

Thèse de Doctorat

Noelle MERCHAK

*Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université de Nantes et de l'Université Saint-Joseph
sous le label de L'Université Nantes Angers Le Mans*

École doctorale : 3MPL

Discipline : Chimie

Spécialité : Chimie Analytique

Unité de recherche : CEISAM- Université de Nantes
LMFI- Université Saint-Joseph

Soutenue le 6 avril 2016

Thèse N° :

Caractérisation et traçabilité des matrices riches en triacylglycérols par des techniques isotopiques et RMN métabolomique

JURY

Président :	Richard MAROUN , Professeur, Université Saint-Joseph
Rapporteurs :	Dominique ROLIN , Professeur, Université de Bordeaux Douglas RUTLEDGE , Professeur, AgroParis Tech
Examineurs :	Gérald REMAUD , Professeur, Université de Nantes Toufic RIZK , Professeur, Université Saint-Joseph
Directeurs de Thèse :	Serge AKOKA , Professeur, Université de Nantes Joseph BEJJANI , Professeur associé, Université Saint-Joseph

Remerciements

Cette thèse a été effectuée grâce à une coopération franco-libanaise entre le Laboratoire de Métrologie et de Fractionnement Isotopique (LMFI, UR : Technologie et Valorisation Agroalimentaire) de l'université Saint-Joseph de Beyrouth et l'Unité Mixte de Recherche CEISAM (Chimie et Interdisciplinarité, Synthèse, Analyse, Modélisation) de l'université de Nantes.

Je tiens à remercier le Conseil National de la Recherche Scientifique libanais et le Conseil de la Recherche de l'Université Saint-Joseph de m'avoir accordé une bourse d'étude pour préparer mon doctorat.

Je remercie vivement M. Dominique ROLIN et M. Douglas RUTLEDGE pour m'avoir fait l'honneur d'être rapporteurs, ainsi que M. Richard MAROUN et M. Gérald REMAUD pour m'avoir fait l'honneur de participer au jury de cette thèse.

J'avoue ma reconnaissance à M. Joseph BEJJANI, mon directeur de thèse qui m'a tant consacré de son temps, apporté son soutien et son appui inaltérables durant ces trois années de thèse. Merci infiniment pour tout ce que vous m'avez appris, et de m'avoir fait partager votre passion de la recherche.

Je tiens à remercier également M. Serge AKOKA qui a accepté de diriger mon travail de recherche en orientant ce travail, pour ses conseils précieux en RMN, pour son temps et surtout pour la confiance qu'il m'a accordée.

Un grand merci à M. Toufic RIZK qui a accepté de co-diriger cette thèse et pour le soutien qu'il m'a apporté.

J'adresse ma profonde gratitude à Virginie SILVESTRE, pour son aide précieuse et ses conseils dans l'utilisation des spectromètres.

Merci également à Anne-Marie SCHIPHORST et Mathilde GRAND pour leurs aides précieuses pour les analyses SMRI.

Merci à Denis LOQUET pour les analyses GC et sa disponibilité.

Des remerciements chaleureux sont adressés à Patrick GIRAUDEAU, Benoît CHARIER, Julie LALANDE et Estelle MARTINEAU.

Un grand merci à Mbaye DIOUM, Laetitia ROUGER, Boris GOUILLEUX, Maxime JULIEN et Didier DIOMANDE pour leur sympathie et la bonne ambiance qu'ils ont toujours créées.

Merci à tous les stagiaires que j'ai connus pendant cette thèse.

Et enfin, je tiens à remercier toute ma famille qui m'a donné tout le soutien moral dont j'ai eu besoin, et à qui je dédie mes travaux.

Abréviations et notations

A	Abondance isotopique
AG	Acide gras
ANOVA	Analysis of variance
B ₁	Champ magnétique radiofréquence transversal
C	Carbone
CAM	Crassulacean Acid Metabolism
CDA	Canonical discriminant analysis
COI	Conseil Oléicole International
D	Deutérium
DAG	Diacylglycérol
DEPT	Distortionless Enhancement by Polarization Transfer
DW	Dwell time
FT-PIR	Proche Infrarouge à transformée de Fourier
GC	Gas Chromatography
GC-FID	Gas Chromatography equipped with a flame ionization detector
GC-MS	Gas Chromatography Mass Spectrometry
HSQC	Heteronuclear Single Quantum Coherence
HPLC	High Performance Liquid Chromatography
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
IR	Infrarouge
LDA	Linear discriminant analysis
MLR	Multiple Linear Regression
N	Azote
NOE	Nuclear Overhauser Effect
NS	Nombre de scans
O	Oxygène
P	Phosphore
PCA	Principal component analysis
PL	Prédiction linéaire

PLS	Partial Least Squares
RF	Radiofréquence
RMN	Résonance Magnétique Nucléaire
RMN 1D	Résonance Magnétique Nucléaire à une dimension
RMN 2D	Résonance Magnétique Nucléaire à deux dimensions
S/B	Rapport Signal sur Bruit
SM	Spectroscopie de Masse
SMRI	Spectrométrie de Masse de Rapports Isotopiques
SW	Spectral width
T ₁	Temps de relaxation longitudinale
TAG	Triacylglycérol
Ta	Temps d'acquisition
Tr	Trace
TR	Temps de répétition
UE	Union Européenne
VIP	Variable Importance Projection

Liste des figures

Figure 1 : Structure d'une molécule de triacylglycérol.....	5
Figure 2 : Un champ d'olivier.....	9
Figure 3 : Répartition du marché de l'huile d'olive selon les pays (moyenne juillet 2006-décembre 2011).....	10
Figure 4 : Répartition de la superficie oléicole libanaise.....	11
Figure 5 : Les isotopomères de la molécule éthanol comportant un isotope lourd et stable....	26
Figure 6 : Séquence INEPT-refocalisée dans laquelle les impulsions RF à 180° d'inversion et de refocalisation sont des impulsions adiabatiques.....	26
Figure 7 : Spectre HSQC ^1H - ^{13}C d'une solution d'ibuprofène dans l'acétone- d_6	26
Figure 8 : a) Séquence d'impulsion HSQC permettant de corréler les déplacements chimiques d'un noyau S (^1H) et d'un noyau I (^{13}C). b) Spectre HSQC d'un échantillon de 2-chlorobutane.....	28
Figure 9 : Séquence HSQC optimisée.....	29
Figure 10 : Carte géographique du Liban.....	42
Figure 11 : Procédure à suivre pour prédire l'origine de l'olive.....	71

Liste des tableaux

Tableau 1 : Nomenclature des différents acides gras.....	6
Tableau 2 : Pourcentage d'acides gras dans quelques matrices alimentaires animales et végétales.....	7
Tableau 3 : Limites de variations (%) des TAG dans des huiles d'olive de différents pays....	12
Tableau 4 : Composition en acides gras de l'huile d'olive (%).	13
Tableau 5 : Précision en fonction du TR/T_1 pour un angle d'impulsion $\theta = 90^\circ$	21
Tableau 6 : Différentes études réalisées pour l'authentification des huiles d'olive.....	31

Sommaire

Introduction générale	8
Partie I - Bibliographie	4
I-1. Triacylglycérols	5
I-1.1. Composition	5
I-1.2. Triacylglycérols « marqueurs d'authenticité d'origine »	7
I-2. Huile d'olive	8
I-2.1. Généralités	8
I-2.2. Oléiculture au Liban.....	10
I-2.3. Composition de l'huile d'olive	11
I-2.3.1. Composition en triacylglycérols.....	11
I-2.3.2. Composition en acides gras	12
I-2.3.3. Composés minoritaires	13
I-2.4. Effets des facteurs géoclimatiques, agronomiques et variétaux sur la composition de l'huile d'olive	14
I.2.4.1. Facteurs géoclimatiques.....	14
I.2.4.2. Facteurs agronomiques	15
I.2.4.3. Variété des olives.....	17
I-3. Analyse métabolomique	17
I-3.1. Définition et applications	17
I-3.2. Résonance magnétique nucléaire	18
I-3.3. Conditions quantitatives.....	19
I-4. Analyse isotopique	21
I-4.1. Isotopes : Définition et notations	21
I-4.2. Techniques analytiques utilisées pour l'analyse isotopique	22
I-4.2.1. Spectrométrie de masse de rapports isotopiques	23
I-4.2.2. RMN ¹³ C isotopique	23
I-4.2.3. RMN à deux dimensions	26
I-4.2.4. Séquences de corrélation hétéronucléaire	27
I-4.2.4.1. Détection directe	27
I-4.2.4.2. Détection indirecte ou inverse	27
I-5. Marqueurs d'authenticité de l'huile d'olive	31
I-6. Analyse statistique	35
I-6.1. Analyse de variance univariée	35
I-6.2. Analyse multivariée	35
I-6.2.1. Analyse en Composantes Principales.....	36
I-6.2.2. Analyse Discriminante Linéaire	36
I-6.2.3. Analyse Canonique Discriminante	37
I-6.2.4. Méthodes prédictives : régression linéaire multivariée	37
I-6.2.4.1. Régression linéaire multiple	37
I-6.2.4.2. Régression des moindres carrés partiels	38

I-6.2.4.3. Évaluation d'un modèle de régression.....	39
<i>Partie II - Effets géoclimatiques, morphologiques et temporels sur la composition des huiles d'olive libanaises et leur classification</i>	41
II-1. Introduction.....	42
II-2. Article	43
II-3. Données complémentaires	65
II-4. Conclusion	70
<i>Partie III - Étude isotopomique des triacylglycérols par RMN ¹³C en utilisant des techniques de transfert de polarisation.....</i>	72
III-1. Introduction	73
III-2. Article	74
III-3. Données complémentaires.....	78
III-4. Conclusion.....	81
<i>Partie IV - Développement d'une nouvelle stratégie pour la quantification rapide des acides gras individuels dans les matrices triacylglycérols par RMN ¹³C en utilisant une technique de transfert de polarisation</i>	82
IV-1. Introduction	83
IV-2. Article.....	83
IV-3. Données complémentaires	98
IV-4. Conclusion.....	105
<i>Partie V - Caractérisation et classification des huiles d'olive en utilisant une technique de transfert de polarisation en RMN ¹³C : comparaison avec la GC et la RMN ¹H....</i>	106
V-1. Introduction	107
V-2. Article	107
V-3. Données complémentaires.....	124
V-4. Conclusion.....	127
<i>Partie VI - Analyse isotopomique rapide et précise en utilisant une séquence ¹H-¹³C HSQC : Application sur les triacylglycérols.....</i>	128
VI-1. Introduction	129
VI-2. Article.....	129
VI-3. Données complémentaires	141
VI-4. Conclusion.....	142
<i>Conclusion générale et perspectives</i>	144
<i>Production scientifique</i>	148
<i>Références bibliographiques</i>	150

Introduction générale

Les questions de fraudes et d'adultérations alimentaires suscitent actuellement un regain d'attention, et la sécurité dans ce domaine est devenue une préoccupation internationale, compte tenu du grand nombre d'altérations commises. Comme exemples récents, nous pouvons citer, la présence non-déclarée de viande de cheval dans des produits à base de viande de bœuf, de la mélamine dans le lait, de l'huile de noisette dans l'huile d'olive, ou encore la vente de nombreux produits étiquetés de manière incorrecte ou trompeuse. Face à ces problèmes, la vérification de l'authenticité des produits alimentaires est devenue non seulement un argument décisif pour la commercialisation, mais aussi une nécessité pour protéger à la fois les consommateurs et les producteurs soucieux de qualité. De ce fait, la demande en produits alimentaires de haute qualité avec des indications claires sur leurs origines géographique et botanique devient croissante. Il devient donc crucial de définir dans les produits alimentaires des molécules sondes porteuses de marqueurs d'authenticité d'origine.

Dans le cadre de cette approche, nous avons choisi d'étudier les triacylglycérols, car ces derniers sont considérés comme des molécules quasi universelles dans les produits alimentaires. Le profil de leurs acides gras, leur distribution sur la chaîne du glycérol, le contenu isotopique spécifique aux différents sites de ces molécules ainsi que les composés minoritaires qui les accompagnent varient en fonction des facteurs climatiques, géographiques, temporels, agricoles, et également selon l'espèce productrice. Ces variables métaboliques et isotopiques constituent donc des marqueurs d'origine géographique, botanique ou animale de la matrice alimentaire en question.

À ce niveau, la Résonance Magnétique Nucléaire (RMN) est une technique spectroscopique robuste pour quantifier ces variables avec fiabilité. Toutefois, cette technique peut être limitée par des contraintes de résolution, de sensibilité et de durée d'expérience dans le cas des molécules complexes comme les triacylglycérols. De plus, la littérature ne fait pas mention de l'utilisation de la RMN pour la quantification individuelle des acides gras dans les triacylglycérols.

Durant cette thèse, nous avons développé et optimisé des approches RMN ^1H et ^{13}C (à une et deux dimensions) pour l'étude des matrices riches en triacylglycérols, dans le but d'obtenir un arsenal complet de méthodes pour l'authentification de ces matrices.

Pour mener nos travaux d'authentification, nous avons principalement utilisé la matrice d'huile d'olive comme support d'investigations. À cet égard, une campagne d'échantillonnage d'olives de différentes régions libanaises a été menée. Des extractions mécaniques en laboratoire ont permis d'obtenir des huiles authentiques, puis de construire une base de

données sur les variables RMN correspondantes en fonction des coordonnées géographiques, morphologiques et temporelles des échantillons.

Dans un deuxième temps, les données accumulées ont été analysées par le biais d'outils statistiques univariés et multivariés afin de classifier les échantillons d'huiles selon leurs coordonnées géographiques et d'identifier les traceurs discriminants. De même, des huiles d'olive de différents pays et des huiles végétales de différentes origines botaniques ont été analysées pour tester le potentiel des méthodes analytiques optimisées.

Parallèlement, des études comparatives ont été réalisées entre les méthodes RMN (^1H et ^{13}C) utilisées et la chromatographie en phase gazeuse (GC) dans le but d'identifier les avantages relatifs de chacune de ces techniques et d'évaluer la possibilité d'obtenir un profil complet des acides gras individuels par RMN.

Le manuscrit est divisé en six parties : la première comprend une synthèse bibliographique générale sur les triacylglycérols, l'huile d'olive, et les différentes approches analytiques utilisées dans cette thèse (métabolomiques et isotopiques). Elle présente également les outils statistiques utilisés pour le traitement des données. Les cinq parties suivantes décrivent et discutent les résultats obtenus. Chacune d'entre elles est constituée d'un article accepté ou soumis à un journal international. Tous les articles sont précédés d'une introduction et suivis d'une conclusion.

La deuxième partie est bâtie sur un article intitulé « Geoclimatic, morphological, and temporal effects on Lebanese olive oils composition and classification : A ^1H NMR metabolomic study ». Ce travail porte sur une approche métabolomique par RMN ^1H qui permet de classifier les huiles d'olive provenant de différentes régions libanaises en fonction de leurs coordonnées géographiques et de la morphologie des olives. Les effets de la date de cueillette et de l'année de production sont aussi étudiés.

La troisième partie expose une publication intitulée « ^{13}C isotopomics of triacylglycerols using NMR with polarization transfer techniques », qui présente l'optimisation d'une séquence de transfert de polarisation (INEPT) pour l'analyse des matrices riches en triacylglycérols afin d'atteindre la précision nécessaire (de l'ordre de pour mille) pour les mesures isotopiques.

L'article qui constitue le corps de la quatrième partie est intitulé « A novel quantification strategy of individual fatty acids in triacylglycerol matrices using optimized ^{13}C -NMR with a polarization transfer technique ». Il rapporte la quantification des acides gras individuels dans les triacylglycérols par spectroscopie RMN couplée à des analyses de régressions linéaires multivariées.

Un article intitulé « Olive oil characterization and classification by ^{13}C -NMR with a polarization transfer technique: A comparison with gas chromatography and ^1H -NMR » fait l'objet de la cinquième partie. Il expose une comparaison entre les techniques RMN (^1H et ^{13}C) et GC, en termes d'informations obtenues, de précision et de potentiels de classification des huiles d'olive.

La sixième partie présente un article intitulé « Precise and rapid isotopomic analysis by ^1H - ^{13}C HSQC: Application to triacylglycerol matrices ». Il comporte l'optimisation de la séquence 2D HSQC dans le but d'obtenir la précision requise pour l'analyse isotopique. Cette séquence a été appliquée pour distinguer différentes huiles végétales commerciales selon leur origine géographique ou botanique.

Enfin, le manuscrit s'achève par une conclusion générale qui synthétise l'ensemble des résultats obtenus et ouvre la voie à d'autres applications et perspectives.

Partie I - Bibliographie

I-1. Triacylglycérols

I-1.1. Composition

Les triacylglycérols (TAG) sont présents dans une très large gamme de produits alimentaires. Ils se trouvent dans les matières premières animales (viande, poisson, lait, œuf) et végétales (grains, fruits et légumes). Ils peuvent donc être considérés comme molécules sondes quasi universelles pour l'authenticité des aliments.

Les TAG résultent de la condensation, de trois molécules d'acides gras (qui ne sont pas forcément identiques) avec une molécule de glycérol (**Figure 1**).

Les principaux acides gras (AG) des matrices alimentaires riches en TAG ainsi que leurs nomenclatures sont donnés dans le **Tableau 1** [1,2,3]. Les acides gras les plus répandus dans les matrices animales et végétales ont de 14 à 20 carbones, avec une prédominance des acides à 16 ou 18 carbones. Ceux dont le nombre de carbones est inférieur à 12 se trouvent dans le beurre et le lait de mammifères.

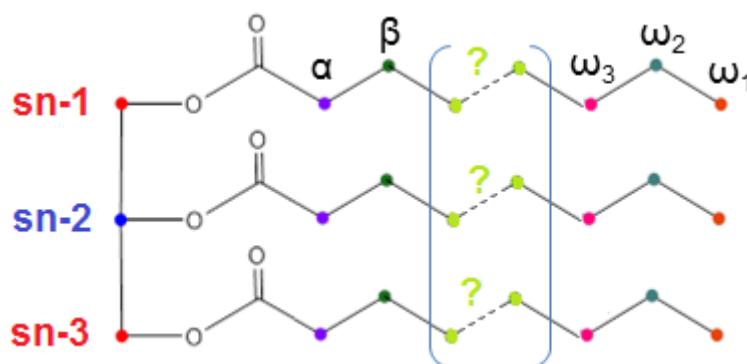


Figure 1 : Structure d'une molécule de triacylglycérol.

sn-1 et **sn-3** indiquent les positions externes des acides gras sur le squelette de glycérol, et **sn-2** la position interne. L'élément α est le carbone 2 (C2) de la chaîne d'acide gras nommé carbone alpha, et β est le carbone 3 (C3) de la chaîne nommé carbone beta. ω_1 est le dernier carbone de la chaîne aliphatique, ω_2 et ω_3 sont respectivement le deuxième et le troisième carbone à partir du dernier carbone de la chaîne aliphatique.

Les TAG sont des mélanges complexes de molécules ; leur profil est défini par la longueur de chaîne des acides gras, leur degré d'insaturation, la position de leurs doubles liaisons, et leur distribution sur le squelette du glycérol.

La composition de ces molécules varie qualitativement et quantitativement en fonction de l'espèce qui les synthétise (**Tableau 2**) [2,4,5,6,7]. Elle est également affectée par des facteurs intrinsèques, tels que l'altitude [8], et la latitude [9] du terroir d'origine, et par des facteurs extrinsèques liés aux pratiques agricoles adoptées (degré de maturité à la cueillette, irrigation, mode d'alimentation dans le cas des produits d'origine animale [10], etc.).

Tableau 1 : Nomenclature des différents acides gras.

	Formule Cn:x (y) ^a	Nom de l'acide gras
<i>AG saturés</i>	C4:0	Butyrique
	C6:0	Caproïque
	C8:0	Caprylique
	C10:0	Caprique
	C12:0	Laurique
	C14:0	Myristique
	C16:0	Palmitique
	C17:0	Margarique
	C18:0	Stéarique
	C20:0	Arachidique
	C22:0	Béhénique
C24:0	Lignocérique	
<i>AG monoinsaturés</i>	C16:1 (7)	Hypogéique
	C16:1 (9)	Palmitoléique
	C17:1 (8)	Margaroléique
	C18:1 (9)	Oléique
	C18:1 (11)	Vaccénique
	C20:1 (11)	Gondoïque
	C22:1 (9)	Érucique
<i>AG polyinsaturés</i>	C18:2 (9,12)	Linoléique
	C18:3 (9,12,15)	Linoléinique
	C18:4 (6,9,12,15)	Stéaridonique
	C20:4 (5,8,11,14)	Arachidonique
	C20:5 (5,8,11,14,17)	Eicosapentaénoïque
	C22:5 (4,7,10,13,16)	d'Osbond
	C22:5 (7,10,13,16,19)	Clupanodonique
	C22:6 (4,7,10,13,16,19)	Cervonique

^a **Cn:x (y)** avec n le nombre de carbones dans la chaîne de l'acide gras, x le nombre de doubles liaisons, et y indique la position de la double liaison en partant du groupe carboxyle (1^{er} carbone).

Les acides gras en position interne (sn-2) dans la molécule de glycérol sont quasiment toujours insaturés. En général, les acides linoléinique et linoléique occupent cette position [11,12,13].

Tableau 2 : Pourcentage d'acides gras dans quelques matrices alimentaires animales et végétales.

		Saturés	Monoinsaturés	Polyinsaturés		
				$n = 2$	$n = 3$	$n > 3$
<i>Matrices végétales</i>	Arachide	20	50	30	Tr	
	Olive	17	72	10	1	
	Colza	7	62	31		
	Soja	15	24	53	7	
	Tournesol	12	19	69	Tr	
	Noisette	8	81	11		
	Blé	20	16	60	4	Tr
	Noix	12	16	61	11	
	Sésame	15	41	44	Tr	
	Maïs	13	28	58	1	
	Riz	28	30	41	1	Tr
	Palme	50	40	10	Tr	
	Noix de coco	91	7	2		
	Pépins de raisin	13	20	67	Tr	
	Concombre	42	3	29	26	
	Pomme	28	7	54	11	
	Banane	46	16	16	22	
	Beurre de cacao	62	35	3		
<i>Matrices animales</i>	Beurre laitier	63	32	4	1	
	Bœuf	51	45	4	Tr	Tr
	Porc	43	48	8	1	Tr
	Poulet	30	49	21	Tr	Tr
	Œuf	38	46	16		
	Huile de poisson	21	52	2	1	24
	Sardine	27	31	2	1	39
	Saumon	28	43	2	1	26

Avec n : nombre de doubles liaisons, Tr < 0,5 %

I-1.2. Triacylglycérols « marqueurs d'authenticité d'origine »

En plus des paramètres métabolomiques des TAG présentées dans le paragraphe I-1.1., les rapports isotopiques D/H, $^{13}\text{C}/^{12}\text{C}$ et $^{18}\text{O}/^{16}\text{O}$ globaux et spécifiques aux différents sites des acides gras et du glycérol sont aussi affectés par les facteurs climatiques, géographiques et agricoles. Le rapport $^{13}\text{C}/^{12}\text{C}$ est surtout affecté par le type de photosynthèse adopté par la plante (C3, C4 ou CAM) [14]. Quant aux rapports D/H et $^{18}\text{O}/^{16}\text{O}$, ils sont reliés aux conditions environnementales et aux fractionnements isotopiques qui se produisent lors de l'évaporation, de la transpiration, de la photosynthèse et d'autres réactions [15].

Ces variables isotopiques ainsi que les variables métabolomiques sont des marqueurs d'authenticité potentiels puisqu'elles permettent de remonter à l'origine géographique, botanique ou animale de la matrice alimentaire en question. Elles peuvent ainsi servir pour la traçabilité de l'origine de l'alimentation des animaux [16].

Dans la littérature, de nombreuses études ont fait appel à ces traceurs pour classifier des produits riches en TAG selon leur origine.

Dans le domaine des produits d'origine animale, nous pouvons citer la classification d'échantillons de saumons d'élevage et sauvages, basée sur des rapports D/H et sur la composition des huiles et des graisses correspondantes [17]. Une autre étude décrit la possibilité de caractériser des saumons issus de l'aquaculture biologique ou d'élevage conventionnel en se référant à la composition des lipides correspondants et au rapport $^{15}\text{N}/^{14}\text{N}$ [18]. La caractérisation du lait de vache [19] et de la viande de bœuf [20] selon la location géographique et l'alimentation est possible par le biais du rapport $^{18}\text{O}/^{16}\text{O}$ de l'eau de la matrice et de variables obtenues par RMN des TAG. La différenciation entre laits bio et conventionnels est obtenue en utilisant la teneur en acides gras (oméga-3) et le rapport $^{13}\text{C}/^{12}\text{C}$ des matières grasses comme descripteurs [21]. D'ailleurs, des différences significatives dans la composition en acides gras permettent la caractérisation de l'origine des œufs frais [22].

Dans le cas des produits d'origine végétale, le profil des acides gras et autres variables (déterminés par RMN ^1H et ^{31}P) permettent de classifier les huiles selon leur origine botanique [23]. Une autre étude a été menée en vue d'identifier l'origine botanique et géographique du blé en se basant sur les variables obtenues par les techniques RMN ^1H et spectrométrie de masse de rapports isotopiques (SMRI) [24].

