani, and N. Saari, “Valuable Nutrients and Functional Bioactives in Different Parts of Olive (Olea europaea L.)—A Review,” Int. J. Mol. Sci., vol. 13, no. 3, p. 3291, 2012, doi: 10.3390/ijms13033291.

[4] International Olive Council, Production techniques in olive growing. Madrid: International Olive Council, 2007.

[5] “olive | Description, Production, & Oil,” Encyclopedia Britannica. https://www.britannica.com/plant/olive-plant (accessed Oct. 14, 2020).

[6] Z. Charrouf and D. Guillaume, “Argan oil: Occurrence, composition and impact on human health,” Eur. J. Lipid Sci. Technol., vol. 110, no. 7, pp. 632–636, Jul. 2008, doi: 10.1002/ejlt.200700220.

[7] R. Vernon, M. Thévenot, P. Bergier, and E. Rousseau, “Argan woodland: an important bird habitat in Morocco,” Afr. Bird Club Bull., vol. 12, pp. 134–146, Jan. 2005.

[8] F. Khallouki, M. Eddouks, A. Mourad, A. Breuer, and R. W. Owen, “Ethnobotanic, Ethnopharmacologic Aspects and New Phytochemical Insights into Moroccan Argan Fruits,” Int. J. Mol. Sci., vol. 18, no. 11, Oct. 2017, doi: 10.3390/ijms18112277.

[9] H. Harhar, S. Gharby, B. Kartah, D. Pioch, D. Guillaume, and Z. Charrouf, “Effect of harvest date of Argania spinosa fruits on Argan oil quality,” Ind. Crops Prod., vol. 56, pp. 156–159, May 2014, doi: 10.1016/j.indcrop.2014.01.046.

[10] N. A. Aabd, A. El Asbahani, Y. El Alem, A. El Finti, F. Msanda, and A. El Mousadik, “Variation in oil content and fatty acid composition in preselected argan trees with morphological characters and geographical localization,” Mediterr. J. Nutr. Metab., vol. 6, no. 3, pp. 217–225, Jan. 2013, doi: 10.3233/s12349-013-0134-2.

[11] Z. Charrouf and D. Guillaume, “Argan oil, the 35-years-of-research product,” Eur. J. Lipid Sci. Technol., vol. 116, no. 10, pp. 1316–1321, 2014, doi: 10.1002/ejlt.201400261.

[12] S. Laaribya, A. Alaoui, and G. Najib, “The Moroccan forest and sustainable development case of the argan tree ((Argania spinosa (L.) Skeels) in Morocco,” vol. 10/2 (S2) (2017) 01–07, pp. 1–7, Feb. 2019.

[13] “Chemistry and Health of Olive Oil Phenolics: Critical Reviews in Food Science and Nutrition: Vol 49, No 3.” https://www.tandfonline.com/doi/abs/10.1080/10408390701856223 (accessed Oct. 31, 2020).

[14] M. Gorzynik-Debicka et al., “Potential Health Benefits of Olive Oil and Plant Polyphenols,” Int. J. Mol. Sci., vol. 19, no. 3, Art. no. 3, Mar. 2018, doi: 10.3390/ijms19030686.

[15] I. Muzzalupo, M. Pellegrino, and E. Perri, “Sensory Analysis of Virgin Olive Oils,” Olive Germplasm - Olive Cultiv. Table Olive Olive Oil Ind. Italy, Dec. 2012, doi: 10.5772/51721.

[16] C. for F. S. and A. Nutrition, “FSMA Rules & Guidance for Industry,” FDA, Oct. 2020, Accessed: Nov. 09, 2020. [Online]. Available: https://www.fda.gov/food/food-safety-modernization-act-fsma/fsma-rules-guidance-industry.

[17] “European Food Safety Authority | Trusted science for safe food,” European Food Safety Authority. https://www.efsa.europa.eu/en (accessed Nov. 09, 2020).

[18] L. Cuadros-Rodríguez, C. Ruiz-Samblás, L. Valverde-Som, E. Pérez-Castaño, and A. González-Casado, “Chromatographic fingerprinting: An innovative approach for food ‘identitation’ and food authentication – A tutorial,” Anal. Chim. Acta, vol. 909, pp. 9–23, Feb. 2016, doi: 10.1016/j.aca.2015.12.042.

[19] S. Medina, J. A. Pereira, P. Silva, R. Perestrelo, and J. S. Câmara, “Food fingerprints – A valuable tool to monitor food authenticity and safety,” Food Chem., vol. 278, pp. 144–162, Apr. 2019, doi: 10.1016/j.foodchem.2018.11.046.

[20] C. P. O’Donnell, Ed., Process analytical technology for the food industry. New York: Springer, 2014.

[21] “Spectroscopic sensors for in-line bioprocess... - Google Scholar.” https://scholar.google.com/scholar_lookup?title=Spectroscopic%20sensors%20for%20in-line%20bioprocess%20monitoring%20in%20research%20and%20pharmaceutical%20industrial%20application&author=J.%20Cla%C3%9Fen&publication_year=2017&pages=1-16 (accessed Nov. 16, 2020).

[22] J. R. Lakowicz, Principles of Fluorescence Spectroscopy, 3rd ed. Springer US, 2006.

[23] “Fluorescence Spectroscopy as a Tool for Ethanol Fermentation On-line Monitoring,” IFAC Proc. Vol., vol. 45, no. 15, pp. 940–945, Jan. 2012, doi: 10.3182/20120710-4-SG-2026.00166.

[24] A. S. Ladokhin, S. Jayasinghe, and S. H. White, “How to Measure and Analyze Tryptophan Fluorescence in Membranes Properly, and Why Bother?,” Anal. Biochem., vol. 285, no. 2, pp. 235–245, Oct. 2000, doi: 10.1006/abio.2000.4773.

[25] A. El Orche, M. Bouatia, and M. Mbarki, “Rapid Analytical Method to Characterize the Freshness of Olive Oils Using Fluorescence Spectroscopy and Chemometric Algorithms,” Journal of Analytical Methods in Chemistry, Jul. 11, 2020. https://www.hindawi.com/journals/jamc/2020/8860161/ (accessed Jul. 14, 2020).

[26] H. Ali, M. Saleem, M. R. Anser, S. Khan, R. Ullah, and M. Bilal, “Validation of Fluorescence Spectroscopy to Detect Adulteration of Edible Oil in Extra Virgin Olive Oil (EVOO) by Applying Chemometrics:,” Appl. Spectrosc., May 2018, doi: 10.1177/0003702818768485.

[27] N. Dupuy, Y. Le Dréau, D. Ollivier, J. Artaud, C. Pinatel, and J. Kister, “Origin of French Virgin Olive Oil Registered Designation of Origins Predicted by Chemometric Analysis of Synchronous Excitation−Emission Fluorescence Spectra,” J. Agric. Food Chem., vol. 53, no. 24, pp. 9361–9368, Nov. 2005, doi: 10.1021/jf051716m.

[28] O. Uncu and B. Ozen, “A comparative study of mid-infrared, UV–Visible and fluorescence spectroscopy in combination with chemometrics for the detection of adulteration of fresh olive oils with old olive oils,” Food Control, vol. 105, pp. 209–218, Nov. 2019, doi: 10.1016/j.foodcont.2019.06.013.

[29] L. Lleó, N. Hernández-Sánchez, F. Ammari, and J.-M. Roger, “3D front-face fluorescence spectroscopy for characterization of extra virgin olive oil and olive oil according to the spectral pattern,” Agric. Eng. Int. CIGR J., vol. 18, no. 4, p. 190, Dec. 2016, Accessed: Feb. 20, 2020. [Online]. Available: https://cigrjournal.org/index.php/Ejounral/article/view/3699.

[30] R. Karoui and C. Blecker, “Fluorescence Spectroscopy Measurement for Quality Assessment of Food Systems—a Review,” Food Bioprocess Technol., vol. 4, no. 3, pp. 364–386, Apr. 2011, doi: 10.1007/s11947-010-0370-0.

[31] A. S. Luna and J. S. de Gois, “Application of Chemometric Methods Coupled With Vibrational Spectroscopy for the Discrimination of Plant Cultivars and to Predict Physicochemical Properties Using R,” in Comprehensive Analytical Chemistry, vol. 80, Elsevier, 2018, pp. 165–194.

[32] I. Gouvinhas, J. M. M. M. de Almeida, T. Carvalho, N. Machado, and A. I. R. N. A. Barros, “Discrimination and characterisation of extra virgin olive oils from three cultivars in different maturation stages using Fourier transform infrared spectroscopy in tandem with chemometrics,” Food Chem., vol. 174, pp. 226–232, May 2015, doi: 10.1016/j.foodchem.2014.11.037.

[33] M. D. Guillén and N. Cabo, “Infrared spectroscopy in the study of edible oils and fats,” J. Sci. Food Agric., vol. 75, no. 1, pp. 1–11, 1997, doi: https://doi.org/10.1002/(SICI)1097-0010(199709)75:1<1::AID-JSFA842>3.0.CO;2-R.

[34] B. Zimmermann and A. Kohler, “Optimizing Savitzky–Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy,” Appl. Spectrosc., vol. 67, no. 8, pp. 892–902, Aug. 2013, doi: 10.1366/12-06723.

[35] J.-M. Roger and M. Ecarnot, “Grain 5 : Pr´etraitements,” p. 17.

[36] “Pattern Recognition,” in Applied Chemometrics for Scientists, John Wiley & Sons, Ltd, 2007, pp. 145–191.

[37] “Practical Guide To Chemometrics | Taylor & Francis Group.” https://www.taylorfrancis.com/books/practical-guide-chemometrics-paul-gemperline/10.1201/9781420018301 (accessed Dec. 26, 2020).

[38] L. A. Berrueta, R. M. Alonso-Salces, and K. Héberger, “Supervised pattern recognition in food analysis,” J. Chromatogr. A, vol. 1158, no. 1–2, pp. 196–214, Jul. 2007, doi: 10.1016/j.chroma.2007.05.024.

[39] I. Stanimirova, M. Daszykowski, and B. Walczak, “Chapter 8 - Robust Methods in Analysis of Multivariate Food Chemistry Data,” in Data Handling in Science and Technology, vol. 28, F. Marini, Ed. Elsevier, 2013, pp. 315–340.

[40] S. Akbarzadeh, A. Paap, S. Ahderom, B. Apopei, and K. Alameh, “Plant discrimination by Support Vector Machine classifier based on spectral reflectance,” Comput. Electron. Agric., vol. 148, pp. 250–258, May 2018, doi: 10.1016/j.compag.2018.03.026.

[41] G. Hanrahan, F. Udeh, and D. G. Patil, “CHEMOMETRICS AND STATISTICS | Multivariate Calibration Techniques,” in Encyclopedia of Analytical Science (Second Edition), P. Worsfold, A. Townshend, and C. Poole, Eds. Oxford: Elsevier, 2005, pp. 27–32.

[42] H. Abdi, “Partial Least Square Regression PLS-Regression,” p. 13.

[43] T. A. Nengsih, F. Bertrand, M. Maumy-Bertrand, and N. Meyer, “Determining the number of components in PLS regression on incomplete data set,” Stat. Appl. Genet. Mol. Biol., vol. 18, no. 6, Dec. 2019, doi: 10.1515/sagmb-2018-0059.

[44] J. Trygg and T. Lundstedt, “Chemometrics Techniques for Metabonomics,” in The Handbook of Metabonomics and Metabolomics, Elsevier, 2007, pp. 171–199.

[45] T. A. Nengsih, F. Bertrand, M. Maumy-Bertrand, and N. Meyer, “Determining the Number of Components in PLS Regression on Incomplete Data,” ArXiv181008104 Stat, Oct. 2018, doi: 10.1515/sagmb-2018-0059.

[46] K.-A. Lê Cao, S. Boitard, and P. Besse, “Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems,” BMC Bioinformatics, vol. 12, no. 1, p. 253, Jun. 2011, doi: 10.1186/1471-2105-12-253.

[47] M. Pérez-Enciso and M. Tenenhaus, “Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach,” Hum. Genet., vol. 112, no. 5, pp. 581–592, May 2003, doi: 10.1007/s00439-003-0921-9.

Part 2: Experimental part, results and discussion

Chapter 1: Geographical traceability of olive oil

Summary of chapter 1

This chapter aims to illustrate how to explore the results of non-targeted fingerprinting technology such as FT-MIR, and UV visible for the determination of geographical origin and classification of olive oil. These results are exploited by chemometric tools in order to associate olive oil to a particular target (geographical origin). An assessment of two spectroscopic techniques and algorithms was performed, taking into account the advantages and inconveniences. These obtained results are mainly based on a number of olive oil samples from the same quality categories (extra virgin or virgin). The olive oil samples are collected from five geographical areas of the Beni Mellal Khenifra region (Boujaad, Fkyh Ben Saleh, Khenifra, Beni Mellal and Qsiba). The spectral profiling of the different olive oils was carried out using mid-infrared spectroscopy and UV-Visible spectroscopy. Chemometric tools (supervised techniques) have been applied to develop classification models regarding the origin of investigated samples. Accurate, robust and reliable models are created for geographic traceability identification and quality control. These results are described in the following document.

Résumé :

Ce chapitre vise à illustrer la manière d'explorer les résultats de la technologie des empreintes digitales non ciblées, telles que la spectroscopie FT-MIR et UV visible, pour la détermination de l'origine géographique et la classification de l'huile d'olive. Ces résultats sont ensuite exploités au moyen d'outils chimiométriques afin d'associer l'huile d'olive à une cible particulière (origine géographique). Une évaluation comprenant deux techniques spectroscopiques et des algorithmes a été réalisée, en tenant compte des avantages et des inconvénients. Les résultats obtenus sont principalement basés sur un certain nombre d'échantillons d'huile d'olive de même catégorie de qualité (vierge extra ou vierge). Les échantillons d'huile d'olive sont prélevés de cinq zones géographiques de la région de Beni Mellal khenifra (Boujaad, Fkyh ben saleh, Khenifra, Beni Mellal et Qsiba). Le profil spectral des différentes huiles d'olive a été réalisé en utilisant la spectroscopie moyenne infrarouge et la spectroscopie UV-Visible. Des outils chimiométriques (techniques supervisées) ont été appliqués pour développer des modèles de classification concernant l'origine des échantillons étudiés. Des modèles précis, robustes et fiables ont été créés pour l'identification de la traçabilité géographique et le contrôle de la qualité. Ces résultats sont décrits dans le document suivant.

Hindawi
Journal of Spectroscopy
Volume 2020, Article ID 9317350, 9 pages
https://doi.org/10.1155/2020/9317350

Research Article

Evaluation of the Capability of Horizontal ATR-FTMIR and UV-Visible Spectroscopy in the Discrimination of Virgin Olive Oils from the Moroccan Region of Beni Mellal-Khenifra

Aimen El Orche,[1] **Mustapha Bouatia**,[2] **Siham Yanisse**,[2] **Houda Labjar**,[3] **Mouad Mouhsin**,[1] **Mohamed Bouha**,[1] **Mustapha Oubenali**,[1] **and Mohamed Mbarki**[1]

[1]*Laboratory of Chemical Processes and Applied Materials, Faculty of Science and Technology, Sultan Moulay Slimane University, Beni-Mellal, Morocco*
[2]*Laboratory of Analytical Chemistry, Faculty of Medicine and Pharmacy, Mohammed V University of Rabat, Rabat, Morocco*
[3]*Mohammedia Faculty of Science and Technology, Hassan II University, Casablanca, Morocco*

Correspondence should be addressed to Aimen El Orche; aimen.elorche@gmail.com

Received 18 January 2020; Revised 14 May 2020; Accepted 5 June 2020; Published 20 June 2020

Academic Editor: Wee Chew

One of the most important challenges in the authentication of olive oil is the determination of the geographical origin of virgin olive oil. In this work, we evaluated the capacity of two spectroscopic techniques, UV-Visible and ATR-FTMIR, coupled with chemometric tools to determine the geographical origin of olive oils. These analytical approaches have been applied to samples that have been collected during the period of olive oil production, in the Moroccan region of Beni Mellal-Khenifra. To develop a rapid analysis tool capable of authenticating the geographical origin of virgin olive oils from five geographical areas of the Moroccan region of Beni Mellal-Khenifra, UV-Visible and ATR-FTMIR spectral data were processed by chemometric algorithms. PCA was applied on the spectral data set to represent the data in a very small space, and then discrimination methods were applied on the principal components synthesized by the PCA. The application of the PCA-LDA method on the spectral data of UV-Visible and ATR-FTMIR shows a good ability to classify olive oils according to their geographical origin with a percentage of correct classification that represents 90.24% and 85.87%, respectively, and the processing of the spectral data of UV-Visible and ATR-FTMIR by PCA-SVM allows differentiating correctly between five olive oils with a correct classification rate of 100% and 97.56, respectively. This study demonstrated the feasibility of UV-Visible and ATR-FTMIR fingerprinting (routine technique) for the geographical classification of olive oils in the Moroccan region of Beni Mellal-Khenifra. Such developed methods can be proposed as alternative and complementary methods to authenticate the geographical origin of virgin olive oil.

I. Introduction

The virgin olive oil is becoming a veritable obsession in the world thanks to its taste qualities and its medicinal and nutritional virtues. Its consumption is increasing every year and has become an essential element in the diet of Mediterranean countries [1], [2]. There are four quality categories of olive oil; extra virgin, virgin, common virgin and lampante olive oil [3], that are determined by physicochemical and sensory analyses as it is described in the guide of the International Olive Council.

Generally the quality of olive oil depends on several factors such as variety, edaphic factors and climatic factors [4], [5]. These last factors play an important role in the chemical composition of olive oils in terms of fatty acids, vitamin E, sterols and polyphenols. For this reason, the determination of geographical origin has now become an important parameter for judging the quality of olive oils, since it is one of the factors causing significant differences in organoleptic properties and chemical composition [4]. So, more and more consumers nowadays are interested in the origin of the food that they consume, in particular olive oils. In order to satisfy this demand, the control authorities have established a food traceability system. In fact, the research of the geographical or varietal origin of olive oils is essential and crucial for the knowledge of their traceability [6], [7]. The traceability must bring answers in terms of identification, localization, authentication and security of olive oils.

Numerous analytical techniques have been developed to determine the geographical origin in some countries, such as High Performance Liquid Chromatography (HPLC), Gas Chromatography (GC) [8], Infrared Spectroscopy (IR) [9], [10], Raman spectroscopy [11], Mass Spectrometry (MS) [12] and Nuclear Magnetic Resonance (NMR) [13], [14].

Several studies have used chromatography as a basic method coupled to chemometric methods (Principal Component Analysis, discriminant factor analysis and other discrimination algorithms). In these cases, to discriminate olive oils coming from various regions, the discrimination may be also based on the composition of fatty acids and triglycerides [15], [16], phenolic compounds [15], [17], [18] and pigments. Principally these basic methods are long and laborious and require the use of more expensive solvents and reagents that can be harmful for the environment and toxic for the analysts. Hence, some techniques known as coupling techniques are emerging and commonly used

like coupled spectroscopic techniques to chemometrics algorithms in order to ensure data processing and discrimination to obtain more precise and complementary information.

Nowadays, Infrared and UV-Visible spectroscopy are widely and frequently used to study and reveal information about the molecular properties of foods [19]–[21]. These two spectroscopic techniques are ideal for rapid, accurate evaluation of raw foods without the use of reagents and solvents [20], [22], [23].These spectroscopic techniques have been used to develop simple and highly effective methods to evaluate the quality parameters of olive oils. They are frequently coupled to multivariate analysis methods for exploitation, classification, discrimination or calibration [20]. To discriminate olive oils after their different geographical regions such a coupling, of IR spectroscopy to chemometrics treatments and sometimes on selected spectral zones, has been applied in previous studies [6], [24]. However, the results remain difficult to be compared because the varieties, the geographical areas, the spectral range and the chemometrics treatments have been different from one study to another.

To the best of our knowledge, it is necessary to carry out studies devoted to the geographical determination, by FTMIR and UV-Visible spectroscopy, of virgin olive oils coming from different provinces of the Moroccan region of Beni Mellal-Khenifra.

The objective of the present work is to develop a rapid spectroscopic methods that could be able to classify virgin olive oils according to their geographical origin. Moreover, our work aimed to evaluate the capability of Mid-Infrared and Visible spectroscopies, when they are coupled to supervised and unsupervised multivariate analysis methods, for the discrimination and classification of virgin olive oils that come from different provinces in the Moroccan Beni Mellal - Khenifra area.

II. Materials and methods

1. Sampling

56 samples of virgin olive oils have been collected at different industrial mills of the Beni Mellal - Khenifra area. The mills are located in different provinces of these area and distributed over the area provinces. The table 1 indicates the geographical origins of the collected samples. The collection is carried on oils that were produced in the November- December 2018 harvest time.

Table 1: Geographical origin of the collected olive oils.

Geographical origin	Number of sample
Boujaad (BJ)	8
Khenifra (KHN)	9
Béni Mellal (BM)	16
Qsiba (QSB)	8
Fkih Ben Saleh (FKB)	15
Total	56

The collected samples were stored at a temperature not exceeding 4°C to avoid alteration of the virgin olive oils. The samples are divided into 41 samples for calibration and 15 samples for analysis.

2. ATR-Mid Infrared Spectroscopy

Analysis of the samples was performed by JASCO 460 plus FTMIR Spectrometer equipped with a horizontal ATR accessory, at a 21°C fixed temperature. Using a micropipette, each sample has been deposited on the crystal surface of the ATR. The spectra were collected between 4000 cm^{-1} to 600 cm^{-1} averaging 130 scans at a resolution of 4 cm^{-1}. For each analysis, the ATR accessory is cleaned using the acetone solution that allows us to dry and clean the ATR accessory.

The spectra have been treated using Spectrum Manager to eliminate the effect of carbon dioxide and then transformed to a JCAMP format.