I-2. Huile d'olive

Dans le cadre de ce travail de thèse, nous avons choisi la matrice de l'huile d'olive pour mener nos investigations. Des informations générales sur la culture de l'olivier, les caractéristiques de l'huile d'olive, sa production et sa consommation mondiale seront abordées brièvement dans cette partie. De plus, l'oléiculture au Liban sera aussi présentée puisque notre étude par RMN ^1H métabolomique est réalisée sur des échantillons libanais (partie II). Les effets des différents facteurs environnementaux et agronomiques sur la composition de l'huile d'olive seront également exposés.

I-2.1. Généralités

La culture de l'olivier est apparue vers 3000 av. J.-C, en Palestine, en Syrie et Phénicie. Elle s'est ensuite développée en Grèce, avant de s'étendre vers la Crète et l'Égypte, pour conquérir l'ensemble du bassin méditerranéen où l'huile d'olive occupe une place très importante dans l'alimentation.

L'olivier est un arbre qui aime le soleil, la chaleur et une pluviométrie moyenne, autrement dit un climat de type méditerranéen. Il nécessite un sol bien drainé et calcaire, les sols filtrants comportant des cailloux conviennent mieux à l'olivier que les terres trop argileuses.



Figure 2 : Un champ d'olivier.

La cueillette qui dépend du climat et de la maturation des olives, varie d'une région à l'autre, et a lieu généralement entre octobre et janvier.

La pratique de cueillette la plus adoptée est le gaulage, qui consiste à faire tomber les olives sur des bâches à l'aide de grandes perches. Cette technique a pour conséquence de casser les branches, ce qui réduit les possibilités de bourgeonnement l'année suivante et conduit à l'alternance de la production. Le coût de la récolte est la composante principale du prix de l'huile d'olive.

Ce produit est bien reconnu pour ses nombreuses vertus. Ses propriétés organoleptiques et ses bienfaits pour la santé sont liés à sa composition en acides gras, où l'acide oléique est le composant primordial, mais aussi à la présence de composants minoritaires, tels que le squalène, l'oléocanthal et les antioxydants naturels.

De nombreuses recherches rapportent une diminution du risque de maladies cardiovasculaires [25], de certains cancers [26,27], d'infarctus du myocarde [28,29], de l'athérosclérose [30,31,32,33] de la mortalité par maladie coronarienne et de la mortalité totale chez les personnes consommant de l'huile d'olive [34].

La haute qualité de l'huile d'olive, ses propriétés gustatives et médicinales lui permettent de se placer sur le marché à des prix élevés et augmentent sa consommation.

La production mondiale d'huile d'olive est répartie entre trois groupes de pays :

- Les pays membres de l'Union Européenne (UE) avec 76% de la production mondiale. Cependant, l'Espagne est l'acteur principal du marché en assurant 62% de la production totale de l'UE et 40 à 45% de la production mondiale.

- Les 16 autres pays membres du Conseil Oléicole International (COI) (Albanie, Algérie, Argentine, Croatie, Égypte, Iran, Iraq, Israël, Jordanie, Liban, Libye, Maroc, Monténégro, Syrie, Tunisie et Turquie) représentent une moyenne de 22%.
- D'autres pays du monde fournissent 2 % de la production mondiale [35].

À l'extérieur de l'UE, la Tunisie et la Syrie ont toujours été des producteurs importants d'huile d'olive, et cherchent à ce titre à augmenter continuellement leur niveau de production (**Figure 3**).

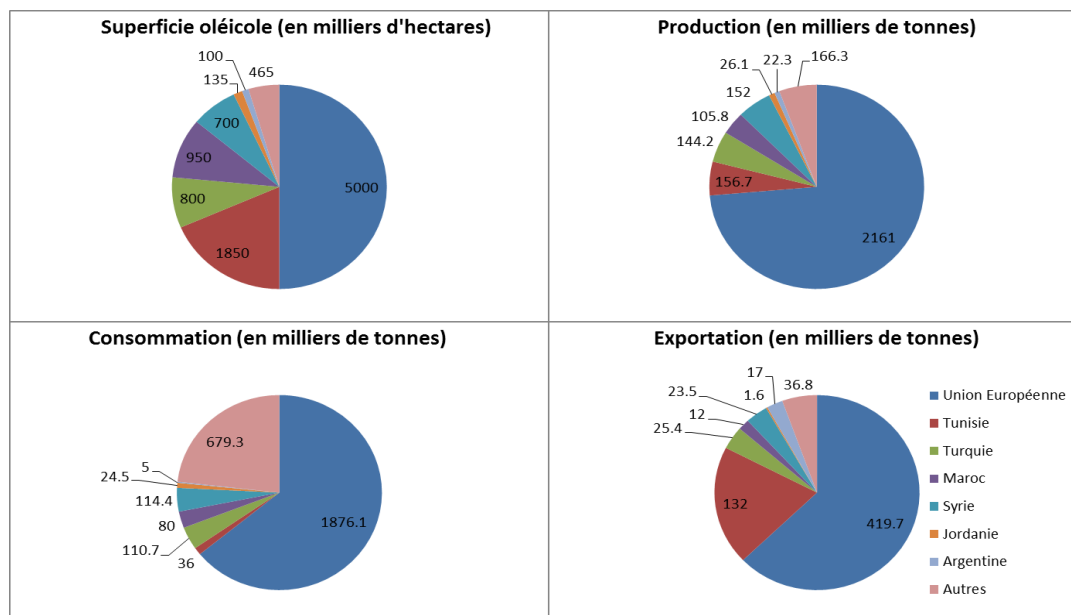


Figure 3 : Répartition du marché de l'huile d'olive selon les pays (moyenne juillet 2006-décembre 2011).

D'après les données statistiques du COI, le commerce de l'huile d'olive a marqué une hausse d'environ 4% par an de 1990 à 2013, bien que l'accroissement de la production mondiale n'ait pas dépassé 2,6%. Ceci montre que la consommation mondiale est en hausse grâce à des pays traditionnellement non consommateurs (États-Unis, Canada, Australie, Brésil, Japon et Chine), et qui représentent une part croissante dans cette évolution [36].

I-2.2. Oléiculture au Liban

Le capital oléicole au Liban repose sur environ 12 millions d'oliviers repartis sur 56,300 ha. Ces derniers sont principalement situés sur les collines voisines de la zone littorale et peuvent culminer jusqu'à 1000 mètres d'altitude dans une zone climatique très favorable à leur culture. La pluviométrie hivernale moyenne y varie entre 600 mm dans le Sud et 1000 mm dans le Nord du pays.

La superficie oléicole est répartie ainsi : environ 40% dans le Nord, 39% dans le Sud, 15% dans le Mont-Liban, et 6% dans la plaine de la Béqaa (**Figure 4**).

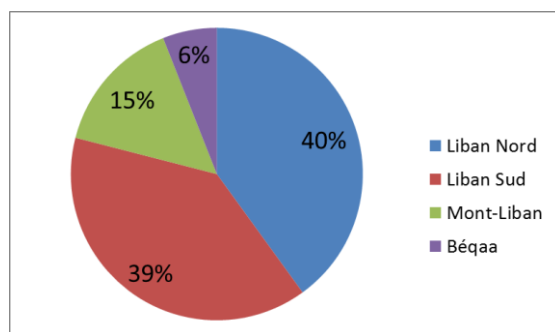


Figure 4 : Répartition de la superficie oléicole libanaise.

La production d'olive au Liban oscille d'une façon très marquée au grès des conditions climatiques entre 15000 et 35000 tonnes/an en moyenne. Le rendement connaît des variations importantes dues principalement aux conditions climatiques, au phénomène de l'alternance et aux techniques culturales [37,38].

En général, la cueillette des olives au Liban commence après les premières pluies d'octobre. Il n'existe pas de données scientifiques qui déterminent le stade de maturité optimale de l'olive pour fixer la date de cueillette.

I-2.3. Composition de l'huile d'olive

L'huile d'olive est composée principalement de TAG (95 à 98%), d'un mélange complexe de composés minoritaires (2 à 5%) et d'une faible quantité d'acides gras libres (< 1%).

I-2.3.1. Composition en triacylglycérols

Vingt TAG ont été identifiés dans l'huile d'olive, dont les principaux sont les suivants : OOO (27,53 – 59,34%), POO + SOL (12,42 – 30,57%), OOL + LnPP (4,14 – 17,46 %), POL + SLL (2,69 – 12,31%) et SOO + OLA (3,17 – 8,39%) (P : palmitique, O : oléique, S : stéarique, L : linoléique, Ln : linoléique, A : arachidique) [39]. De nombreuses recherches ont montré que la composition de l'huile d'olive en TAG varie en fonction de l'origine géographique. Le **Tableau 3** représente les limites des variations (ou la moyenne) de chaque TAG pour des huiles d'olive provenant de différents pays [40].

Tableau 3: Limites de variations (%) des TAG dans des huiles d'olive de différents pays.

<i>Origine</i>	<i>France</i>	<i>Portugal</i>	<i>Tunisie</i>	<i>Italie</i>	<i>Espagne</i>
TAG	Moy (Min - Max)	Min - Max	Moy	Moy	Moy
LLL	0,13 (0,01 - 0,90)	0,02 - 0,06	0,8	0,09	
OLnL + PoLL*	0,24 (0,02 - 0,85)	0,1 - 0,6	0,6	0,26	0,9
PLnL	0,06 (Tr - 0,29)	0,01 - 0,3			
LOL	1,90 (0,13 - 6,20)	0,0 - 5,1	5,8		0,3
OLnO + PoOL*	1,36 (0,52 - 2,46)	2,1 - 3,2	1,5	1,04	1,0
PLL + PoPoO*	0,55 (0,05 - 2,17)	0,4 - 1,2	2,8	0,27	0,5
PLnO + PPoL + PPoPo*	0,64 (0,25 - 1,35)	0,1 - 1,4	1,1		0,3
LOO + LnPP*	13,93 (7,48 - 23,27)	4,2 - 16,7	18,2	16,48	10,4
PoOO	1,10 (0,14 - 3,21)	1,1 - 4,0			1,1
PLO + SLL*	5,57 (2,16 - 11,71)	2,3 - 11,8	12,3	4,41	4,5
PoOP + SPoL + SOLn + SPoPo*	0,66 (0,17 - 1,63)	0,1 - 2,2	1,2		0,4
PLP	0,46 (Tr - 1,53)	0,2 - 1,1	2,1	0,12	0,7
OOO + PoPP*	44,69 (27,32 - 58,76)	25,6 - 57,2	21,8	44,05	43,1
SLO	0,52 (Tr - 1,77)			1,12	
POO	20,03 (14,69 - 27,65)	19,1 - 32,2	20,0	17,68	23,1
POP	3,08 (0,45 - 5,38)	2,1 - 5,4		1,77	2,9
SOO	3,72 (0,49 - 7,22)	2,6 - 8,0	3,7	4,47	3,6
POS + SLS*	0,85 (0,37 - 3,47)	0,3 - 2,7	1,2	0,9	0,4
PPS	0,52 (0,23 - 1,03)	0,4 - 1,9	0,5		0,6

*TAG présent en faible quantité

I-2.3.2. Composition en acides gras

L'huile d'olive se caractérise par une teneur élevée en acides gras monoinsaturés, en particulier l'acide oléique, par une teneur relativement faible en saturés et par une teneur acceptable d'acides gras polyinsaturés. Ces derniers ne peuvent pas être biosynthétisés par l'homme, ils sont apportés par le régime alimentaire. Le **Tableau 4** représente les limites des variations (et la moyenne) des AG présents dans l'huile d'olive [41].

Tableau 4 : Composition en acides gras de l'huile d'olive (%).

Acides gras	Limites de variations (moyenne %)
Palmitique	7,5-20 (11,8)
Hypogéique	0,09-0,20 (0,12)
Palmitoléique	0,3-3,5 (0,81)
Margarique	0,03-0,20 (0,08)
Margaroléique	0,06-0,36 (0,15)
Stéarique	0,5-5 (2,26)
Oléique	55-83 (72,6)
Vaccénique	1,2-3,9 (2,3)
Linoléique	3,5-21 (7,9)
Linoléinique	<1,5 (0,65)
Arachidique	<0,6 (0,37)
Gondoïque	0,21-0,40 (0,28)
Béhénique	<0,3 (0,11)
Lignocérique	<0,2 (0,05)

I-2.3.3. Composés minoritaires

Les composés minoritaires de l'huile d'olive sont essentiellement le squalène, les terpènes, les stérols, les composés phénoliques, et les dérivés du tocophérol.

- Le squalène est un hydrocarbure et un triterpène. Relativement aux autres huiles végétales, l'huile d'olive est la plus riche en squalène (400-450 mg/100 g) [27]. Ce composé est le précurseur des stéroïdes. Parfois, il peut être responsable de l'augmentation du cholestérol sanguin [42]. Des études ont montré que le squalène diminue toutefois le risque de cancer [43,44,45].
- Les composés phénoliques les plus répandus dans l'huile d'olive sont l'acide vanillique, l'acide gallique, l'acide caféique, le tyrosol, l'hydroxytyrosol, le 1-acétoxypinorésinol et l'oleuropéine [46,47]. Les phénols possèdent une activité antioxydante [48,49]. Des études ont rapporté que la consommation d'huile d'olive diminue le stress oxydatif dans l'organisme [50,51,52,53] et peut protéger contre l'apparition de certaines maladies, y compris le cancer [26,27] et les troubles cardiovasculaires [50].

- Les tocophérols existent sous forme α , β , γ et δ . Dans l'huile d'olive, il y a une prédominance du α -tocophérol. Ces composés ont une activité vitaminique et jouent le rôle d'antioxydants [54].
- Les stérols sont des composés tétracycliques. Ils ont des effets bénéfiques sur les concentrations sériques de cholestérol [55]. Le β -sitostérol est le principal stérol de l'huile d'olive, il a une action anticarcinogène [56,57,58,59].
- Les terpènes se trouvent en très faible quantité dans l'huile d'olive [60].

I.2.4. Effets des facteurs géoclimatiques, agronomiques et variétaux sur la composition de l'huile d'olive

La variation de la composition et des propriétés de l'huile d'olive provient d'une interaction complexe entre les facteurs géoclimatiques, les facteurs agronomiques et la variété des olives. Les effets de ces différents facteurs sont présentés dans les paragraphes suivants.

I.2.4.1. Facteurs géoclimatiques

De nombreuses études ont rapporté l'influence du climat et de l'altitude de la parcelle d'oliviers sur la composition chimique de l'huile qui en résulte. La même variété d'olive est bien affectée par l'origine géographique et les variations climatiques [61].

Les analyses statistiques appliquées à des échantillons d'huile d'olive d'Italie [8] et d'Espagne [62] ont montré des différences significatives entre les huiles provenant des régions d'altitudes différentes.

Il a été constaté que la teneur en acide oléique est plus élevée dans les huiles provenant des hautes altitudes, alors que celles des acides linoléique et linoléique sont plus basses. En effet, la quantité d'acide oléique diminue lorsque la quantité d'acide linoléique augmente [9,63,64,65]. De plus, il a été démontré qu'il existe une corrélation négative entre l'acide oléique et l'humidité de l'atmosphère [69,66].

Il a été également observé que la teneur en acides oléique et linoléique ne varie pas dans les huiles d'olive provenant de deux régions grecques ayant respectivement 100 m et 400 m d'altitude [67].

De même, l'acide linoléique, le β -sitostérol et certains alcools terpéniques sont des marqueurs qui permettent de distinguer entre les huiles provenant des altitudes inférieures à 400 m et supérieures à 700 m [68].

D'autres recherches ont montré que l'acide palmitoléique [69] et l'acide palmitique [70] diminuent en montant en altitude, et que les acides insaturés augmentent en passant d'une région chaude à une autre froide [71].

D'autre part, il a été remarqué que les huiles provenant des basses altitudes se caractérisent par des teneurs plus élevées en stérols, en composés phénoliques, en α -tocophérol et par des teneurs plus faibles en chlorophylles et en acides gras insaturés que dans les huiles provenant des hautes altitudes [63,72,73].

Certains auteurs ont rapporté que les huiles d'olive italiennes de hautes altitudes se caractérisent par un faible contenu en terpènes et par un contenu en squalène élevé ; et que les huiles du nord sont pauvres en terpènes et celles du sud sont plus riches en squalène [74].

D'autres recherches effectuées sur des huiles italiennes ont constaté que la quantité de l'acide oléique diminue à des latitudes plus élevées, et celle de l'acide linoléique augmente [9].

En revanche, des études sur des huiles tunisiennes ont montré que les huiles du sud contiennent une teneur en acide palmitique plus élevée que celles du nord, tandis qu'une teneur élevée en acide oléique est obtenue dans les huiles du nord [75].

Les acides palmitique, palmitoléique, linoléique et linoléique ont tendance à être beaucoup plus élevés dans les huiles provenant des régions très proches de l'équateur (comme l'Argentine), où les étés sont très chauds et la température élevée. En général, les températures élevées conduisent à des huiles pauvres en acide oléique, tandis que, les huiles des régions froides comme l'Australie peuvent avoir un profil d'acide gras très différent, puisqu'elles sont très riches en oléique et stéarique [76].

D'après d'autres expériences réalisées sur les huiles d'olive portugaises, il a été confirmé que les acides palmitique et palmitoléique augmentent en allant du nord au sud [77].

D'après ces données, on déduit que les huiles provenant des régions de hautes altitudes sont plus riches en acide oléique et en squalène. Celles de basses altitudes sont plus riches en acides linoléique, linoléique, palmitique et palmitoléique, en β -sitostérol, en terpènes, en composés phénoliques et en tocophérols. De plus, les effets de l'altitude et de la latitude sont généralement cohérents.

I.2.4.2. Facteurs agronomiques

L'ajout d'engrais conventionnels, en particulier l'azote et le phosphore, peut affecter la composition des acides gras dans l'huile d'olive. Lorsque le niveau de phosphore est élevé, la teneur en acides oléique et linoléique augmente. En revanche, lorsque le niveau d'azote augmente au-delà du niveau optimal, l'acide oléique diminue, tandis que l'acide linoléique augmente, ce qui diminue la stabilité de l'huile. L'excès de fertilisation en azote permet d'abaisser la quantité totale de polyphénols dans l'huile d'olive.

Cependant, les données rapportées dans la littérature ne montrent pas que le type de sol a un effet marqué sur la composition de l'huile d'olive [76].

D'un autre côté, l'irrigation a un effet significatif sur la croissance, la production des olives, le rendement en huile [78] et la synthèse des phénols. Ces derniers diminuent avec l'irrigation [79,80]. En effet, l'activité de phénylalanine ammonia-lyase, responsable de la synthèse des composés phénoliques, augmente sous le stress hydrique ; ce phénomène a été démontré dans plusieurs études précédentes [81,82,83,84].

L'effet de l'irrigation sur le profil des acides gras présente des contradictions. Certains chercheurs ont rapporté que le profil de l'huile en acides gras n'est pas affecté par l'irrigation [78,85]. Tandis que d'autres auteurs ont trouvé que la teneur en acide oléique augmente ; celle des acides saturés et des acides polyinsaturés diminue avec l'irrigation [74,86].

La maturité des olives est un autre facteur qui affecte la composition de l'huile. Il a été confirmé que les contenus en stérols [87], en squalène [88,89], en composés phénoliques [90] et en tocophérols [91] diminuent au fur et à mesure que l'olive mûrit. Certains auteurs ont trouvé que la teneur en acide oléique reste constante durant la maturité ou présente une légère augmentation et que les acides gras saturés (palmitique et stéarique) et l'acide linoléique diminuent, tandis que l'acide linoléique augmente. L'huile devient plus riche en acides polyinsaturés et donc moins stable. Ces variations sont dues aux différences de température et de quantité de précipitations entre les stades de maturité [88,92,93,94,95,96,97].

Dans d'autres études, l'effet de la maturité des olives sur les acides gras est moins clair, aucune différence significative n'a été mise en évidence [90,98,99].

En revanche, d'autres recherches effectuées en Espagne, ont montré que le profil en acide gras varie en fonction de la maturité, mais cette observation ne semble pas identique pour toutes les variétés. Pour les variétés Picudo, Chetoui et Picholine, la quantité d'acides gras monoinsaturés augmente avec la maturité tandis que la quantité des acides saturés et polyinsaturés diminue. Par contre, pour les variétés Picual et Arbequina, l'effet inverse est observé. Il a été montré également que la variété Barnea peut être récoltée dans les derniers stades de maturité, car l'huile tend à s'accumuler sans aucun changement significatif de sa qualité. À l'inverse, la variété Souris doit être récoltée tôt, car avec la maturité, l'acide oléique diminue, l'acide linoléique et les acides gras libres augmentent, les olives tombent de l'arbre, et le niveau de polyphénols diminue ce qui fait que l'huile devient moins stable à l'oxydation [100]. Il a été remarqué que la maturation des olives des hautes altitudes est plus lente que celle des basses altitudes [63,101].

I.2.4.3. Variété des olives

La composition de l'huile résulte d'une interaction entre le génotype de l'olive et son environnement (origine géographique, climat, etc.).

Des études ont montré que la composition en acides gras des huiles d'olive tunisiennes est nettement influencée par le facteur variétal. De même, il a été noté que les proportions des diverses fractions triglycériques varient selon les variétés [102]. Le ratio entre les acides saturés et insaturés peut également varier d'une variété à une autre. Une autre étude sur les huiles italiennes a montré que les proportions d'acides gras sont influencées par les variétés, et que les plus discriminants dans cette analyse sont les acides palmitique, palmitoléique, margarique, oléique et linoléique [103].

I-3. Analyse métabolomique

L'un des objectifs de cette étude était de caractériser les huiles d'olive libanaises et de déterminer les variations de leur composition en fonction de différents facteurs géographiques, variétaux et temporels. Cela nécessite une analyse métabolomique de cette matrice.

Une brève définition de l'analyse métabolomique, ses applications, et les principales techniques utilisées pour l'analyse des TAG sont présentées dans les paragraphes suivants.

I-3.1. Définition et applications

L'analyse métabolomique consiste en l'identification et la quantification exhaustive et non sélective des différents métabolites de faibles poids moléculaires (< 1500 Da) présents dans un échantillon biologique dans des conditions données [104,105]. La métabolomique permet de faire le lien entre génotypes et phénotypes [106] et permet d'éclaircir l'influence de l'environnement sur le phénotype exprimé.

De nombreuses études ont montré l'apport de l'approche métabolomique dans différents secteurs tels que : la santé, la nutrition, et l'environnement. À titre d'exemple, cette approche a été utilisée pour la détection des marqueurs précoces du développement des maladies (cancer, maladies cardio-vasculaires, Alzheimer, etc.) et le diagnostic des perturbations associées [107,108], pour constater les effets de l'alimentation et des médicaments sur l'organisme [109], pour détecter certains polluants [110], pour étudier la variation de la composition d'une matrice donnée en fonction des facteurs géo-climatiques [64].

Dans le domaine de la métabolomique, deux stratégies ont été développées, la première est dite « profilage métabolique » qui permet la quantification des composés appartenant à une voie métabolique donnée, cela nécessite principalement l'utilisation de standards ce qui

constitue une limitation. La deuxième approche est dite « empreinte métabolique », elle permet de comparer qualitativement et quantitativement l'ensemble de métabolites dont la teneur change en réponse à différents traitements [111].

De nombreuses techniques analytiques [112] telles que la chromatographie en phase gazeuse (GC) [113,114,115], la chromatographie liquide haute performance (HPLC) [39], la spectrométrie de masse (SM) [40], l'infrarouge (IR) [116,117] et la résonance magnétique nucléaire (RMN) [118] sont utilisées pour l'analyse métabolomique des TAG dans le but d'identifier et de quantifier leur composition en acides gras, ce qui permet de remonter à leurs origines géographique et botanique [119].

Dans le cadre de cette thèse, nous avons adopté la technique RMN pour l'analyse des triacylglycérols parce qu'elle peut fournir à la fois des informations métabolomiques et isotopiques.

I-3.2. Résonance magnétique nucléaire

La résonance magnétique nucléaire (RMN) utilise l'absorption d'ondes radiofréquences par les noyaux atomiques placés dans un champ magnétique. C'est une technique quantitative, non-destructive, non sélective, robuste et très reproductible; de plus la préparation de l'échantillon est très simple et rapide [120]. La RMN permet de détecter et de quantifier les différents métabolites dans des mélanges complexes [118]. De même, elle fournit des informations très précises sur les structures moléculaires des composés détectés [121]. Les inconvénients de cette technique sont son manque de sensibilité (10^{-5} mol/L) qui ne permet pas de détecter les métabolites présents en faible quantité et un coût d'équipement élevé.

Les techniques RMN ^1H , ^{13}C et ^{31}P ont été largement utilisées pour l'analyse de l'huile d'olive, ils donnent des informations complémentaires sur l'identité des constituants majeurs et mineurs de l'huile. Le ^{31}P a un intérêt particulier pour l'analyse des phospholipides [122]. La RMN ^{31}P a été utilisée comme une technique complémentaire pour l'analyse de l'huile d'olive [123,124].

En général, les spectres RMN ^1H des mélanges sont beaucoup plus compliqués que les spectres ^{13}C et ^{31}P , cela est dû à la gamme restreinte de déplacement chimique des protons (environ 15 ppm). Une meilleure discrimination des signaux est donc observée en RMN ^{13}C et ^{31}P grâce à des gammes spectrales beaucoup plus étendues (environ 250 ppm et 1000 ppm pour le ^{13}C et le ^{31}P respectivement). De plus, les spectres ^{13}C et ^{31}P sont généralement obtenus suite à un découplage des protons, ce qui permet d'obtenir des signaux simples et par la suite des spectres plus propres.