3. UV-Visible Spectroscopy

The olive oil samples were analyzed by Perkin Elmer UV-Visible spectroscopy at the 350 to 800 nm range. In fact, the analysis of the olive oil samples was carried out, without centrifugation, using a spectrophotometer and a quartz cell of 1 cm optical path then the spectra are saved directly at an Excel table format.

4. Multivariate analysis

In this study different statistical techniques have been applied for the processing and evaluation of spectral data that have been obtained from ATR-FTMIR and UV-Visible spectroscopy. In order to

ensure exploration and representation of the data set, we started processing the results by Principal Component Analysis (PCA) that consists of searching for the directions of greater dispersion to find new synthetic variables and represents the data in a reduced dimensional space. In addition to reducing the dimensionality of the data set this method is often used for data cleaning by identifying outliers. It also serves as an effective tool for the identification of similar groups of individuals that behave similarly concerning the measured variables.

These principal components can also be used in turn for many different applications that are, in our case, support vector machine (SVM) and linear discriminant analysis (LDA).

The SVM method is one of the most commonly used methods for the classification of groups, the objective of this algorithm is to find a hyperplane in N-dimensional space (N - the number of lines) that distinctly classifies the data points. To separate classes of data points, many hyperplanes have been generated and the objective is to find a plane that has the maximum margin that separates well between the classes in the data point space. Maximizing the margin distance provides some strengthening for future data points to be classified with more confidence [25].

The LDA method is also considered among the effective methods used to discriminate groups or classes of individuals. It consists of finding linear combinations of the p variables of a data X matrix. These p variables make it possible to have representations of the K groups that are as compact as possible but also as far away from each other as possible. This separation is provided by hyperplanes so that the total variability is decomposed into inter-group and intra-group variability and the groups presenting high inter-group variability are those that are well separated and well discriminated [26]

The performance parameters of the models built by PCA-SVM and PCA-LDA, such as sensitivity, specificity, correct classification rate (CCR) and accuracy, are used to characterize the classification performance of the analytical method. The best performance of any classification method is to minimize false positives (FP; the number of positive samples that are correctly identified as positive samples) and false negatives (FN; the number of positive samples that are misclassified as negative samples). Evaluation criteria, for a classifier method can be obtained from statistical measures[27].

The statistical processing of the data was carried out using The Unscramble Version X 10.4 software.

III. Results and discussion

The absorption spectra of the samples at the Fourier-transformed mid-infrared are shown in Figure 1. The MIR spectra show intense bands of high intensity up to a 1.7 absorbance due to the ATR accessory. The intensity of the bands gives information on the concentration of functional groups that characterizes olive oil compounds.

According to the observation of the FTMIR spectra, we have observed that different bands characterize the usual functional groups of olive oil at wavelengths 720, 968, 1159, 1377, 1464, 1743, 2852, 2920 and 3004 cm^{-1} that corresponds respectively to the functional groups: CH_2 (CH sp^3); C=C-H trans (CH sp^2); C-OH;CH_3 (CH sp^3); CH_2 (CH sp^3); C=O; CH_2, CH_3 (CH sp^3); CH_2, CH_3 (CH sp^3) et C=C-H cis (CH sp^2) [28].

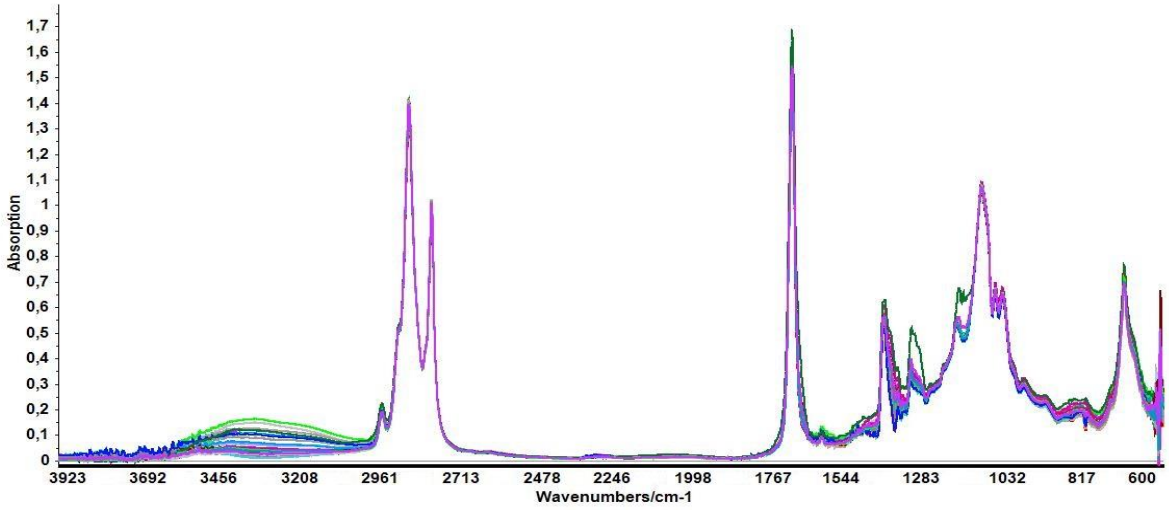

Figure 1: ATR-FT-MIR spectra of the collected olive oil samples.

The visual observation of the ATR-FTMIR spectra of these 41 samples cannot be used to determine the similarities between individuals to differentiate between olive oils coming from different geographical origins.

The spectral absorption of olive oils at the UV-Visible presents spectral bands that correspond to reliable information on the compounds of olive oils, in particular the pigments because they control the coloration of olive oils. From the observation of the UV-Visible spectra (figure 2) it can be seen

that there is an additive effect due to the effect of suspended particles since the olive oils have not been previously treated by centrifugation to eliminate the suspended particles. Light-scattering phenomenon are generated that introduce additive effects on the spectral database. After the mathematical correction of the UV-Visible spectra by base line correction .The spectra are processed by base line correction using the method (Weighted Least Squares). This method is generally used in spectroscopic applications, it iteratively adapts a base line to each spectrum and determines the variables that are clearly above the base line (i.e. the signal) and those that are below the base line. The points under the baseline are supposed to be more significant in adjusting the baseline to the spectrum. This method is also referred to as asymmetric weighted least squares method. The clear effect is the automatic suppression of the background while avoiding the creation of very negative spectral peaks [29].

After the mathematical correction of the UV-Visible spectra by base line correction it can be seen from figure 3 that there is a difference in terms of the pigments. The spectra show an intense absorption between 400 nm and 500 nm. This spectral range corresponds to the absorption of blue light. In addition, absorptions at wavelengths 530, 615 and 670 nm correspond, respectively, to the following pigments: lutein, β-carotene, pheophytin a and pheophytin b, chlorophyll a and chlorophyll b and other pigments [30], [31].

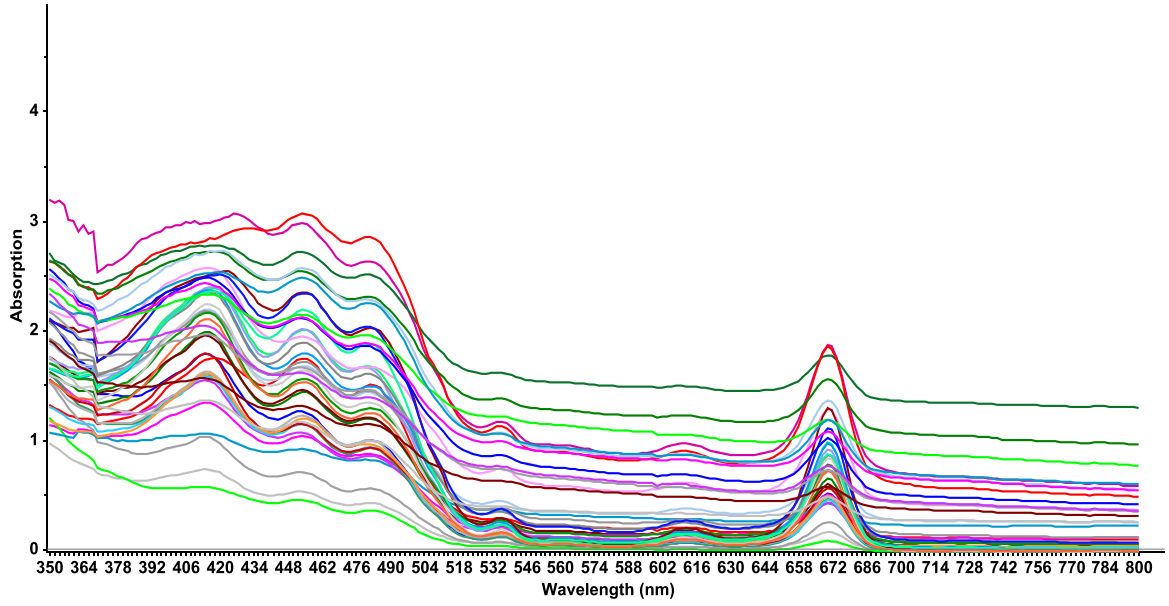

Figure 2: Olive oil UV-Visible spectra at wavelengths between 350-800nm

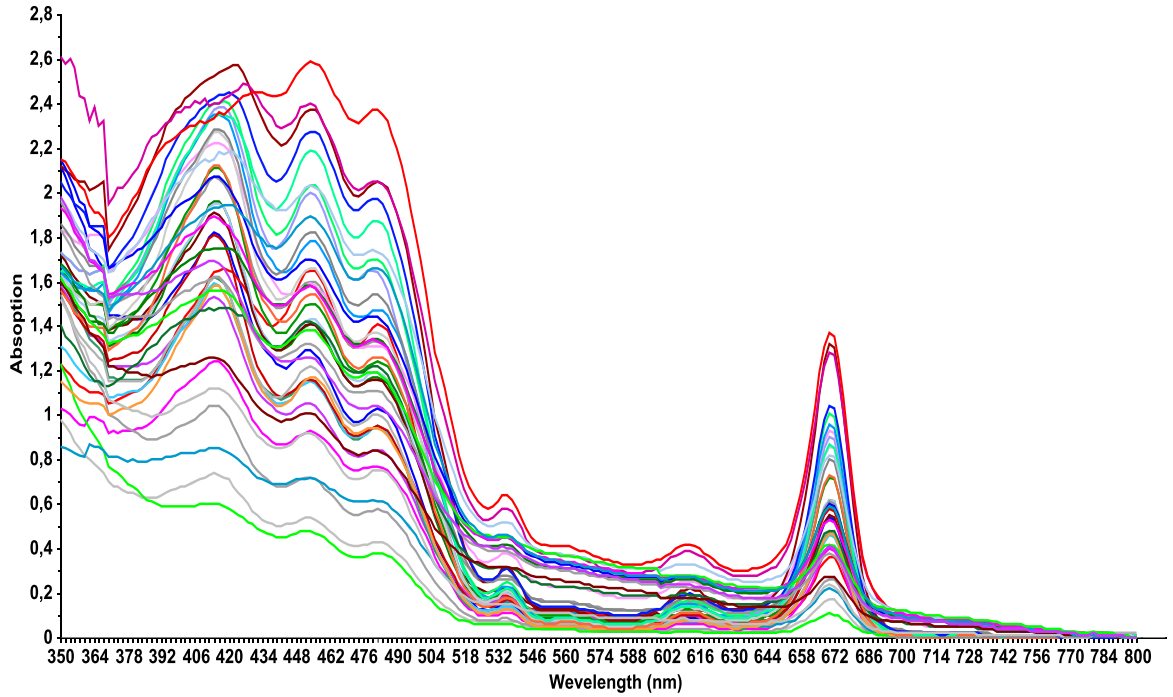

Figure 3: Olive oil UV-Visible spectra corrected by base line correction

1. Principal Component Analysis

A preliminary examination of the spectral data collected was carried out by principal component analysis using the spectral range of 600-4000 cm-1 of FT-MIR, in order to represent the data in a reduced dimensional space.

The PCA results (figure 4) show that the first two main components account for 81% of the total variance of the spectral data, representing 53% and 28% of the total variance of the raw data, respectively. It is clear that there was information on the varietal origin in the MIR spectra of virgin olive oil VOO samples, because the observation of score plot shows that there is a clustering of VOOs according to their geographical origin. The PCA also shows that the five groups of VOOs show small inter-group variability because these oils have the same varietal origin (Moroccan Picholine) and from geographical origins inside the same geographical area. In addition, climatic and edaphic conditions do not show significant differences in the Beni Mellal-Khenifra area.

The first principal component (PC1), representing the total variability, allow to classify between BJD oils and other oils while the second principal component (PC2) allow to classify between BM and FKB oils on the one hand and BM and KHN oils on the other hand.

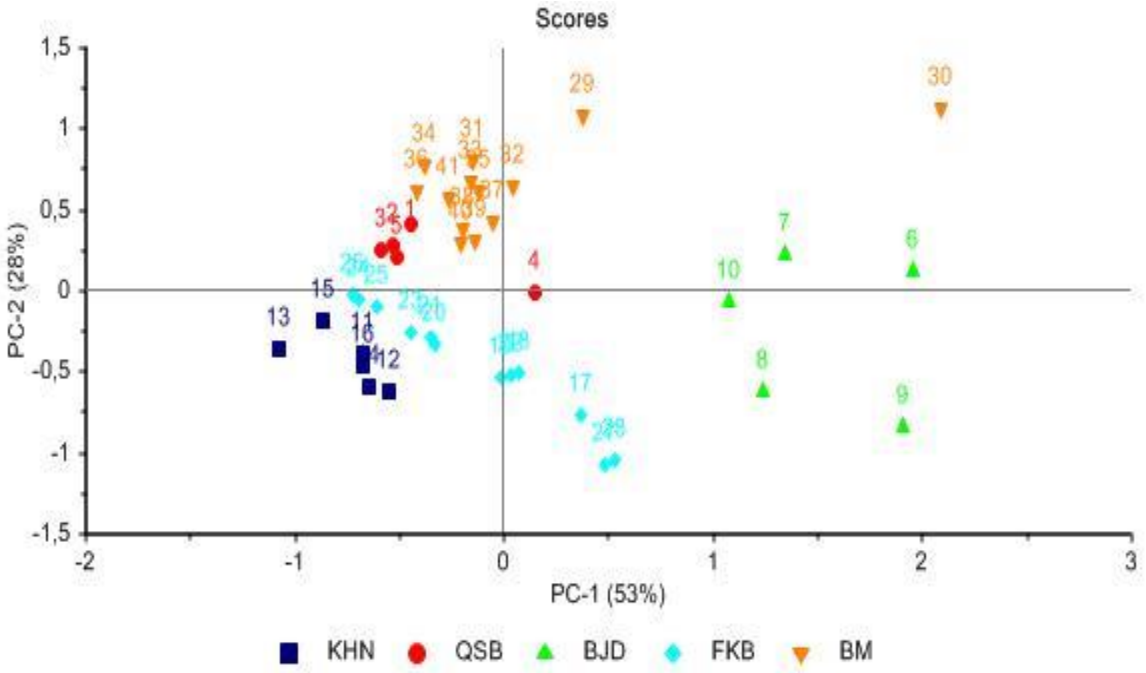

Figure 4: The PCA score plot of ATR-FTMIR spectra at the PC1-PC2 plan

Analysis of the UV-Visible spectra by PCA shows that the total variability of the data is explained by the first two components PC1 and PC2 that represent respectively 86% and 12% shown in figure 5. According to the PC1-PC2 plan of the score plot, there is a clustering of the five groups of olive oils according to their geographical origin. This clustering is mainly carried out by the first main component which contains 86% of the information available in the UV-Visible spectral database.

Mathematical processing of the spectral data by the Baseline correction algorithm can remove the additive effects caused by suspended particles in olive oil samples and improve the clustering between the groups. The PCA of the UV-Visible spectra also shows that olive oils of BM, FKB and BJD have high intra-group variability as shown in figure 6.

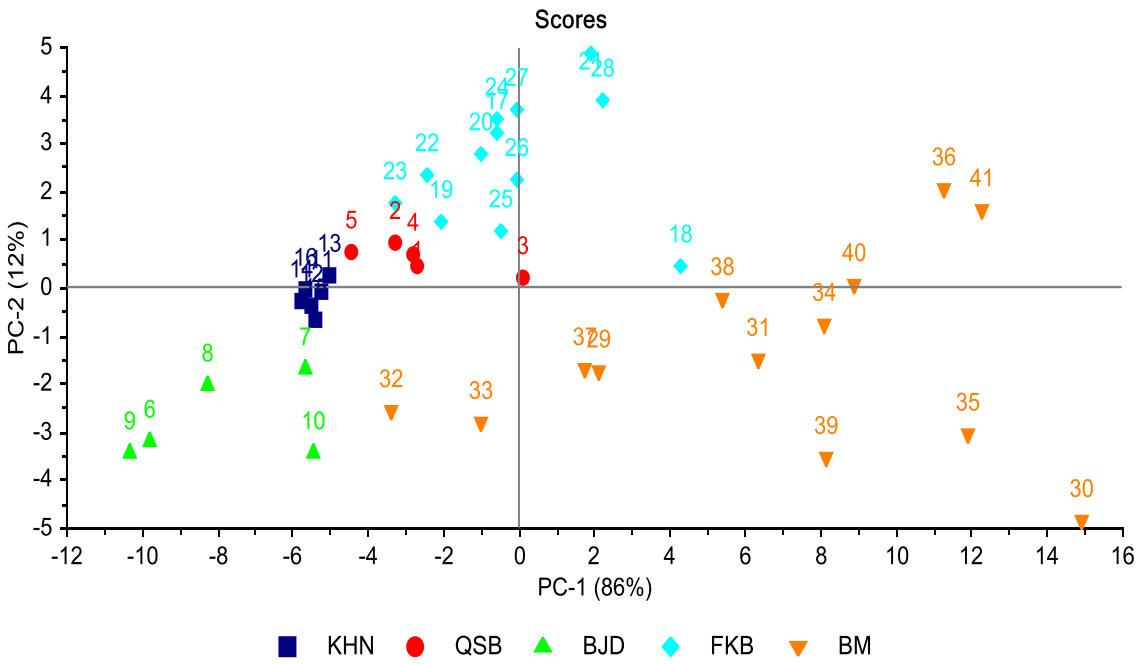

Figure 5: Score plot of UV-Visible spectral data without spectral correction at the PC1-PC2 plan

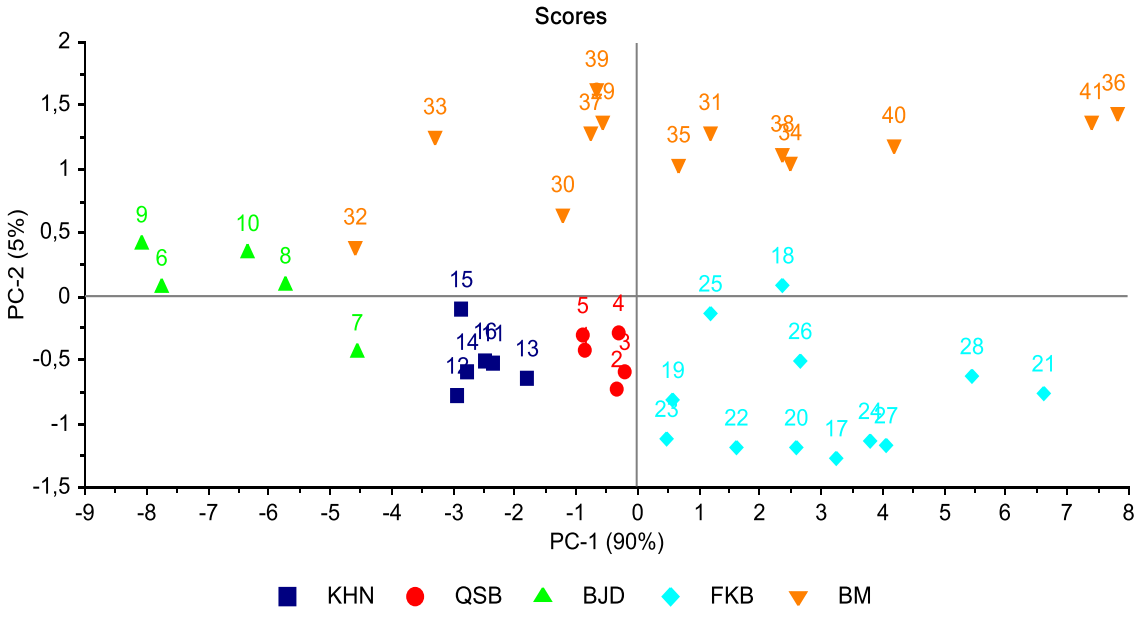

Figure 6: Score plot of UV-Visible spectral data processed with baseline correction at the PC1-PC2 plan.

2. Linear discriminant analysis and support vector machine.

PCA is used to reduce the dimensionality of the data, especially the spectral data because they contain a high number of the variables. Thanks to PCA we have generated independent synthetic variables that can then be used for the application of linear discriminant analysis and support vector machine classification.

To evaluate the discriminatory capability of olive oils that come from the Béni Mellal-Khenifra area using UV-Visible and ATR-FTMIR spectroscopy, the linear discriminant analysis and the Support Vector Machine method were applied to synthetic variables that have been generated by the PCA.

The use of the PCA-LDA method on the spectral data of the UV-Visible and FT-MIR shows a good discrimination capacity of the 5 groups of olive oil according to their geographical origin, this discrimination capacity is represented by the CCR coefficient of the training data which represents 90.24% in the case of the UV-Visible and 85.37% for the results of FT-MIR.

In order to evaluate the predictive performance of these classification models an external validation was performed using external samples (3 samples for each class). This validation procedure shows that 86.67% of the samples were correctly classified using the UV-Visible and FT-MIR spectroscopic techniques.