Pendant de nombreuses années, la RMN ^1H a été la méthode analytique la plus utilisée, du fait de l'abondance naturelle de 99,98% du ^1H dans toute molécule organique. Cette méthode permet de quantifier les AG. Mais contrairement à la GC, elle ne permet pas de déterminer les AG individuels en raison de chevauchements de signaux.

De ce fait, la quantification de la somme des acides gras saturés, monoinsaturés et polyinsaturés peut être obtenue au moyen d'équations mathématiques et ceci en utilisant les aires des signaux appropriés. Seuls, les acides linoléique et linoléique peuvent être déterminés individuellement [125,126].

De plus, la sensibilité élevée de la RMN ^1H permet la détection et la quantification des composés mineurs de l'huile d'olive, tels que les alcools, les aldéhydes, les diacylglycérols, les stérols et le squalène.

Pratiquement toutes les analyses effectuées par RMN ^1H peuvent être également effectuées par RMN ^{13}C qui permet de plus d'obtenir des informations sur la distribution des acides gras saturés, des acides oléique, linoléique, et linoléique sur le squelette du glycérol [127,128,129]. Ces résultats sont obtenus à partir des régions des carbonyles, des carbones oléfiniques et des carbones α (C2) du spectre ^{13}C . Toutefois la RMN ^{13}C étant beaucoup moins sensible que la RMN ^1H , les temps de mesure sont beaucoup plus longs et des composés mineurs (en faible concentration) ne peuvent pas être détectés.

I-3.3. Conditions quantitatives

Au cours d'une expérience RMN, les signaux détectés sont proportionnels à la concentration des composés correspondants. La mesure de l'aire (en RMN 1D) ou du volume (en RMN 2D) des pics permettra d'accéder à cette concentration. Cependant, plusieurs facteurs peuvent modifier ce rapport de proportionnalité, tels que les conditions de mesures, la température de la sonde et l'état du spectromètre.

Pour pouvoir se placer dans les conditions quantitatives, il faut que la mesure des surfaces des pics soit la plus précise possible [130]. De ce fait, la même séquence d'impulsions est répétée NS fois et les signaux recueillis sont accumulés afin d'améliorer le rapport signal sur bruit (S/B).

$S = NS.s$, avec s le signal recueilli lors d'une expérience unitaire, et NS le nombre de scans.

Par contre, le bruit s'additionne d'une façon incohérente et devient $B = b\sqrt{NS}$, où b est le bruit produit par une expérience unitaire.

Le rapport S/B est donc proportionnel à \sqrt{NS} au bout de NS scans. Ce rapport doit être ajusté pour atteindre la précision souhaitée.

La relation entre précision et rapport S/B est donnée par :

$$\frac{\Delta A}{A} \geq \frac{1}{2 \left(\frac{S}{B}\right)} \quad (1)$$

avec A : la surface du pic considéré et ΔA : l'erreur maximale commise pour déterminer A . Ainsi, un rapport S/B d'au moins 50 est essentiel pour obtenir une précision de 1%.

D'un autre côté, deux autres paramètres doivent être pris en compte pour que l'expérience RMN soit juste : l'angle d'impulsion et le temps de répétition (TR) qui correspond au délai de retour à l'équilibre entre deux scans.

Après une impulsion, l'aimantation est basculée d'un angle θ . Elle doit pouvoir revenir à l'équilibre suivant l'axe z pendant la durée TR qui précède l'impulsion suivante.

L'aimantation longitudinale atteint une valeur d'équilibre (M_{eq}) après une impulsion d'un angle de bascule θ et un délai TR :

$$M_{eq} = \frac{(1-E_1)M_0}{1-E_1 \cos \theta} \quad (2)$$

$$\text{avec } E_1 = e^{-\frac{TR}{T_1}} \quad (3)$$

M_0 : valeur initiale de l'aimantation

θ : angle d'impulsion

T_1 : temps de relaxation longitudinale

Il faut que $TR \gg T_1$ pour permettre à l'aimantation un retour complet à l'équilibre $M_{eq} = M_0$ entre deux impulsions. On peut définir ainsi l'erreur commise (ε) par :

$$\varepsilon = \frac{M_0 - M_{eq}}{M_0} = \frac{E_1 - (1 - \cos \theta)}{1 - E_1 \cos \theta} \quad (4)$$

Il est ainsi possible de déterminer la valeur de TR, pour un θ donné, permettant d'obtenir la précision souhaitée. Il a été montré que le meilleur rapport S/B est obtenu pour θ proche de 90° [131].

Par ailleurs, l'angle d'impulsion peut être affecté par les instabilités de l'appareillage qui engendrent des variations de la valeur de cet angle. Le signal détecté étant proportionnel à $\sin(\theta)$, la répercussion de ces instabilités sur la précision sera donc minimale pour θ proche de 90° .

Le **Tableau 5** montre les valeurs de précisions obtenues pour différents rapports TR/ T_1 pour un angle d'impulsion $\theta = 90^\circ$ où $\varepsilon = E_1$.

Tableau 5 : Précision en fonction du TR/T_1 pour un angle d'impulsion $\theta = 90^\circ$.

TR/T_1	Erreur (ϵ)
3	5%
4,6	1%
5	0,70%
7	0,1%
9	0,01%

Le rapport TR/T_1^{\max} doit être alors supérieur de 4,6 pour avoir une précision meilleure que 1%, avec T_1^{\max} : le temps de relaxation longitudinale le plus long de la molécule.

I-4. Analyse isotopique

De nos jours, l'analyse isotopique occupe une place très importante dans la caractérisation de l'origine géographique des produits alimentaires [132,133,134,135,136,137] et la détection d'adultération, comme la chaptalisation des vins par exemple [138]. Les rapports isotopiques stables ($^{13}C/^{12}C$) ont permis de classifier les huiles d'olive en fonction de leur origine géographique [139,140,141] et de leurs variétés [142].

Dans les paragraphes suivants, des informations sur les isotopes stables seront présentées. De plus, nous rappellerons brièvement les principes des techniques analytiques (SMRI, RMN ^{13}C et HSQC) utilisées pour l'analyse isotopique.

I-4.1. Isotopes : Définition et notations

Les isotopes se divisent en deux catégories, certains montrent une nature instable et se désintègrent au cours du temps en émettant des radiations, ce sont les isotopes radioactifs. Tandis que les autres sont énergétiquement stables et ne subissent aucun changement de leur structure nucléaire, ce sont les isotopes stables. Durant cette étude, nous abordons seulement les isotopes stables.

Les isotopes se caractérisent par des propriétés chimiques identiques car ils ont le même nombre d'électrons. En revanche, leur nombre de neutrons n'est pas le même, ce qui engendre des masses atomiques et des forces de liaisons chimiques différentes qui sont à l'origine du processus de fractionnement isotopique.

Les isotopomères possèdent un isotope lourd de l'un des éléments présents dans la molécule mais à des positions différentes (**Figure 5**).

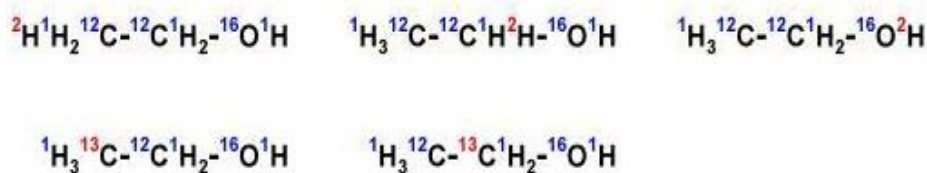


Figure 5 : Les isotopomères de la molécule éthanol comportant un isotope lourd et stable [143].

Durant cette thèse, seuls les isotopes stables du carbone (C) ont été étudiés, l'isotope ^{12}C dont l'abondance naturelle moyenne est de 98,9% et l'isotope ^{13}C qui est beaucoup plus rare (1,1%).

La répartition isotopique peut être exprimée de différentes façons [144] :

- **Le rapport isotopique (R)** représente le rapport du nombre de moles de l'élément lourd sur celui de l'élément léger. Dans le cas du carbone, R s'exprime par :

$$R(^{13}\text{C}) = \frac{^{13}\text{c}}{^{12}\text{c}} \quad (5)$$

- **L'abondance isotopique (A)** est la fraction molaire de l'élément le plus lourd et s'exprime par :

$$A(^{13}\text{C}) = \frac{^{13}\text{c}}{^{13}\text{c} + ^{12}\text{c}} \quad (6)$$

Les variations du rapport et de l'abondance isotopique étant relativement faible, le paramètre de composition isotopique δ (exprimé en ‰) est préféré. Ce dernier est déterminé pour étudier la variation de la teneur relative en ^{13}C entre l'échantillon et le standard international (pour le ^{13}C , la référence est le Vienna Pee Dee Belemnite ou PDB).

$$\delta(^{13}\text{C}) = \frac{R(^{13}\text{C})_{\text{échantillon}} - R(^{13}\text{C})_{\text{PDB}}}{R(^{13}\text{C})_{\text{PDB}}} \times 1000 \quad (7)$$

avec $R_{\text{PDB}} = 0,0112372$

Plus un composé est riche en isotope lourd, plus $\delta(^{13}\text{C})$ est élevé. Les valeurs de $\delta^{13}\text{C}$ mesurées sont généralement négatives.

I-4.2. Techniques analytiques utilisées pour l'analyse isotopique

De nombreuses techniques analytiques permettent de déterminer les rapports isotopiques, nous présenterons ici les plus utilisées pour le carbone : la SMRI et la RMN ^{13}C (à une et deux dimensions).

I-4.2.1. Spectrométrie de masse de rapports isotopiques

La SMRI est la technique analytique la plus couramment utilisée pour la détermination des rapports $^{13}\text{C}/^{12}\text{C}$. Le principe de cette technique consiste à oxyder un échantillon organique en gaz simples, suivi par une étape de réduction (pour transformer les oxydes d'azote en N_2 dans le cas où la molécule contient de l'azote). Puis, après avoir éliminé l'eau, le CO_2 formé est séparé du N_2 avant d'être introduit dans la source du spectromètre de masse. Ces opérations se produisent d'une façon automatique dans un analyseur élémentaire (AE) couplé au spectromètre de masse. Les gaz ainsi formés sont transformés en ions dans la source du spectromètre. Les isotopomères du CO_2 seront ensuite déviés par un champ magnétique en fonction de leur masse (m/z), recueillis et quantifiés à l'aide d'un collecteur de Faraday.

La SMRI se caractérise par une très bonne précision (écart-type 0,2‰ sur l'échelle des compositions isotopiques δ pour une analyse $^{13}\text{C}/^{12}\text{C}$), elle bénéficie de références internationales certifiées pour le calibrage, et la quantité d'échantillon requise pour l'analyse est très faible (de l'ordre du 1 mg pour l'analyse ^{13}C).

En revanche, cette technique est destructive, et elle est limitée à l'obtention d'une valeur moyenne de $\delta^{13}\text{C}$ pour l'ensemble des carbones du composé étudié. Des réactions de dégradations et de purifications sont nécessaires pour obtenir les rapports isotopiques pour chaque site ce qui peut fausser les mesures.

I-4.2.2. RMN ^{13}C isotopique

Contrairement à la SMRI, la technique RMN ^{13}C isotopique est un outil qui permet d'obtenir les rapports isotopiques site par site. Toutefois, il faut pour cela se placer dans des conditions de mesures particulières afin d'atteindre un niveau de précision élevée.

Outre les conditions générales de la RMN quantitative, mentionnées dans le paragraphe I-3.3., des exigences supplémentaires doivent être appliquées pour la RMN isotopique ^{13}C . Ces dernières ont été décrites dans de nombreuses études au sein du laboratoire CEISAM et seront présentées brièvement dans ce paragraphe.

La variation des teneurs isotopiques $\delta^{13}\text{C}$ peut atteindre environ 50‰. À titre d'exemple, pour les carbones C1, C2 et C3 du glycérol cette variation est comprise entre 1,6 et 40,9‰ [145]. De ce fait, une précision de l'ordre de 1‰ est nécessaire afin de garantir une interprétation correcte des mesures isotopiques. Un tel niveau de précision nécessite un rapport S/B d'au moins 500.

De plus, il faut utiliser une séquence de découplage en créneau inverse ou « inverse-gated decoupling » pour l'élimination de l'effet Overhauser nucléaire (NOE), ceci implique un délai $TR = 10 \times T_1 (^{13}\text{C})$ le plus long de la molécule étudiée.

L'efficacité du découplage large bande proton est également un paramètre essentiel. Pour une mesure des teneurs isotopiques ^{13}C ; le point le plus important étant l'homogénéité à l'intérieur de la gamme de déplacements chimiques découplés. Les méthodes utilisant des impulsions composites, comme WALTZ-16, ne permettent pas une quantification avec une précision de l'ordre de quelques pour mille. Seules les impulsions adiabatiques ont des performances suffisantes pour atteindre cet objectif. Et même dans ce cas, il est indispensable d'optimiser la forme (forme cosinus et balayage en fréquence calculé numériquement pour une adiabaticité indépendante de l'offset), la durée de l'impulsion et le cycle de phase [146].

Hormis les conditions d'acquisition, il est indispensable de lisser les signaux le plus précisément possible afin d'atteindre la précision et la justesse nécessaires à l'analyse isotopique ^{13}C . Les meilleurs résultats sont obtenus par ajustement du spectre à des fonctions théoriques (lorentzienne ou combinaison de lorentzienne et de gaussienne) en optimisant jusqu'à cinq paramètres par raie : position, hauteur, largeur à mi-hauteur, phase et proportion de gaussienne.

De ce fait, il a été confirmé que la séquence RMN ^{13}C conventionnelle mono-impulsionnelle avec un découplage adiabatique permet d'acquérir une précision élevée (1-2%) [146,147] demandée pour l'analyse isotopique et qui pourrait être utilisée dans le domaine de l'authentification des produits naturels [148].

Les rapports position-spécifiques $^{13}\text{C}/^{12}\text{C}$ ont permis une discrimination complète des échantillons de la vanilline en fonction de leur origine [149,150].

D'autre part, la RMN ^{13}C a été utilisée pour fournir le profil isotopique de chaque site des principes actifs de certains médicaments permettant par exemple de comparer des échantillons d'aspirine et de paracétamol provenant d'un même fournisseur mais de différents lots.

En conséquence, cette technique apparaît comme un outil de grand intérêt pour l'authentification des produits et pour la détection des fraudes [151].

Bien qu'une précision élevée (1-2%) soit obtenue avec la séquence RMN ^{13}C mono-impulsionnelle, cette technique souffre d'un long temps d'expérience en raison de la faible abondance du ^{13}C (1,1%), son petit rapport gyromagnétique (1/4 celui du proton) et des temps de relaxation longitudinale (T_1) longs pour le ^{13}C . Ce qui fait que la sensibilité de la RMN ^{13}C est beaucoup plus faible que celle de la RMN ^1H .

Les techniques multi-impulsionnelles DEPT (Distortionless Enhancement by Polarization Transfer) et INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) transfèrent la polarisation d'un noyau sensible (^1H) vers un noyau à faible rapport gyromagnétique (^{13}C) et sont donc potentiellement très intéressantes. Elles permettent une augmentation significative de la sensibilité, ainsi qu'une réduction importante du temps d'expérience puisque le temps de répétition dépend alors des T_1 des protons qui sont beaucoup plus courts que ceux des ^{13}C [152].

Dans la littérature, de nombreux auteurs ont rapporté l'utilisation de la séquence DEPT pour une analyse quantitative rapide. Elle a été par exemple utilisée pour l'analyse de la lignine [153]. Ce type de techniques a également permis de classifier les huiles d'olive en fonction de leur variété [127] et de leur région d'origine [154,155,156]. De plus, cette séquence a été utilisée pour détecter l'adultération de l'huile d'olive par de l'huile de soja [157].

Cependant, dans les séquences multi-impulsionnelles, de nombreux facteurs participent aux imperfections des impulsions RF comme : les erreurs de calibrage, l'inhomogénéité du champ B_1 et les effets d'« off-résonance ». Ces erreurs sont cumulées au cours de la séquence, et conduisent à une instabilité du signal RMN [158].

En conséquence, les séquences multi-impulsionnelles ne permettent généralement pas d'atteindre une précision de quelques pour mille indispensable pour les analyses isotopiques.

Des travaux réalisés notamment au sein du laboratoire CEISAM ont montré que l'incorporation des impulsions composites et adiabatiques dans les séquences de transfert de polarisation (INEPT et DEPT) permet de compenser les imperfections d'impulsions RF dues à l'inhomogénéité du champ RF ou à l'effet d'« off-résonance ». La précision nécessaire à l'analyse isotopique ^{13}C peut alors être obtenue [159].

D'autre part, dans une autre étude, deux séquences adiabatiques (DEPT et INEPT) ont été utilisées sur des échantillons d'ibuprofène afin d'évaluer leur robustesse et leur capacité à fournir le profil isotopique des échantillons dans une courte durée. La séquence DEPT s'est révélée peu stable à long terme, tandis que la séquence INEPT adiabatique (**Figure 6**) a permis d'atteindre une précision et une reproductibilité de l'ordre du pour mille et ceci dans un temps d'expérience réduit par rapport à la séquence mono-impulsionnelle.

Ces résultats ont donc démontré qu'une séquence de transfert de polarisation comme l'INEPT constitue une méthode de quantification robuste et stable lorsque les impulsions RF à 180° sont correctement optimisés [160]. C'est pourquoi nous allons exploiter les potentialités de la

séquence INEPT optimisée pour l'analyse des matrices riches en TAG, notamment les huiles d'olive.

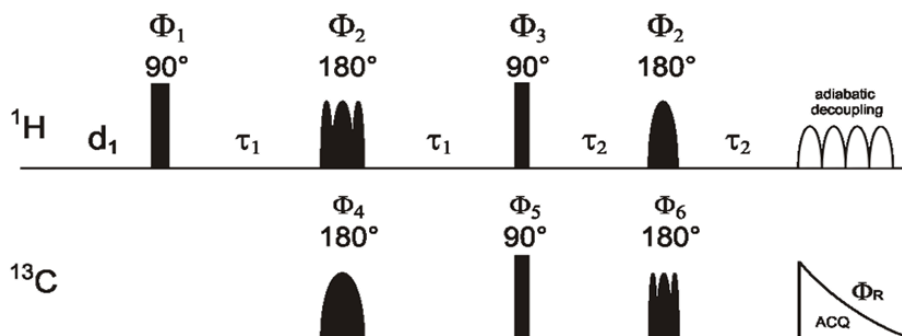


Figure 6 : Séquence INEPT-refocalisée dans laquelle les impulsions RF à 180° d'inversion et de refocalisation sont des impulsions adiabatiques [159].

I-4.2.3. RMN à deux dimensions

Les techniques RMN à une dimension (1D) sont donc reconnues pour l'analyse quantitative, et font l'objet de nombreuses applications dans des domaines très diversifiés comme l'étude métabolique [161,162,163] et la caractérisation de l'origine des produits [150,164,165]. Ces méthodes ont toutefois leurs limites et ne permettent pas de résoudre complètement le problème de chevauchements de signaux surtout dans le cas des mélanges complexes. Même dans le cas des petites molécules comme l'ibuprofène, malgré la large gamme de déplacement chimique du ^{13}C , des chevauchements sont observés, ce qui a réduit la précision de la détermination de la teneur isotopique. Cette limitation peut être surmontée en utilisant la RMN à deux dimensions (2D) qui permet une meilleure discrimination des résonances (**Figure 7**) [166].

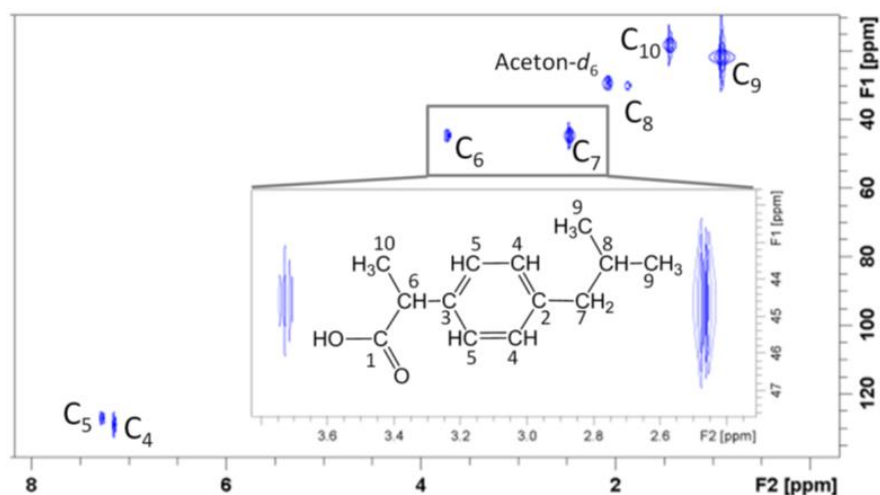


Figure 7 : Spectre HSQC ^1H - ^{13}C d'une solution d'ibuprofène dans l'acétone- d_6 . Le zoom sur la région des carbones C6-C7 montre que le recouvrement partiel observé en 1D sur ces deux raies est supprimée par l'utilisation de la RMN à deux dimensions [166].

La littérature sur les applications déjà publiées de la RMN 2D pour l'analyse quantitative met en évidence un certain nombre de limitations :

- les séquences RMN 2D sont très sensibles aux imperfections d'impulsions en raison de leur caractère multi-impulsionnel ;
- les volumes des pics sont affectés par les temps de relaxation et par les couplages homo- ou hétéronucléaires [167] ;
- la longue durée des expériences RMN 2D (plusieurs heures) les rend très sensibles aux instabilités temporelles et réduit la précision (le temps d'acquisition étant proportionnel au nombre de scans, ainsi qu'au nombre d'incrémentes dans la dimension F_1) [168].

I-4.2.4. Séquences de corrélation hétéronucléaire

Dans certains cas, pour résoudre des problèmes de recouvrement spectral, il est concevable d'utiliser des séquences de corrélation hétéronucléaire permettant de quantifier les composés dont les déplacements chimiques sont trop proches et de mettre en évidence l'interaction entre les noyaux. Il existe deux types de détection :

I-4.2.4.1. Détection directe

Dans le cas de détection directe, le FID est détecté à la fréquence du ^{13}C . Les spectres ^1H et ^{13}C sont enregistrés dans les dimensions F_1 et F_2 respectivement. Les séquences de détection directe ont une faible gamme de fréquences à échantillonner selon F_1 , puisque la gamme de déplacement chimique des ^1H (environ 4000 Hz) est plus faible que celle du ^{13}C (environ 20000 Hz). De ce fait, les techniques de détection directe exigent moins d'incrémentes de t_1 . Cependant, en raison de la faible abondance naturelle du ^{13}C (noyau détecté), ces techniques sont peu sensibles. Par la suite, la durée des expériences est longue, car un grand nombre d'accumulations est nécessaire pour aboutir à un rapport S/B correct. C'est pourquoi la détection directe n'a pas été utilisée pour l'analyse quantitative dans le cadre de cette thèse.

I-4.2.4.2. Détection indirecte ou inverse

Contrairement à la détection directe, les séquences de détection inverse permettent d'obtenir dans la dimension F_1 les déplacements chimiques du ^{13}C , noyau peu sensible, tout en observant le signal du noyau le plus sensible ^1H . De ce fait, les résonances ^{13}C sont détectées avec la sensibilité du ^1H . L'avantage de ce type de détection, c'est qu'un nombre d'accumulations inférieur est requis pour obtenir le même rapport S/B qu'en détection directe, ce qui permet de gagner de temps. Par contre, la gamme d'échantillonnage dans la dimension F_1 (celle du ^{13}C) est beaucoup plus importante.

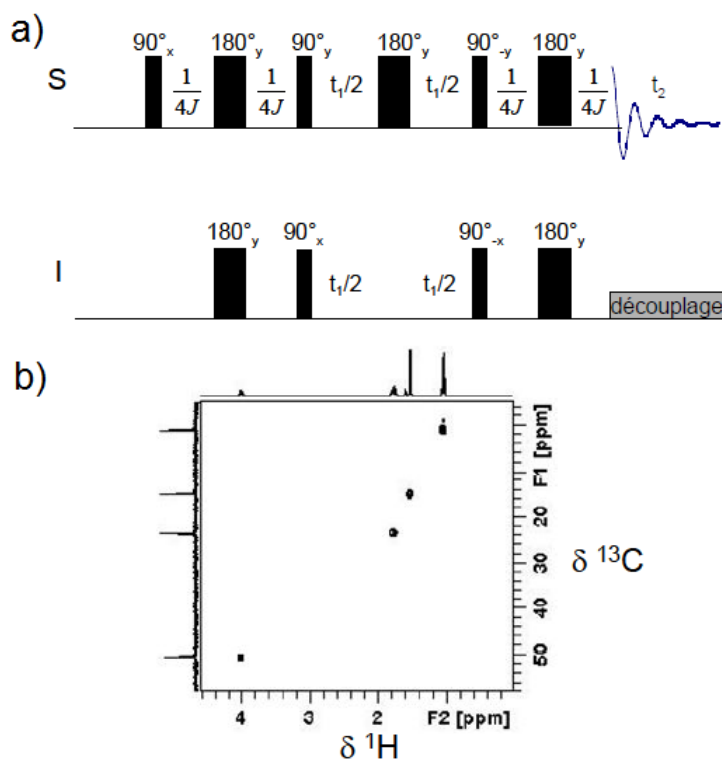


Figure 8 : a) Séquence d'impulsion HSQC permettant de corrélérer les déplacements chimiques d'un noyau S (^1H) et d'un noyau I (^{13}C). b) Spectre HSQC d'un échantillon de 2-chlorobutane [169].

La séquence de corrélation hétéronucléaire la plus utilisée pour les applications quantitatives est sans doute l'HSQC (Heteronuclear Single Quantum Correlation) et elle utilise la détection inverse. Cette séquence est composée d'un motif INEPT qui transfère la polarisation du spin S, plus sensible (^1H), vers le spin I, le moins sensible (^{13}C), et d'un motif symétrique retro-INEPT qui réalise un transfert de I vers S (**Figure 8**). Ce double transfert, couplé à un cycle de phase adéquate, permet d'éliminer le signal des protons liés aux ^{12}C et de ne retenir que le signal des ^1H portés par des ^{13}C .

L'emploi de la séquence HSQC pour l'analyse quantitative pose des problèmes de précision et de justesse. De ce fait, il est intéressant d'optimiser cette séquence pour la quantification. Des travaux, effectués au sein du laboratoire CEISAM [166], ont cependant montré qu'une excellente précision peut être atteinte avec la séquence HSQC représentée sur la **Figure 9**. Cela est probablement dû à la symétrie intrinsèque de la séquence qui a permis de compenser les imperfections dues aux inhomogénéités du champ et à l'effet « off-resonance ». Dans ce travail, le temps d'expérience a été optimisé afin d'obtenir une précision élevée. Ainsi, cette séquence modifiée a permis d'atteindre une bonne répétabilité, mais doit encore être améliorée pour permettre une application à l'analyse isotopique ^{13}C [166].

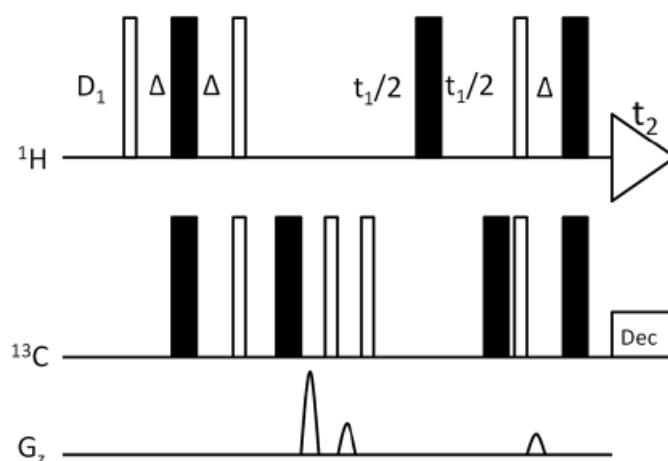


Figure 9 : Séquence HSQC optimisée. D_1 indique le délai entre deux acquisitions, $\Delta = \frac{1}{4} \cdot {}^1J_{CH}$, avec ${}^1J_{CH} = 139$ Hz la moyenne de la constante de couplage ${}^1H-{}^{13}C$, les angles d'impulsion 90° (rectangles vides), les angles 180° (rectangles remplis). L'amplitude des 3 gradients est 80:30:20,1 successivement [166].

Dans les paragraphes suivants, nous allons présenter les stratégies utilisées dans le but de réduire le temps d'expérience 2D.

En RMN 2D, le temps d'acquisition (T_a) est donné par :

$$T_a = NS \cdot N_f \cdot TR \quad (8)$$

Avec NS : le nombre de scans, N_f : le nombre d'incrément dans la dimension 1, et TR : temps de répétition qui correspond au délai de retour à l'équilibre entre deux scans.

Par la suite, il est essentiel d'ajuster les paramètres ci-dessus au plus juste en 2D pour obtenir un temps d'acquisition minimal.

TR doit être choisi suffisamment élevé pour préserver les conditions quantitatives, un $TR = 5 \times T_1^{\max}$ (avec T_1^{\max} le plus long T_1 de la molécule étudiée) est suffisant pour obtenir une exactitude de 1 pour mille.

$NS = 2$ est la valeur minimale pour préserver la sélection du chemin de cohérence en HSQC, qui consiste à filtrer le signal indésirable issu du 1H lié au ${}^{12}C$.

N_f doit être réduit autant que possible pour diminuer la durée d'expérience tout en préservant une bonne résolution dans la dimension 1.

Dans le but de réduire le temps d'expérience tout en préservant une précision élevée, les différentes stratégies décrites brièvement dans les paragraphes ci-dessous peuvent être mises en œuvre :

a- Repliement de spectre ou « spectral aliasing »

La largeur spectrale (SW) est reliée à l'écart DW qui sépare 2 points du FID par la relation :

$$SW = \frac{1}{2.DW} \quad (9)$$

Pour une durée d'acquisition fixe, le FID est alors d'autant mieux digitalisé que la valeur SW est plus petite. De ce fait, la SW dans la dimension F_1 joue un rôle très important dans l'optimisation du temps d'acquisition.

Quand la valeur de SW est trop faible pour regarder l'ensemble des résonances, la violation du critère de Nyquist entraîne un repliement des signaux. Dans la dimension F_2 , des filtres, électroniques et numériques, sont appliqués pour supprimer tout signal à l'extérieur du SW et permettent d'éviter ce repliement. Par contre, dans la dimension F_1 , les filtres évoqués plus haut ne sont pas disponibles, et les signaux à l'extérieur de SW peuvent donc se replier. La valeur de SW_1 doit être alors choisie avec précaution pour contrôler le repliement [170,171]. Toutefois, une plus petite SW permet de diminuer le nombre d'incrément N_1 . Il faut noter, que les déplacements chimiques de la dimension F_1 ne sont pas corrects dans le cas de repliement de spectre, mais ce n'est pas un problème lors de l'analyse d'une molécule bien connue, où les vrais déplacements chimiques des pics 2D peuvent être facilement déterminés [172,173].

b- Échantillonnage non-uniforme « non-uniform sampling »

L'échantillonnage non-uniforme (NUS) [174] a commencé à être utilisé pour les mélanges complexes [175,176]. Il est possible de l'utiliser pour réduire le temps d'acquisition des expériences RMN 2D. De plus, le NUS permet aussi d'augmenter la résolution spectrale et de réduire le recouvrement le plus possible sans augmenter la durée d'expérience. Cette approche consiste à acquérir un nombre réduit de points répartis de manière aléatoire dans la dimension F_1 [177,178,179,180]. Afin d'obtenir des spectres propres, le traitement des expériences NUS doit être associé à une procédure appropriée telle que la « recursive multidimensional decomposition » (R-MDD). L'étude déjà mentionnée sur la molécule d'ibuprofène [166], a montré que cette approche peut diminuer la précision ce qui ne correspond plus aux exigences de l'analyse isotopique ^{13}C , mais reste raisonnable pour certaines applications quantitatives [159].

c- Prédiction linéaire

La prédiction linéaire (PL) est une autre approche largement utilisée pour réduire le nombre d'incrément N_1 , tout en préservant la résolution dans la dimension F_1 . Seuls les premiers points du FID sont recueillis, les points ultérieurs sont eux prédits à partir de combinaisons

linéaires des points initiaux. La PL permet de diviser par deux le nombre d'incrémentations dans la dimension F_1 tout en gardant une résolution identique, ou bien tout en doublant la résolution de F_1 pour une même durée d'expérience [181,182].

Dans le cadre de cette thèse, nous avons développé une nouvelle version de la séquence HSQC qui permet d'éliminer les signaux relatifs aux ^1H liés aux ^{12}C et d'améliorer la précision de la méthode afin de répondre aux exigences de l'analyse isotopique.

I-5. Marqueurs d'authenticité de l'huile d'olive

La caractérisation de l'origine géographique et variétale de l'huile d'olive est devenue de plus en plus importante. Les analyses métabolomiques et isotopiques des huiles fournissent des traceurs au sein de ces matrices qui permettent de classer les huiles et de remonter à leur origine. Le **Tableau 6** présente les différentes études réalisées pour l'authentification des huiles d'olive. Ce tableau comporte des informations sur la technique utilisée, les marqueurs adoptés, la variété et l'origine géographique des olives, ainsi que les résultats de la classification.

Tableau 6 : Différentes études réalisées pour l'authentification des huiles d'olive.

Techniques	Marqueurs	Variété des olives	Origine géographique	Résultats de la classification	N° Réf
GC-FID	AG	Non précisée	Libie, Tunisie, Turquie, Italie	Classification des huiles en fonction de l'origine géographique	[183]
GC	AG	Non précisée	Espagne, Italie, Portugal	Classification des huiles en fonction de leur origine géographique	[184]
GC	AG	Non précisée	Différentes régions de Grèce	Classification géographique des huiles	[185]
RMN ^{13}C	Informations positionnelles des acides saturés, oléique et linoléique	Leccino	5 régions d'Italie	Caractérisation des huiles en fonction de l'origine géographique	[186]
RMN ^{13}C	Variables du spectre ^{13}C	Coratina, Dritta, Grossa di Cassano, Moraiolo, Picholine	Abruzzes, Israël, Puglia, Toscane	Classification en fonction de la variété et de la région (70-100%)	[187]
RMN ^{13}C	Variables du spectre ^{13}C	Leccino, Moraiolo, I-77, Dritta	Abruzzes, Puglia, Toscane	Classification en fonction de la variété et de l'origine (100%)	[127]

SMRI	$\delta^{13}\text{C}$ (pour l'huile, sterols et alcools aliphatiques) $\delta^{18}\text{O}$ (pour l'huile)	Non précisée	Grèce, Maroc, Espagne, Italie, Tunisie, Turquie	Classification géographique	[139]
GC RMN ^1H	AG, phénols	Coratina, Leccino, Oliarola, Olivastro, Simone	3 régions différentes d'Apulia	Les AG ont contribué à une classification de 100% en fonction de la variété les phénols ont contribué à une classification de 100% en fonction de l'origine	[188]
GC RMN ^1H , ^{13}C	AG, composés minoritaires, informations positionnels	17 variétés	Italie, Argentine	Classification (100%) en fonction de l'origine géographique	[189]
RMN ^1H	Empreinte des TAG, composés mineurs	Frantoio, Leccino, Moraiolo, Nerino, Quercetana	Seggiano, Lucca, Florence, Arezzo	Classification en fonction de l'origine géographique	[190]
RMN ^{13}C	Variables du spectre ^{13}C	12 variétés	13 régions d'Italie	Classification en fonction de la variété	[154]
RMN ^{13}C GC	Variables du spectre ^{13}C , AG	Nocellara del Belice, Tonda Iblea, Cerasuola, Biancollila	Castel Vetrano, Trapani, Delia, Paceco, Caltabellota, Ragusa	Les positions sn-1,3 ont joué un rôle important dans la classification variétale (100%)	[191]
RMN ^{13}C GC HPLC	Variables du spectre ^{13}C , AG, TAG	Coratina, Leccino, Peranzana, Oliarola	Apulia	La GC et la RMN ^{13}C ont permis 92% et 88% de la classification variétale respectivement	[192]
Proche-IR	AG TAG	Non précisée	Crête, Péloponnèse, Grèce	Classification des huiles en fonction de leur origine géographique	[193]
RMN ^{13}C	Variables du spectre ^{13}C	Coratina, Ogliarola, Peranzana	Terra di Barri, Colline di Brindisi, Foggia	Origine : 72% (Dauno), 90% (Terra di Barri et Colline di Brindisi). L'erreur dans la classification est due aux échantillons qui ont la même variété mais provenant de région différente	[156]
GC HPLC	AG, phénols, tocophérols	Frantoio, Leccino	Mengibar, Cabra	La composition de l'huile d'olive en AG et ses propriétés sensorielles dépendent de leur origine	[64]

géographique					
GC-MS HPLC	AG TAG squalène	Aglандаu, Cailletier, Picholine, Salonenque, Tanche, Vallée des Baux	Aix en Provence	Classification en fonction de la variété (100%)	[40]
HPLC GC-FID	TAG AG AG en position sn-2	Cornicabra	Espagne	Classification en fonction de la variété (90%) TAG sont plus significatifs pour la classification variétale que les AG en position sn-2	[194]
GC-FID HPLC	AG, TAG	Bien précisée	6 Appellations d'origine contrôlée françaises	Classification des échantillons en fonction de leur origine géographique	[195]
GC-FID HPLC UV-Visible	AG, TAG, phénols, stérois, chlorophylles	Bien précisée	4 régions de Grèce	Classification (97,7%) en fonction de l'origine géographique	[196]
RMN ¹ H	Empreinte des TAG	Non précisée	Grèce, Italie, Espagne, Tunisie, Turquie	Une classification de 35-92% dépendant de l'origine et de l'année de production	[197]
RMN ¹ H	Empreinte des TAG	Casaliva, multi variétés	Garda, Veneto	Classification des huiles en fonction de leur origine géographique	[198]
RMN ¹ H	Empreinte des TAG	Casaliva, multi variétés	Garda Lake, Veneto, Lombardia banks	Classification des huiles en fonction de leur origine géographique	[199]
GC-FID HPLC	AG TAG	Bien précisée	Aix-en- Provence, Haute- Provence, Nyons, Nice, Vallée des Baux de Provence	Classification des huiles en fonction de leur origine géographique	[200]
GC	AG	22 variétés	Sicile	Classification des huiles en fonction de leur variété, l'année de production et des conditions climatiques	[103]
RMN ¹ H, ¹³ C	Composés minoritaires (squalène,	22 variétés	Lazio	Classification des huiles en fonction de leur origine géographique en étudiant l'effet de l'altitude et de	[74]

	stéroïls et terpènes), variables du spectre ^{13}C			l'irrigation	
Proche-IR	AG TAG	Non précisée	Aix-en-Provence, Haute-Provence, Nice, Nyons, Vallée des Baux	Classification des huiles en fonction de leur origine géographique	[201]
SMRI	$\delta^{13}\text{C}$, $\delta^{18}\text{O}$	Bien précisée	8 régions italiennes	$^{18}\text{O}/^{16}\text{O}$ est un paramètre qui dépend de la latitude et du climat, il a permis de remonter à l'origine géographique des huiles d'Italie. $^{13}\text{C}/^{12}\text{C}$ a permis de classer les échantillons d'huiles d'olive en fonction de leur variété	[142]
SMRI, ICP-MS	$\delta^{13}\text{C}$ et $\delta^{18}\text{O}$ (de l'huile et du glycérol), éléments traces des minéraux	Multivariétés	Différentes régions italiennes	Classification des huiles selon leur origine géographique	[140]
RMN ^1H , ^{31}P	AG DAG Phénols	Koroneiki	Crète (Heraklion, Sitia, Chania), Peloponnesus (Messinia, Lakonia), Zakynthos	87% pour la classification des huiles d'olive en fonction de 3 régions, 74% pour 6 régions, 92% si l'année de production est prise en compte	[124]
RMN ^1H	Empreinte TAG, composés minoritaires	Non précisée	Italie, Ligurie, Espagne, France, Grèce, Turquie	Différenciation des huiles liguriennes des non liguriennes	[202]
RMN ^1H , IRMS	Empreinte TAG, composés minoritaires (composés terpéniques, composés phénoliques), $\delta^{13}\text{C}$, $\delta^2\text{H}$	Non précisée	Grèce, Espagne, Italie, Turquie, Syrie, Chypre	Classification des huiles selon leur origine géographique	[141]
GC-FID FT-PIR	AG	Picholine marocaine	Régions Marocaines (Tadla, Azilal)	La composition de l'huile en AG, l'acidité et l'indice de peroxyde varient en fonction de l'origine	[65]
GC-FID	AG,	Bien précisée	Différentes	Classification géographique	[203]

	composés minoritaires (stéroïdes, terpènes, alcools)		régions d'Espagne		
SM	AG	7 variétés tunisiennes	Tunisie	Discrimination des huiles en fonction de la variété et l'indice de maturité	[204]

I-6. Analyse statistique

L'application de la chimiométrie à la spectroscopie RMN a permis de classer avec succès les huiles d'olive en fonction de leurs origines géographique et variétale [74,103,156,154,198,199].

Dans le cadre de cette thèse, des analyses univariées et multivariées (supervisées et non-supervisées) [205], ont été employées pour le traitement et l'interprétation des données. Les méthodes utilisées durant ce travail sont brièvement décrites dans les paragraphes suivants.

I-6.1. Analyse de variance univariée

L'analyse de variance univariée (ANOVA) permet de détecter si une variable est à la base de la séparation des groupes d'échantillons selon le facteur étudié. Cette analyse consiste à trouver le rapport (F) entre la variance inter-groupe (la somme moyenne des carrés des écarts entre chaque moyenne de groupe et la grande moyenne) et la variance intra-groupe (la somme moyenne des carrés des écarts entre chaque observation et la moyenne du groupe) et de comparer le résultat avec une distribution F d'échantillonnage connue [206]. Plus la valeur du rapport F est grande, plus la variable possède un pouvoir discriminant élevé et plus son effet est significatif ; en d'autres termes, il y a une bonne séparation des groupes et il est d'autant moins probable qu'elle soit due au hasard.

Cependant, l'ANOVA, qui considère une seule variable à la fois, ne tient pas compte des interactions existant entre les différentes variables de la matrice initiale ce qui impose l'utilisation de l'analyse multivariée.

I-6.2. Analyse multivariée

L'objectif des analyses multivariées est de réduire la complexité des informations fournies par des variables indépendantes (explicatives) utilisées pour expliquer des variables dites « réponses ». Elles permettent de diminuer le nombre des variables explicatives ce qui élimine le « bruit » tout en laissant les informations significatives. Il existe un grand nombre de méthodes de classifications multivariées qu'on peut grouper en analyses non-supervisées et supervisées. Les méthodes non-supervisées permettent d'avoir une vision globale du jeu de

données sans aucun *a priori*. Dans ce cas, l'appartenance des échantillons n'est pas connue, il s'agit justement de la retrouver à partir de descripteurs disponibles en maximisant la variance sur l'ensemble de données. Dans le paragraphe qui suit, nous décrivons uniquement l'Analyse en Composantes Principales (PCA) que nous avons utilisée dans ce travail.

Par contre, en analyse supervisée, le but est d'optimiser la séparation des différents échantillons dont le groupe est connu *a priori* et de déterminer un nombre réduit de paramètres discriminants. Ce type de traitement de données cherche à maximiser la covariance entre les groupes et la minimiser au sein d'un même groupe. Dans cette catégorie d'analyses, nous présentons plus loin l'Analyse Canonique Discriminante (CDA) et l'Analyse Discriminante Linéaire (LDA) qui sont étroitement liées.

De même, les statistiques multivariées englobent des méthodes de régressions qui visent à construire des modèles pour prédire les valeurs des variables réponses (dépendantes). À cet égard, nous présentons la Régression Linéaire Multiple (MLR) et la Régression des moindres carrés partiels (PLS) qui ont servi dans notre étude.

I-6.2.1. Analyse en Composantes Principales

L'Analyse en Composantes Principales (PCA) est une technique descriptive généralement utilisée en première approche multivariée et permet d'avoir les relations existant entre les variables (p) et les échantillons (n). Elle consiste à remplacer les variables d'origine par des variables synthétiques, appelées composantes principales ou facteurs, contenant la quasi-totalité de l'information et classées par ordre décroissant de variance. Plus la variance associée au facteur est élevée, plus ce dernier est important. Cette nouvelle série des variables maximise la dispersion entre les échantillons.

La représentation graphique obtenue par cette analyse permet de rassembler les échantillons similaires et de séparer les échantillons disposant d'une grande variabilité. Les données résultantes de cette analyse peuvent servir de variables de base pour d'autres traitements statistiques tels que l'analyse discriminante ou la régression [207].

I-6.2.2. Analyse Discriminante Linéaire

Les méthodes non supervisées ne donnent pas toujours la meilleure discrimination des échantillons en fonction de leur groupe d'appartenance. À cet égard, l'Analyse Discriminante Linéaire (LDA) permet de trouver les combinaisons linéaires multivariées qui peuvent décrire le mieux possible les individus et ainsi conduire à la meilleure séparation des différents groupes.

Le paramètre « wilks' Lambda » (λ) est utilisé comme indicateur qui reflète cette séparation ; il varie entre 0 à 1, plus sa valeur est petite, plus les individus des différents groupes sont distinguables. Wilk' λ peut servir aussi pour la sélection des variables indépendantes lors de la construction du modèle de classification. L'élimination d'une variable significative du modèle conduira à une augmentation importante de ce paramètre.

En outre, la LDA fournit des fonctions de classification propres aux différents groupes qui servent pour classer un échantillon inconnu. L'échantillon sera attribué au groupe dont la fonction fournit le score le plus élevé.

Les tests validation croisée et « Leave-one-out » sont utilisés pour évaluer le potentiel du modèle à prédire la classe d'une variable préalablement éliminée pendant la construction du modèle (ces tests sont expliqués dans le paragraphe I-6.2.4.3.).

I-6.2.3. Analyse Canonique Discriminante

L'Analyse Canonique Discriminante (CDA), connue sous l'appellation « Analyse Factorielle Discriminante » ou aussi « Analyse Discriminante Descriptive », élabore de nouvelles variables (variables latentes) qui optimisent la séparation des groupes et qui permettent de les représenter graphiquement.

Les coefficients standardisés des fonctions canoniques relatives aux variables latentes permettent de comparer l'importance des variables indépendantes dans le modèle. Un coefficient élevé signifie que la variable correspondante joue un rôle important dans la discrimination des groupes.

Les coefficients non-standardisés permettent de projeter un nouvel échantillon dans le nuage de points [208]. Les techniques CDA et LDA sont complémentaires.

I-6.2.4. Méthodes prédictives : régression linéaire multivariée

Les méthodes de régression ont pour but de construire un modèle pour prédire le mieux possible la valeur d'une variable quantitative Y (variable dépendante) à partir des variables X (variables indépendantes) [209]. Nous présentons dans ce paragraphe la régression linéaire multiple (MLR) et la régression des moindres carrés partiels (PLS) utilisées dans ce travail.

I-6.2.4.1. Régression linéaire multiple

La régression linéaire multiple (MLR) vise à prédire une variable cible quantitative Y (endogène) à partir de p variables indépendantes explicatives (exogènes) X_1, \dots, X_p , quantitatives ou qualitatives [210].

Lorsqu'un modèle de MLR contient un grand nombre de variables indépendantes explicatives très corrélées entre elles, il est intéressant d'observer si la sélection d'un sous-ensemble de ces

variables peut donner un meilleur résultat que le modèle entier. Dans ce cadre, deux approches peuvent être adoptées.

- **Méthode ascendante (Forward selection)**

La méthode ascendante consiste à introduire une à une les variables explicatives, et ceci en sélectionnant la variable qui conduit à la corrélation la plus élevée avec la variable dépendante. Il faut cependant écarter les variables non significatives.

- **Méthode descendante (Backward selection)**

La méthode descendante consiste à commencer avec toutes les variables explicatives et retirer à chaque fois la variable la moins significative du modèle (qui a la plus grande valeur de p).

Cette approche présente un avantage par rapport à la méthode ascendante. Pour un ensemble de variables, il est possible d'avoir une capacité prédictive considérable même si les variables individuelles ne sont pas significatives.

- **Méthode bidirectionnelle**

La méthode bidirectionnelle est similaire à la méthode ascendante, mais les variables sont supprimées du modèle si elles deviennent non significatives lorsque d'autres variables explicatives sont ajoutées.

I-6.2.4.2. Régression des moindres carrés partiels

La régression des moindres carrés partiels ou « Partial Least Squares regression » (PLS) permet de construire un modèle prédictif quand les variables explicatives sont très corrélées et nombreuses. C'est une technique supervisée dans le sens qu'elle cherche à maximiser la covariance entre les variables d'entrée (prédicteurs) et la variable réponse. Elle permet de créer un ensemble de h composantes, appelées variables latentes qui sont responsables de la variation des variables explicatives, à partir de n observations décrites par p variables (avec $h \leq p$).

Le nombre de composantes à garder est en général déterminé en mettant en jeu une validation croisée utilisant le paramètre Q^2 (expliqué dans le paragraphe I-6.2.4.3.) comme indicateur de prédictibilité. Les composantes retenues sont celles qui permettent d'obtenir un minimum d'écarts entre les valeurs prédites et les valeurs mesurées.

Dans une deuxième étape, les modèles PLS sont construits en éliminant les variables non pertinentes tout en se basant sur leur coefficient de régression standardisé et/ou sur leur importance dans la projection (paramètre VIP) dans le modèle [211].

Les coefficients de régression sont utilisés pour la prédiction de la valeur d'un nouvel échantillon analysé [212].

I-6.2.4.3. Évaluation d'un modèle de régression

Dans le cas des méthodes de régression, PLS et MLR, le coefficient de détermination (R^2) est un paramètre qui permet de mesurer la concordance entre le modèle et les données observées. Il permet de juger la qualité d'une régression linéaire, sa valeur est comprise entre 0 et 1. Cependant, R^2 présente un défaut primordial, il augmente avec le nombre de variables explicatives utilisées pour construire le modèle. Un grand nombre de variables conduit à des modèles non robustes. De ce fait, le coefficient de détermination ajusté (R^2 -ajusté) qui tient compte du nombre de variables lui est préféré. R^2 -ajusté n'est pas un vrai « carré » et peut même être négatif. Il s'exprime par :

$$R^2\text{-ajusté} = 1 - \frac{(n-1)}{(n-m-1)} (1-R^2) \quad (10)$$

avec n : taille d'échantillons, m : nombre de variables utilisées pour construire le modèle sans considérant aussi la constante, et R^2 : coefficient de détermination multiple.