Then, the sensitivity and specificity of the training and validation datasets were calculated to evaluate the classification performance of these algorithms, and the results are presented in Table 2 and 3.

Table 2: Confusion matrix of PCA-LDA results on the first two principal components of UV-Visible spectral data.

PCA-LDA UV-visible										
	Actual training set						Sensitivity (%)	Specificity (%)	Classification (%) accuracy	CCR (%)
Predicted		1-QSB	2-BJD	3-KHN	4-FKB	5-BM				
	1-QSB	5	0	0	2	0	71	100	100	90.24
	2-BJD	0	4	0	0	1	80	97	80	
	3-KHN	0	1	6	0	0	86	100	100	
	4-FKB	0	0	0	10	0	100	93	83	
	5-BM	0	0	0	0	12	100	96	92	
	Actual validation set						Sensitivity (%)	Specificity (%)	Classification (%) accuracy	CCR (%)
Predicted		1-QSB	2-BJD	3-KHN	4-FKB	5-BM				
	1-QSB	3	0	0	1	0	75	100	100	86.67
	2-BJD	0	2	0	0	0	100	92	67	
	3-KHN	0	1	3	0	0	75	100	100	
	4-FKB	0	0	0	2	0	100	92	67	
	5-BM	0	0	0	0	3	100	100	100	

Table 3: Confusion matrix of PCA-LDA results on the first four principal components of ATR-FT-MIR spectral data.

PCA-LDA FT-MIR										
	Actual training set						Sensitivity (%)	Specificity (%)	Classification (%) accuracy	CCR (%)
Predicted		1-QSB	2-BJD	3-KHN	4-FKB	5-BM				
	1-QSB	5	0	0	1	4	50	100	100	85.67
	2-BJD	0	5	0	0	0	100	100	100	
	3-KHN	0	0	6	1	0	86	100	100	
	4-FKB	0	0	0	10	0	100	93	83	
	5-BM	0	0	0	0	9	100	87	69	
	Actual validation set						Sensitivity (%)	Specificity (%)	Classification (%) accuracy	CCR (%)
Predicted		1-QSB	2-BJD	3-KHN	4-FKB	5-BM				
	1-QSB	3	0	0	1	1	60	100	100	86.67
	2-BJD	0	3	0	0	0	100	100	100	
	3-KHN	0	0	3	0	0	100	100	100	
	4-FKB	0	0	0	2	0	100	92	67	
	5-BM	0	0	0	0	2	100	92	67	

The ATR-FT-MIR and UV-Visible spectral databases were processed by the SVM method using a radial basis function algorithm. The application of this method was carried out on the first synthetic variables by the PCA.

The application of the PCA-SVM method on the two spectroscopic techniques UV-Visible and FT-MIR shows a good classification capacity of the 5 groups of olive oil according to their geographical origins, the percentage of correct classification calculated by the CCR coefficient reaches 100% and 97.56% for the training data using the two spectroscopic techniques UV-Visible and FT-MIR respectively.

The evaluation of the predictive performance of these classification models shows a high predictive capacity of the five groups of virgin olive oil, this classification capacity was represented by the CCR coefficient which reaches 100% and 93.33% using UV-Visible and FT-MIR respectively.

The sensitivity and specificity of the training and validation datasets were calculated to evaluate the classification performance of these algorithms, and the results are presented in Table 4 and 5.

Table 4: Confusion matrix of PCA-SVM results on the first two principal components of UV-Visible spectral data.

PCA-SVM UV-visible										
	Actual training set						Sensitivity (%)	Specificity (%)	Classification (%) accuracy	CCR (%)
Predicted		1-QSB	2-BJD	3-KHN	4-FKB	5-BM				
	1-QSB	5	0	0	0	0	100	100	100	100
	2-BJD	0	5	0	0	0	100	100	100	
	3-KHN	0	0	6	0	0	100	100	100	
	4-FKB	0	0	0	12	0	100	100	100	
	5-BM	0	0	0	0	13	100	100	100	
	Actual validation set						Sensitivity (%)	Specificity (%)	Classification (%) accuracy	CCR (%)
Predicted		1-QSB	2-BJD	3-KHN	4-FKB	5-BM				
	1-QSB	3	0	0	0	0	100	100	100	100
	2-BJD	0	3	0	0	0	100	100	100	
	3-KHN	0	0	3	0	0	100	100	100	
	4-FKB	0	0	0	3	0	100	100	100	
	5-BM	0	0	0	0	3	100	100	100	

Table 5: Confusion matrix of PCA-SVM results on the first four principal components of ATR-FT-MIR spectral data

PCA-SVM FT-MIR										
	Actual training set						Sensitivity (%)	Specificity (%)	Classification (%) accuracy	CCR (%)
Predicted		1-QSB	2-BJD	3-KHN	4-FKB	5-BM				
	1-QSB	4	0	0	0	0	100	97	80	97.56
	2-BJD	0	5	0	0	0	100	100	100	
	3-KHN	0	0	6	0	0	100	100	100	
	4-FKB	0	0	0	12	0	100	100	100	
	5-BM	1	0	0	0	13	93	100	100	
	Actual validation set						Sensitivity (%)	Specificity (%)	Classification (%) accuracy	CCR (%)
Predicted		1-QSB	2-BJD	3-KHN	4-FKB	5-BM				
	1-QSB	2	0	0	0	0	100	92	67	93.33
	2-BJD	0	3	0	0	0	100	100	100	
	3-KHN	0	0	3	0	0	100	100	100	
	4-FKB	0	0	0	3	0	100	100	100	
	5-BM	1	0	0	0	3	75	100	100	

The observation of the statistical parameters (CCR, specificity, sensitivity) of the two classification methods PCA-SVM and the PCA-LDA applied on the two spectroscopic techniques shows an efficient classification of the 5 groups of virgin olive oils. These results also show that the

application of the PCA-SVM method gives better results than the one obtained by the PCA-LDA because the PCA-SVM allows to classify the 5 groups with a high percentage of classification using the training data and the external validation.

These results are compared with two study that constitutes to determine the geographical origin of the Moroccan olive oils the first one in the region of Fes-Meknes in which they found a percentage of correct classification of 100% for the training data and 94.23% for the cross-validation using chromatography coupled with mass spectroscopy [17], which is an efficient, expensive, laborious and time consuming method. The other study demonstrate 100% of correct classification using electronic nose and tongue combination coupled with SVM [32].

This study has an important advantage compared to the other methods because it allows to classify between the 5 groups of olive oils using UV-Visible spectroscopy and TF-MIR which are fast methods and does not require the use of reagents.

IV. Conclusion.

This work demonstrates the capability of UV-Visible and TF-MIR spectroscopy combined with PCA-LDA and PCA-SVM classification techniques for the rapid detection of Moroccan origin olive oils VOOs from the Beni-Mellal Khenifra region. This study was carried out on 41 samples of the Picholine VOO variety taken from five different geographical areas. The application of the two methods on the spectral data of UV-Visible and TF-MIR shows that the PCA-SVM method allows a better classification of the 5 olive oil groups with a higher sensitivity, specificity and correct classification rate. In view of the results obtained by the external validation, this new approach provides more reliable information to predict the geographical origin of Moroccan olive oils. Last but not least, it should be noted that the proposed method is environmentally friendly, fast and easy to use.

For a rapid and reliable process of assessment and authentication of virgin olive oils on the basis of their geographical origin, the development of robust spectral databases is encouraged as far as possible.

References

[1] J. López-Miranda *et al.*, "Olive oil and health: Summary of the II international conference on olive oil and health consensus report, Jaén and Córdoba (Spain) 2008," *Nutr. Metab. Cardiovasc. Dis.*, vol. 20, no. 4, pp. 284–294, May 2010, doi: 10.1016/j.numecd.2009.12.007.

[2] A. Allalout, D. Krichène, K. Methenni, A. Taamalli, D. Daoud, and M. Zarrouk, "BEHAVIOR OF SUPER-INTENSIVE SPANISH AND GREEK OLIVE CULTIVARS GROWN IN NORTHERN TUNISIA: SPANISH AND GREEK OLIVE CULTIVARS," *J. Food Biochem.*, vol. 35, no. 1, pp. 27–43, Feb. 2011, doi: 10.1111/j.1745-4514.2010.00364.x.

[3] P. Vossen, "INTERNATIONAL OLIVE COUNCIL (IOC) and CALIFORNIA TRADE STANDARDS for OLIVE OIL," p. 11.

[4] A. Bakhouche, J. Lozano-Sánchez, R. Beltrán-Debón, J. Joven, A. Segura-Carretero, and A. Fernández-Gutiérrez, "Phenolic characterization and geographical classification of commercial Arbequina extra-virgin olive oils produced in southern Catalonia," *Food Res. Int.*, vol. 50, no. 1, pp. 401–408, Jan. 2013, doi: 10.1016/j.foodres.2012.11.001.

[5] I. Gouvinhas, J. M. M. M. de Almeida, T. Carvalho, N. Machado, and A. I. R. N. A. Barros, "Discrimination and characterisation of extra virgin olive oils from three cultivars in different maturation stages using Fourier transform infrared spectroscopy in tandem with chemometrics," *Food Chem.*, vol. 174, pp. 226–232, May 2015, doi: 10.1016/j.foodchem.2014.11.037.

[6] D. L. García-González, G. Luna, M. T. Morales, and R. Aparicio, "Stepwise geographical traceability of virgin olive oils by chemical profiles using artificial neural network models," *Eur. J. Lipid Sci. Technol.*, vol. 111, no. 10, pp. 1003–1013, 2009, doi: 10.1002/ejlt.200900015.

[7] P. Zunin, R. Boggia, P. Salvadeo, and F. Evangelisti, "Geographical traceability of West Liguria extravirgin olive oils by the analysis of volatile terpenoid hydrocarbons," *J. Chromatogr. A*, vol. 1089, no. 1, pp. 243–249, Sep. 2005, doi: 10.1016/j.chroma.2005.07.005.

[8] G. Gurdeniz, B. Ozen, and F. Tokatli, "Classification of Turkish olive oils with respect to cultivar, geographic origin and harvest year, using fatty acid profile and mid-IR spectroscopy," *Eur. Food Res. Technol.*, vol. 227, no. 4, pp. 1275–1281, Aug. 2008, doi: 10.1007/s00217-008-0845-7.

[9] F. Aouidi *et al.*, "Discrimination of five Tunisian cultivars by Mid InfraRed spectroscopy combined with chemometric analyses of olive Olea europaea leaves," *Food Chem.*, vol. 131, no. 1, pp. 360–366, Mar. 2012, doi: 10.1016/j.foodchem.2011.08.041.

[10] M. Casale, C. Casolino, P. Oliveri, and M. Forina, "The potential of coupling information using three analytical techniques for identifying the geographical origin of Liguria extra virgin olive oil," *Food Chem.*, vol. 118, no. 1, pp. 163–170, Jan. 2010, doi: 10.1016/j.foodchem.2009.04.091.

[11] R. Korifi, Y. Le Dréau, J. Molinet, J. Artaud, and N. Dupuy, "Composition and authentication of virgin olive oil from French PDO regions by chemometric treatment of Raman spectra," *J. Raman Spectrosc.*, vol. 42, no. 7, pp. 1540–1547, Jul. 2011, doi: 10.1002/jrs.2891.

[12] S. Portarena, C. Baldacchini, and E. Brugnoli, “Geographical discrimination of extra-virgin olive oils from the Italian coasts by combining stable isotope data and carotenoid content within a multivariate analysis,” *Food Chem.*, vol. 215, pp. 1–6, Jan. 2017, doi: 10.1016/j.foodchem.2016.07.135.

[13] D. Rongai *et al.*, “1H NMR and Multivariate Analysis for Geographic Characterization of Commercial Extra Virgin Olive Oil: A Possible Correlation with Climate Data,” *Foods*, vol. 6, no. 11, p. 96, Nov. 2017, doi: 10.3390/foods6110096.

[14] G. Vlahov, P. Del Re, and N. Simone, “Determination of Geographical Origin of Olive Oils Using ^{13}C Nuclear Magnetic Resonance Spectroscopy. I − Classification of Olive Oils of the Puglia Region with Denomination of Protected Origin,” *J. Agric. Food Chem.*, vol. 51, no. 19, pp. 5612–5615, Sep. 2003, doi: 10.1021/jf0207124.

[15] H. B. Hlima, R. B. Ayed, K. Ennouri, and S. Smaoui, “Geographical Discrimination of Virgin Olive Oils from the Tunisian Coasts by Combining Fatty Acids and Phenolic Acids Profiles within a Multivariate Analysis,” *J. Oleo Sci.*, vol. 66, no. 9, pp. 963–971, Sep. 2017, doi: 10.5650/jos.ess17080.

[16] D. N. Vera, A. M. Jiménez-Carvelo, L. Cuadros-Rodríguez, I. Ruisánchez, and M. P. Callao, “Authentication of the geographical origin of extra-virgin olive oil of the Arbequina cultivar by chromatographic fingerprinting and chemometrics,” *Talanta*, vol. 203, pp. 194–202, Oct. 2019, doi: 10.1016/j.talanta.2019.05.064.

[17] A. Bajoub, A. Carrasco-Pancorbo, E. A. Ajal, N. Ouazzani, and A. Fernández-Gutiérrez, “Potential of LC–MS phenolic profiling combined with multivariate analysis as an approach for the determination of the geographical origin of north Moroccan virgin olive oils,” *Food Chem.*, vol. 166, pp. 292–300, Jan. 2015, doi: 10.1016/j.foodchem.2014.05.153.

[18] A. Bajoub, E. Hurtado-Fernández, E. A. Ajal, A. Fernández-Gutiérrez, A. Carrasco-Pancorbo, and N. Ouazzani, “Quality and chemical profiles of monovarietal north Moroccan olive oils from ‘Picholine Marocaine’ cultivar: Registration database development and geographical discrimination,” *Food Chem.*, vol. 179, pp. 127–136, Jul. 2015, doi: 10.1016/j.foodchem.2015.01.101.

[19] R. Aroca-Santos, M. Lastra-Mejías, J. C. Cancilla, and J. S. Torrecilla, “Intelligent modelling to monitor the evolution of quality of extra virgin olive oil in simulated distribution conditions,” *Biosyst. Eng.*, vol. 172, pp. 49–56, Aug. 2018, doi: 10.1016/j.biosystemseng.2018.05.007.

[20] V. Lagouri, V. Manti, and T. Gimisis, “Optical Nondestructive UV-Vis-NIR-MIR Spectroscopic Tools and Chemometrics in the Monitoring of Olive Oil Functional Compounds,” in *Olives and Olive Oil as Functional Foods*, F. Shahidi and A. Kiritsakis, Eds. Chichester, UK: John Wiley & Sons, Ltd, 2017, pp. 221–247.

[21] O. Maria Porcu, “The Importance of UV-Vis Spectroscopy: Application in Food Products Characterization,” *Sch. J. Food Nutr.*, vol. 1, no. 3, Jun. 2018, doi: 10.32474/sjfn.2018.01.000111.

[22] C. P. O’Donnell, Ed., *Process analytical technology for the food industry*. New York: Springer, 2014.

[23] L. N. Rosa, A. Coqueiro, P. H. Março, and P. Valderrama, “Thermal rice oil degradation evaluated by UV–Vis-NIR and PARAFAC,” *Food Chem.*, vol. 273, pp. 52–56, Feb. 2019, doi: 10.1016/j.foodchem.2018.03.065.

[24] S. Caetano *et al.*, “Geographical classification of olive oils by the application of CART and SVM to their FT-IR,” *J. Chemom.*, vol. 21, no. 7–9, pp. 324–334, Jul. 2007, doi: 10.1002/cem.1077.

[25] H. P. Bhavsar and M. H. Panchal, “A Review on Support Vector Machine for Data Classification,” 2012.

[26] Q. Mai, “A review of discriminant analysis in high dimensions: Discriminant analysis in high dimensions,” *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 5, no. 3, pp. 190–197, May 2013, doi: 10.1002/wics.1257.

[27] “Confusion Matrix - an overview | ScienceDirect Topics.” https://www.sciencedirect.com/topics/engineering/confusion-matrix (accessed May 08, 2020).

[28] M. Safar, D. Bertrand, P. Robert, M. F. Devaux, and C. Genot, “Characterization of edible oils, butters and margarines by Fourier transform infrared spectroscopy with attenuated total reflectance,” *J. Am. Oil Chem. Soc.*, vol. 71, no. 4, p. 371, Apr. 1994, doi: 10.1007/BF02540516.

[29] “Advanced Preprocessing: Noise, Offset, and Baseline Filtering - Eigenvector Research Documentation Wiki.” http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Noise,_Offset,_and_Baseline_Filtering (accessed May 09, 2020).

[30] J. Ayuso, M. R. Haro, and D. Escolar, “Simulation of the visible spectra for edible virgin oils: potential uses,” *Appl. Spectrosc.*, vol. 58, no. 4, pp. 474–480, Apr. 2004, doi: 10.1366/000370204773580347.

[31] J. Shi *et al.*, “Stimulated Brillouin scattering in combination with visible absorption spectroscopy for authentication of vegetable oils and detection of olive oil adulteration,” *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.*, vol. 206, pp. 320–327, Jan. 2019, doi: 10.1016/j.saa.2018.08.031.

[32] Z. Haddi, H. Alami, N. El Bari et al., “Electronic nose and tongue combination for improved classification of Moroccan virgin olive oil profiles,” Food Research International, vol. 54, no. 2, pp. 1488–1498, 2013.View at: Publisher Site | Google Scholar

Chapter 2: Shelf-life evaluation of olive oil

Summary of chapter 2:

This chapter aims to illustrate how to explore the results of non-targeted fingerprinting fluorescence spectroscopy for the characterization of olive oil freshness and classification of olive oils according to their shelf-life. These results processed by chemometric tools in order to associate olive oil to a particular target (freshness). The assessment of fluorescence spectroscopy techniques through various algorithms such as PLS-DA, SVM and LDA was performed, taking into account the advantages and inconveniences. These obtained results are mainly based on a number of olive oil samples. The olive oil samples are collected from five geographical areas of the Beni Mellal khenifra region (Boujaad, Fkyh ben saleh, Khenifra, Beni Mellal and Qsiba). The spectral profiling of the different olive oils was carried out using fluorescence emission spectroscopy. Chemometric tools (supervised techniques) have been applied to develop classification models regarding the shelf-life of investigated samples. Accurate, robust and reliable classification models are created for the classification of olive oils and quality control. These results are described in the following document.

Résumé :

Ce chapitre vise à illustrer comment explorer les résultats de la spectroscopie de fluorescence à empreinte digitale non ciblée pour la caractérisation de la fraîcheur de l'huile d'olive et la classification des huiles d'olive en fonction de leur durée de conservation. Ces résultats sont traités par des outils chimiométriques afin d'associer l'huile d'olive à une cible particulière (fraîcheur). L'évaluation des techniques de spectroscopie de fluorescence par le biais de divers algorithmes tels que PLS-DA, SVM et LDA a été effectuée tout en tenant compte des avantages et des inconvénients. Les résultats obtenus sont principalement basés sur un certain nombre d'échantillons d'huile d'olive. Les échantillons d'huile d'olive sont recueillis auprès de cinq zones géographiques de la région de Beni Mellal khenifra (Boujaad, Fkyh ben saleh, Khenifra, Beni Mellal et Qsiba). Le profil spectral des diverses huiles d'olive a été établi par spectroscopie d'émission de fluorescence. Des outils chimiométriques (techniques supervisées) ont été appliqués pour développer des modèles de classification concernant la durée de conservation des échantillons étudiés. Des modèles de classification précis, robustes et fiables ont été créés pour la classification des huiles d'olive et le contrôle de la qualité. Ces résultats sont décrits dans le document suivant.

Hindawi
Journal of Analytical Methods in Chemistry
Volume 2020, Article ID 8860161, 9 pages
https://doi.org/10.1155/2020/8860161

Research Article

Rapid Analytical Method to Characterize the Freshness of Olive Oils Using Fluorescence Spectroscopy and Chemometric Algorithms

Aimen El Orche,[1] **Mustapha Bouatia**,[2] **and Mohamed Mbarki**[1]

[1]*Laboratory of Chemical Processes and Applied Materials, Faculty of Science and Technology, Sultan Moulay Slimane University, Beni-Mellal, Morocco*
[2]*Laboratory of Analytical Chemistry & Bromatology, Faculty of Medicine and Pharmacy, Mohammed V University in Rabat, Rabat, Morocco*

Correspondence should be addressed to Aimen El Orche; aimen.elorche@gmail.com

Received 11 April 2020; Revised 24 May 2020; Accepted 23 June 2020; Published 11 July 2020

Academic Editor: Eduardo Dellacassa

One of the most important issues in the field of quality assurance of olive oils is the detection of the freshness of olive oil. In this study, 400 nm laser-induced fluorescence spectroscopy was used with supervised and unsupervised multivariate analysis methods to develop a rapid method able to discriminate between freshly produced olive oils and oil that has been stored for a period of time ranging from 12 to 24 months. The fluorescence spectral data were firstly processed by the PCA. This method shows strong discrimination of the three oil classes using the first three components which present 96% of the total variability of the initial data, and then supervised classification models were constructed using the discriminant partial least square regression PLS-DA, support vector machine SVM, and linear discriminant analysis LDA. These methods show a high capacity in the classification of the three classes of olive oil. The validation of these classification models by external samples shows a high capacity of classification of the samples in their class with an accuracy of 100%. This study demonstrated the feasibility of the fluorescence spectroscopy fingerprint (routine technique) for the classification of olive oils according to their freshness and storage time.

I. Introduction

Olive oil is an important vegetable oil in the Mediterranean countries, currently this nutrient is attracting the attention of many consumers around the world thanks to its nutraceutical, sensory properties and for contributing to the protection of the human well-being [1]. These proprieties are especially related to its composition rich in fatty acids especially oleic and linoleic acid [2] , and its high level of minor compounds that have bioactive characteristics, principally phenolic compounds and tocopherol [3], [4].