De plus, le test F est utilisé pour l'évaluation statistique du modèle. À ce niveau, les paramètres $F(df1, df2)$ et p sont utilisés comme indicateurs. F est donné par l'équation suivante :

$$F = \frac{\frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{m-1}}{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-m}} \quad (11)$$

avec y_i : valeurs obtenues à l'aide de la méthode de référence ; \hat{y}_i : valeurs calculées par le modèle ; \bar{y} : valeur moyenne des y_i ,

$df1 = m-1$, $df2 = n - m$, et p est la probabilité d'avoir $F \leq Fisher$ (valeur critique de la distribution F).

D'un autre côté, la capacité prédictive du modèle est idéalement évaluée en utilisant des échantillons de validation externe. Le paramètre R^2 prédictif ($Pred-R^2$) est un indicateur communément utilisé à cette fin. Plus $Pred-R^2$ est proche de 1, plus la capacité prédictive du modèle est élevée. $Pred-R^2$ est défini comme suit [213] :

$$Pred-R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

avec y_i : valeurs obtenues à l'aide de la méthode de référence pour les échantillons de validation externe ; \hat{y}_i : valeurs prédites pour les échantillons de validation externe ; \bar{y} : valeur

moyenne des valeurs obtenues à l'aide de la méthode de référence des échantillons utilisés pour construire le modèle.

Puisque le $\text{Pred-}R^2$ dépend de la valeur moyenne (\bar{y}), donc du choix des échantillons utilisés pour construire le modèle, une valeur élevée de cet indicateur ne garantit pas une bonne capacité prédictive du modèle. À cet égard, Roy et al. ont proposé l'utilisation d'un autre paramètre noté R_m^2 , donné par [213] :

$$R_m^2 = r^2 \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (13)$$

r^2 et r_0^2 sont respectivement les coefficients de corrélation entre les valeurs observées et celles prédites des échantillons de validation externe, avec ou sans l'ordonnée à l'origine. Un R_m^2 supérieur à 0,5 indique une bonne capacité prédictive du modèle.

Lorsque le nombre d'échantillons initial est petit, une validation interne s'impose. Dans un tel cas, la totalité des échantillons disponibles est utilisée pour construire le modèle, puis sa robustesse est évaluée avec des tests de validation croisée comme « Leave-one-out » et « Leave-many-out ».

Le test « Leave-one-out » consiste à exclure un échantillon à la fois du jeu de taille n , puis à construire le modèle sur les $n-1$ échantillons restants. Ce modèle est ensuite testé sur l'échantillon retiré du jeu en comparant sa valeur obtenue à celle observée. Puis cet échantillon est réintégré dans le jeu et un autre échantillon est écarté. Cette procédure est répétée sur tous les échantillons.

Le test « Leave-many-out » consiste à diviser les échantillons en K blocs, construire le modèle sur les $(K - 1)$ blocs et enfin le tester sur le K -ième bloc. Ce processus est répété un grand nombre de fois [214,215].

Dans le cas des régressions, La robustesse du modèle est évaluée en mesurant Q^2 , qui s'exprime par :

$$Q^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{(-i)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (14)$$

avec y_i : valeurs obtenues à l'aide de la méthode de référence pour les échantillons exclus de la phase d'apprentissage du modèle ; $\hat{y}_{(-i)}$: valeurs prédites par le modèle de régression construit après l'exclusion des échantillons ; \bar{y} : moyenne des valeurs de tous les échantillons utilisés pour construire le modèle initial.

Un Q^2 supérieur ou égal à 0,5 indique une bonne capacité prédictive du modèle.

*Partie II - Effets géoclimatiques,
morphologiques et temporels sur la
composition des huiles d'olive
libanaises et leur classification*

II-1. Introduction

Comme nous l'avons montré dans la première partie, de nombreuses études ont été menées afin de caractériser et classer les huiles d'olives Italiennes, Grecques et d'autres provenant de différentes régions du Moyen-Orient, et ceci en utilisant la technique RMN ^1H et la chimométrie.

Cependant, il faut signaler qu'une structure professionnelle, autorisée à accorder un certificat d'origine pour les huiles d'olive et à assurer le contrôle de la conformité du label, est absente au Liban. De même, il n'existe pas d'études sur l'identification de l'origine des huiles d'olives libanaises. Le consommateur libanais reste donc toujours sceptique et douteux concernant l'indication précisée sur les étiquetages et les pratiques frauduleuses.

Dans le cadre de cette problématique, l'élaboration d'une approche métabolomique rapide, par la technique RMN ^1H , pour l'identification de l'origine des huiles d'olive est devenue primordiale afin de protéger à la fois le consommateur et le producteur soucieux de qualité.

Dans cette étude, 187 échantillons d'olive ont été collectés en 2012 au nord et au sud du pays ainsi que dans la région du Mont-Liban, et cela à différentes altitudes.

Nous avons regroupé ces échantillons en fonction de leurs coordonnées géographiques.

Trois groupes d'altitude ont été créés : basse altitude (< 350 m), moyenne altitude (entre 350-700 m) et haute altitude (> 700 m) ; ainsi que deux groupes de latitude (voir **Figure 10**) : sud (à partir de la frontière sud du Liban jusqu'à 90 km vers le nord) et nord (plus que 90 Km de la frontière sud du pays).

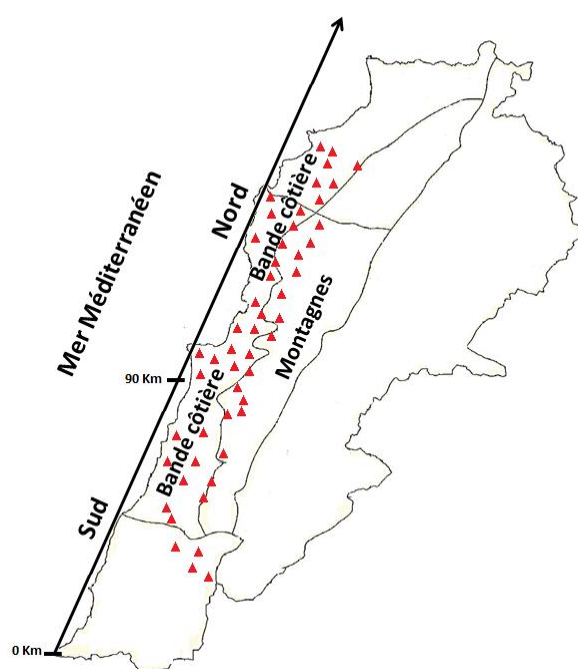


Figure 10 : Carte géographique du Liban.

Les triangles correspondent aux différentes régions libanaises d'où les échantillons d'olive ont été collectés.

La difficulté de cette classification réside dans le fait que les conditions géoclimatiques du nord et du sud du Liban ne sont pas tellement différentes.

Par ailleurs, une deuxième complication provient du fait que les variétés des olives libanaises ne peuvent pas être déterminées avec précision en raison des changements morphologiques qui peuvent avoir lieu suite à des variations dans les conditions géoclimatiques et les pratiques agricoles. Pour surmonter cette difficulté, les échantillons ont été triés en fonction de la couleur (noire ou verte), la taille (petite ou grande) et la forme des olives (ronde ou longue).

Les olives ont été récoltées sur le terrain et les huiles obtenues, suite à des extractions mécaniques en laboratoire, ont été analysées par la technique RMN ^1H . Nous avons utilisé cette méthode car elle est rapide et permet de fournir la quantité totale des acides gras saturés, insaturés, monoinsaturés, des acides linoléique et linoléique, ainsi que la quantité de quelques composants minoritaires (squalène, composés phénoliques, terpènes et autres).

L'ensemble des variables obtenu a ensuite été traité par le biais des outils statistiques, notamment l'analyse en composantes principales (PCA), l'analyse discriminante linéaire (LDA) et l'analyse canonique discriminante (CDA) afin de caractériser les huiles d'olive libanaises en fonction de leurs coordonnées géographiques (altitude et latitude du terroir), et en fonction de la morphologie (couleur, taille et forme) des olives.

Dans une seconde étape, nous avons étudié les effets temporels, tels que la date de récolte et l'année de production afin de déterminer leur influence sur la classification géographique des huiles. À cette fin, des échantillons d'olive ont été collectés de manière à permettre d'étudier ces effets. En 2013, des échantillons de 6 oliviers d'un même terroir, situé dans la région du Mont-Liban à une altitude de 520 m, ont été collectés à 4 reprises espacées de 7 jours (24 échantillons) ; le même type d'échantillonnage, 3 fois à des intervalles de 10 jours, a été effectué à 30 m d'altitude au nord du Liban (5 oliviers, 15 échantillons).

En 2014, 8 échantillons d'olive ont été recueillis dans les mêmes terroirs qu'en 2013 et 2012 (4 échantillons de chaque champ). Par la suite, des modèles robustes de classification basés sur les données RMN ^1H correspondantes ont été construits.

La méthodologie employée, les résultats obtenus et leur interprétation ont fait l'objet d'un article intitulé « Geoclimatic, morphological, and temporal effects on Lebanese olive oils composition and classification : A ^1H NMR metabolomic study ». Cet article est évalué par la revue Food Chemistry, il est révisé et resoumis.

II-2. Article

Geoclimatic, morphological, and temporal effects on Lebanese olive oils composition and classification: A ^1H NMR metabolomic study

Noelle Merchak ^{a, b}, Elias El Bacha ^a, Rola Bou Khouzam ^c, Toufic Rizk ^a, Serge Akoka ^b, Joseph Bejjani ^{a, *}

^a Research Unit: Technologies et Valorisation Agroalimentaire (TVA), Laboratory of Metrology and Isotopic Fractionation, Faculty of Science, Saint-Joseph University, P.O. Box 11-514 Riad el Solh, Beirut 1107 2050.

^b EBSI team, Interdisciplinary Chemistry: Synthesis, Analysis, Modelling (CEISAM), University of Nantes-CNRS UMR 6230, 2 rue de la Houssinière, BP 92208, F-44322 Nantes cedex 3, France.

^c Lebanese Atomic Energy Commission, P.O. Box 11-8281 Beirut.

Abstract

Two hundred and forty-two Lebanese olive samples were collected from different regions and the corresponding oils were analyzed by ^1H NMR spectroscopy. Thus, variables related to fatty acids and minor components were used as inputs in univariate and multivariate analyses aiming to characterize and classify the oils according to geographical, morphological, and temporal factors. Samples were sorted according to the colour, size, and shape of olives, which allowed significant classifications to be reached. A sequential strategy was developed to discriminate among samples from different altitudes and latitudes. Following this strategy, obvious trends and classifications were obtained at subregional level. Furthermore, the shift in the harvest date within a range of three weeks was considered and its effect on the classification models was investigated. Likewise, the effect of the production year was evaluated; the precipitation level in April and May plays a significant role.

Keywords:

Olive oil; Authentication; ^1H NMR metabolomics; Chemometrics; Fatty acids profile; Minor components; Harvest date effect; Production year effect

1. Introduction

Olive oil is a fundamental food ingredient recognized for its nutritional qualities and potential health benefits owed not only to its high oleic acid and antioxidant contents but also its uniquely high level of squalene, which is known to have an anticancer activity (Waterman, & Lockwood, 2007), as well as oleocanthal, a secoiridoid phenolic compound with potential therapeutic properties against inflammation cancer, and neurodegenerative diseases (Parkinson, & Russell, 2014). It is well established that the organoleptic properties of olive oil, which are strongly correlated to its geographical and varietal origins (Perri, Benincasa, & Muzzalupo, 2012), are behind its wide commercialization and elevated market value. All these features of olive oil are affected by several factors, such as variety of olives, geoclimatic conditions of the soil, and agricultural practices. In this context, the designation of olive oils according to their terroir (e.g., “protected designation of origin” and “protected geographical indication”) has become a primary labeling information prompting the need for an affordable analytical method of authentication (Perri et al., 2012). ^1H NMR spectroscopy can be used as a powerful tool for rapid quantification of fatty acids as well as minor components (e.g., squalene and phenolic compounds) without the need of any sample treatment (Dais, & Hatzakis, 2013). Many studies have been performed to characterize and classify olive oils using ^1H NMR spectroscopy in combination with statistical tools. Olive oils have been discriminated according to olive varieties grown in the same environment and collected at the same ripening stage based on minor components in the oil (Sacchi et al., 1996); geographical origin across Italy (Sacchi et al. 1998; Sacco et al., 2000, Longobardi et al., 2012; Mannina, Patumi, Proietti, Bassi, & Segre, 2001; D'imperio et al., 2007); and authentication of Greece olive oils (Petraakis, Agiomyrgianaki, Christophoridou, Spyros, & Dais, 2008; Alonso-Salces et al., 2015) as well as other olive oils from different Mediterranean areas (Mannina, Marini, Gobbino, Sobolev, & Capitani, 2010; Alonso-Salces, Moreno-Rojas, Holland, Reniero, Guillou, & Héberer, 2010). However, geographical and varietal characterizations of Lebanese olive oils, based on composition profiles, have not been investigated. Lebanon has a moderate Mediterranean climate, creating ideal conditions for olive trees cultivation. The Lebanese olive groves are almost exclusively based on non-irrigated native varieties that cover approximately 56,300 hectares, or 5.4% of the country land mass, with a productivity ranging from 1.5 to 3 Metric tonnes of olive per hectare, depending on the production year, and an oil content varying from 18 to 25% (Lebanese Ministry of Agriculture, 2003; Investment Development Authority of Lebanon, 2014). Olive production is spread across the country but can be grouped mainly into northern and southern regions, each generating around 40% of the

total crop. Thus, the two main regions are not readily differentiated in terms of their geoclimatic conditions, which complicate the regional classification of oils. Furthermore, an additional complication may arise due to the interference from varietal factors since the oils, to be discriminated based on geographical considerations cannot be accurately sorted according to varieties beforehand. In some cases, a given variety could not be determined due to morphological changes in response to different geoclimatic conditions and agricultural practices (Rotondi, Magli, Ricciolini, & Baldoni, 2003; Hagidimitriou, Katsiotis, Menexes, Pontikis, & Loukas, 2005). In this study, olive samples from northern, southern, and central regions of Lebanon were collected at different altitudes and the oils were analyzed using ^1H NMR. Metabolomic profiling of the oil samples was used to obtain characteristic variables related to saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), linoleic and linolenic acids, as well as minor components. The set of variables thus obtained was then treated by means of multivariate statistical procedures, notably the principal component analysis (PCA) and the linear discriminant analysis (LDA), to characterize Lebanese olive oils according to their geographical origin (altitude and latitude) as a function of olive morphology (color, size, and shape), taking into account the effects of the harvest date and the production year.

2. Materials and methods

2.1. Samples

Green and black olive samples (187), each weighing 300 g, were collected during the 2012 harvest from either the same tree or from several similar trees within the same plot. The samples were obtained from different olive-growing regions across Lebanon at various altitudes and latitudes. The olive regions were designated according to their altitude as low (Lw, 0-350 m), mid (Md, 350-700 m), and high (Hh, above 700 m), and their latitude as south (Sh, from the south boundary of Lebanon to 90 km toward the north) and north (Nh, above 90 km from the south boundary of Lebanon). Eleven sets of three or four samples were collected at 10- or 7-day intervals, respectively, during the 2013 harvest. Five of these sets were from northern Lebanon at an elevation of 30 m above sea level and six sets were from Mount Lebanon (central region) at an elevation of 520 m. During the 2014 harvest, eight olive samples were collected from the same olive groves as in 2013 (4 samples from each).

2.2. Olive oil extraction

Oils were obtained solely by mechanical methods in the laboratory. Olives were crushed, pitted and then ground to obtain a paste. The paste was centrifuged and the resulting oil layer was decanted. Oils were centrifuged to eliminate residual particles and then stored at -18 °C.

2.3. ^1H NMR analyses

^1H NMR spectra of olive oil samples were recorded on a Bruker Avance II spectrometer operating at 400.13 MHz. Samples were carefully prepared in 5 mm NMR tubes and the oil concentration was optimized in order to obtain the best compromise between sensitivity and resolution. A 25% (v/v) concentration of oil (180 μL) in chloroform-d (560 μL , purchased from Deutero GmbH) was used (Fauhl, Reniero, & Guillou, 2000). The longitudinal relaxation times (T_1) were determined using the inversion-recovery method. The longest T_1 (3.19 s) was observed for the methyl group of α -linolenic acid. For each sample, 7 spectra were recorded with the following conditions for acquisition: sample temperature 298 K; pulse angle 30°; pulse width 27.4 μs ; spectral width 9 ppm; time domain size 64 K; acquisition time 9.1 s; relaxation delay 1 s; 4 dummy scans and 32 transients. FIDs were zero-filled to 128 K and exponentially apodized with $\text{LB} = 0.3$ Hz. ^1H NMR spectra were manually phased. The baseline was corrected using the Cubic Spline technique of Bruker TOPSPIN 1.3 software. Chemical shifts were calibrated by setting the peak with the highest chemical shift corresponding to the glycerol moiety (CH_2 protons) at 4.299 ppm (**Figure S1** and **S2** in SM). Spectral resolution, expressed as intensity of the minimum between the signals at 4.288 and 4.299 ppm relative to intensity of the signal at 4.299 ppm (**Figure S1** in SM), was determined for each replicate spectrum; it never exceeds 25%. Spectral alignment was systematically checked in comparison with an olive oil spectrum chosen as reference. Peak areas were determined using fixed chemical shift intervals optimized in order to obtain the best repeatability. However, in rare cases, slight chemical shift deviations were noted and integration intervals were corrected accordingly. Integrals were normalised relative to the signal of protons at position 2 of acyl chains. The same protons were taken as reference when measuring intensities of signals (**Table 1** and **Figure S2** in SM). For each olive oil sample, one discrete and thirteen continuous variables (**Table 1**) related to fatty acids and minor components were retained for the statistical analysis (variables *Allylic CH₂*, *CH₂ position 2*, and *diallylic CH₂* were not considered; they were only used to calculate SFA, MUFA, and linoleic acid percentages).

Table 1: Olive oils variables obtained by ^1H NMR spectroscopy.

Variable ^a (abbreviation)	δ ppm	S_{RW} ^b %	σ ^c %	Mean	Range	
					Min	Max
Unknown compound (X)	0.512-0.528	11.0	62.9	0.12	0.02	0.63
β -Sitosterol (Ss)	0.655-0.672	7.5	27.1	0.74	0.34	1.57
Linolenic acid CH_3 (Ln) ^d	0.929-0.979	3.5	24.1	0.52	0.32	1.17
CH_2 intensity (I-1) ^e	1.238	0.8	8.8	4.37	3.13	5.71
CH_2 intensity (I-2) ^e	1.249	0.5	3.1	3.32	2.97	3.56
CH_2 intensity (I-3) ^e	1.283	0.5	2.6	3.49	3.17	3.74
Squalene (Sq)	1.640-1.666	1.7	25.5	8.0	2.5	14.6
Allylic CH_2 ^f	1.891-2.099	0.3	2.8	1022	925	1096
CH_2 position 2 (α) ^g	2.182-2.409					
Diallylic CH_2 ^h	2.653-2.866	0.5	22.0	74.0	34.0	128.1
1,2-Diglycerides (1,2-dig)	3.678-3.718	3.2	24.5	4.17	2.26	7.97
Terpenes (Trp-1)	4.563-4.579	18.3	54.4	0.07	0.00	0.29
Terpenes (Trp-2) ⁱ	4.6412; 4.6919				Absent ⁱ	High ⁱ
Phenolic compounds (Ph)	6.740-6.764	12.6	68.2	0.12	0.00	0.71
Linoleic acid (L) ^j		0.6	23.1	11.30	4.83	20.16
Monounsaturated fatty acids (MUFA) ^k		0.4	5.7	71.1	58.5	81.2
Saturated fatty acids (SFA) ^l		2.1	13.5	17.0	10.1	25.0

^a Variables were normalized relative to the signal of methylene at position 2 of fatty acids (its integral is set to 600); They are expressed as peak areas, unless specified otherwise. ^b Relative standard deviation of measurement corresponding to within-laboratory reproducibility (based on control charts over a one-year period). ^c Relative standard deviation of 187 Lebanese olive oil samples. ^d Expressed in molar percentage (Table S2 in SM). ^e Intensity of the signal at the mentioned chemical shift relative to the highest peak (2.285 ppm) of methylene signal at position 2 (Figure S2-B in SM). ^f Only used to calculate molar percentages of monounsaturated and saturated fatty acids. ^g Only used as a reference signal for peak areas, its value (600) divided by two represents the total number of fatty acid moles (as tri-, di-, monoglycerides, and as free fatty acids). ^h Only used to calculate molar percentages of monounsaturated and linoleic fatty acids. ⁱ Semiquantitated relative to the ^{13}C satellite signal of the glycerol moiety (4.435-4.489 ppm) and reported as: absent, faint, moderate or high (Figure 2). ^j Expressed in molar percentage; calculated from diallylic CH_2 and linolenic acid CH_3 signals (Table S2 in SM). ^k Expressed in molar percentage; calculated from allylic CH_2 , diallylic CH_3 , and squalene (1.653 ppm, 6H) signals; the last one was used in order to correct the fatty acids allylic CH_2 signal which overlaps with that of the 20 allylic protons of squalene (Table S2 in SM). ^l Expressed in molar percentage; calculated from allylic CH_2 and squalene (1.653 ppm) signals (cf. the precedent note and Table S2 in SM).

2.4. Statistical analysis

NMR data were submitted to statistical univariate and multivariate analyses, performed with TANAGRA data mining software (Rakotomalala, 2005). The applied chemometric tools were principal component analysis (PCA), canonical discriminant analysis (CDA), linear discriminant analysis (LDA), analysis of variance (ANOVA), and multivariate analysis of variance (MANOVA). A backward elimination approach was used as a guide to conduct variable selection for CDA classification models. When the number of instances did not permit to consider the whole set of variables, several subsets were tested and comprehensively

evaluated in order to select the most appropriate model. In each case, cross validation tests were performed in order to assess the robustness of the model.

3. Results and discussion

3.1. Variables determined by ^1H NMR and used in the classification of olive oils

Fatty acid profile of olive oil is affected by varietal and geoclimatic factors (Sacco et al., 2000, D'Imperio, Dugo, Alfa, Mannina, & Segre, 2007b). In a preliminary study to evaluate the impact of these factors on Lebanese olive oils, thirty-six authentic oil samples from various cultivars and olive regions were analyzed by gas chromatography. Fourteen fatty acids and squalene were detected and quantified. The results, summarized in **Table S1** in SM, showed significant differences between the analyzed samples which imply that fatty acids and squalene can be used as origin predictors of Lebanese olive oils at subregional level.

In this work, sampling was extended (187 samples) and ^1H NMR spectroscopy was used as a rapid metabolomic tool. Thirteen explanatory variables related to fatty acid composition and to minor components of olive oil were obtained and used in multivariate analyses (**Table 1**). Signal assignments were based on literature data (Alonso-Salces et al., 2010b). Measurement precisions (S_{Rw}), mean values, relative standard deviations of variables (σ) and their ranges are reported in **Table 1**. Mathematical equations allowing to calculate the molar percentages of SFA, MUFA, linoleic (L) and linolenic (Ln) acids as well as those of β -sitosterol (Ss), squalene (Sq), and 1,2-diglycerides (1,2-dig) are presented in **Table S2** in SM. Variables I-1, I-2, and I-3 (**Table 1**) are highly correlated to saturated (I-1 = $0.161\text{SFA} + 1.63$, Pearson's $r(187) = 0.97$, $p = 0$), monounsaturated (I-2 = $0.0237\text{MUFA} + 1.64$, Pearson's $r(187) = 0.91$, $p = 0$), and unsaturated (I-3 = $0.0360\text{UFA} + 0.509$, Pearson's $r(187) = 0.91$, $p = 0$) fatty acids, respectively. Nevertheless, they may afford additional information since they should be affected by the ratios of individual fatty acids (e.g., the ratio palmitic acid / stearic acid may affect I-1). Finally, a significant inverse correlation exists between squalene and β -sitosterol signals ($Ss = -0.0668\text{Sq} + 1.28$, Pearson's $r(187) = 0.68$, $p = 0$), which can be explained by the fact that squalene is the precursor of sterols (Seo et al., 1988).

3.2. Classification of oils according to the morphological characteristics of olive drupes

In this study, the classification of oils according to the morphological aspects of olives was investigated aiming to improve their geographical discrimination. Here we mention that olive samples were not genetically identified and their variety cannot be accurately determined

based on their morphological traits which may change under different geoclimatic conditions and agricultural practices (Hagidimitriou et al., 2005). First, samples were sorted into different categories according to the colour of olive drupes [black (Bk) or green (Gn)], their size [big (Bg, drupes > 2.2 cm), average (Av, 1.8 cm < drupes < 2.2 cm), or small (Sl, drupes < 1.8 cm)], and their shape [long (Lg) if the majority of drupes have the length/width ratio ≥ 1.25 , otherwise samples were considered round (Rd)]. In a second step, multivariate analyses, using variables reported in **Table 1**, were performed in order to discriminate between samples from these categories. The corresponding results are grouped in **Figure 1** and discussed in the appropriate sections below.