These natural biochemical compounds of virgin olive oil, are able to delay the effects of oxidation by deactivating the singular oxygen [5], [6]. The greenish coloration of olive oil is attributed to the chlorophyll pigments formed essentially from chlorophylls and their derivatives product [7]. The quantification of these compounds in olive oils is considered to be very important for determining the quality of olive oil because the decrease in chlorophyll levels during storage indicates the presence of oxidation processes that affect the quality of olive oil [4], [8]. Its concentration in olive oil depends on several factors such as geography, edaphic factors, climate, storage conditions, ripening stage and type of extraction [9]–[12]

Moreover, these compounds are significantly decreased during the storage of olive oil, although new products appear due to the oxidation process [7], [13]. In many markets, the storage of olive oil can vary between 6 and 24 months so that it causes an alteration in the quality of olive oil.

Nowadays, the authentication of olive oils is still a major problem. Virgin olive oil, due to its high price compared to other edible oils, can be the object of more or less sophisticated fraudulent practices. The most common ones consist of adulterating virgin olives oils with lower-priced oils (seed oils, refined olive oil, or olive pomace oil). These practices have been the subject of numerous studies aimed at combating fraud that disrupts the market and damages the importance of virgin olive oil (VOO) [14]–[16]. There is also another type of fraud that consists of falsifying the freshness of olive oil and presents to the consumer non-fresh olive oils, that have been stored for a period of time, as freshly produced.

Authentication of the VOO belonging to a designation of origin often constitutes a real analytical challenge. For this reason, a great deal of researchers has been devoted to answering this authentication problem, in order to develop robust and reliable analytical tools able to retrieve all the information on the quality, safety and the origin of olive oil and other oil[17]. These analytical tools can be classified in two main categories, those based on the analysis of chemical compounds

of olive oils, Gas Chromatography GC [18], [19], High Performance Liquid Chromatography HPLC [19]–[21], and those based on spectroscopic techniques, such as Infrared spectroscopy IR [22]–[24], Ultraviolet-Visible spectroscopy UV-Visible [24], [25], Magnetic Nuclear Resonance MNR [26] and Fluorescence spectroscopy [27], [28]. Which have been used for adulteration detection, origin geographic determination, variety determination and the examination of the oxidative stability of olive oils.

The HPLC and GC, as reference methods, are generally time-consuming, sometimes require the use of expensive and polluting reagents, and are only performed by qualified operators. Moreover, these methods are not sufficiently efficient to cover the growing demand for an analytical procedure that requires several hours. The use of spectroscopic methods, such as fluorescence combined with chemometrics tools, make possible the realization of these evaluations in a few time without using reagents.

Fluorescence spectroscopy is a specific, nondestructive and rapid analytical tool for food authentication study [29]. It provides information on the presence of fluorescent molecules and the fluorescence properties of fluorophores. Recently, the application of fluorescence spectroscopy in combination with chemometric tools to evaluate the quality of olive oil has been increased in the majority of research papers [30], because the obtained fluorescence signal corresponds to specific fluorophores such as vitamin E and chlorophyll [31], after having defined the excitation or emission wavelength [32].

This analytical method is combined usually with chemometric approaches using multivariate data processing to extract information from spectroscopic data. Chemometric methods can be supervised or unsupervised. The applications of fluorescence spectroscopy coupled to multivariate analysis with more or less complex preprocessing and sometimes with different excitation and emission wavelengths have been developed by several authors. However, the obtained results in different studies are difficult to compare since the performance criteria and reference value ranges are different.

The present study aims to develop a rapid method based on fluorescence spectroscopy coupled to supervised and unsupervised chemometric algorithms to determine the membership of virgin olive oil in a group of olive oils. The first aim of the work is to know if these olive oils are freshly produced or are stored for a period of time, since the storage of olive oil during period leads to the

loss in the quality of olive oils. The second aim is to evaluate the effectiveness of the chemometric classification tools that we have used for the determination and prediction of the olive oil category.

II. Materials and methods.

1. Sampling

This study was carried out on 81 samples of monovarietal (*Picholine*) virgin olive oil from Morocco. These oils were stored in the dark at a temperature range of 10 ± 1°C. To preserve the molecular qualities of olive oils for a shelf life of 0 and 24 months, as shown in Table 1. During the storage period, the olive oil did not undergo any freezing.

To carry out this study 63 samples were used for calibration and 18 for external validation of the models built.

Table 1: Storage conditions for virgin olive oil.

Number of Samples	Origin	Type of Mills	Variety	Light condition	Temperature condition	Storage Time (month)
3	Beni Mellal province	Traditional mill	Picholine	Darkness	10 ± 1	0
4		Modern mill	Picholine	Darkness	10 ± 1	0
7	Khenifra province	Modern mill	Picholine	Darkness	10 ± 1	0
2	Khouribga province	Traditional mill	Picholine	Darkness	10 ± 1	0
5		Modern mill	Picholine	Darkness	10 ± 1	0
6	Fquih Ben Salah province	Modern mill	Picholine	Darkness	10 ± 1	0
3	Beni Mellal province	Traditional mill	Picholine	Darkness	10 ± 1	12
4		Modern mill	Picholine	Darkness	10 ± 1	12
7	Khenifra province	Modern mill	Picholine	Darkness	10 ± 1	12
2	Khouribga province	Traditional mill	Picholine	Darkness	10 ± 1	12
5		Modern mill	Picholine	Darkness	10 ± 1	12
6	Fquih Ben Salah province	Modern mill	Picholine	Darkness	10 ± 1	12
3	Beni Mellal province	Traditional mill	Picholine	Darkness	10 ± 1	24
4		Modern mill	Picholine	Darkness	10 ± 1	24
7	Khenifra province	Modern mill	Picholine	Darkness	10 ± 1	24
2	Khouribga province	Traditional mill	Picholine	Darkness	10 ± 1	24
5		Modern mill	Picholine	Darkness	10 ± 1	24
6	Fquih Ben Salah province	Modern mill	Picholine	Darkness	10 ± 1	24

2. Spectral fluorescence acquisition.

The fresh and stored virgin olive oils are directly analyzed by fluorescence spectroscopy, using the FluoroMax-4 (Jobin Yvon) spectrophotometer. These fluorescence measurements are carried out using fluorescence cuvette with Polytetrafluoroethylene (PTFE) cover, UV quartz with light path of 10 mm.

The acquisition of emission spectra of olive oil has been made at an excitation wavelength of 400 nm and emission wavelength ranged from 415nm to 785nm with a step of 0.5nm. Some fluorescent molecules of the olive oil have been excited following the absorption of photons at this wavelength which allows them to enter into an electronically excited state, these molecules will return to their fundamental state by emitting photons with a wavelength greater than the excitation wavelength.

3. Multivariate data analysis.

Multivariate data analysis is a group of statistical methods that focus on the simultaneous observation, exploitation and processing of several statistical variables in order to extract relevant synthetic information. These chemometric tools are generally divided into two groups, unsupervised methods such as PCA and supervised methods such as PLS-DA, LDA, SVM. Generally these supervised methods are part of the discriminant analysis that consists in determining the belonging of an individual to a predefined group according to the observation of predictive qualitative variables. These discriminant analysis can provide additional details to the obtained results, such as the identification of the variables that led to the creation of the typology groups. The visualization of the results of this analysis can take the form of a mapping similar to the PCA score plot, where the different individuals are grouped together according to their group affiliation.

Principal Component Analysis (PCA) is an extremely powerful unsupervised method of synthesizing information, very useful when there is a large amount of quantitative data to be processed and interpreted. As a basic tool in chemometrics, PCA serves different purposes; exploration and description of a dataset, preparation and cleaning of data, identification of individual groups and preliminary step for another chemometric treatment LDA, SVM.

The supervised partial least squares discriminant (PLS-DA) method, is a use of the PLS2 regression method, where the response variable is a categorical variable expressing the membership class of the units. This response is coded to contain only two whole numbers. In general, 0 and 1 are used to indicate "outside the group" and "within the group" respectively [33]. The components of this

method are constructed by trying to find an adequate compromise between two main purposes: to describe the whole set of explanatory variables and to predict the response variables [34].

Linear discriminant analysis (LDA)is one of the most important methods of discrimination, it consists in finding linear combinations of the discriminating variables, making it possible to discriminate the most compact and distant groups by using hyperplans. In the case of spectral data this method was often preceded by selections of variables because the model produced is often difficult to interpret in the absence of initial variable selections and the results are unstable in the case of correlations between variables, as it is always the case with spectral data [35].

Support Vector Machine (SVM) is a method that belongs to the family of automatic learning algorithms that solve both classification and regression problems. However, it is commonly applied in classification objectives. It consists of finding n-hyperplane with the maximum margin distance between the points through the use of techniques called kernel trick. The most used algorithms are Linear Kernel, Polynomial Kernel, Radial Basis Function Kernel, and Sigmoid Kernel [36].

4. Software

All data processing of fluorescence spectra and applications of chemometric methods, principal component analysis and partial least squares discriminant analysis, machine vector support and linear discriminant analysis have been realized thanks to The Unscrambler software, version 10.4 camo analytic.

III. Results and discussion

Figure 1 shows the emission spectra of fresh olive oils produced and of stored olive oils as shown in table 1, these spectra present differences in the spectral intensity of the band corresponding to the maximum emission at 675 nm. This band corresponds notably to the emission of some fluorescent molecules in olive oil; chlorophyll and pheophytin [30], these molecules are responsible for the green coloration of the olive oil and represent an important parameter of olive oils quality.

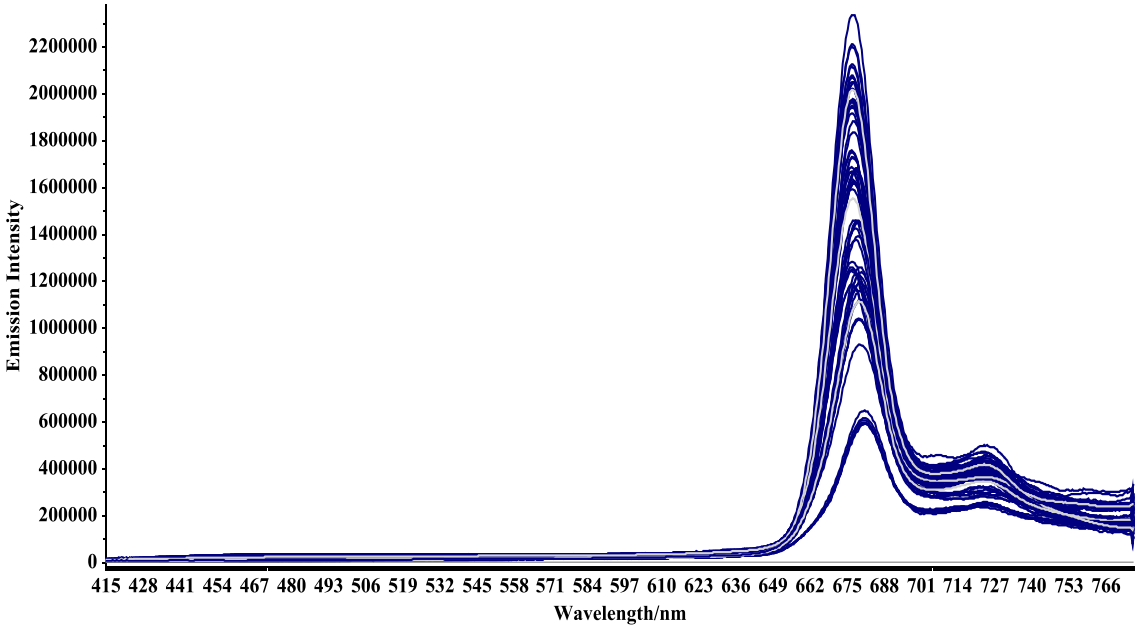

Figure 1: Fluorescence emission spectrum of fresh virgin olive oil and stored olive oil.

The fluorescence spectra show that there is a decrease of the spectral intensity during the storage time due to the degradation of chlorophyll [37]. In fact, the spectra also show that the behavior of these oils is varying because the samples belong to different origins as geographical areas and mills. Consequently, different contents of chlorophyll pigments [7]. The average spectrum representation of each group of oil allows representing the behavior of the oils during the storage time as shown in figure 2.

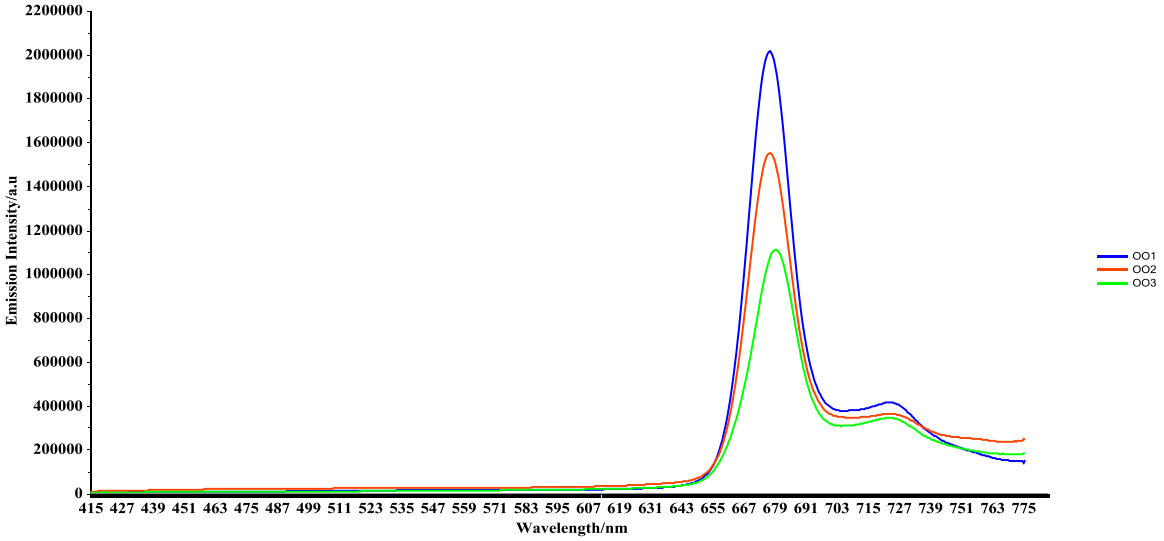

Figure 2: Average emission spectra of each group of olive oil (OO1=fresh olive oil, OO2= stored olive oil during 12 months and OO3= stored olive oil during 24 months).

- **Principal component analysis.**

To describe the data in a very small dimensional space, a PCA has been firstly performed on the 81 spectra of olive oils, to exploit the dataset and getting pieces information on the distribution and the behavior of the samples concerning the measured variables that represent the wavelengths of the fluorescence spectral data. The following figure 3 illustrates the PCA 3D score plot.

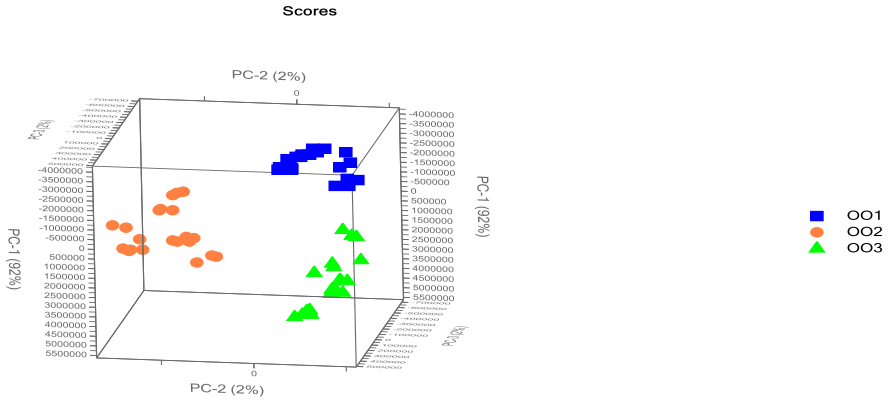

Figure 3: PCA 3D score plot of the first tree principal components PC1-PC2-PC3, (OO1=Fresh olive oil, OO2= stored olive oil during 12 months, OO3= stored olive oil during 24 months).

PCA shows that the first three principal components explain 96% of the total variability in the data. 92% for the first component and 4% for others components. Moreover, PCA shows that there is discrimination between the three groups of oil according to storage time, it also shows that there is intra-group variability for each group. This classification is ensured essentially by the first component which represents the majority of the spectral information. The study of the loading figure 4 associated with the first PC shows that all weights are negative, which is characteristic of chemical or biochemical effects on the spectra, and not of physical characterization of the spectra. This remark allows us to show that the first axis represents chlorophyll pigment content.

The separation tendency of olive oils according to the storage time was evident on the 3D-score plot PC1-PC2-PC3, which demonstrated the capability to use PCA on fluorescence data to identify the freshness and the storage time of the virgin olive oils.

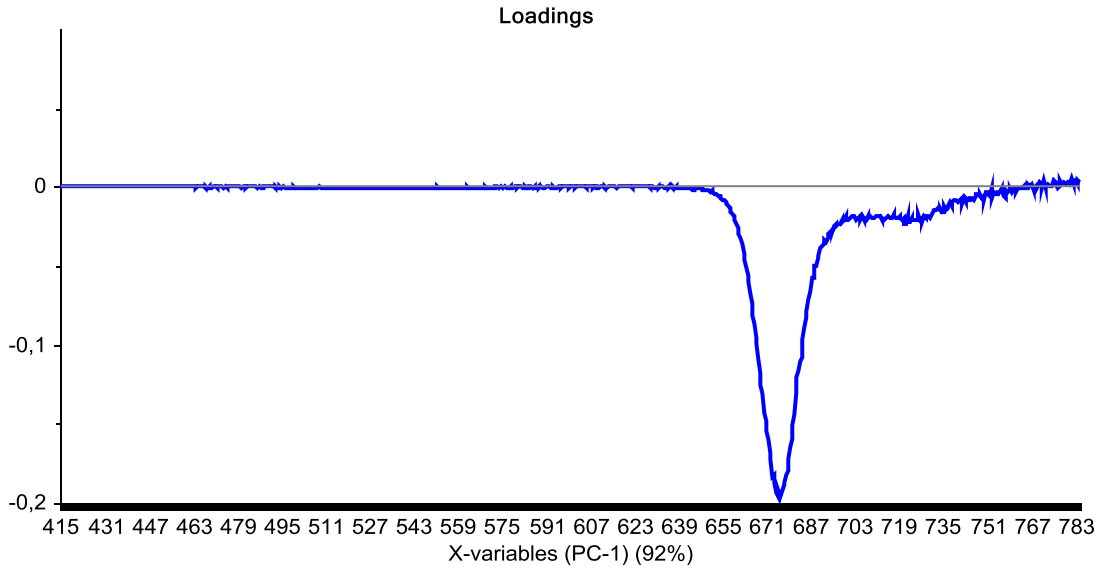

Figure 4: PCA loading plot of the first principal component PC1.

- **Partial least squares discriminant analysis PLS-DA.**

In order to develop a supervised classification method capable of classifying and authenticating virgin olive oils according to their shelf life, the PLS-DA discrimination model has been developed for the three olive oil groups on 63 calibration samples using NIPALS algorithm. The performance of the constructed models was evaluated using the root mean square error of calibration (RMSEC), the root mean square error of cross-validation (RMSECV), the root mean square error of prediction (RMSEP) obtained by external validation and the slope of the regression R^2.

The application of the discriminant PLS shows a high capability in the discrimination of the three groups of olive oils as shown in the score plot figure 5.

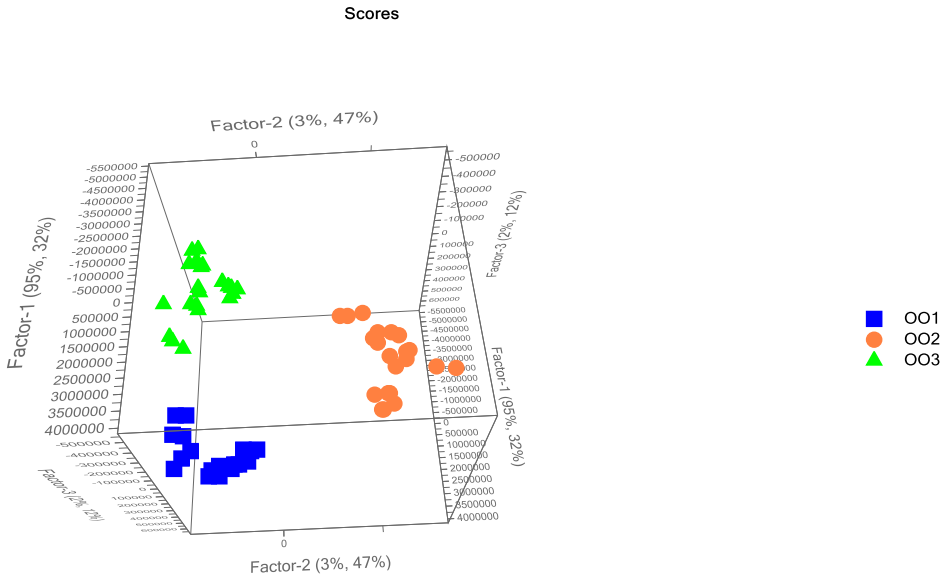

Figure 5: 3D-PLS-DA score plots for the Fluorescence spectra of olive oils groups, (OO1=Fresh olive oil, OO2= stored olive oil during (12 months), OO3= stored olive oil during 24 months).

The discrimination quality of the constructed model is summarized in the following table 2, The performance evaluation of the built models shows that the correlation coefficient ranges between 94% and 89% in the case of the calibration results, and between 94% and 86% in the case of cross-validation results, while the mean square error of the calibration ranges between 0.11 and 0.16 and for the cross-validation ranges between 0.12 and 0.18.

Table 2: Statistical parameters of the built models with and without data preprocessing (PLS-DA).

Label	Preprocessing	Number of latent variable	Calibration		Cross Validation	
			R-square (%)	RMSEC	R-square (%)	RMSECV
OO1	Without preprocessing	3 LV	91	0.14	89	0.16
OO2			94	0.12	93	0.13
OO3			89	0.16	89	0.17
OO1	Smoothing (Savitzky and Golay)	3 LV	90	0.15	89	0.16
OO2			94	0.12	94	0.12
OO3			89	0.16	87	0.17
OO1	Detrend (polynomial 1)	3 LV	91	0.15	86	0.18
OO2			95	0.11	90	0.15
OO3			89	0.16	87	0.17

The predictive performance of the constructed calibration models have been evaluated by external validation using external samples, (6 samples of each class). The predicted y-value of a new sample near to 1 (or greater than 0.5) allocates the sample to a specific category, while a sample with a predicted y-value less than 0.5 is allocated outside the category [33].

The results of external samples prediction by the constructed models mentioned in the following tables 3 show that these samples have been clearly assigned to their respective classes with perfect accuracy of 100%.

Table 3: External validation of the classification of PLS-DA models for the fluorescence spectra of the three categories of olive oil.

Confusion matrix	label	OO1	OO2	OO3	Accuracy of External validation
Predicted external set	OO1	6	0	0	100%
	OO2	0	6	0	
	OO3	0	0	6	

- **Support vector machine classification SVM.**

SVM (type C-SVC) has been applied on the fluorescence spectral data of the three groups of olive oils, using a linear Kernel algorithm. The reported results in table 4, show that the model has provided a good classification performance for the three classes of oils according to their

membership (freshness and storage time). The calibration model has been validated using firstly cross-validation that shows a significant accuracy of classification that reached 100%. Finally we used an external validation by a new set of samples (6 samples of each class) to evaluate the predictive performance of the constructed model. The 18 samples of the sample set have clearly been attributed to their respective classes with perfect accuracy of 100%. The results that we have obtained by the SVM model confirm the predictive capability to classify the different classes of samples according to their freshness and storage time.

Table 4: Confusion matrix for the classification of training and external dataset using the SVM method.

Confusion matrix	Actual				accuracy	
	label	OO1	OO2	OO3	Calibration	Cross-validation
Predicted training set	OO1	21	0	0	100%	100%
	OO2	0	21	0		
	OO3	0	0	21		
		OO1	OO2	OO3	External validation	
Predicted external set	OO1	6	0	0	100%	
	OO2	0	6	0		
	OO3	0	0	6		

- **Linear discriminant analysis LDA.**

The supervised discrimination method was also used, LDA has been applied on the three synthetic variables generated by the PCA. This method is not applicable on the data where the variables have a co-linearity among themselves, for this reason it is necessary to combine this method with methods of variable selection like the PCA method, because the PCA allows to generate independent synthetic variables from the initial variables. The application of the LDA method on the first three components of the PCA shows a very high capacity of discrimination between the three classes of olive oil as shown in table 5. This classification model provides a high discrimination performance of the three classes according to their membership. The results of calibration and cross-validation show that this model can correctly classify the three classes with an accuracy that reaches 100%.

The predictive assessment of this model through external validation by a new set of samples (6 samples of each category). The 18 samples of the test set are clearly assigned to their respective categories ensuring a perfect accuracy of 100% as it is reported in table 5.

Table 5: Confusion matrix for the classification of training and external dataset using the PCA-LDA method.

Confusion matrix	Actual				accuracy	
	label	OO1	OO2	OO3	Calibration	Cross-validation
Predicted training set	OO1	21	0	0	100%	100%
	OO2	0	21	0		
	OO3	0	0	21		
		OO1	OO2	OO3	External validation	
Predicted external set	OO1	6	0	0	100%	
	OO2	0	6	0		
	OO3	0	0	6		

It is clear that the ideal situation occurs when all VOO samples arrive at the diagonal cells of the matrix. That is to say, each olive oil class was correctly classified by the SVM, PLS-DA and ACP-LDA models, which led to a 100% success rate in the classification of the three Moroccan oil groups according to their freshness. This success rate was also higher than that of Sinelli et al [38], who found 87% by combining physicochemical data (Acidity (%), Peroxide Value, K_{232} and K_{270}) with linear discriminant analysis and 98% by using Mid-Infrared spectroscopy.

The improvement of this method with a wide range of olive oils by the introduction of several varieties of olive oil of different freshness allows to increase the analytical performance of this method and to use it as a routine method for the authentication of the freshness of olive oils in analytical laboratories. Such a process allows many control authorities to check the freshness of olive oils on the market in order to protect the consumer against fraudulent actions.

IV. Conclusion.

The present study shows the capability of fluorescence spectroscopy coupled to supervised and unsupervised methods for the classification and the prediction of freshly produced virgin olive oils and virgin olive oils that have been stored during a time.

The obtained results by PCA as an unsupervised method of exploitation and grouping of individuals show that there is discrimination between the three groups of olive oils concerning the variables measured by fluorescence spectroscopy.

The application of the supervised classification methods PLS-DA, SVM and LDA, shows a very high capacity in the discrimination between these three categories of oil. They also show a very accurate capacity for the prediction and correct classification of external samples in its class.

For a reliable process of rapid evaluation and authentication of virgin olive oils in the market to identify the freshness of olive oils, the development of robust spectral databases is encouraged as much as possible.

References.

[1] M. N. Criado, M. J. Motilva, M. Goñi, and M. P. Romero, "Comparative study of the effect of the maturation process of the olive fruit on the chlorophyll and carotenoid fractions of drupes and virgin oils from Arbequina and Farga cultivars," *Food Chem.*, vol. 100, no. 2, pp. 748–755, Jan. 2007, doi: 10.1016/j.foodchem.2005.10.035.

[2] "Monitoring of fatty acid composition in virgin olive oil by Fourier transformed infrared spectroscopy coupled with partial least squares," *Food Chem.*, vol. 114, no. 4, pp. 1549–1554, Jun. 2009, doi: 10.1016/j.foodchem.2008.11.029.

[3] D. Giuffrida, F. Salvo, A. Salvo, L. Cossignani, and G. Dugo, "Pigments profile in monovarietal virgin olive oils from various Italian olive varieties," *Food Chem.*, vol. 124, no. 3, pp. 1119–1123, Feb. 2011, doi: 10.1016/j.foodchem.2010.07.012.

[4] "(PDF) Pigments in Extra-Virgin Olive Oils Produced in Tuscany (Italy) in Different Years," *ResearchGate*. https://www.researchgate.net/publication/316676046_Pigments_in_Extra-Virgin_Olive_Oils_Produced_in_Tuscany_Italy_in_Different_Years (accessed Feb. 15, 2020).

[5] R. Yamauchi and S. Matsushita, "Quenching Effect of Tocopherols on the Methyl Linoleate Photooxidation and Their Oxidation Products," *Agric. Biol. Chem.*, vol. 41, no. 8, pp. 1425–1430, Aug. 1977, doi: 10.1080/00021369.1977.10862693.

[6] K. Bouarroudj, A. Tamendjari, and R. Larbat, "Quality, composition and antioxidant activity of Algerian wild olive (Olea europaea L. subsp. Oleaster) oil," *Ind. Crops Prod.*, vol. 83, pp. 484–491, May 2016, doi: 10.1016/j.indcrop.2015.12.081.

[7] A. Giuliani, L. Cerretani, and A. Cichelli, "Chlorophylls in olive and in olive oil: chemistry and occurrences," *Crit. Rev. Food Sci. Nutr.*, vol. 51, no. 7, pp. 678–690, Aug. 2011, doi: 10.1080/10408391003768199.

[8] B. Gandul-Rojas, M. R.-L. Cepero, and M. I. Mínguez-Mosquera, "Use of chlorophyll and carotenoid pigment composition to determine authenticity of virgin olive oil," *J. Am. Oil Chem. Soc.*, vol. 77, no. 8, pp. 853–858, Aug. 2000, doi: 10.1007/s11746-000-0136-z.

[9] S. El-Gharbi *et al.*, "Effects of Geographical Location on Chemical Properties of *Zarazi* Virgin Olive Oil Produced in the South of Tunisia," *Am. J. Food Sci. Technol.*, vol. 6, no. 6, pp. 228–236, Sep. 2018, doi: 10.12691/ajfst-6-6-1.

[10] M. Noorali, M. Barzegar, and M. A. Sahari, "Sterol and Fatty Acid Compositions of Olive Oil as an Indicator of Cultivar and Growing Area," *J. Am. Oil Chem. Soc.*, vol. 91, no. 9, pp. 1571–1581, 2014, doi: 10.1007/s11746-014-2497-z.

[11] D. Arslan, Y. Karabekir, and M. Schreiner, "Variations of phenolic compounds, fatty acids and some qualitative characteristics of Sarıulak olive oil as induced by growing area," *Food Res. Int.*, vol. 54, no. 2, pp. 1897–1906, Dec. 2013, doi: 10.1016/j.foodres.2013.06.016.

[12] D. Tura, C. Gigliotti, S. Pedò, O. Failla, D. Bassi, and A. Serraiocco, "Influence of cultivar and site of cultivation on levels of lipophilic and hydrophilic antioxidants in virgin olive oils (Olea Europea L.) and

correlations with oxidative stability," *Sci. Hortic.*, vol. 112, no. 1, pp. 108–119, Mar. 2007, doi: 10.1016/j.scienta.2006.12.036.

[13] C. M. Kalua, D. R. Bedgood, A. G. Bishop, and P. D. Prenzler, "Changes in Virgin Olive Oil Quality during Low-Temperature Fruit Storage," *J. Agric. Food Chem.*, vol. 56, no. 7, pp. 2415–2422, Apr. 2008, doi: 10.1021/jf073027b.

[14] R. Garrido-Delgado, Ma. Eugenia Muñoz-Pérez, and L. Arce, "Detection of adulteration in extra virgin olive oils by using UV-IMS and chemometric analysis," *Food Control*, vol. 85, pp. 292–299, Mar. 2018, doi: 10.1016/j.foodcont.2017.10.012.

[15] H. Azizian, M. M. Mossoba, A. R. Fardin-Kia, S. R. Karunathilaka, and J. K. G. Kramer, "Developing FT-NIR and PLS1 Methodology for Predicting Adulteration in Representative Varieties/Blends of Extra Virgin Olive Oils," *Lipids*, vol. 51, no. 11, pp. 1309–1321, Nov. 2016, doi: 10.1007/s11745-016-4195-0.

[16] Y. Li, Y. Xiong, and S. Min, "Data fusion strategy in quantitative analysis of spectroscopy relevant to olive oil adulteration," *Vib. Spectrosc.*, vol. 101, pp. 20–27, Mar. 2019, doi: 10.1016/j.vibspec.2018.12.009.

[17] M. Meenu, Q. Cai, and B. Xu, "A critical review on analytical techniques to detect adulteration of extra virgin olive oil," *Trends Food Sci. Technol.*, vol. 91, pp. 391–408, Sep. 2019, doi: 10.1016/j.tifs.2019.07.045.

[18] M. del M. Contreras, N. Arroyo-Manzanares, C. Arce, and L. Arce, "HS-GC-IMS and chemometric data treatment for food authenticity assessment: Olive oil mapping and classification through two different devices as an example," *Food Control*, vol. 98, pp. 82–93, Apr. 2019, doi: 10.1016/j.foodcont.2018.11.001.

[19] D. N. Vera, A. M. Jiménez-Carvelo, L. Cuadros-Rodríguez, I. Ruisánchez, and M. P. Callao, "Authentication of the geographical origin of extra-virgin olive oil of the Arbequina cultivar by chromatographic fingerprinting and chemometrics," *Talanta*, vol. 203, pp. 194–202, Oct. 2019, doi: 10.1016/j.talanta.2019.05.064.

[20] L. Cecchi *et al.*, "Authentication of the geographical origin of virgin olive oils from the main worldwide producing countries: A new combination of HS-SPME-GC-MS analysis of volatile compounds and chemometrics applied to 1217 samples," *Food Control*, vol. 112, p. 107156, Jun. 2020, doi: 10.1016/j.foodcont.2020.107156.

[21] N. Carranco, M. Farrés-Cebrián, J. Saurina, and O. Núñez, "Authentication and Quantitation of Fraud in Extra Virgin Olive Oils Based on HPLC-UV Fingerprinting and Multivariate Calibration," *Foods*, vol. 7, no. 4, Mar. 2018, doi: 10.3390/foods7040044.

[22] "Review: Near Infrared Spectroscopy for Analysing Olive Oils - Monica Casale, Remo Simonetti, 2014." https://journals.sagepub.com/doi/10.1255/jnirs.1106 (accessed Feb. 20, 2020).

[23] A. El Orche, M. Bouatia, H. Labjar, M. Maaouni, and M. Mbarki, "Coupling Mid Infrared Spectroscopy to mathematical and statistical tools for automatic classification, qualification and quantification of Argan oil adulteration," in *2020 IEEE 6th International Conference on Optimization and Applications (ICOA)*, Apr. 2020, pp. 1–4, doi: 10.1109/ICOA49421.2020.9094456.

[24] A. El Orche *et al.*, "Evaluation of the Capability of Horizontal ATR-FTMIR and UV-Visible Spectroscopy in the Discrimination of Virgin Olive Oils from the Moroccan Region of Beni Mellal-Khenifra," *J. Spectrosc.*, vol. 2020, pp. 1–9, Jun. 2020, doi: 10.1155/2020/9317350.

[25] J. Shi *et al.*, "Stimulated Brillouin scattering in combination with visible absorption spectroscopy for authentication of vegetable oils and detection of olive oil adulteration," *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.*, vol. 206, pp. 320–327, Jan. 2019, doi: 10.1016/j.saa.2018.08.031.

[26] P. Dais and E. Hatzakis, "Quality assessment and authentication of virgin olive oil by NMR spectroscopy: A critical review," *Anal. Chim. Acta*, vol. 765, pp. 1–27, Feb. 2013, doi: 10.1016/j.aca.2012.12.003.

[27] E. Sikorska *et al.*, "Fluorescence spectroscopy in monitoring of extra virgin olive oil during storage," *Int. J. Food Sci. Technol.*, vol. 43, no. 1, pp. 52–61, 2008, doi: 10.1111/j.1365-2621.2006.01384.x.

[28] L. Lleó, N. Hernández-Sánchez, F. Ammari, and J.-M. Roger, "3D front-face fluorescence spectroscopy for characterization of extra virgin olive oil and olive oil according to the spectral pattern," *Agric. Eng. Int. CIGR J.*, vol. 18, no. 4, p. 190, Dec. 2016, Accessed: Feb. 20, 2020. [Online]. Available: https://cigrjournal.org/index.php/Ejounral/article/view/3699.

[29] E. Sikorska, I. Khmelinskii, and M. Sikorski, "19 - Fluorescence spectroscopy and imaging instruments for food quality evaluation," in *Evaluation Technologies for Food Quality*, J. Zhong and X. Wang, Eds. Woodhead Publishing, 2019, pp. 491–533.

[30] E. Sikorska, I. Khmelinskii, and M. Sikorski, "Analysis of Olive Oils by Fluorescence Spectroscopy: Methods and Applications," *Olive Oil - Const. Qual. Health Prop. Bioconversions*, Feb. 2012, doi: 10.5772/30676.

[31] "Evaluation of the overall quality of olive oil using fluorescence spectroscopy," *Food Chem.*, vol. 173, pp. 927–934, Apr. 2015, doi: 10.1016/j.foodchem.2014.10.041.

[32] R. Karoui and C. Blecker, "Fluorescence Spectroscopy Measurement for Quality Assessment of Food Systems—a Review," *Food Bioprocess Technol.*, vol. 4, no. 3, pp. 364–386, Apr. 2011, doi: 10.1007/s11947-010-0370-0.

[33] L. C. Lee, C.-Y. Liong, and A. A. Jemain, "Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps," *Analyst*, vol. 143, no. 15, pp. 3526–3539, Jul. 2018, doi: 10.1039/C8AN00599K.

[34] R. G. Brereton and G. R. Lloyd, "Partial least squares discriminant analysis: taking the magic away," *J. Chemom.*, vol. 28, no. 4, pp. 213–225, 2014, doi: 10.1002/cem.2609.

[35] A. L. B. Brito, L. R. Brito, F. A. Honorato, M. J. C. Pontes, and L. F. B. L. Pontes, "Classification of cereal bars using near infrared spectroscopy and linear discriminant analysis," *Food Res. Int.*, vol. 51, no. 2, pp. 924–928, May 2013, doi: 10.1016/j.foodres.2013.02.014.

[36] S. Zomer, R. G. Brereton, J. F. Carter, and C. Eckers, "Support vector machines for the discrimination of analytical chemical data: application to the determination of tablet production by pyrolysis-gas chromatography-mass spectrometry," *The Analyst*, vol. 129, no. 2, p. 175, 2004, doi: 10.1039/b312982a.

[37] “Changes occurring in chemical composition and oxidative stability of virgin olive oil during storage |OCL-Oilseeds and fats, Crops and Lipids.” https://www.ocl journal.org/articles/ocl/full_html/2018/06/ocl180023/ocl180023.html (accessed Mar. 15, 2020).

[38] N. Sinelli, M. S. Cosio, C. Gigliotti, and E. Casiraghi, “Preliminary study on application of mid infrared spectroscopy for the evaluation of the virgin olive oil ‘freshness,’” *Anal. Chim. Acta*, vol. 598, no. 1, pp. 128–134, Aug. 2007, doi: 10.1016/j.aca.2007.07.024.

Chapter 3: Detection and quantification of argan oil adulteration.

Summary of chapter 3:

Food adulteration has emerged as a global problem. Adulteration represents not only a denial of the human right to safe food, but also a major menace to public health, with numerous acute and chronic illnesses. The physical and mental growth has been affected by food adulteration. The aim of this chapter is to develop methods suitable for rapid screening of argan oil adulteration by olive oil. These analytical methods are essentially based on the combination of spectroscopic sensors (fluorescence spectroscopy, UV-Visible and FT-MIR) and chemometric algorithms for discrimination and quantification. These obtained results are mainly based on a number of argan oil and olive oil samples from the same quality categories (extra virgin or virgin). The spectral approaches adopted during this study were validated through the accuracy profile. The results show that these models are efficient, accurate, and reliable for the detection of argan oil adulteration.

Résumé:

La falsification des aliments est devenue un problème mondial. L'adultération représente non seulement un déni du droit de l'homme à une alimentation saine, mais aussi une menace majeure pour la santé publique, avec de nombreuses maladies aiguës et chroniques. L'adultération alimentaire a affecté la croissance physique et mentale. L'objectif de ce chapitre est de développer des méthodes adaptées au dépistage rapide de l'adultération de l'huile d'argan par l'huile d'olive. Ces méthodes d'analyse sont essentiellement basées sur la combinaison de capteurs spectroscopiques (spectroscopie de fluorescence, UV-Visible et FT-MIR) et des algorithmes chimiométriques pour la discrimination et la quantification de l'adultération. Les résultats obtenus sont principalement basés sur un certain nombre d'échantillons d'huile d'argan et d'huile d'olive de mêmes catégories de qualité (vierge extra ou vierge). Les approches spectrales adoptées au cours de cette étude ont été validées par le profil d'exactitude. Les résultats montrent que ces modèles sont efficaces, précis et fiables pour la détection de l'adultération de l'huile d'argan.

Research Article

Received: 8 March 2021 Revised: 18 April 2021 Accepted article published: 25 May 2021 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/jsfa.11335

Comparative study of three fingerprint analytical approaches based on spectroscopic sensors and chemometrics for the detection and quantification of argan oil adulteration

Aimen El Orche,[a*] Omar Elhamdaoui,[b] Amine Cheikh,[c] Brahim Zoukeni,[a] Miloud El Karbane,[b] Mohamed Mbarki[a] and Mustapha Bouatia[b]

Abstract

BACKGROUND: Argan oil is one of the purest and rarest oils in the world, so that the addition of any further product is strictly prohibited by international regulations. Consequently, it is necessary to establish reliable analytical methods to ensure its authenticity. In this study, three multivariate approaches have been developed and validated using fluorescence, UV–visible, and attenuated total reflectance Fourier transform mid-infrared (FT-MIR) spectroscopies.

RESULTS: The application of a partial least squares discriminant analysis model showed an accuracy of 100%. The quantification of adulteration have been evaluated using partial least squares (PLS) regression. The PLS model developed from fluorescence spectroscopy provided the best results for the calibration and cross-validation sets, as it showed the highest R^2 (0.99) and the lowest root mean square error of calibration and cross-validation (0.55, 0.79). The external validation of the three multivariate approaches by the accuracy profile shows that these approaches guarantee reliable and valid results of 0.5–32%, 7–32%, and 10–32% using fluorescence, FT-MIR and UV–visible spectroscopies respectively.

CONCLUSION: This study confirmed the feasibility of using spectroscopic sensors (routine technique) for rapid determination of argan oil falsification.