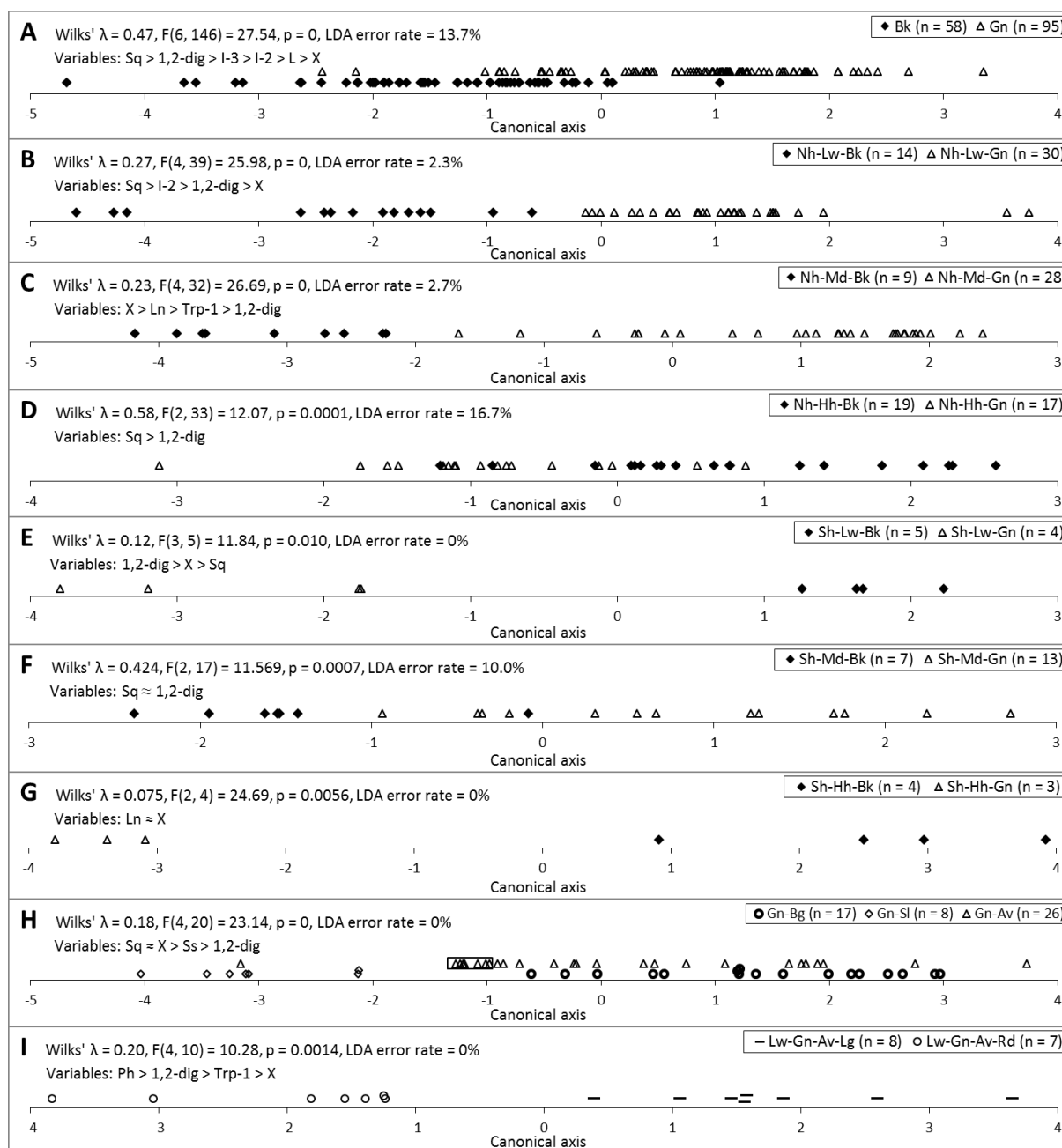


Figure 1: Classification of oils according to the morphological traits of olives. Variables used in the classifications are reported in decreasing order of importance based on Wilks' λ parameter.

3.2.1. Classification according to the colour of olives

A significant separation (Wilks' $\lambda = 0.469$, $p = 0$) between oils from black and green olives was achieved independently of their regional origin by means of the canonical discriminant analysis (CDA), the error rate given by the linear discriminant analysis (LDA) as classifier was 13.7% (**Figure 1-A**).

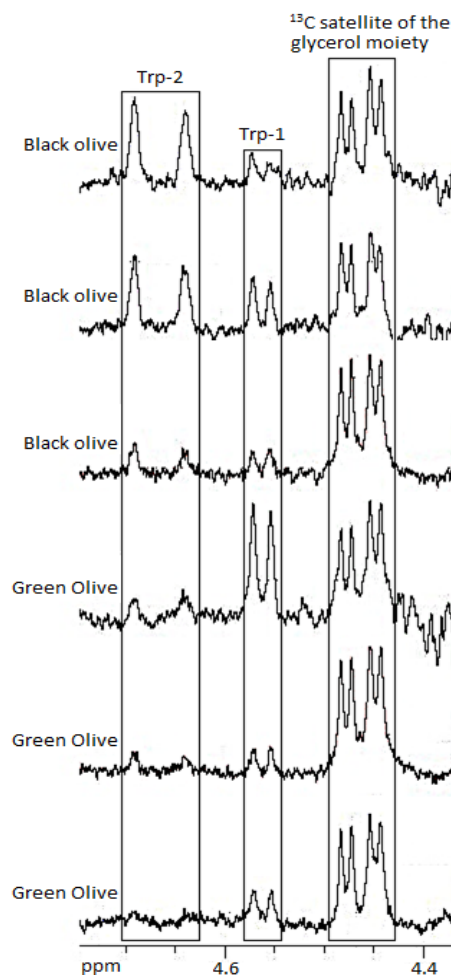


Figure 2: Differentiation of oils from black and green olives using variable Trp-2.

In order to assess whether the geographical origin of olives interferes with this classification, samples were grouped in a manner allowing to study the effect of one parameter at a time (i.e., olives colour effect). Oil samples were sorted according to their geographical coordinates (latitude and altitude) into north-low (Nh-Lw), north-mid (Nh-Md), north-high (Nh-Hh), south-low (Sh-Lw), south-mid (Sh-Md), and south-high (Sh-Hh). Afterwards, a CDA classification based on olives colours was carried out in each group (**Figure 1-B to G**). Following this procedure, the global LDA error rate decreased to 6.5% (based on weighted average of error rates reported in **Figure 1-B to G**). The T-test performed according to the

colour of olives (58 Bk and 95 Gn samples from the different geographical groups) showed significant variations within the oil matrix. L, Ln, 1,2-dig, Ss, and X were generally higher in oils from black olives compared with those from green ones. In contrast, Sq and Trp-1 were lower in black olives (see **Table 2** for variation percentages of variables and corresponding p values). Nevertheless, using solely the discrete variable Trp-2 (**Table 1**), a classification with an error rate of 9.8% was reached independently of the regional origin. As shown in **Figure 2**, Trp-2 signals were absent or very faint and smaller than Trp-1 signals in oils from green olives.

3.2.2. Classification according to the size of olives

The oils from Gn group were classified according to the size [big (Bg) or small (Sl)] of their corresponding olives by using the CDA (**Figure 1-H**). The oils from average-size green olives (Gn-Av) were not used to construct the model. However, when these samples were projected onto the canonical axis, 18 samples out of 26 fell into group Gn-Bg, 7 samples (shown in a rectangle in **Figure 1-H**) were between the two groups, but according to the LDA functions were classified as Gn-Sl, and 1 sample was clearly assigned as Gn-Sl. As a general trend, Gn-Bg oils were higher in 1,2-dig and Sq, lower in Ln and Ss in comparison with Gn-Sl oils (**Table 2**). In the case of Bk oils, it was not possible to classify the samples according to the size of olives.

3.2.3. Classification according to the shape of olives

As shown in (**Figure 1-I**), the oil samples from low altitude regions and from green average size olives (Lw-Gn-Av) were well classified according to the shape (Lg or Rd) of olives. As a general trend, Lw-Gn-Av-Lg oils have higher Ph values and lower X values than Lw-Gn-Av-Rd ones (**Table 2**). The factors colour, size, and altitude interfered with the classification of oils according to the shape of olives.

3.3. Classification of oils according to their geographical coordinates at subregional level

The classification of the Lebanese olive oils according to their geographical origin is a challenging task since the main olive regions in Lebanon (the North and the South) are not clearly differentiated in terms of their climatic conditions. Other challenges include the interference of altitude and varietal factors with this classification. The two main regions producing olive oil, which are on average 120 km apart, are coastal foothills ranging from sea level to 1200 m. They are characterized by an average annual precipitation of 900 mm with generally higher rainfall rates (around 150 mm) in the North than in the South. Taking into

account the geographical characteristics of the Lebanese territory, the olive regions were designated according to their altitude as Lw, Md, and Hh; and according to their latitude as Sh and Nh (cf. section 2.1). The discrimination of oils according to these subregions was investigated while considering the aforementioned interferences. Results are discussed in the following sections (corresponding graphs and statistics are presented in **Figure 3**).

Table 2: Variations of the olive oils variables in response to the studied geographical, morphological, and temporal factors.

Variable	Gn to Bk ^{a,b} %, <i>p</i>	Sl to Bg ^{a,c} %, <i>p</i>	Rd to Lg ^{a,d} %, <i>p</i>	Lw to Md ^{a,e} %, <i>p</i>	Md to Hh ^{a,e} %, <i>p</i>	Lw to Hh ^{a,e} %, <i>p</i>	Sh to Nh ^a %	Harvest date shift ^{a,f} %, <i>p</i>	Precipitation rate decrease ^{a,g} %, <i>p</i>
X	+ 67.3 0		- 43.4 0.09					- (Md) 18.1 0.03	
Ss	+ 25.5 0	- 32.4 0.001			- (Bk) 19.5 0.01			+ (Md) 6.8 0.008	
Sq	- 32.2 0	+ 27.95 0.01			+ (Gn) 9.1 0.08	+ (Gn) 12.7 0.01		- (Md) 8.1 0.01	
1,2-dig	+ 14.3 0.0002	+ 31.6 0.005			- (Gn) 16.9 0.01	- (Gn) 17.8 0.007		- (Lw) 14.7 0.002	
					- (Bk) 21.2 0.05			+ (Md) 29.6 0.003	
Trp-1	- 21.4 0.08				- (Gn) 40.6 0.0008			+ (Md) 29.2 0	
Trp-2 ^h	+								
Ph			+ 38.6 0.04	- (Gn) 65.0 0		- (Gn) 46.1 0.002		- (Md) ⁱ 51.0 0.01	+ 89.3 ^k 0.0007
				- (Bk) 50.8 0.05				+ (Md) ^j 45.6 0.02	+ 63.0 ^l 0.06
Ln	+ 9.9 0.01	- 32.6 0.005		- (Gn) 12.0 0.06		- (Gn) 18.5 0.03		- (Lw) 9.3 0.02	
						- (Bk) 19.7 0.02			
L	+ 12.3 0.001				- (Gn) 20.1 0.005	- (Gn) 27.7 0		+ (Md) 10.4 0.0001	
					- (Bk) 33.5 0.0006				
SFA				- (Gn) 8.8 0.0006	- (Gn) 12.8 0	- (Gn) 21.6 0	- 8.7 ^m	- (Md) 3.4 0.0005	
				- (Bk) 6.7 0.05	- (Bk) 9.0 0.04	- (Bk) 15.7 0.0001			
MUFA				+ (Gn) 3.6 0.006	+ (Gn) 5.5 0	+ (Gn) 9.1 0	+ 3.9 ^m	- (Md) 0.9 0.005	
					+ (Bk) 7.7 0.0002				

^a (+) and (-) signs represent, respectively, the increases and decreases of oils variables in response to the different factors, e.g., variations between oils from green and black olives (Gn to Bk); the first number represents the percentage of variation of the corresponding variable [(|average group1 – average group 2|/(av. gr. 1 + av. gr. 2)/2)*100]; the second number represents the *p* value obtained by using the T-test. ^b 58 Bk samples and 95 Gn samples. ^c 8 Sl and 17 Bg samples from Gn group. ^d 7 Rd and 8 Lg samples from GN-Av-Lw group. ^e 31 Gn-Lw, 28 Gn-Md, 16 Gn-Hh, 16 Bk-Lw, 9 Bk-Md, and 19 Bk-Hh samples from Nh group; the colour of olives is indicated in each case (i.e., Bk or Gn). ^f Variations were determined after normalization of the variables (section 3.4.1). Two sets of samples were used: the first set contains samples from 6 trees in a field at an elevation of 520 m in the Nh group at 4 different harvest dates (24 samples denoted by Md); the second set contains samples from 5 trees in a field at an elevation of 30 m in the Nh group at 3 different harvest dates (15 samples denoted by Lw). For the first set (Md), variations were determined between harvest dates 1 and 4, unless otherwise specified; for the second set (Lw), variations were determined between harvest dates 1 and 3. ^g Variations of variables according to the sum of precipitation rates in April and May of years 2012 (7.2 mm), 2013 (130.5 mm), and 2014 (18.4 mm); 12 samples from 2012, 11 samples from 2013, and 8 samples from 2014 obtained from the same olive fields at elevations of 30 and 520 m within Nh region. ^h Since Trp-2 is a discrete variable (Table 1), no percentages of variation and *p* values can be calculated. ⁱ Variation between dates 1 and 3. ^j Variation between dates 3 and 4. ^k Variation between years 2013 (130.5 mm) and 2012 (7.2 mm). ^l Variation between years 2013 (130.5 mm) and 2014 (18.4 mm). ^m Average of variations between Nh and Sh samples within the groups Bk-Lw, Bk-Md, Bk-Hh, Gn-Lw, and Gn-Md.

3.3.1. Classification according to the altitude of the olive grove

Statistical analysis NMR data related to the studied samples showed that the altitude of the olive grove is the most obvious factor affecting the classification of olive oils. Based on the PCA of samples from Nh group, the Lw/Hh distinction was more prominent than the Gn/Bk one, whereas the Md/Hh, Md/Lw, and Gn/Bk discriminations were approximately at the same level (**Figure S3-A to F** in SM). By using the CDA, a good classification of the oil samples as Lw, Md, and Hh was achieved independently of the colour of olives and the latitude of the grove (**Figure 3-A**). The performance of the classification models, expressed as LDA error rates, was as follows: 2.5% (Lw/Hh), 18.3% (Lw/Md), 12.3% (Md/Hh), and 22.4% (global error). In order to eliminate possible interferences from the colour (Bk or Gn) and latitude (Nh or Sh) factors, samples were grouped as Bk-Nh, Bk-Sh, Gn-Nh, and Gn-Sh, then a classification according to the altitude was performed within each group. A classification improvement was obtained in the Gn-Nh category (**Figure 3-B**) where the LDA error rates became: 0% (Lw/Hh), 5.1% (Lw/Md), 15.9% (Md/Hh), and 13.3% (global error). However, no enhancement was noted in Bk and Sh categories. This phenomenon could be due to the early ripening of black olives which may decrease the disparities in oil composition related to the differences in temperature and precipitation rate between low and high regions. Moreover, the southern regions are less elevated with relatively less steep foothills, higher temperatures, and lower precipitation rates than the northern regions which may exert similar effects (i.e., early ripening and decreased disparity between oils from low and high regions) to those suspected in the case of black olives (Motilva, Tovar, Romero, Alerge, & Girona, 2000; Oteros, García-Mozo, Vázquez, Mestre, Domínguez-Vilches, & Galán, 2013).

3.3.2. Variations within the olive oil matrix related to the altitude factor

Gradual decrease in temperature and increase in precipitation rate are noted starting from the coastal zone to the Mount-Lebanon peaks. Olive trees, spread over foothills till 1200 m above the sea level, and the corresponding oils should be affected by these climatic variations. Statistical treatment (T-test) of data related to fatty acids and minor components in oils revealed the impact of the altitude factor on the composition of this matrix. Significant trends were detected within Bk-Nh and Gn-Nh categories. These categories were considered separately in order to eliminate interferences of the colour and latitude factors and to verify whether Bk and Gn oils are affected in the same manner. Variations associated with the elevation in altitude were as follows for the Bk-Nh group: decrease of SFA (from Lw to Md and Md to Hh), Ph (from Lw to Md), Ss, L, 1,2-dig (from Md to Hh), and Ln (from Lw to

Hh); increase of MUFA (from Md to Hh). Variations observed in the Gn-Nh group were as follows: decrease of SFA (from Lw to Md and Md to Hh), Ln, Ph (from Lw to Md), Trp-1, L, and 1,2-dig (from Md to Hh); increase of MUFA (from Lw to Md and Md to Hh), and Sq (from Md to Hh). Variation percentages related to these variables are reported in **Table 2** along with the corresponding p values. The trends observed for SFA and MUFA are in good accordance with those obtained previously by Stefanoudaki, Kotsifaki, & Koutsaftakis (1999).

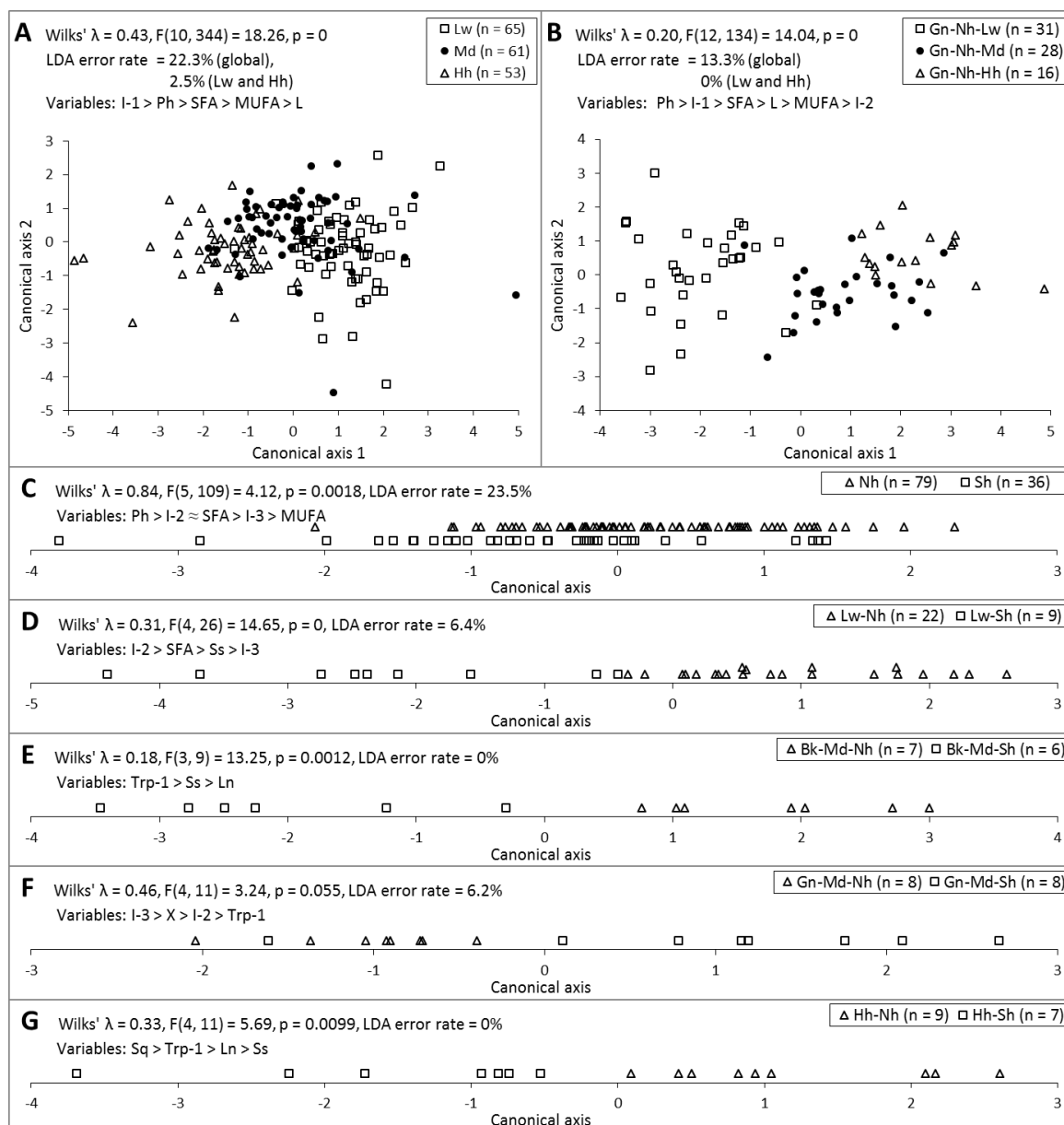


Figure 3: Classification of olive oils according to the altitude and latitude of the grove. Variables used in the classifications are reported in decreasing order of importance based on Wilks' λ parameter.

3.3.3. Classification according to the latitude of the olive grove

The effect of latitude on the olive oil matrix is less pronounced compared with those of altitude and olive colour. Samples in a population comprising oils from different groups (Bk, Gn, Lw, Md, and Hh) were poorly classified according to the latitude (**Figure 3-C**). However, a dramatic improvement in the classification was noted within the same group (**Figure 3-D to G**). Going from south to north across the studied regions, an increase in MUFA and a decrease in SFA were noted. The same trends were observed between low and high altitudes. The effect of an elevation of 290 m was found to be approximately equivalent to that of a 75 km shift towards the north, with an average increase of 3.9% for MUFA and an average decrease of 8.7% for SFA. This finding is consistent with the existing equivalence between the increase in precipitation rate with altitude and that with latitude.

3.4. Study of the effects of temporal factors on the classification of oils

The shift of the olive harvest date and the climatic changes between years may affect the classification of olive oils. Such effects were investigated in this study and the corresponding results are presented in the next sections. Related graphs and statistics are reported in **Figure 4**.

3.4.1. Effect of the harvest date

Several studies report significant variations in oil composition through the stages of growth and ripening of olives (Gutiérrez, Jiménez, Ruíz, & Albi, 1999; Matos et al., 2007; Beltrán, Del Rio, Sánchez, & Martínez, 2004; Imperio, Gobbino, Picanza, Costanzo, Corte, & Mannina, 2010). In this work, we studied the changes in the olive oil matrix due to shifts in the harvest date within an interval of three weeks between October and November, as it occurs in an actual situation in Lebanon. Besides, the impact of this temporal factor on the classification of oils according to their origin was evaluated. Fifteen olive samples from North Lebanon (Nh-Lw group) at an elevation of 30 m (5 different trees harvested three times at 10-day intervals) and twenty-four samples from Mount-Lebanon region (considered in the Nh-Md group in our study) at an elevation of 520 m (6 different trees harvested four times at 7-day intervals) were collected and the corresponding oils extracted then analyzed by ^1H NMR. The sampling was performed at 30 and 520 m in order to verify whether or not the harvest date effect is dependent on the altitude of the olive grove. The latitude factor was not considered since its effect is faint compared with that of altitude. In a first step, an unsupervised multivariate analysis (PCA), using the whole set of ^1H NMR variables as inputs,

was performed aiming to assess whether a classification based on the harvest date could be detected. As shown in **Figure 4-A**, it was possible to discriminate samples according to their olive tree and geographical origin, but no harvest date effect was noted. In a second step, a supervised multivariate analysis (CDA) was used in order to evaluate whether any subset of variables may permit to distinguish the oil samples according to the harvest date. For the low altitude (30 m) set of samples, it was not possible to detect any trend. However, in the case of the mid altitude (520 m) set, samples were significantly classified with only small overlaps between those from two consecutive weeks (**Figure 4-B**). Based on Wilks' λ parameter, Trp-1 was the most important variable in the classification. Aiming to understand what happens within the olive matrix during the last stage of ripening, the data was analyzed using ANOVA. Again, at low altitude (30 m) no significant difference was found, but at mid altitude (520 m) increases in Trp-1 (variation of 28.9%, $p = 0.004$), 1,2-dig (variation of 28.3%, $p = 0.02$), and L (variation of 10.0%, $p = 0.09$) were detected between harvest dates 1 and 4. It is likely that the difference between samples from Lw and Md regions is due to the decrease in temperature and increase in precipitation rate encountered while moving towards higher altitudes. As such, olives in Md regions undergo a slower ripening leading to olive samples in a stage where their composition is still changing (Motilva et al., 2000; Oteros et al., 2013). In order to improve the detection of differences related to the harvest date of olives, variations due to the varietal factor were eliminated through normalization of all variables in dividing them by the corresponding mean values of the 4 harvest dates (i.e., x_i is divided by the mean value of x_1 , x_2 , x_3 , and x_4 , where x_i is a variable of a given oil at a given harvest date with $i = 1, 2, 3$, and 4). After this normalization, the following trends were detected during the olive ripening (i.e., from date 1 to date 3 or 4): at low altitude (30 m), decrease in Ln and 1,2-dig; at mid altitude (520 m), increase in L, 1,2-dig, Trp-1, Ss, decrease in SFA, MUFA (very slight), X, and Sq (see **Table 2** for variation percentages and corresponding p values). The detected decrease in squalene content is in agreement with the conclusion reported by D'Imperio et al. (2010) and Baccouri et al. (2008). The increase in L and decrease in SFA coincide with the findings reported by Baccouri et al (2008), Gutiérrez, et al. (1999), and Beltrán, et al. (2004); however, the latter reference points out a disagreement in the trend of MUFA. Concerning the Ph variable in samples from the grove at 520 m, significant decrease from date 1 to 3 and increase from date 3 to 4 were noted (**Table 2**); a similar trend was reported by Motilva et al. (2000) for olive trees under regulated deficit irrigation, which is a situation somewhat like that of the rain-fed trees in the present study. Nevertheless, independently of the harvest date, all samples from the field at 520 m were found within the appropriate group (Nh-Md) when projected onto the classification model according to altitude (**Figure 4-C**). This finding

suggests that the effect of a shift in the harvest date within an interval of three weeks does not impair the geographical and varietal classifications of olive oils.