Keywords: argan oil; adulteration; spectroscopy; FT-MIR; UV–visible; fluorescence; chemometrics; accuracy profile

INTRODUCTION

Argan oil is one of the rarest vegetable oils in the world. This oil is extracted mainly from the argan tree, which is an endemic tree (*Argania spinosa*) that exists in Morocco.[1] In the south of Morocco the argan tree covers an area of 3200 km^2 and plays an important economic role for this region of Morocco. Traditionally, argan oil is producing by grinding the kernels by hand; however, modern mechanical presses are often used these days.[2] This oil is considered to be an additional source of income for the local population since most of the population is involved in agriculture. Generally, the various activity sectors of argan oil have today become an opportunity for sustainable development.[3]

Currently, argan oil is considered to be among the most expensive vegetable oils in the world, thanks to its rarity, protective cardiovascular properties and richness in vitamins (vitamin E), polyunsaturated fatty acids (omega-6) and antioxidants, which make it much requested in the cosmetics and pharmaceutical sectors.[3,4] The spread of argan oil has recently crossed the frontiers of Morocco and reached many countries.[5] This evolution is strongly encouraged currently by the scientific recognition of the potential pharmaceutical properties and the continuous discovery of new anticancer substances in argan oil.[6,7]

Argan oil has become a product of great interest because of its pharmaceutical, cosmetic and nutritional properties. These qualities increase its value as an export product,[8] and for these reasons argan oil is subject to increased cases of falsification. Most common are adulteration with lower-priced oils such as soybean oil or sunflower oil.[9,10] These practices have been the subject of numerous studies aimed at combating fraud that disrupts the market and deteriorates the positive image of argan oil. These studies can be divided into two general categories: those involving the analysis of chemical compounds in the oil by gas chromatography,[5] high-performance liquid chromatography[11,12]

* *Correspondence to: A El Orche, Laboratory of Organic and Analytical Chemistry, University of Sultan Moulay Slimane, Beni-Mellal, Morocco. E-mail: aimen.elorche@gmail.com*

a *Laboratory of Organic and Analytical Chemistry, University of Sultan Moulay Slimane, Beni-Mellal, Morocco*

b *Laboratory of Analytical Chemistry, Faculty of Medicine and Pharmacy, Mohammed V University, Rabat, Morocco*

c *Faculty of Medicine, Abulcasis University, Rabat, Morocco*

I. Introduction.

Argan oil is one of the rarest vegetable oils in the world, this oil is extracted mainly from the argan tree which is an endemic tree (argania spinosa) that exists in Morocco [1]. In the south of Morocco the argan tree covers an area of 3200 km^2 which plays an important economic role for this region of Morocco. Traditionally, argan oil is producing by grinding the kernels by hand, however, modern mechanical presses are often used these days [2]. This oil is considered as an additional source of income for the local population since most of the population is involved in agriculture. Generally, the various activity sectors of argan oil have become today an opportunity for sustainable development [3].

Currently, argan oil is considered to be among the most expensive vegetable oil in the world thanks to its rarity, its protective cardiovascular properties and its richness in vitamins (vitamin E), polyunsaturated fatty acids (omega 6) and antioxidants which makes it very requested in the cosmetic and pharmaceutical sectors [4, 5]. The spread of argan oil has recently crossed the frontiers of Morocco and reached many countries [6]. This evolution is strongly encouraged currently by the scientific recognition of the potential pharmaceutical properties and the continuous discovery of new anti-cancerous substances in argan oil [7, 8].

Argan oil has become a product of great interest due to its pharmaceutical, cosmetic and nutritional properties, these qualities increase its value as an export product [9], or these reasons argan oil subject to an increase in cases of falsification. Most common are the adulteration of argan oil with lower-priced oils as soybean oil or sunflower oil [10, 11]. These practices have been the subject of numerous studies aimed at combating fraud that disrupts the market and deteriorates the positive image of argan oil. These studies can be divided into two general categories: those involving the analysis of chemical compounds in the oil by gas chromatography (GC) [6], high performance liquid chromatography (HPLC) [12, 13], and inductively coupled with plasma optical emission spectrometry (ICP-OES) [14], and those based on the combination of spectroscopic techniques with chemometrics tools, as mid-infrared spectroscopy [11, 15], and visible-near infrared [16].

The methods based on the analysis of chemical composition are generally considered time-consuming, expensive for routine use in the food industry, require competent personnel and can have a high environmental impact. The use of fast, accurate and robust analytical methods, when properly applied to verify the authenticity of argan oils, constitutes a valuable and essential tool for

authorities aiming to control food products on the market. For this reason the main recommendations made by the participants in the first edition of the international congress of argan tree, emphasizes the need to develop rapid methods able to detect the adulteration of argan oil [17].

Fourier-transform infrared spectroscopy (FT-IR) has been successfully used with chemometric techniques to monitor food adulteration. It has been used in many different authentication studies to classify argan oil according to its quality and for the detection of food adulteration [18–21], and the quantification of oils that has been adulterated by low cost oil such as argan oil by sunflower and soybean oils, or olive oil by soybean, sunflower, canola, corn, peanut, sesame, and camellia oils [22].

Ultraviolet-visible spectroscopy (UV-vis) comes as a fairly easy to use technique. However, there are no studies in the literature concerning the detection of argan oil adulteration by UV-visible spectroscopy, but there are relatively few studies in the literature on the adulteration of olive oils. This technique has been used to quantify low-quality of old olive oil in extra virgin olive oil as well as different mixtures of olive oil with corn, soybean, and sunflower oil [23].

Fluorescence spectroscopy is a non-destructive method that requires no sample preparation and is reliable enough for an authentic and accurate analysis. This technique is gaining the attention of many industries in the application of food authentication, especially in the case of oils, because it reveals the presence of intrinsic fluorophores such as carotenoids, tocopherol, phenols, oxidation products of fatty acids and in particular chlorophyll, which have made it a choice for the detection of virgin olive oil adulteration [22, 24]. Fluorescence spectroscopy has been also implemented for the characterization of edible oils during the oxidation [25, 26].

In the literature there are only few studies that investigate the argan oil adulteration based on spectroscopic methods. However, there are no studies that compare the performance of these spectroscopic sensors on this emerging issue.

The aim and novelty of the current study were to investigate the capacity of these spectroscopic sensors to classify argan oils according to their purity and the prediction of the adulteration rate. This study demonstrates also the feasibility of using the accuracy profile as a reliable approach for the validation of PLS models. The combination of spectroscopy with multivariate analysis will

allow a quantitative comparison of fluorescence, UV-visible, and FT-MIR spectroscopy for the authentication of argan oil.

II. Material and methods.

1. Sample preparation

The samples of argan oil were taken directly from farmers located in the region of Tafraout (southwestern Morocco) with the guarantee of their geographical origin and purity. The quality of olive oil used for adulteration was also guaranteed. The samples of argan oil and olive oil were maintained in the dark at a temperature of 10 ± 3°C until their analysis.

In order to study the falsification, different sets of adulterated samples were provided. The samples were prepared by mixing argan oil with an adulterant (olive oil) at different levels of adulteration. The samples were stirred and analyzed directly by fluorescence spectroscopy, UV-visible and FT-MIR. All levels of adulteration involved are recorded as w/w percentages like described in the following equation:

$$\%\ \mathbf{Adulteration} = \frac{(\mathbf{mass\ of\ Extra\ virgin\ olive\ oil\ in\ Argan\ oil})}{\mathbf{total\ mass\ of\ sample}} * \mathbf{100} \qquad \mathbf{(1)}$$

To perform the discrimination study, a total of 55 samples were used, and are randomly divided into calibration and validation samples. 45 samples were selected for calibration (15 corresponded to pure argan oil and 30 to falsified argan oil at different levels: 0.5, 0.8, 1.2, 3.2, 4.2, 5.6, 8, 10.6, 13, 16, 18, 22, 24, 26.6, 30, and 32%) and 10 samples were selected for external validation (5 corresponded to pure argan oil and 5 to falsified argan oil at different levels 0.5, 8, 10, 20 and 32%).

For the quantification model, a total of 42 samples were used for the construction of calibration models. In general two samples were prepared for each level of concentration 0.5, 0.8, 1.2, 2.3, 3.2, 4.2, 5.6, 6.6, 8, 8.9, 10.6, 13, 16, 18, 19.6, 22, 24, 26.6, 30, and 32%, then a cross-validation was used on the basis of the leave-one-out cross-validation procedure for the selection of the optimal number of latent variables needed for a good prediction.

For the full validation of these multivariate approaches developed for the quantification of argan oil adulteration, accuracy profile was applied, for this reason samples of argan oil adulterated at 4 concentration levels of 0.5, 10, 20 and 32% were chosen. Accuracy and intermediate precision were determined by analyzing these samples on three different days: three replicates on each day

for each level of concentration. The total number of samples used for the validation of the PLS quantification models was 36. (3(days)*3(replicate)*4(concentration levels)).

2. Spectral acquisition

2.1. Fluorescence spectroscopy

The samples of adulterated and non-adulterated argan oil are directly analyzed by fluorescence spectroscopy, using the FluoroMax-4 spectrophotometer (Jobin Yvon). These measurements are carried out using a cuvette with a polytetrafluoroethylene (PTFE) cover, a quartz UV with a light path of 10 mm. The acquisition of the emission spectra of argan oil was carried out at an excitation wavelength of 400 nm and the emission wavelength was recorded between 415 nm to 785 nm with a step of 0.5 nm and slit of 0.5.

2.2. FT-MIR spectroscopy

The samples of adulterated and non-adulterated Argan oils are analyzed by mid-infrared spectrophotometer (JASCO FTIR 460 PLUS (Pike Technologies, Madison, USA)) using the attenuated total reflectance (ATR) accessory with Germanium crystal. The resolution was adjusted at $4cm^{-1}$. Finally, the spectra are recorded in JCAMP format between 4000 and 600 cm^{-1}.After each use the ATR accessory is cleaned with acetonitrile and then dried with a mild paper and a new background spectra has been taken.

2.3. UV–visible spectroscopy

The analysis of the adulterated and non-adulterated argan oil was carried out using UV-visible spectrophotometer of Perkin Elmer type and a quartz cell with an optical path of 1cm. The spectrum obtained is recorded between 250 nm and 800 nm region. Duplicate spectra were collected for each sample of adulterated and non-adulterated argan oil and those replicates were treated as different samples.

3. Multivariate data analysis

In order to properly process the spectral data corresponding the three spectroscopic techniques; fluorescence, mid-infrared and UV-visible, multivariate data analysis were used to build classification and quantification models. For the validation of different quantification models of adulteration, the accuracy profile approach based on the calculation of β-content tolerance limits was used [27].

3.1. Partial least squares regression (PLS-R)

PLS regression is a recent technique in chemometrics that extends and combines the features of principal component analysis and multiple regression. Its goal is to predict or investigate a set of dependent variables from a set of independent variables or predictors. This prediction is made by extracting from the predictors a set of orthogonal factors known as latent variables that have the best predictive power [28–30].

For the quantification of the different levels of adulteration (0.5% to 32% v/v) the partial least squares regression (PLS) was performed. Basically, the PLS regression has been used to correlate the spectral intensity of each adulterated and non-adulterated sample (block X) with the percentages of argan oil adulteration (block Y) [31]. The predictive capability of the generated PLS models has been studied with several statistical performance parameters such as R-square of calibration and R-square of cross-validation. Error-values such as RMSEC (root mean square error of calibration), RMSECV (root mean square error of cross-Validation) were also used in the evaluation of the predictive performance of the built models. R2 values must be close to 1, while the error values must be small and close to each other so as to reduce the error as much as possible by maintaining the balance between the error values created in terms of amplitude and to obtain a reliable prediction model.

$$RMSEC = \sqrt{\frac{\sum(Yi-\hat{Y}i)^2}{A-1}} \quad (1.a)$$

$$RMSECV = \sqrt{\frac{\sum(Yi-\hat{Y}i)^2}{B-1}} \quad (1.b)$$

Where, Yi and $\hat{Y}$i indicate the actual and predicted values. While A and B indicate the number of samples used in the calibration and cross-validation data sets.

3.2. Partial least squares-discriminant analysis (PLS-DA)

Partial least squares-discriminant analysis (PLS-DA) is one of the most frequently applied classification methods in chemistry [32]. PLS-DA is a linear classification technique that integrates the properties of partial least squares regression and the discrimination capability of a classification approach. PLS-DA is based on the PLS regression algorithm (PLS1 when dealing with a single dependent Y variable and PLS2 when dealing with several dependent Y variables), this algorithm

looks for latent variables with maximum covariance with the Y variables [33]. The principal advantage of PLS-DA is that the relevant sources of data variability are modeled by the latent variables (LV), which represent the linear combinations of the original variables, and consequently, it allows the visualization and comprehension of the different data patterns and relationships through LV scores and loads. The optimal number of LVs is generally determined by a cross-validation method, which minimizes the error of classification.

In order to evaluate the discriminating ability of the three chemometrics approaches of PLS-DA, external validation was applied and several parameters have been calculated such as specificity, sensitivity and accuracy [33, 34]. A new sample with a predicted y-value close to 1 or greater than 0.5 assigns the sample to a specific category, while a sample with a predicted y-value less than 0.5 is assigned outside the category [35].

Sensitivity also known as the true positive rate, is defined as the ability of the model to correctly identify samples that resulted in the assignment of true positives.

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

Specificity measures the correctly classified negative observations with respect to the sum of all negative observations:

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

Accuracy indicates the proportion of the observations that are correctly classified in relation to the total number of observations:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Were TP, FP, TN and FN are:

TP: number of true positive classifications
FP: number of false positive classifications
TN: number of true negative classifications
FN: number of false negative classifications

3.3. β-content tolerance interval

To properly demonstrate the reliability of this method developed in routine application and to demonstrate the accuracy (i.e. accuracy and precision) of the obtained results with an acceptable guarantee. We have been using a full validation methodology by introducing a simple and efficient graphical decision tool so called accuracy profile based on statistical tolerance intervals such as β-content, γ-confidence tolerance interval (βγ-CCTI) [36, 37].

The principle of this validation strategy can be translated by the equation Probability (X-Z < λ) ≥ β which stipulates that the difference between a measurement (X: the value found by the developed model) and its true value (Z: reference value) must be less than the acceptance limit (λ) defined a priori, and β represents the probability of the dispersion/tolerance interval [37].

The tolerance interval used in this validation approach, defined as an interval within which it is able to predict that on average a known proportion of the measurements are within. This interval can be calculated by several methods. In general its determination is made by the calculation method proposed by Mee [38].

The tolerance interval (TI) is calculated for each concentration level according to the following formula based on the parameters calculated previously; bias (%), S_W^2 (Standard deviation intra-days), S_B^2 (Standard deviation inter-days), and RSD_{IP} (Relative standard deviation of intermediate precision).

$$Tolerance\ interval = Bias \pm t_{(df,\frac{1+\beta}{2})} \times \sqrt{\frac{1}{p \times n \times B^2}} \times RSD_{IP} \quad (5)$$

Where p, n, t are respectively the numbers of days, repetition for each concentration level, and the Student's statistical test in relation to the number of degrees of freedom (df) and the expected probability (β) of the tolerance interval. The coefficient B2 and df are calculated by the following equations.

$$B^2 = \frac{\left(\frac{S_B^2}{S_W^2}\right)+1}{\left(\frac{S_B^2}{S_W^2}\right)\times n+1} RSD_{IP} \quad (6)$$

$$df = \frac{\left(\left(\frac{S_B^2}{S_W^2}\right)+1\right)^2}{\left(\frac{\left(\frac{S_B^2}{S_W^2}\right)+\frac{1}{n}}{p-1}\right)\left(\frac{1-\frac{1}{n}}{p\times n}\right)} RSD_{IP} \quad (7)$$

3.4. Data pre-processing

To enhance the performance and to build optimal chemometric models, spectral pre-treatments were applied to the data. The data is mean-centered. Then spectral smoothing was carried out to minimize noise randomness. For the smoothing, the Savitzky-Golay polynomial fitting algorithm was used, having a second polynomial order [39]. For the correction of the systemic baseline deviation, slope and curve-linearity, which could occur due to the variation in oils properties and viscosity, the detrend correction was applied using first and second polynomial, resulting in improved spectral resolution [40]. Standard normal variation correction (SNV) has been also applied to reduce interferences or scattering variation between samples [40].

3.5. Software

The spectral processing of the data and the chemometric analyses (PLS-DA and PLS) were applied using the Unscrambler software 10.4. For the validation, an Excel table was used to calculate all statistical parameters needed for the construction of the accuracy profile.

III. Results and discussion

1. Spectral evaluation

Typical spectra, of all pure and adulterated argan oil samples studied, obtained using the spectroscopic techniques, are presented in Figure. 1. The FT-MIR spectra of the samples (Figure. 1a) are characterized by bands at different wavenumbers of 2924, 2852, 1743, 1463, 1377, 1238, 1163, 1114, 1099 and 721 cm^{-1} [41]. The absorptions at 2924 and 2852 cm^{-1} are respectively related to asymmetrical and symmetrical stretching vibrations of $-CH_2$. The major bands at 1743, 1463 and 1377 cm^{-1} are respectively associated with C=O stretching, CH_2 and CH_3 scissor vibrations. The remaining bands at 1238, 1163, 1114, 1099 cm^{-1} are correlated with C-O stretching vibrations, while a small band at 721 cm^{-1} is associated to CH_2 rocking model [41].

The UV-vis spectra of pure and adulterated argan oil samples are shown in figure 1b. The absorption spectra of argan oil samples show specific bands around 250-270 nm indicating the presence of conjugated dienes and trienes of unsaturated fatty acids, in addition the band between 300-400 nm is correlated with a variety of polyphenols compounds [42]. The Low absorption observed in the visible zone between 400 and 550 nm is associated to carotenoids [22].

Fluorescence emission spectra of the adulterated and non-adulterated argan oil samples are shown in Figure 1c. These spectra indicate two regions of interest around 400–625 nm, and 650-750 nm. The emission bands between 650 and 750 nm showed a well-known relationship with chlorophylls a and b and pheophytins a and b. The emission spectral bands range from 400 nm to 600 nm could be attributed to vitamin E (Tocopherol) and carotenoids, as well as oxidation products of fatty acids, especially conjugated hydroperoxides that are found in the range of 440–480 nm [24, 25, 43, 44].

The spectra obtained by these three spectroscopic methods were further examined in order to observe any visual trace differences between a pure argan sample and adulterated samples. The differences in the FT-MIR spectra were difficult to recognize visually. On the other hand, visual inspection revealed a weak visible differences between the spectra of adulterated and non-adulterated argan oil samples obtained by UV-visible and a strong difference for the fluorescence spectroscopy.

The fluorescence emission spectra of argan oil samples at different levels of adulteration are provided in Figure 1c. Upon the addition of olive oil on argan oil, it was observed a decrease in the intensity of the emission signal over the spectral region (425-525 nm), this decrease could be correlated with the decrease of fatty acid oxidation products such as hydroperoxides emitted around 450 nm on the one hand, and a decrease of vitamin E emitted around 525 nm on the other hand [45] . However, samples of argan oil adulterated with olive oil have a higher intensity at 650-750 nm compared to samples of pure argan oil which presents a low intensity, this difference could be attributed to the change in chlorophyll content having a negative linear relationship with the oxidation products because olive oil is very rich in chlorophyll [45].

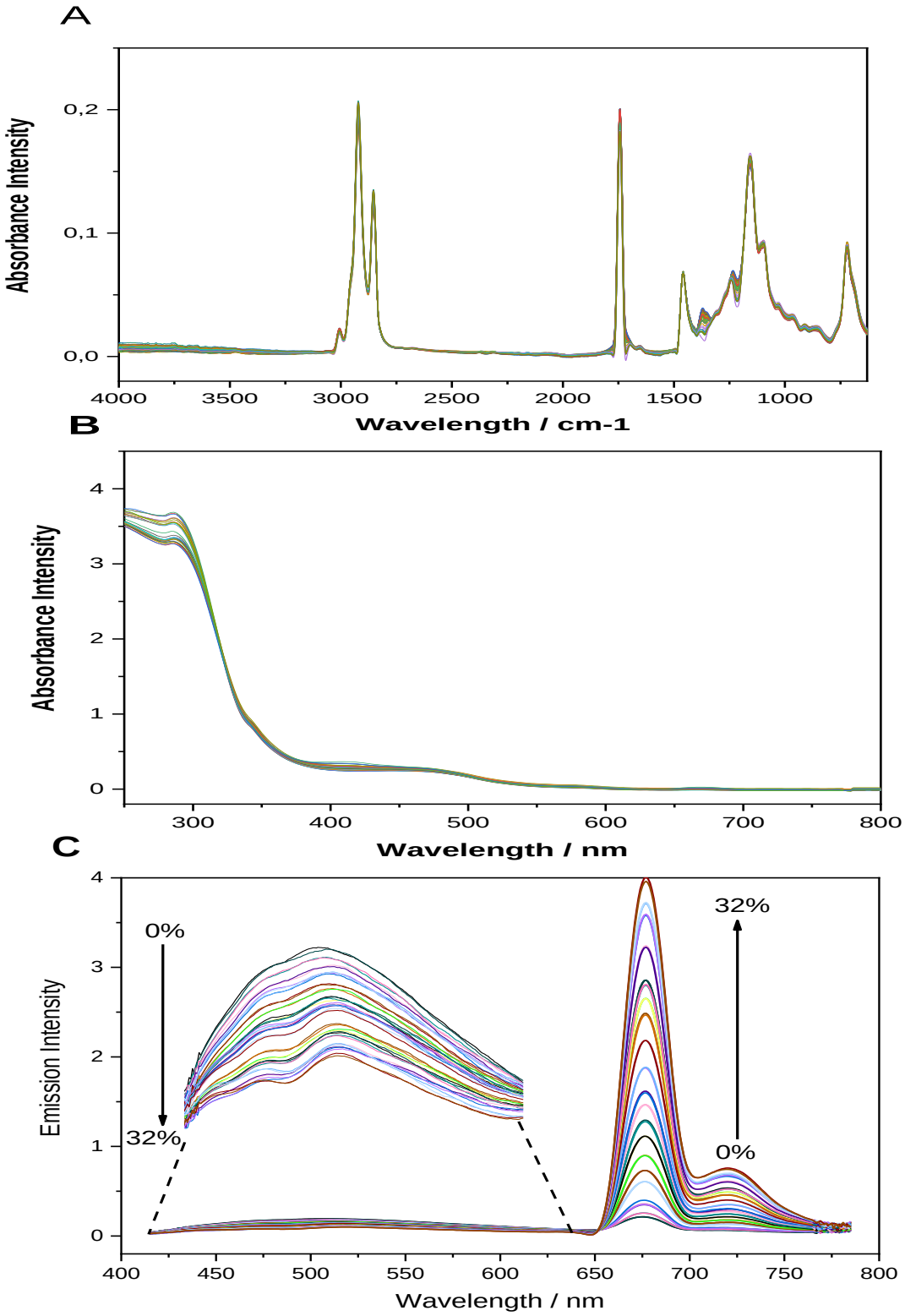

Figure 1: visualization of FT-MIR (A), UV-Visible (B) and fluorescence (C) spectra of pure argan oil (0%) and adulterated argan oil at different levels (from 0.5% to 32%).