3.4.2. Effect of the harvest year

The harvest year is another factor that may affect the geographical classification of olive oils due to climatic variations observed across years. In order to test the robustness of the classification models in response to this factor, olive samples were collected during 2012, 2013, and 2014 harvest seasons from two olive fields in the Nh-Lw and Nh-Md subregions at elevations of 30 and 520 m, respectively. Based on the unsupervised PCA procedure, it was not possible to classify the corresponding oils according to their harvest year, but the geographical effect was obvious as the samples from the two different regions were clearly separated (**Figure 4-D**). In a second attempt to classify the samples based on their harvest year, the supervised CDA procedure was used, and thus a significant classification was reached (**Figure 4-E**) in which I-2 and MUFA were the most important variables based on Wilks' λ parameter, followed by Ph then Sq. It is noteworthy that I-2 and MUFA are highly correlated to the oleic acid content of the olive oil which is the fatty acid with the greatest dependence on the harvest year (Beltrán, et al., 2004). However, by means of ANOVA statistics, it was found that only the Ph variable is significantly affected by the production year; the other minor components and fatty acids variables were not statistically different. Moreover, the highest Ph value (0.12) corresponds to year 2012 characterized by the lowest rainfall level (7.2 mm) during the months of April and May (no precipitation were noted during the months of June, July, and August of the considered years). In 2013 and 2014, these variables (Ph value, rainfall level during April and May) were (0.046, 130.5 mm) and (0.088, 18.4 mm), respectively. This result suggests an inverse correlation between Ph and the precipitation level of April and May independently of the annual rate of precipitation (e.g., the annual precipitation rate in 2014 was extremely low, only 398 mm compared with 1095 and 1080 mm in 2012 and 2013, respectively; nevertheless, the Ph variable was not remarkably affected). A similar conclusion was reported by Motilva et al. (2000) regarding the phenolic compounds and the profile of fatty acids. In fact, the activity of phenylalanine ammonia-lyase responsible for phenolics synthesis is greater under higher water deficit; this phenomenon was proved in several previous studies (Patumi, D'andria, Fontanazza, Morelli, Giorio, & Sorrentino, 1999; Parr, & Bolwell, 2000; Romero, Tovar, Ramo, & Motilva, 2003; Morello, Romero, Ramo, & Motilva, 2005). However, as shown in **Figure 4-F**, the geographical classification of samples from the three harvest years considered in this study, which include one year with extreme climatic conditions (year 2014), remains significant.

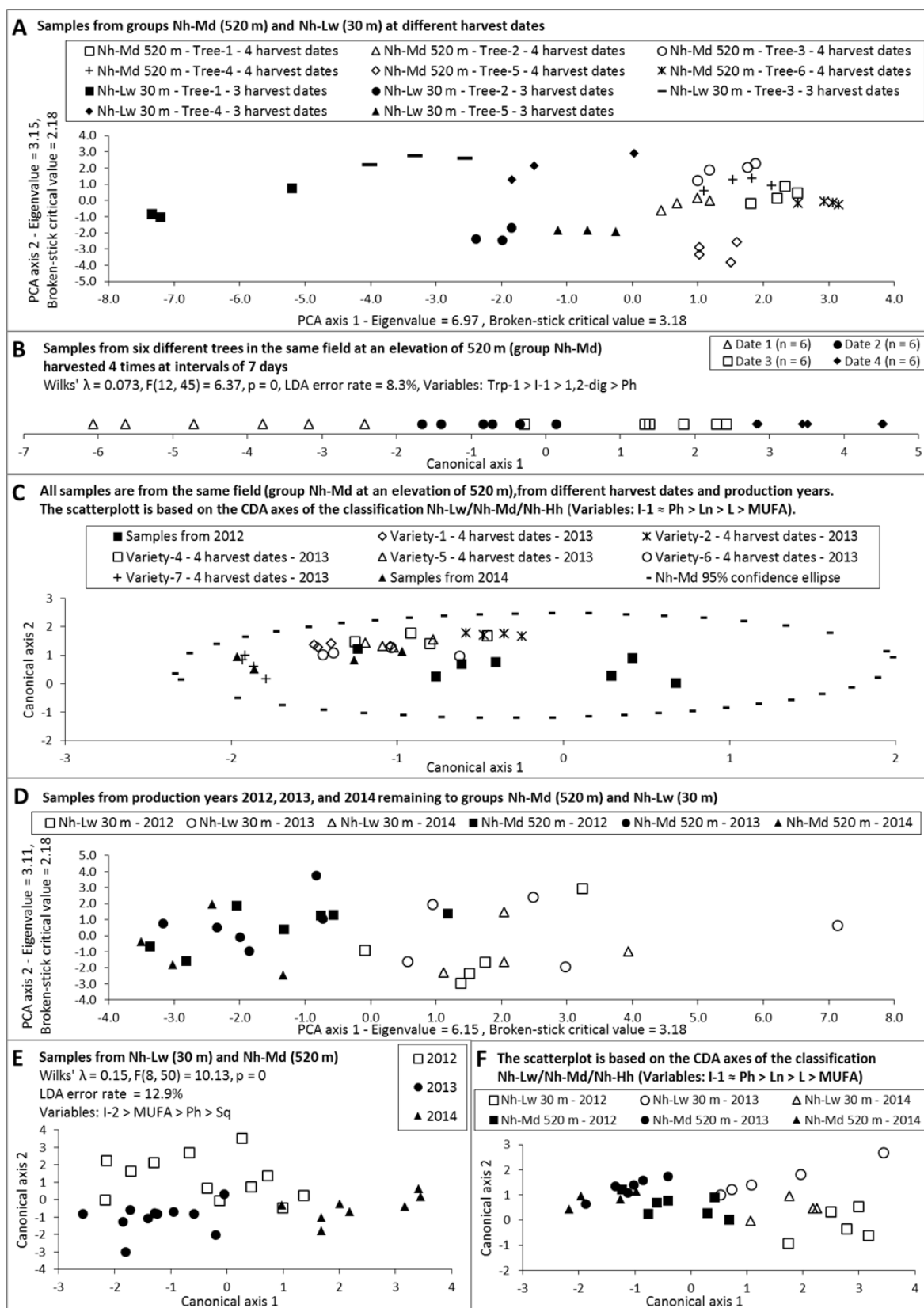


Figure 4: Harvest date and year effects on the classification of olive oils according to their origin. Variables used in the classifications are reported in decreasing order of importance based on Wilks' λ parameter.

It should be noted that the scatterplot in **Figure 4-F** is based on the general classification model according to the altitude in Nh group (constructed by using the samples of year 2012) and not on a specific classification for the samples in question. Moreover, in order to test whether the quantification of the harvest year effect (precipitation rate of April and May) on the Ph variable may help to improve the geographical classification of oils, Ph values of 2013 and 2014 samples from were incremented with ΔPh_{2013} (average Ph_{2012} – average Ph_{2013}) and ΔPh_{2014} (average Ph_{2012} – average Ph_{2014}), respectively; then the adjusted Ph values thus obtained were used as inputs (instead of the original ones) in the same classification model of **Figure 4-F**. Following this procedure, a clear improvement in the classification was observed (**Figure S4** in SM).

4. Conclusion

In this study, a sequential strategy, based on a set of olive oil variables rapidly obtained with sufficient precision by means of ^1H NMR metabolomics, was developed in order to construct statistical models allowing to classify the Lebanese olive oils according to some morphological traits of olives and to verify their geographical origin at subregional level. The related procedure first considers the prediction of the olives colour corresponding to the oil in question, which can be performed by projecting the oil sample into the adequate CDA model, or by using solely a discrete variable associated to an unidentified terpenic compound in olive oil. Once the colour of olives is determined, the verification of the altitude group to which the sample belongs becomes more accurate via appropriate CDA models. In a last step, the verification of the latitude group (North or South) of the sample can be performed in a similar manner as for altitude. Moreover, the effects of the main geographical, morphological, and temporal factors on the olive oil composition were quantified, which offered a better understanding of their mutual interferences, and thus allowed improving the classifications of olive oils.

Acknowledgements

N.M. acknowledges the financial support of the Lebanese National Council for Scientific Research and the Research Council of Saint-Joseph University. The CORSAIRE platform from Biogenouest is also acknowledged. The authors are grateful to Mrs. Adele Semaan for the language revision of the manuscript.

References

- Alonso-Salces, R.M., Moreno-Rojas, J.M., Holland, M.V., Reniero, F., Guillou, C., & Héberger, K. (2010a). Virgin olive oil authentication by multivariate analyses of ^1H NMR fingerprints and $\delta^{13}\text{C}$ and $\delta^2\text{H}$ data. *J. Agr. Food Chem.*, *58*, 5586-5596.
- Alonso-Salces, R.M., Héberger, K., Holland, M.V., Moreno-Rojas, J.M., Mariani, C., Bellan, G., Reniero, F., & Guillou, C. (2010b). Multivariate analysis of NMR fingerprint of the unsaponifiable fraction of virgin olive oils for authentication purposes. *Food Chem.*, *118*, 956-965.
- Alonso-Salces, R.M., Segebarth, N., Garmón-Lobato, S., Holland, M.V., Moreno-Rojas, J.M., Fernández-Pierna, J.A., Baeten, V., Fuselli, S.R., Gallo, B., Berrueta, L.A., Reniero, F., Guillou, C., & Héberger, K. (2015). ^1H -NMR and isotopic fingerprinting of olive oil and its unsaponifiable fraction: geographical origin of virgin olive oils by pattern recognition. *Eur. J. Lipid Sci. Tech.*, DOI: 10.1002/ejlt.201400243.
- Baccouri, O., Guerfel, M., Baccouri, B., Cerretani, L., Bendini, A., Lercker, G., Zarrouk, M., & Daoud Ben Miled, D. (2008). Chemical composition and oxidative stability of Tunisian monovarietal virgin olive oils with regards to fruit ripening. *Food Chem.*, *109*, 743-754.
- Beltrán, G., Del Rio, C., Sánchez, S., & Martínez, L. (2004). Influence of harvest date and crop yield on the fatty acid composition of virgin olive oils from Cv. Picual. *J. Agr. Food Chem.*, *52*, 3434-3440.
- Dais, P., & Hatzakis, E. (2013). Quality assessment and authentication of virgin olive oil by NMR spectroscopy: A critical review. *Anal. Chim. Acta*, *765*, 1-27.
- D'Imperio, M., Mannina, L., Capitani, D., Bidet, O., Rossi, E., Bucarelli, F.M., Quaglia, G.B., & Serge, A. (2007a). NMR and statistical study of olive oils from Lazio: A geographical, ecological and agronomic characterization. *Food Chem.*, *105*, 1256-1267.
- D'Imperio, M., Dugo, G., Alfa, M., Mannina, L., & Segre, A.L. (2007b). Statistical analysis on Sicilian olive oils. *Food Chem.*, *102*, 956-965.
- D'Imperio, M., Gobbino, M., Picanza, A., Costanzo, S., Corte, A.D., & Mannina, L. (2010). Influence of harvest method and period on olive oil composition: an NMR and statistical study. *J. Agr. Food Chem.*, *58*, 11043-11051.
- Fauhl, C., Reniero, F., & Guillou, C. (2000). ^1H NMR as a tool for the analysis of mixtures of virgin olive oil with oils of different botanical origin. *Magn. Reson. Chem.*, *38*, 436-443.
- Gutiérrez, F., Jiménez, B., Ruíz, A., & Albi, M. A. (1999). Effect of olive ripeness on the oxidative stability of virgin olive oil extracted from the varieties Picual and Hojiblanca and on the different components involved. *J. Agr. Food Chem.*, *47*, 121-127.
- Hagidimitriou, M., Katsiotis, A., Menexes, G., Pontikis, C., & Loukas, M. (2005). Genetic diversity of major Greek olive cultivars using molecular (AFLPs and RAPDs) markers and morphological traits. *J. Amer. Soc. Hort. Sci.*, *130*, 211-217.

- Investment Development Authority of Lebanon (2014). Olive oil industry in Lebanon.
- Lebanese Ministry of Agriculture (2003). L'olivier: Contraintes et Potentialités.
- Longobardi, F., Ventrella, A., Napoli, C., Humpfer, E., Schütz, B., Schäfer, H., Kotominas, M.G., & Sacco, A. (2012). Classification of olive oils according to geographical origin by using ^1H NMR fingerprinting combined with multivariate analysis. *Food Chem.*, *130*, 177-183.
- Mannina, L.; Patumi, M.; Proietti, N.; Bassi, D.; Segre, A. L. (2001). Geographical characterization of Italian extra virgin olive oil using high field ^1H NMR spectroscopy. *J. Agr. Food Chem.*, *49*, 2687-2696.
- Mannina, L., Marini, F., Gobbino, M., Sobolev, A.P., & Capitani, D. (2010). NMR and chemometrics in tracing European olive oils: the case study of Ligurian samples. *Talanta*, *80*, 2141-2178.
- Matos, L.C., Cunha, S.C., Amaral, J.S., Pereira, J.A., Andrade, P.B., Seabra, R.M., & Oliveira, B.P.P. (2007). Chemometric characterization of three varietal olive oils (Cvs. Cobrançosa, Madural and Verdeal Transmontana) extracted from olives with different maturation indices. *Food Chem.*, *102*, 406-414.
- Morello, J. R., Romero, M. P., Ramo, T., & Motilva, M. J. (2005). Evaluation of L-phenyllanine ammonia-lyase activity and phenolic profile in olive drupe from fruit setting period to harvesting time. *Plant Sci.*, *168*, 65–72.
- Motilva, M.J., Tovar, M.J., Romero, M.P., Alerge, S., & Girona, J. (2000). Influence of regulated deficit irrigation strategies applied to olive trees (Arbequina cultivar) on oil yield and oil composition during the fruit ripening period. *J. Sci. Food Agric.*, *80*, 2037-2043.
- Oteros, J., García-Mozo, H., Vázquez, L., Mestre, A., Domínguez-Vilches, E., & Galán, C. (2013). Modelling olive phenological response to weather and topography. *Agric. Ecosyst. Environ.*, *179*, 62-68.
- Parkinson, L., & Russell, K. (2014). Oleocanthal, a Phenolic Derived from Virgin Olive Oil: A Review of the Beneficial Effects on Inflammatory Disease. *Int. J. Mol. Sci.*, *15*, 12323-12334.
- Parr, A.J., & Bolwell, G.P. (2000). Phenols in the plant and in man. The potential for possible nutritional enhancement of the diet by modifying the phenols content or profile. *J. Sci. Food Agr.*, *80*, 985-1012.
- Patumi, M., D'andria, R., Fontanazza, G., Morelli, G., Giorio, P., & Sorrentino, G. (1999). Yield and oil quality of intensively trained trees of three cultivars of olive (*Olea europaea* L.) Under different irrigation regimes. *J. Hortic. Sci. Biotech.*, *74*, 729-737.
- Perri, E., Benincasa, C., & Muzzalupo, I. (2012). Olive oil traceability. *Olive Germplasm - The Olive Cultivation, Table Olive and Olive Oil Industry in Italy*, Ed. I. Muzzalupo, InTech, Ch. 13, 265-286.

- Petrakis, P.V., Agiomyrgianaki, A., Christophoridou, S., Spyros, A., & Dais, P. (2008). Geographical Characterization of Greek Virgin Olive Oils (Cv. Koroneiki) Using ^1H and ^{31}P NMR Fingerprinting with Canonical Discriminant Analysis and Classification Binary Trees. *J. Agr. Food Chem.*, *56*, 3200-3207.
- Rakotomalala, R. (2005). TANAGRA: a free software for research and academic purposes. In: Proceedings of EGC'2005, RNTI-E-3, vol. 2, 697-702.
- Romero, M.P., Tovar, M.J., Ramo, T., & Motilva, J. (2003). Effect of crop season on the composition of virgin olive oil with Protected Designation of Origin "Les Garrigues". *J. Am. Oil Chem. Soc.*, *80*, 423-430.
- Sacchi, R., Patumi, M., Fontanazza, G., Barone, P., Fiodiponti, P., Mannina, L., Rossi, E., & Serge, A.L. (1996). A High-Field ^1H Nuclear Magnetic Resonance Study of the minor components in virgin olive oils. *J. Am. Oil Chem. Soc.*, *73*, 747-758.
- Sacchi, R., Mannina, L., Fiordiponti, P., Barone, P., Paolillo, L., Patumi, M., & Serge, A. (1998). Characterization of Italian extra virgin olive oils using ^1H -NMR spectroscopy. *J. Agr. Food Chem.*, *46*, 3947-3951.
- Sacco, A., Brescia, M.A., Liuzzi, V., Reniero, F., Guillou, C., Ghellid S., & van der Meer, P. (2000). Characterization of Italian olive oils based on Analytical and Nuclear Magnetic Resonance determinations. *J. Am. Oil Chem. Soc.*, *77*, 619-625.
- Seo, S., Uomori, A., Yoshimura, Y., Takeda, K., Seto, H., Ebizuka, Y., Noguchi, H., & Sankawa, U. (1988). Biosynthesis of sitosterol, cycloartenol, and 24-methylenecycloartanol in tissue cultures of higher plants and of ergosterol in yeast from $[1, 2-^{13}\text{C}_2]$ - and $[2-^{13}\text{C}^2\text{H}_3]$ -acetate and $[5-^{13}\text{C}^2\text{H}_2]$ MVA. *J. Chem. Soc., Perkin Trans. 1*, *8*, 2407-2414.
- Stefanouadaki, E., Kotsifaki, F., & Koutsaftakis, A. (1999). Classification of virgin olive oils of the two major Cretan cultivars based on their fatty acid composition. *J. Sci. Food Agr.*, *76*, 623-626.
- Waterman, E., & Lockwood, B. (2007). Active components and clinical applications of olive oil. *Alternat. Med. Rev.*, *12*, 331-342.

II-3. Données complémentaires

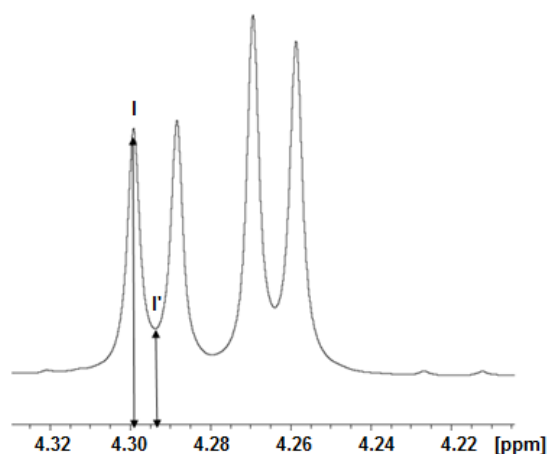


Figure S1: Intensities used to determine the spectral resolution ($100 \cdot I'/I$).

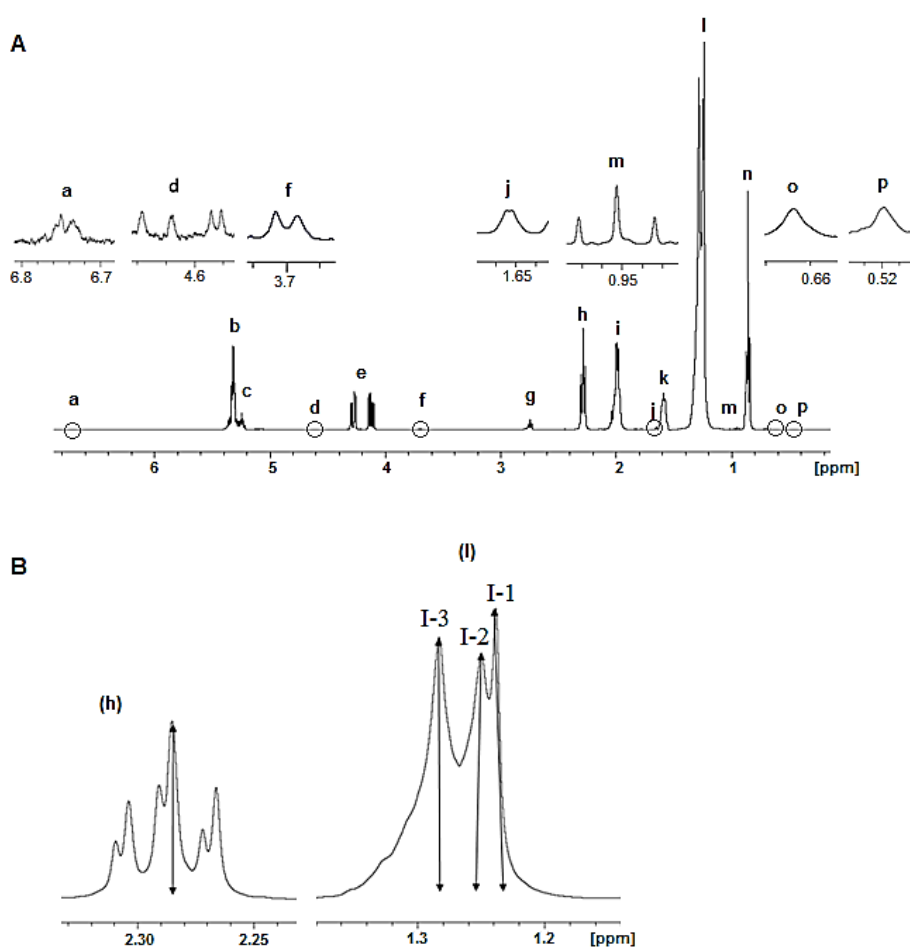


Figure S2: A) ^1H NMR spectrum of olive oil: phenolic compounds (a); olefinic protons (b); sn-2 glycerol proton (c); terpenes (d); sn-1,3 glycerol protons (e); 1,2-diglycerides (f); diallylic protons (g); CH_2 at position 2 (h); allylic protons (i); squalene (j); CH_2 at position 3 (k); aliphatic CH_2 protons (l); methylenic protons of linolenic acid (m); methylenic protons of fatty acids chains (n); β -sitosterol (o); unknown compound X (p). B) Peaks used to calculate intensities I-1, I-2 and I-3

Table S1: Fatty acids ranges in Lebanese olive oils

	Fatty acid ^a	Fatty acid range (%)	Mean (%) ^b	Sd ^c
Saturated	Palmitic (P, C16:0)	8.5 – 21.9	13.7	2.9
	Margaric (M, C17:0)	nd ^d – 0.34	0.17	0.07
	Stearic (S, C18:0)	1.6 – 4.7	3.5	0.81
	Arachidic (A, C20:0)	0.25 – 0.58	0.46	0.08
	Behenic (B, C22:0)	0.058 – 0.14	0.11	0.02
	Lignoceric (Lg, C24:0)	nd ^d – 0.065	0.047	0.01
Monounsaturated	Palmitoleic (Po, <i>c</i> -9-C16:1)	0.34 – 3.9	0.81	0.69
	Margaroleic (Ma, <i>c</i> -9-C17:1)	0.067 – 0.32	0.21	0.07
	Oleic (O, <i>c</i> -9-C18:1)	50.9 – 78.3	67.2	6.3
	Vaccenic (V, <i>c</i> -11-C18:1)	1.2 – 4.8	1.8	0.72
	Gondoic (G, <i>c</i> -11-C20:1)	0.19 – 0.39	0.27	0.04
Polyunsaturated	Linoleic (L, <i>c</i> -9,12-C18:2)	4.5 – 19.6	11.4	3.4
	Linolenic (Ln, <i>c</i> -9,12,15-C18:3)	0.42 – 0.94	0.64	0.13

^a Fatty acids composition was determined by gas chromatography. Fatty acids methyl esters were prepared by shaking 200 mg of oil in 3mL of hexane with 0.4 mL of 2N methanolic potassium hydroxide [^{1,2}]. An Agilent Technologies chromatograph equipped with a flame ionization detector and a fused-silica capillary column (SGE-054616, 30m × 0.32mm i.d., BPX70 0.25µm) was used in this study with helium as carrier gas at a flow of 1.0 mL/min. Oven temperature was programmed as follows: at 160 °C for 15 min, to 200 °C at a rate of 10 °C/min, at 200 °C for 10 min. A 1:10 dilution in hexane was performed. A volume of 1µL was injected with a split ratio of 50:1 and an injector temperature of 200 °C. The detector temperature was set at 270 °C. Two injections were performed for each sample. Retention times were determined by using standard fatty acid methyl esters. Molar percentages of palmitic, palmitoleic, margaric, margaroleic, hypogeic, stearic, oleic, vaccenic, linoleic, linolenic, arachidic, gondoic, behenic, and lignoceric acids in olive oils were determined relatively to their total number of moles by dividing each peak area by the molar weight of the corresponding fatty acid methyl ester. Squalene was also quantified in the same way. ^b Mean of 36 analyzed samples. ^c Standard deviation of samples. ^d Non-detected.

¹ European Economic Community (1977). European Community regulation no. 72/77. Amending Regulation (EEC) No 1470/68 on the drawing and reduction of samples and the determination of oil content, impurities and moisture in oil seeds.