2. Discrimination of argan oil adulterated and non-adulterated.

The PLS-DA classification method has been generally applied to the spectral data obtained for each spectroscopic method. Therefore, several models have been generated using raw data without preprocessing and data that undergo mathematical correction to improve predictive capability. Based on the obtained data in Figure 2, we notice a good discrimination between adulterated and non-adulterated argan oil. The statistical parameters for each approach summarized in the table 1 show that the application of the PLS-DA method provides good results using the data transformed by SNV in the case of fluorescence in which we observe an R2 of 0. 99 for calibration and cross-validation with a low RMSE of 0.01 for calibration and cross-validation. For the approach developed by UV-visible and FT-MIR, the good results were obtained using the raw spectral data without mathematical transformation as shown in table 1.

The results of the external validation cited in Table 2 show that all classes belong to their group with a specificity, sensitivity and accuracy of 100% for all chemometric approaches. Even with a small number of samples, the results obtained show that these approaches developed have a high capacity in the authentication of argan oil against falsification frauds by olive oils.

This discrimination study represents the first study that involves the PLS-DA approach using spectroscopic analysis in order to classify argan oils according to their purity, as there are no studies in the literature that use this approach to achieve this classification target.

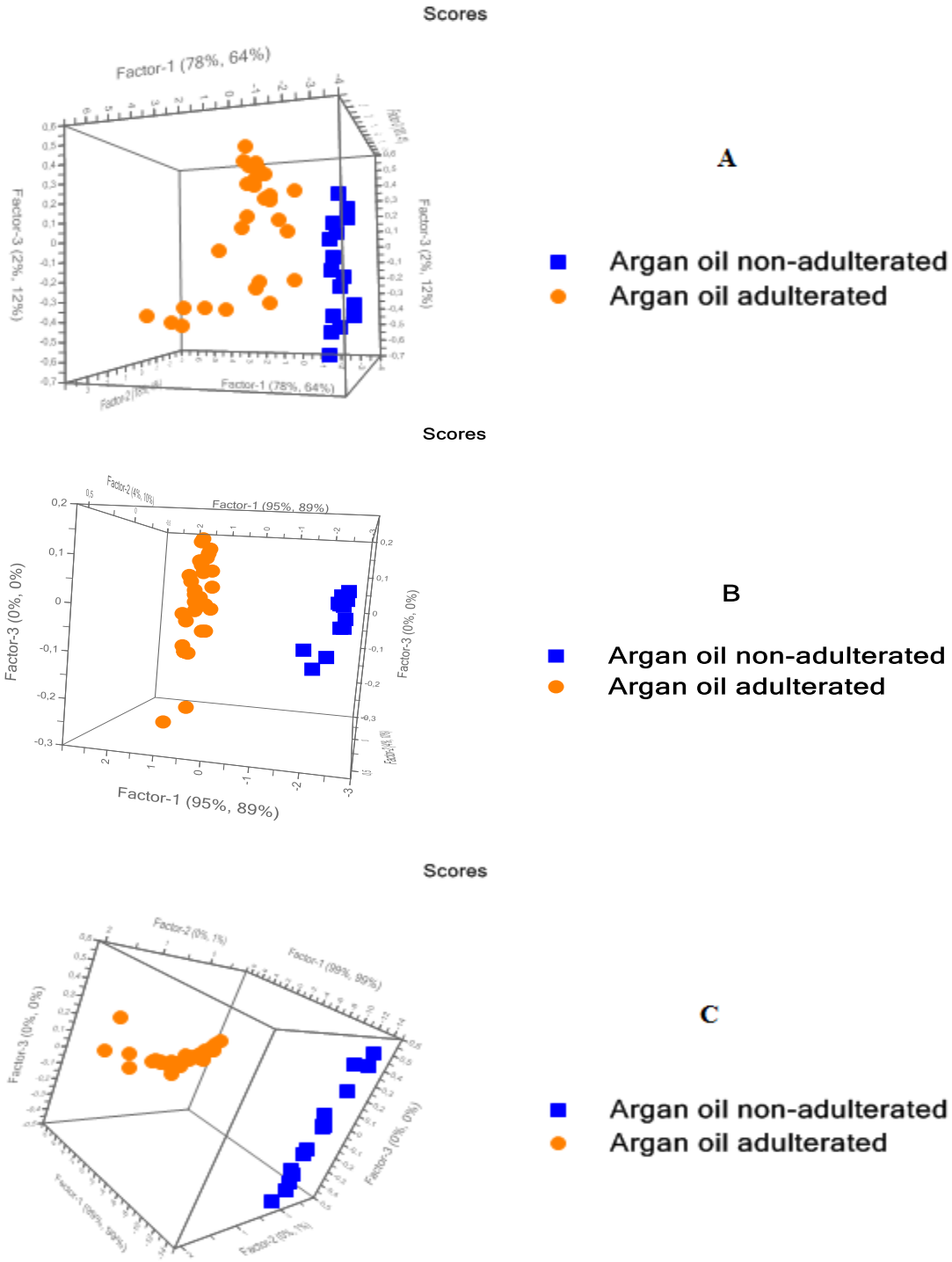

Figure 2: PLS-DA score plot of the models developed by FT-MIR (A), UV-visible (B) and fluorescence (C) for the discrimination of pure argan oil and adulterated argan oil.

Table 1: PLS-DA results of different spectroscopic methods for the classification of pure and adulterated argan oil.

Preprocessing	LV	$R^2_{Calibration}$	RMSEC	$R^2_{Cross-Validation}$	RMSECV
	FT-MIR				
Without	7	0.98	0.06	0.93	0.13
SG	7	0.90	0.15	0.75	0.27
SG-SNV	7	0.93	0.13	0.70	0.21
SNV	7	0.97	0.08	0.80	0.20
Detrend polynomial 1	7	0.95	0.10	0.78	0.22
Detrend polynomial 2	7	0.96	0.03	0.82	0.12
	UV-Visible				
Without	5	0.99	0.05	0.98	0.07
SG	5	0.99	0.05	0.97	0.08
SG-SNV	5	0.97	0.07	0.93	0.12
SNV	5	0.98	0.06	0.93	0.12
Detrend polynomial 1	5	0.99	0.05	0.96	0.09
Detrend polynomial 2	5	0.99	0.04	0.96	0.09
	Fluorescence				
without	6	0.99	0.02	0.79	0.17
SG	4	0.96	0.09	0.81	0.15
SG-SNV	2	0.99	0.01	0.99	0.01
SNV	2	0.99	0.01	0.99	0.01
Detrend polynomial 1	6	0.99	0.02	0.87	0.17
Detrend polynomial 2	2	0.99	0.03	0.80	0.21

Table 2: Confusion matrix of the external validation of the PLS-DA models of the three spectroscopic methods (PAO=pure Argan oil; AAO=Adulterated Argan oil) for the three spectroscopic approaches.

Confusion matrix		FT-MIR					
		Actual validation set					
		PAO	AAO	Sensitivity	Specificity	Accuracy	%CCR
Predicted set	PAO	5	0	100%	100%	100%	100%
	AAO	0	5	100%	100%	100%	
		UV-Visible					
	PAO	5	0	100%	100%	100%	100%
	AAO	0	5	100%	100%	100%	
		Fluorescence					
	PAO	5	0	100%	100%	100%	100%
	AAO	0	5	100%	100%	100%	

3. Quantitative Analysis using PLS-R.

The development of a quantification model based on PLS can therefore extend the potential of the approach proposed in this work to quantify the rate of adulteration in argan oil using spectroscopy sensors.

Quantification of adulteration level (0.5% to 32%) in samples of pure argan oil was performed by implementing the PLS1 algorithm to the datasets of each spectroscopic method. Table 3

summarizes all the statistical parameters obtained for each spectroscopic technique. Different appropriate pre-processing techniques were used for the development of quantitative models.

The evaluation of the spectral results found by FT-MIR using PLS regression demonstrates high performance for the quantification of the adulteration rate, this performance was demonstrated by the high values of R2 and the low values of RMSE for calibration and cross-validation. It was found that the model built on FT-MIR data pre-processed by SNV using 7 LV has a good performance with respect to other models, in which we observe an R2 value of 0.99, 0.98 and RMSE of 0.81%, 1.59% for calibration and cross-validation respectively as shown in table 3. There is only one preliminary study in the literature that predicts successfully a low grade of different edible oils (sunflower or soybean oils) in pure argan oil using FT-MIR spectroscopy and PLS regression with an R2 of 0.99 and RMSE less than 1% [11].

The PLS model developed on the UV-visible spectral data showed high performance in calibration and cross-validation, including 6 LVs, as well as acceptable values of R2 and RMSE were obtained in which the R2 is greater than 0.98 while the RMSE values range from 0.54% to 1.06% for calibration and 0.54% to 1.75% for cross-validation (Table 3). According to these results the best model was constructed using data preprocessing by detrend using polynomial degree 2. In the literature, there are no studies using UV-visible spectroscopy for the prediction of this type of adulteration except only one study based on the use of visible/near-infrared that has performed to determine the level of adulteration in argan oil with cheap vegetable oils [16].

For fluorescence spectroscopy, the application of PLS shows a strong ability in quantifying the adulteration rate, expressed by the statistical values of R2 and RMSE and the number of LVs used for the construction of PLS models. The PLS model built on non-preprocessed fluorescence data using 2 LVs is considered the best PLS model, as it provides an R2 of 0.99 and an RMSE of 0.55%, 0.79% respectively for calibration and cross-validation. Generally, these RMSE values are considered to be the lowest values obtained compared to those obtained using FT-MIR and UV-Visible as shown in table 3. The evaluation of the PLS regression plot (Figure 3) reveals a good fit and suggests that this analysis could be used to determine the addition of olive oil in argan oil at levels below 1%. This capacity is explained by the high specificity of this technique regarding the fluorescence compounds (chlorophyll, pheophytin and tochopherol) present in olive oil [46]. This study results are consistent with a very recent study using a laser beam at 532 nm [47].

Table 3: PLS results of different spectroscopic methods for the quantification of olive oil in pure argan oil.

Preprocessing	LV	$R^2_{Calibration}$	RMSEC	$R^2_{Cross\text{-}validation}$	RMSECV
	FT-MIR				
Without	7	0.98	1.40	0.96	2.19
SG	7	0.98	1.60	0.96	2.31
SG-SNV	7	0.99	0.96	0.97	1.92
SNV	7	0.99	0.81	0.98	1.59
Detrend polynomial 1	7	0.99	0.96	0.97	1.81
Detrend polynomial 2	7	0.99	0.85	0.97	1.71
	UV-Visible				
Without	6	0.99	0.75	0.98	1.75
SG	6	0.98	1.06	0.98	1.44
SG-SNV	6	0.99	0.93	0.98	1.38
SNV	6	0.99	0.54	0.99	1.03
Detrend polynomial 1	6	0.99	0.59	0.98	1.32
Detrend polynomial 2	6	0.99	0.55	0.98	0.99
	Fluorescence				
without	2	0.99	0.55	0.99	0.79
SG	2	0.99	0.82	0.99	0.81
SG-SNV	3	0.99	2.64	0.99	2.27
SNV	5	0.98	1.35	0.83	3.98
Detrend polynomial 1	2	0.99	0.62	0.99	0.78
Detrend polynomial 2	2	0.99	0.62	0.99	0.78

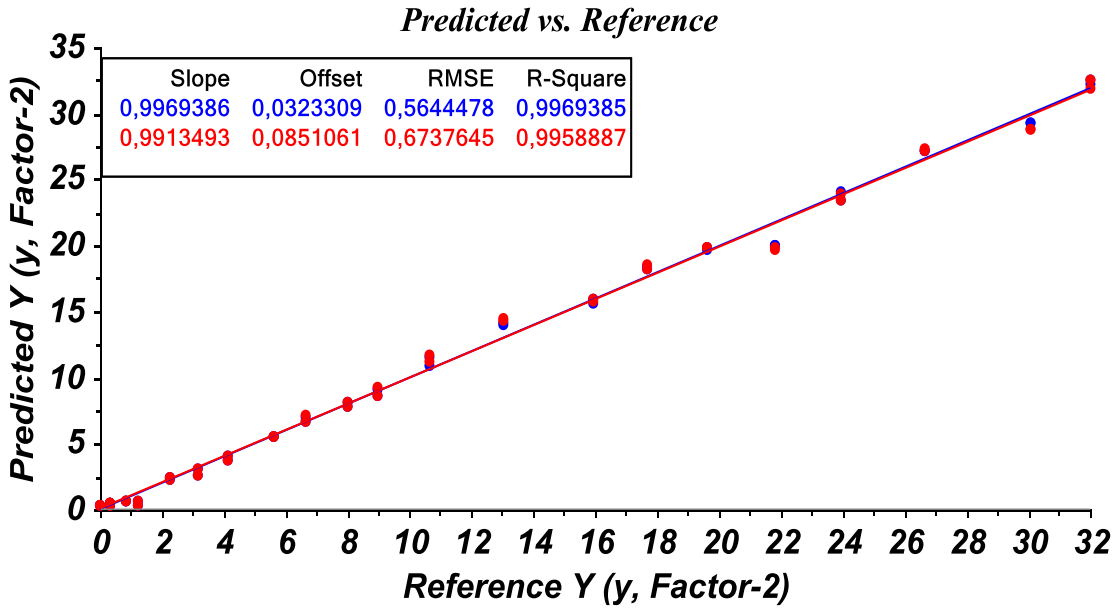

Figure 3: Actual vs. predicted adulteration levels of the best PLS model obtained using fluorescence spectroscopy for the quantification of argan adulteration by olive oil. The model was constructed on two latent variables.

4. Validation of the quantification models using accuracy profile.

In order to evaluate the predictive capacity of the three approaches developed, external validation was used. On the basis of the predicted concentration values obtained for the samples of the validation set, the accuracy profiles were calculated for each chemometric approach. In this approach of validation the limit of quantification (LOQ) is determined by the intercept between the calculated tolerance interval and the acceptance limits defined a priori. The performance and validation results of the models selected for each approach have been presented in Table 4.

From the accuracy profile figure 4, it can be seen that in the case of FT-MIR (A) the upper and lower β expectation tolerance limit exceed the upper and lower limit of acceptability settled to 15% at the adulteration level of 7%. As well as the accuracy profile of UV-visible (B) demonstrate that the β expectation tolerance limit exceed the upper acceptable limit at 10% .This means that for a percentage of adulteration below 7% and 10% of adulteration rate using respectively FT-MIR and UV-visible spectroscopy the analyst cannot guarantee that the method is routinely capable of producing an average probability β (90%) of acceptable results. Whereas the accuracy profile used for the validation of fluorescence spectroscopy approach (C) shows that the upper and lower β expectation tolerance limits are included within the acceptance limits set at 15%. This means that the method is routinely able of producing an acceptable and valid result in an average of 0.5% to 32%.

Therefore, the range of linearity has been limited for all approaches between 0.5 % and 32 % range. Furthermore, it can be seen that the precision expressed by the relative standard deviation of intermediate precision (RSD) varies according to the concentration since its RSD varies from 0.70% to 4.43%, 0.19% to 43.31%, and 0.70% to 27.09%. As well as the accuracy is also varying since the accuracy bias varies from 0.35% to 2.65%, 1.70% to -16.83, and 0.52% to -3.81 for fluorescence, UV-visible, and FT-MIR respectively. The high value of the RSD of precision obtained in this study corresponds to the first concentration level of 0.5% for both methods UV-visible and FT-MIR. Although for the fluorescence method the RSD values obtained are less than 5%. This remark underlines the interest in performing calculations level by level.

Table 4: results of validation of the three spectroscopic methods developed by PLS regression using accuracy profile.

	Fluorescence		UV-Visible		FT-MIR	
Range of linearity	0.5%-32%		0.5%-32%		0.5%-32%	
Intercept	0.045		0.1353		0.062	
Slope	0.99		0.98		0.99	
R^2	0.99		0.99		0.99	
LOQ	0.5%		10%		7%	
Predicted average concentration						
Level 1 (0.5%)	0.51		0.42		0.48	
Level 2 (8%)	8.04		8.41		8.04	
Level 3 (20%)	20.07		21.12		20.10	
Level 4 (32%)	32.38		32.54		32.38	
RSD of precision (%)						
Level 1 (0.5%)	3.42		43.31		27.09	
Level 2 (8%)	4.43		5.68		4.47	
Level 3 (20%)	1.26		0.19		1.35	
Level 4 (32%)	0.70		1.12		0.70	
Relative bias (%)						
Level 1 (0.5%)	2.65		-16.83		-3.81	
Level 2 (8%)	0.48		5.14		0.50	
Level 3 (20%)	0.35		5.62		0.52	
Level 4 (32%)	1.20		1.70		1.20	
Expectation tolerance limit (%)	**Low**	**Upper**	**Low**	**Upper**	**Low**	**Upper**
Level 1 (0.5%)	95.74	109.57	12.18	154.17	44.84	147.54
Level 2 (8%)	91.71	109.25	93.37	116.91	91.65	109.36
Level 3 (20%)	97.86	102.84	105.22	106.01	97.84	103.20
Level 4 (32%)	99.80	102.59	99.45	103.94	99.80	102.59

According to this validation approach, it can be stated that the range of validity of spectroscopic methods combined with PLS regression ranges from 0.5% to 32%, 7% to 32%, and 10% to 32% using fluorescence, FT-MIR, and UV-visible respectively. As the validity range corresponds to the concentrations for which the tolerance interval is within the limits of acceptability. This shows that the fluorescence spectroscopy coupled with PLS regression is suitable to quantify adulteration rates of argan oil by olive oil at small quantities starting from 0.5%.

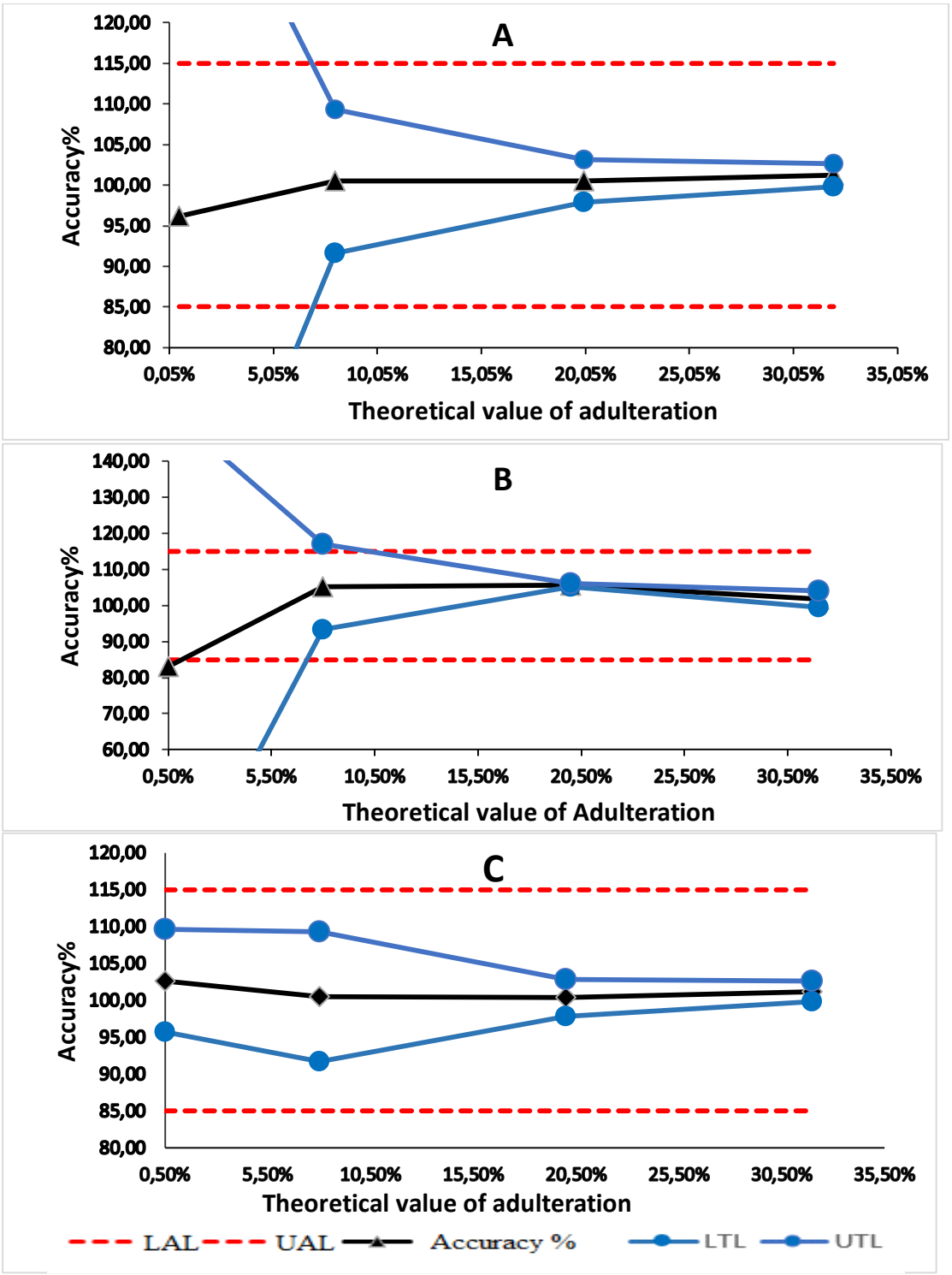

Figure 4: : Accuracy profiles obtained for the validation of the three spectroscopic approaches FT-MIR (A), UV-visible (B), and Fluorescence (C) to quantify the adulteration rate of argan oil by olive oil considering PLS regression. The red dashed lines are the upper and lower acceptance limits(UAL and LAL) defined at 15%, the solid blue lines are the upper and lower calculated tolerance limits(UTL and LTL) of β expectation (with β= 90%), and the solid black line is the recovery rate (Accuracy).

IV. Conclusion

The present study demonstrated the application of three spectroscopy sensors combined with multivariate analysis (chemometrics) to determine the adulteration of argan oil by olive oil. Pure and adulterated samples were classified by means of PLS-DA. Clear discrimination of adulterated and non-adulterated samples was achieved for all spectroscopic methods. PLS-R calibration models were applied to quantify the percentage of adulteration of argan oil by olive oil. Calibration and cross-validation results indicate that all three approaches are credible for the determination of adulteration due to their lower error values and higher regression coefficients. The validation of these approaches by the accuracy profile shows that the best results are obtained using fluorescence spectroscopy followed by FT-MIR and UV-Visible spectroscopy respectively. From these results, it can be deduced that the approach developed by fluorescence allows guaranteeing reliable results in all the validation domains between 0.5% to 32% while the other approaches allow guaranteeing reliable results only in the interval 7% to 32% and 10% to 32% for FT-MIR and UV-Visible spectroscopy respectively.

For a robust assessment and rapid authentication of virgin Argan oils on the market in order to recognize their falsification, the development of reliable spectral information databases is encouraged as much as possible.

References

1. Z. Charrouf and D. Guillaume: “Argan oil: Occurrence, composition and impact on human health.” *Eur. J. Lipid Sci. Technol.* vol. 110, no. 7, pp. 632–636, 2008.
2. Z. Charrouf and D. Guillaume: “Argan oil, the 35-years-of-research product.” *Eur. J. Lipid Sci. Technol.* vol. 116, no. 10, pp. 1316–1321, 2014.
3. H.E. Monfalouti, D. Guillaume, C. Denhez, and Z. Charrouf: “Therapeutic potential of argan oil: a review.” *J. Pharm. Pharmacol.* vol. 62, no. 12, pp. 1669–1675, 2010.
4. F. Henry, L. Danoux, G. Pauly, and Z. Charrouf: “Use of an extract from the plant argania spinosa,” https://patents.google.com/patent/US20060083794/en, (2006).
5. H.E. Monfalouti, D. Guillaume, C. Denhez, and Z. Charrouf: “Therapeutic potential of argan oil: a review: Therapeutic potential of argan oil.” *J. Pharm. Pharmacol.* vol. 62, no. 12, pp. 1669–1675, 2010.
6. M. Hilali, Z. Charrouf, A.E.A. Soulhi, L. Hachimi, and D. Guillaume: “Detection of Argan Oil Adulteration Using Quantitative Campesterol GC-Analysis.” *J. Am. Oil Chem. Soc.* vol. 84, no. 8, pp. 761–764, 2007.
7. H. Berrougui, M. Alvarez de Sotomayor, C. Pérez-Guerrero, A. Ettaib, M. Hmamouchi, E. Marhuenda, and M.D. Herrera: “Argan (Argania spinosa) oil lowers blood pressure and improves endothelial dysfunction in spontaneously hypertensive rats.” *Br. J. Nutr.* vol. 92, no. 6, pp. 921–929, 2004.
8. A.B. Awad and C.S. Fink: “Phytosterols as anticancer dietary components: evidence and mechanism of action.” *J. Nutr.* vol. 130, no. 9, pp. 2127–2130, 2000.
9. M. Cherki, H. Berrougui, A. Drissi, A. Adlouni, and A. Khalil: “Argan oil: which benefits on cardiovascular diseases?” *Pharmacol. Res.* vol. 54, no. 1, pp. 1–5, 2006.
10. Y. Gunning, A.J. Jackson, J. Colmer, F. Taous, M. Philo, R.M. Brignall, T.E. Ghali, M. Defernez, and E.K. Kemsley: “High-throughput screening of argan oil composition and authenticity using benchtop 1H NMR.” *Magn. Reson. Chem.* vol. 58, no. 12, pp. 1177–1186, 2020.
11. A. Oussama, F. Elabadi, and O. Devos: “Analysis of Argan Oil Adulteration Using Infrared Spectroscopy.” Presented at the (2012).
12. R. Salghi, W. Armbruster, and W. Schwack: “Detection of argan oil adulteration with vegetable oils by high-performance liquid chromatography–evaporative light scattering detection.” *Food Chem.* vol. 153, pp. 387–392, 2014.
13. I. Ourrach, M. Rada, M.C. Pérez-Camino, M. Benaissa, and Á. Guinda: “Detection of argan oil adulterated with vegetable oils: new markers.” *Grasas Aceites*. 2012.
14. A. Gonzálvez, S. Armenta, and M. de la Guardia: “Adulteration detection of argan oil by inductively coupled plasma optical emission spectrometry.” *Food Chem.* vol. 121, no. 3, pp. 878–886, 2010.
15. A. El Orche, M. Bouatia, H. Labjar, M. Maaouni, and M. Mbarki: “Coupling Mid Infrared Spectroscopy to mathematical and statistical tools for automatic classification, qualification and quantification of Argan oil adulteration.” 2020 IEEE 6th International Conference on Optimization and Applications (ICOA). pp. 1–4 (2020).
16. S. Farres, L. Srata, F. Fethi, and A. Kadaoui: “Argan oil authentication using visible/near infrared spectroscopy combined to chemometrics tools.” *Vib. Spectrosc.* vol. 102, pp. 79–84, 2019.
17. “Recommandations,” http://www.congresarganier.ma/recommandations.php.
18. W. Sohng, Y. Park, D. Jang, K. Cha, Y.M. Jung, and H. Chung: “Incorporation of two-dimensional correlation analysis into discriminant analysis as a potential tool for improving discrimination accuracy: Near-infrared spectroscopic discrimination of adulterated olive oils.” *Talanta*. vol. 212, 2020.

19. M. Kharbach, R. Kamal, M. Bousrabat, M. Alaoui Mansouri, I. Barra, K. Alaoui, Y. Cherrah, Y. Vander Heyden, and A. Bouklouze: “Characterization and classification of PGI Moroccan Argan oils based on their FTIR fingerprints and chemical composition.” *Chemom. Intell. Lab. Syst.* vol. 162, pp. 182–190, 2017.
20. M. Kharbach, I. Marmouzi, R. Kamal, H. Yu, I. Barra, Y. Cherrah, K. Alaoui, Y.V. Heyden, and A. Bouklouze: “Extra virgin Argan oils’ shelf-life monitoring and prediction based on chemical properties or FTIR fingerprints and chemometrics.” *Food Control*. vol. 121, pp. 107607, 2021.
21. O. Elhamdaoui, A. El Orche, A. Cheikh, B. Mojemmi, R. Nejjari, and M. Bouatia: “Development of Fast Analytical Method for the Detection and Quantification of Honey Adulteration Using Vibrational Spectroscopy and Chemometrics Tools,” https://www.hindawi.com/journals/jamc/2020/8816249/.
22. O. Uncu and B. Ozen: “A comparative study of mid-infrared, UV–Visible and fluorescence spectroscopy in combination with chemometrics for the detection of adulteration of fresh olive oils with old olive oils.” *Food Control*. vol. 105, pp. 209–218, 2019.
23. J.S. Torrecilla, E. Rojo, J.C. Domínguez, and F. Rodríguez: “A Novel Method To Quantify the Adulteration of Extra Virgin Olive Oil with Low-Grade Olive Oils by UV−Vis.” *J. Agric. Food Chem.* vol. 58, no. 3, pp. 1679–1684, 2010.
24. H. Ali, M. Saleem, M.R. Anser, S. Khan, R. Ullah, and M. Bilal: “Validation of Fluorescence Spectroscopy to Detect Adulteration of Edible Oil in Extra Virgin Olive Oil (EVOO) by Applying Chemometrics:” *Appl. Spectrosc.* 2018.
25. A. El Orche, M. Bouatia, and M. Mbarki: “Rapid Analytical Method to Characterize the Freshness of Olive Oils Using Fluorescence Spectroscopy and Chemometric Algorithms,” https://www.hindawi.com/journals/jamc/2020/8860161/.
26. D. Üçüncüoğlu and V.A. Küçük: “A multivariate data approach for FTIR-ATR monitoring of virgin olive oils: Lipid structure during autoxidation.” *OCL*. vol. 26, pp. 42, 2019.
27. P. Jitaru, R. Chekri, P. Fisicaro, and T. Guérin: “METHOD VALIDATION BY MEANS OF THE ACCURACY PROFILE: A REALISTIC APPROACH FOR ROUTINE CHEMICAL LABORATORIES.” pp. 7, 2016.
28. I. Barra, M. Kharbach, E.M. Qannari, M. Hanafi, Y. Cherrah, and A. Bouklouze: “Predicting cetane number in diesel fuels using FTIR spectroscopy and PLS regression.” *Vib. Spectrosc.* vol. 111, pp. 103157, 2020.
29. A.-L. Boulesteix and K. Strimmer: “Partial least squares: a versatile tool for the analysis of high-dimensional genomic data.” *Brief. Bioinform.* vol. 8, no. 1, pp. 32–44, 2007.
30. S. Wold, M. Sjöström, and L. Eriksson: “PLS-regression: a basic tool of chemometrics.” *Chemom. Intell. Lab. Syst.* vol. 58, no. 2, pp. 109–130, 2001.
31. G. Gurdeniz and B. Ozen: “Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data.” *Food Chem.* vol. 116, no. 2, pp. 519–525, 2009.
32. U.G. Indahl, H. Martens, and T. Næs: “From dummy regression to prior probabilities in PLS-DA.” *J. Chemom.* vol. 21, no. 12, pp. 529–536, 2007.
33. D. Ballabio and V. Consonni: “Classification tools in chemistry. Part 1: linear models. PLS-DA.” *Anal. Methods*. vol. 5, no. 16, pp. 3790, 2013.
34. A. El Orche, M. Bouatia, S. Yanisse, H. Labjar, M. Mouhsin, M. Bouha, M. Oubenali, and M. Mbarki: “Evaluation of the Capability of Horizontal ATR-FTMIR and UV-Visible Spectroscopy in the Discrimination of Virgin Olive Oils from the Moroccan Region of Beni Mellal-Khenifra.” *J. Spectrosc.* vol. 2020, pp. 1–9, 2020.
35. L.C. Lee, C.-Y. Liong, and A.A. Jemain: “Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps.” *Analyst*. vol. 143, no. 15, pp. 3526–3539, 2018.

36. Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, and L. Valat: "Harmonization of strategies for the validation of quantitative analytical procedures: A SFSTP proposal—part I." *J. Pharm. Biomed. Anal.* vol. 36, no. 3, pp. 579–586, 2004.
37. M. Feinberg: "Validation of analytical methods based on accuracy profiles." *J. Chromatogr. A*. vol. 1158, no. 1, pp. 174–183, 2007.
38. R.W. Mee: "β-Expectation and β-Content Tolerance Limits for Balanced One-Way ANOVA Random Model." *Technometrics*. vol. 26, no. 3, pp. 251–254, 1984.
39. B. Zimmermann and A. Kohler: "Optimizing Savitzky–Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy." *Appl. Spectrosc.* vol. 67, no. 8, pp. 892–902, 2013.
40. J. Luypaert, S. Heuerding, Y.V. Heyden, and D.L. Massart: "The effect of preprocessing methods in reducing interfering variability from near-infrared measurements of creams." *J. Pharm. Biomed. Anal.* vol. 36, no. 3, pp. 495–503, 2004.
41. N. Sinelli, M.S. Cosio, C. Gigliotti, and E. Casiraghi: "Preliminary study on application of mid infrared spectroscopy for the evaluation of the virgin olive oil 'freshness.'" *Anal. Chim. Acta*. vol. 598, no. 1, pp. 128–134, 2007.
42. A. Grazia, L. Ciaccheri, A. Azelio, and A. Cimato: "Optical Absorption Spectroscopy for Quality Assessment of Extra Virgin Olive Oil." In: Boskou, D. (ed.) Olive Oil - Constituents, Quality, Health Properties and Bioconversions. *InTech* (2012).
43. N. Dupuy, Y. Le Dréau, D. Ollivier, J. Artaud, C. Pinatel, and J. Kister: "Origin of French Virgin Olive Oil Registered Designation of Origins Predicted by Chemometric Analysis of Synchronous Excitation−Emission Fluorescence Spectra." *J. Agric. Food Chem.* vol. 53, no. 24, pp. 9361–9368, 2005.
44. Y.G.M. Kongbonga, H. Ghalila, M.B. Onana, Y. Majdi, Z.B. Lakhdar, H. Mezlini, and S. Sevestre-Ghalila: "Characterization of Vegetable Oils by Fluorescence Spectroscopy." *Food Nutr. Sci.* vol. 02, no. 07, pp. 692–699, 2011.
45. L. Lleó, N. Hernández-Sánchez, F. Ammari, and J.-M. Roger: "3D front-face fluorescence spectroscopy for characterization of extra virgin olive oil and olive oil according to the spectral pattern." *Agric. Eng. Int. CIGR J.* vol. 18, no. 4, pp. 190, 2016.
46. I. Durán Merás, J. Domínguez Manzano, D. Airado Rodríguez, and A. Muñoz de la Peña: "Detection and quantification of extra virgin olive oil adulteration by means of autofluorescence excitation-emission profiles combined with multi-way classification." *Talanta*. vol. 178, pp. 751–762, 2018.
47. S. Addou, F. Fethi, M. Chikri, and A. Rrhioua: "Detection of argan oil adulteration with olive oil using fluorescence spectroscopy and chemometrics tools." pp. 10, 2016.

Part 3: Conclusion and perspectives

In this thesis quality control methods have been developed and validated to determine the geographical traceability of olive oils from the region of Beni Mellal-khenifra, the classification of olive oils according to their freshness, and the detection of argan oils adulteration, which are widely used and commercially available especially in the Moroccan market.

In addition to the traditional criteria used to evaluate the results of classification and quantification models, based only on the number of samples correctly or incorrectly classified, RMSEC, RMSECV, and R-square, other parameters have been used to obtain more information when analyzing the results. A clear set of measurement criteria was established for comparing the effectiveness of the different classification models constructed. In particular, the overall classification success rate, as it is a parameter that allows to evaluate the overall efficiency of the models. A statistical approach was also used to assess the accuracy of the quantification model.

The proposed classification models have been tested with an external sample, providing information on the performance of the different methods used for the classification of olive oils according to their geographical origin and their freshness using spectroscopic sensors. External validation has also been used to evaluate the performance of the models built to quantify the adulteration level accurately and reliably, in this case, a statistical approach called "accuracy profile" has been applied to demonstrate the validity of these quantification methods in various levels of adulteration. .

Firstly, the capability of UV-Visible and MIR spectroscopy combined with PCA-LDA and PCA-SVM classification techniques was evaluated for the determination of the geographical origin of virgin olive oils produced in the region of Beni Mellal-Khenifra in Morocco. This study was carried out on 41 samples collected from five different geographical areas. The exploitation of UV-Visible and MIR spectroscopic data by chemometric classification tools shows that the PCA-SVM and

PCA-LDA method provides a good classification of the five groups of olive oil according to their geographical origin with high sensitivity, specificity, and correct classification rate.

Secondly an evaluation of fluorescence spectroscopy for the classification of three classes of olive oil with different shelf life. The emission spectrum data were firstly processed by PCA, which reveals a good classification of the three groups of olive oil according to their shelf life. Application of discrimination approach such as LDA, SVM, and PLS-DA shows a high performance in the classification of olive oils according to their shelf-life. External validation of these multivariate approaches demonstrates high performance represented by an accuracy of 100%.

Thirdly, rapid analytical methods have been established for the investigation of argan oil adulteration by olive oil, the combination of spectral data provided by MIR, UV-Visible, and fluorescence spectroscopy with the PLS-DA method shows a strong ability to discriminate between pure and adulterated argan oil at different levels of adulteration. The validation of these discrimination approaches shows a specificity, sensitivity, and accuracy of 100%. For the quantification of adulteration rate, the application of PLS provides good results, represented by the low RMSE, and high R-square. The validation of these three multivariate approaches by the accuracy profile shows that the fluorescence approach guarantees reliable results over the whole range of validity 0.5%-32% while the MIR and UV-Visible spectroscopy guarantees reliable results only in the range of validity 7% to 32% and 10% to 32% respectively.

In this thesis, spectroscopic techniques were used to construct a spectral matrix that was used as a fingerprint for food quality control. However, the simultaneous use of spectral information contained in a data matrix is an area of opportunity for the development of rapid and advanced methods for food quality control using chemometric methods.

This represents a breakthrough for food authorities to control the traceability of foodstuffs quickly and reliably

The future perspectives of research on olive and argan oil involve the application of other spectroscopic fingerprinting techniques such as Raman spectroscopy, nuclear magnetic reasoning, and near-infrared spectroscopy. Hence, chromatographic fingerprints (UPLC/DAD and UPLC/MS), measuring the phenolic fraction and fatty acid content, which can be combined with chemometric tools, to control argan and olive oils. The objective is, therefore, to study the possibility of using the phenolic and fatty acid profiles of argan oil, in combination with pattern recognition and quantification techniques, to identify the characteristic marker compounds that distinguish pure and adulterated argan oil, and also to classify the oils according to their geographical and varietal origin. Chromatographic fingerprints can also be used to classify samples according to their geographical and varietal origin, for authentication, and quality control of the oils.

In addition, quantification studies of vitamin E (tocopherol) and polyphenol content in olive and argan oils can be evaluated using fluorescence spectroscopy imagery and chemometric modeling tools such as PLS, SVM, and ANN. This type of study could allow industry and control laboratories to predict the levels of these chemical compounds (markers present in argan oil and argan).

Scientific Contribution

Articles

- **El Orche A**, Bouatia M, Yanisse S, Labjar H, Mouhsin M, Bouha M, et al. Evaluation of the Capability of Horizontal ATR-FTMIR and UV-Visible Spectroscopy in the Discrimination of Virgin Olive Oils from the Moroccan Region of Beni Mellal-Khenifra. J Spectrosc. 2020Jun 20 [cited 2020 Jun 21];2020:1–9.

- **El Orche A**, Bouatia M, Mbarki M. Rapid Analytical Method to Characterize the Freshness of Olive Oils Using Fluorescence Spectroscopy and Chemometric Algorithms. Vol. 2020, Journal of Analytical Methods in Chemistry. Hindawi; 2020 [cited 2020 Jul 14]. p. e8860161.

- Elhamdaoui O, **El Orche A**, Cheikh A, Mojemmi B, Nejjari R, Bouatia M. Development of Fast Analytical Method for the Detection and Quantification of Honey Adulteration Using Vibrational Spectroscopy and Chemometrics Tools [Internet]. Vol. 2020, Journal of Analytical Methods in Chemistry. Hindawi; 2020 [cited 2020 Dec 26]. p. e8816249.

- **El Orche A**, Elhamdaoui O, Cheikh A, Zoukeni B, El Karbane M, Mbarki M, and Bouatia M. Comparative study of three fingerprint analytical approaches based on spectroscopic sensors and chemometrics for the detection and quantification of argan oil adulteration. J Sci Food Agric :jsfa.11335 (2021).

- **El Orche A**, Adade CA, Mefetah H, Cheikh A, Karrouchi K, El Karbane M, and Bouatia M. Chemometric Analysis of UV-Visible Spectral Fingerprints for the Discrimination and Quantification of Clinical Anthracycline Drug Preparation Used in Oncology. Deniz NG, ed. BioMed Res Int 2021:1–8 (2021).

- Elhamdaoui O, **El Orche A**, Cheikh A, Karrouchi K, Laarej K, and Bouatia M. Assessment of a Non-Destructive Method for Rapid Discrimination of Moroccan Date Palm Varieties via Mid-Infrared Spectroscopy Combined with Chemometric Models. J AOAC Int (2021).

Communications

- Elhamdaoui O, **El Orche A**, Bouchafra H, Karbane ME, Cheikh A, Bouatia M. The development of green analytical methods to monitor adulteration in honey by UV-visible spectroscopy and chemometrics models. E3S Web Conf [Internet]. 2020 [cited 2020 Dec 6];211:02011.

- **El Orche A**, Bouatia M, Labjar H, Maaouni M, Mbarki M. Coupling Mid Infrared Spectroscopy to mathematical and statistical tools for automatic classification, qualification and quantification of Argan oil adulteration. In: 2020 IEEE 6th International Conference on Optimization and Applications (ICOA). 2020. p. 1–4.

- **El Orche A**, Mbarki M, Cheikh A, Bouatia M. Assessment of non-destructive spectroscopy and chemometrics tools for the development of green analytical methods to determine the shelf-life of olive oils. E3S Web Conf [Internet]. 2020 [cited 2020 Dec 6];183:04003.

- CA Adade, **El Orche A**, H Attjioui, A Cheikh, H Mefetah, M El Karbane, M Bouatia. Development and validation of a discriminative method for anthracyclines used in oncology by visible spectrometer | European Journal of Hospital Pharmacy [Internet]. [cited 2021Jan 30].

- **El Orche A**, Mbarki M. Couplage des méthodes spectroscopique et mathématicostatistique pour valoriser la qualité des huiles d'olive marocaine, Congrès International 'Valorisation des Ressources Naturelles: De la Recherche Scientifique à la Faisabilité Socio-économique' -VARENA 2019.