² International Olive Oil Council (2001). Preparation of the fatty acid methyl esters from olive oil and olive-pomace oil. <http://www.internationaloliveoil.org/estaticos/view/222-standards>. (Accessed 02.06.14).

Table S2: Equations for percentages of some fatty acids and minor components in olive oil

Component (abbreviation)	Equation ^a
% Linolenic acid CH ₃ (Ln)	$[(m/3)/(h/2)]*100$
% Linoleic acid	$[((g/2) - (2*m/3))/(h/2)]*100$
% Unsaturated fatty acids (UFA)	$[((i - (j*20/6))/4)/(h/2)]*100$
% Monounsaturated fatty acids (MUFA)	$[((i - (j*20/6)) /4 - (g/2) + (m/3))/(h/2)]*100$
% Saturated fatty acids (SFA)	100 – UFA
% β -sitosterol (Ss)	$[(o/3)/(h/2)]*100$
% Squalene (Sq)	$[(j/6)/(h/2)]*100$
% 1,2-Diglycerides (1,2-dig)	$[(f/2)/(h/2)]*100$

^a f, g, h, i, j, l, m, and o are areas of signals presented in Fig. S2-A; equations to calculate molar percentages relatively to the total number of fatty acid moles (as tri-, di-, monoglycerides, and as free fatty acids).

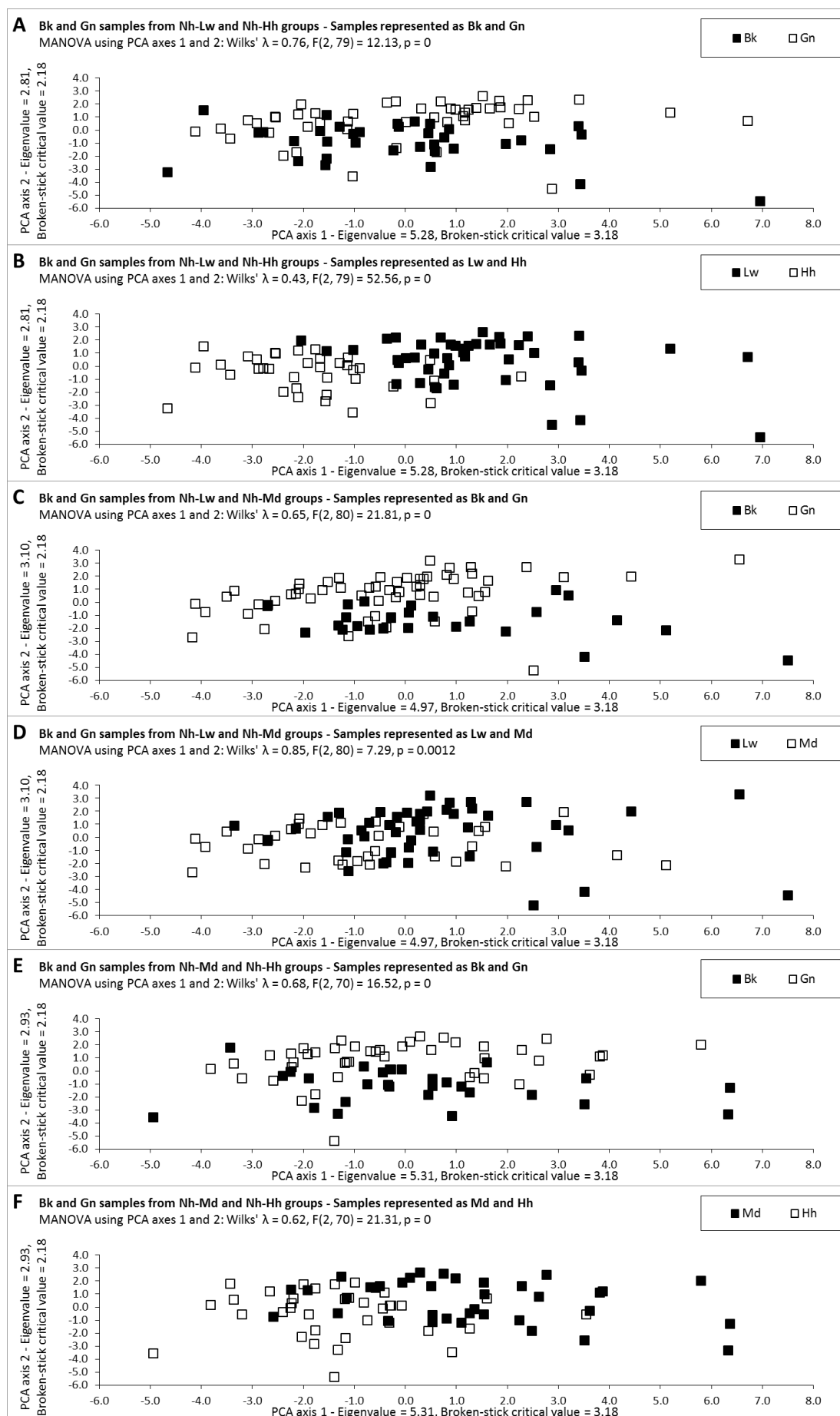


Figure S3: Comparison of the altitude and colour of olives effects on the classification of olive oils.

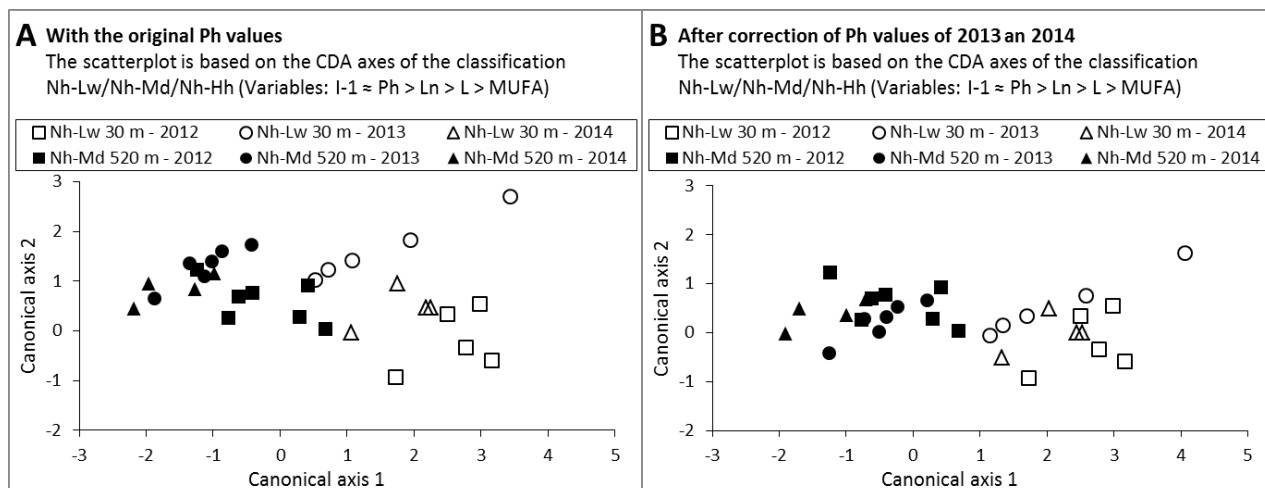


Figure S4: Classifications of olive oils according to the harvest year before (A) and after (B) Ph values adjustment.

II-4. Conclusion

Dans le cadre de cette étude, nous avons développé une stratégie séquentielle, basée sur l'ensemble des variables obtenues par l'analyse RMN ^1H , dans le but de construire des modèles statistiques permettant de discriminer les huiles d'olive libanaises selon leurs caractéristiques morphologiques et de vérifier leur origine géographique au niveau sous-régional.

Cette procédure consiste tout d'abord à prédire la couleur des olives, tout en projetant l'échantillon d'huile en question dans le modèle de l'analyse canonique discriminante (CDA) adéquat. Nous avons constaté que les olives vertes sont plus riches en squalène que les noires. De plus, nous avons découvert que des composés terpéniques, ayant des déplacements chimiques à 4,65 et 4,70 ppm, peuvent être considérés comme des marqueurs capables de discerner nettement les huiles selon la couleur des olives indépendamment du terroir d'origine. La présence de ces composés est un indicateur efficace de la couleur noire des olives.

Une fois la couleur des olives prédite, la vérification du groupe d'altitude, auquel appartient l'échantillon, devient plus précise et cela en utilisant les modèles d'analyse discriminantes appropriés. À cet égard, les acides gras saturés ainsi que les acides linoléique et linoléique diminuent lorsque l'altitude augmente. En revanche, les acides gras monoinsaturés augmentent. Ces résultats sont en accord avec ceux obtenus dans de nombreuses études [63-65]. De même, les huiles provenant des basses altitudes se caractérisent par des teneurs plus élevées en composés phénoliques, cette observation a été confirmée également par d'autres chercheurs [63].

Dans une dernière étape, la vérification de la latitude du terroir, d'où provient l'échantillon, peut être effectuée d'une manière similaire à celle de l'altitude.

Les huiles provenant du nord sont plus riches en monoinsaturés tandis que les huiles du sud contiennent plus d'acides saturés. Ces effets ont été observés dans des études précédentes sur les huiles d'olive tunisiennes parmi lesquelles les huiles du sud contiennent une teneur en acide palmitique plus élevée que celles du nord, tandis que la teneur élevée en acide oléique est observée dans les huiles du nord [75].

D'autre part, les effets temporels tels que la date de récolte et l'année de production n'affectent pas la classification géographique des échantillons, celle-ci se caractérise par un pouvoir discriminant beaucoup plus puissant que les effets temporels. Néanmoins, la quantité de précipitation au cours des mois d'avril et de mai affecte la production des composés phénoliques. En effet, l'activité de la phénylalanine ammonia-lyase, responsable de la synthèse des composés phénoliques, augmente en cas de déficit hydrique; ce phénomène a été démontré dans plusieurs études antérieures [81-83].

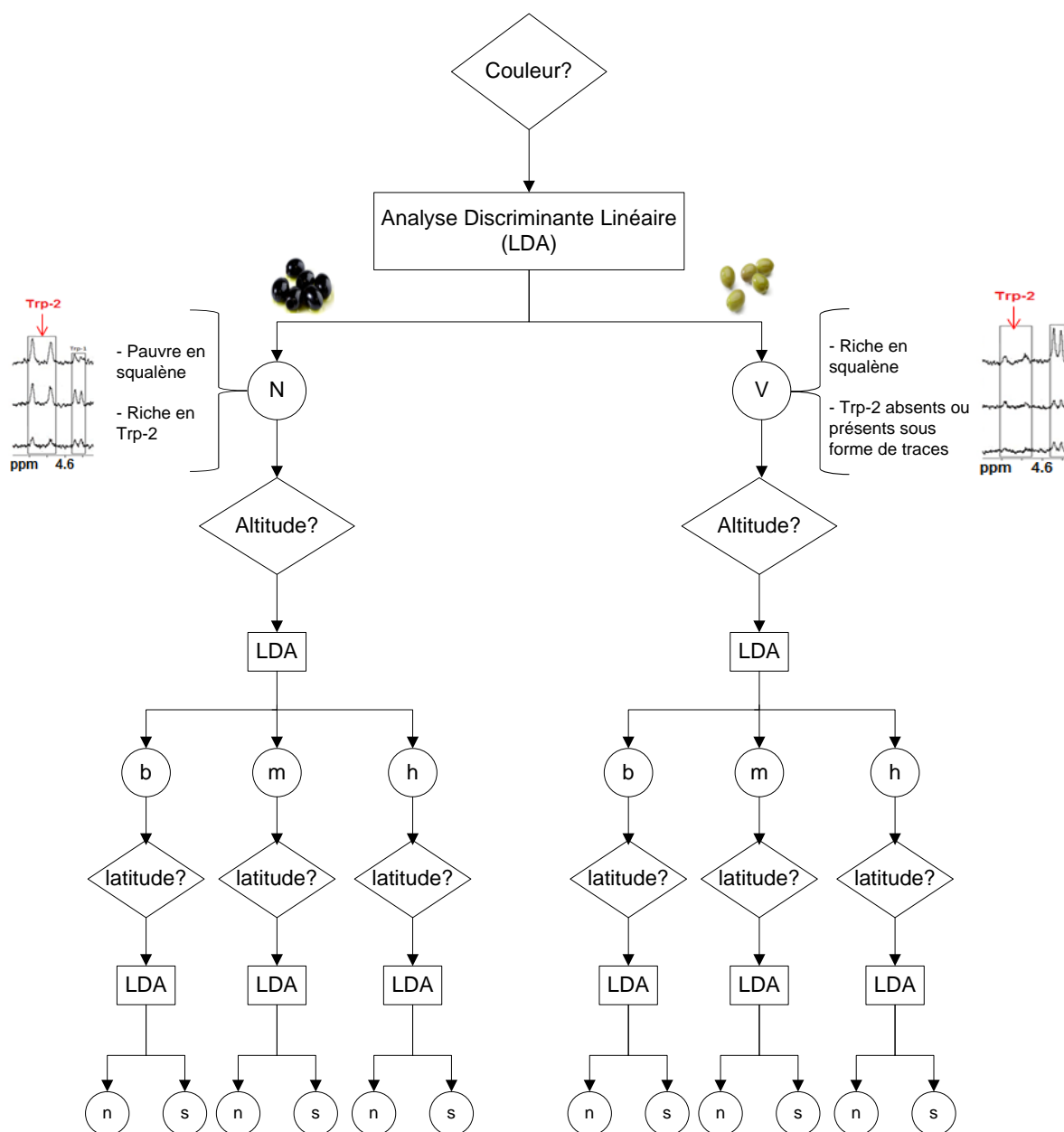


Figure 11 : Procédure à suivre pour prédire l'origine de l'olive.

Par ailleurs, en étudiant l'effet de la maturité, des variations ont été observées dans le cas des huiles provenant du terroir à 520 m et non de celui à 30 m d'altitude. Cela peut être dû à une maturité plus tardive pour ces échantillons.

En conclusion, les effets des principaux facteurs géographiques, morphologiques, et temporels sur la composition des huiles d'olive libanaises ont été étudiés, ce qui a permis une meilleure compréhension de leurs interférences, et une amélioration de la classification des huiles.

*Partie III - Étude isotopomique des
triacylglycérols par RMN ¹³C en
utilisant des techniques de transfert de
polarisation*

III-1. Introduction

Comme nous l'avons déjà mentionné, un recouvrement important de signaux est observé au niveau des spectres RMN ^1H des triacylglycérols. Cela est dû à la gamme restreinte de déplacement chimique des protons (environ 15 ppm). La quantification de la somme des acides gras saturés, monoinsaturés et polyinsaturés peut être obtenue au moyen d'équations mathématiques et ceci en utilisant les intensités des signaux appropriés. Mais seuls les acides linoléique et linoléique peuvent être déterminés individuellement. En revanche, une meilleure discrimination des signaux est observée en RMN ^{13}C grâce à une gamme spectrale beaucoup plus étendue (environ 250 ppm). Toutefois, la RMN ^{13}C est beaucoup moins sensible que la RMN ^1H ce qui induit des temps d'acquisition longs qui expliquent le fait que le noyau ^{13}C soit beaucoup moins exploité pour l'analyse des mélanges de TAG.

D'autre part, les rapports isotopiques $^{13}\text{C}/^{12}\text{C}$ peuvent également être utilisés pour la caractérisation des matrices alimentaires. Ils sont classiquement déterminés en utilisant la SMRI. Cette technique est destructive, et reste limitée à l'obtention d'une valeur moyenne de $\delta^{13}\text{C}$ pour l'ensemble des carbones de la molécule étudiée. Des réactions de dégradation chimique et de fragmentations moléculaires longues sont obligatoires pour obtenir les rapports isotopiques spécifiques $^{13}\text{C}/^{12}\text{C}$ de certains sites. Ces réactions sont susceptibles d'entraîner un fractionnement isotopique indésirable et peuvent fausser les mesures. Contrairement à la SMRI, la RMN ^{13}C isotopique est un outil qui permet d'obtenir les rapports isotopiques site par site.

Cette technique se caractérise donc par sa capacité à donner, simultanément, les profils métabolomique et isotopique, c'est ce qu'on appelle l'analyse isotopomique. Cependant, une précision de l'ordre de 1‰ est indispensable afin de garantir une interprétation correcte des mesures isotopiques, puisque la gamme de déviation en abondance naturelle $^{13}\text{C}/^{12}\text{C}$ est très restreinte (*e.g.*, $\Delta\delta^{13}\text{C}$ entre les C1, C2 et C3 du glycérol est comprise entre 1,6 et 40,9 pour mille). Une précision élevée (1-2‰) peut effectivement être obtenue avec une séquence RMN ^{13}C mono-impulsionnelle, mais seulement au prix d'un très long temps d'expérience, en raison : de la faible abondance du ^{13}C (1,1%), de son petit rapport gyromagnétique (1/4 celui du proton) et de ses temps de relaxation longitudinale (T_1) longs.

Face à ce problème, les méthodes de transfert de polarisation d'un noyau sensible (^1H) vers un noyau à faible rapport gyromagnétique (^{13}C), DEPT et INEPT, offrent un potentiel intéressant en permettant une augmentation significative de la sensibilité, et donc une réduction importante du temps d'expérience. Cette réduction est encore amplifiée par le fait que le

temps de répétition dépend alors des T_1 des protons qui sont beaucoup plus courts que ceux des ^{13}C .

Cependant, les séquences multi-impulsionnelles conventionnelles ne permettent pas d'atteindre une précision de quelques pour mille, pourtant essentielle pour les analyses isotopiques [158]. Cela est principalement dû aux imperfections des impulsions RF provenant des erreurs de calibration, de l'inhomogénéité du champ B_1 et des effets d'« off-resonance ».

Des travaux réalisés notamment au sein du laboratoire CEISAM ont montré que l'incorporation des impulsions composites et adiabatiques dans les séquences de transfert de polarisation (INEPT et DEPT) permet de compenser les imperfections d'impulsions RF et la précision nécessaire à l'analyse isotopique ^{13}C est ainsi obtenue [159]. D'autre part, dans une étude ultérieure, contrairement à la séquence DEPT, l'INEPT adiabatique (**Figure S1** dans la partie III-3) s'est révélée plus stable à long terme ; elle a permis d'aboutir, dans un temps d'expérience réduit par rapport à la séquence mono-impulsionnelle, à une précision de l'ordre du pour mille nécessaire pour les analyses isotopiques [160].

Dans le cadre de cette thèse, nous avons adapté la technique INEPT adiabatique aux matrices TAG afin d'obtenir des informations métabolomiques, positionnelles (concernant la distribution des acides gras sur le squelette du glycérol) et isotopiques (sur les carbones dont la fraction molaire ne varie pas avec le profil métabolomique). Puis, nous avons exploité les potentialités de cette technique dans le domaine de l'authentification des huiles d'olive selon leur origine géographique.

Les résultats obtenus suite à cette étude ont fait l'objet d'un article intitulé « ^{13}C isotopomics of triacylglycerols using NMR with polarization transfer techniques » : Noelle Merchak, Joseph Bejjani, Toufic Rizk, Virginie Silvestre, Gerald Remaud and Serge Akoka. *Analytical Methods*, 2015,7, 4889-4891.

III-2. Article

CrossMark
click for updatesCite this: *Anal. Methods*, 2015, 7, 4889Received 13th May 2015
Accepted 19th May 2015

DOI: 10.1039/c5ay01250c

www.rsc.org/methods

^{13}C isotopomics of triacylglycerols using NMR with polarization transfer techniques†

Noelle Merchak,^{ab} Joseph Bejjani,^{*b} Toufic Rizk,^b Virginie Silvestre,^a
Gerald S. Remaud^a and Serge Akoka^a

A rapid and very precise method is proposed to simultaneously determine the position-specific ^{13}C content and the metabolomic profile of triacylglycerols (isotopomics). ^{13}C -NMR spectra are obtained by using an adiabatic INEPT sequence; the experiment time is reduced by a factor of 7 with respect to a one-pulse acquisition.

Nuclear Magnetic Resonance (NMR) is highly recognized as a fast and accurate technique for identification and quantitation of tracers in food matrices.¹ In this respect, ^1H NMR spectroscopy allows precise determination of the compositional profile.² However, ^{13}C NMR offers an interesting alternative as it provides a much better discrimination of resonances,³ but it suffers from lower sensitivity.

Isotopic ratios can also be used to characterize food matrices. They are routinely determined using mass spectrometry (irm-MS: isotope ratios monitoring by Mass Spectrometry).⁴ However, this technique, based on the complete combustion of the molecules, only provides average values of the heavy isotope content of each element.⁵ Valuable information on the isotope intramolecular distribution is therefore inevitably lost. NMR is the only analytical technique which allows a site-by-site isotopic ratio determination. The isotope ^{13}C profile determined by ^{13}C NMR should reflect the history of the molecule, including the raw materials used and the chemical and/or biochemical pathways.

^{13}C NMR permits therefore simultaneous determination of both metabolomic and isotopic profiles (isotopomics). But in the case of ^{13}C , a precision of about 1 per mil is necessary for isotopic analysis since the range of ^{13}C isotopic compositions ($\delta^{13}\text{C}$) within a given molecule could be highly restricted, e.g., $\Delta\delta^{13}\text{C}$ between C1, C2, and C3 of glycerol is between 1.6 and 40.9 per mil.⁶ The determination of the intramolecular isotopic profile with a high degree of accuracy is therefore challenging and is a lengthy experiment due to the low natural abundance of ^{13}C , its small gyromagnetic ratio and long longitudinal relaxation.

A substantial reduction in the acquisition time of ^{13}C NMR can be achieved by sensitivity enhancement using polarization transfer from ^1H to ^{13}C nuclei.^{7,8} Besides, the repetition time of the experiment becomes $^1\text{H-T}_1$ dependent, thus allowing further time reduction. DEPT (Distortionless Enhancement by Polarization Transfer)⁸ and INEPT (Insensitive Nuclei Enhanced by Polarization Transfer)⁷ pulse sequences are widely used to this end. However, the regular versions of these experiments do not satisfy the isotopic analysis requirement in terms of precision. A low repeatability, approximately 5%, is obtained due to off-resonance effects and radiofrequency inhomogeneity in the 180° pulses applied to ^1H and ^{13}C nuclei.⁹ A dramatic improvement in precision (repeatability of per mil) was reached *via* the incorporation of adiabatic 180° ^1H and ^{13}C pulses into the sequences.^{7,9} Nevertheless, it was proved in a previous study that the long time stability of DEPT is not efficient for isotopic measurements at natural abundance, while the use of the INEPT sequence for ^{13}C isotopic measurements affords a substantial reduction in the experiment time without deterioration of short and long time stability.¹⁰ Such a high precision offers extended potentialities in the position-specific isotopic composition analysis at natural abundance, provided that the exact stoichiometry of the different carbon sites to be compared is available. An isotopic profile can therefore be obtained from the peaks for which the concentrations are precisely known and have a signal-to-noise ratio of ≈ 500 , while the other peaks of the ^{13}C NMR spectrum can be used for metabolomic profiling of the samples.

^aEESI Team, Interdisciplinary Chemistry: Synthesis, Analysis, Modelling (CEISAM), University of Nantes-CNRS UMR 6230, 2 rue de la Houssinière, BP 92208, F-44322 Nantes cedex 3, France

^bResearch Unit: Technologies et Valorisation Alimentaire (TVA), Laboratory of Metrology and Isotopic Fractionation, Faculty of Science, Saint-Joseph University, P.O. Box 11-514 Riad el Solh, Beirut 1107 2050, Lebanon. E-mail: joseph.bejjani@usj.edu.lb

† Electronic supplementary information (ESI) available: Details of ^{13}C -INEPT spectra acquisition and processing, pulse sequence scheme for adiabatic INEPT, designation of carbon atoms in triacylglycerol molecules, and enlargements of different regions in the ^1H and ^{13}C NMR spectra of olive oil for a resolving power comparison. See DOI: 10.1039/c5ay01250c

In this paper, we report the optimization of the adiabatic INEPT sequence for the metabolomic and isotopic characterization of triacylglycerols, which are quasiuniversal components of vegetable and animal matrices. Thus, metabolomic and isotopic profiling of these molecules are very powerful tools for food authentication issues and for metabolic studies at natural isotopic abundance. The composition in fatty acids of an oil sample cannot be known with a precision as high as a few per mil; therefore, most of the peaks observed in the ^{13}C NMR spectrum can only be used as compositional profile elements and not for the measurement of the intramolecular isotopic distribution. However, the NMR signals related to the carbons of the glycerol moiety (C1,3 and C2) and the first part of the fatty acids (C2 and C3) are always present and are common to all triacylglycerol molecules which permits their use for isotopic purposes. The other signals will be used for the profiling.

To illustrate the high potential of the adiabatic INEPT experiment in the origin classification of triacylglycerols, olive oil samples (bought from ethical and fair-trade shops) from different geographic origins (Crete, Lebanon, Spain, and Tunisia) were analyzed. Here we have to mention that the global ^{13}C content measured by irm-MS does not allow any classification of the samples.

Description of experimental parameters used for sample preparation, NMR acquisition and data processing is given in the ESI.† However, we want to describe here the way by which the INEPT sequence was optimized for the characterization of triacylglycerols.

In the refocused INEPT sequence (see Fig. S1 in the ESI†), initial longitudinal ^1H magnetization is converted by the 90° ^1H pulse into in-phase ^1H coherence in the transverse plane. The first spin-echo aims to convert this coherence into antiphase ^1H coherence. $\tau_1 = 1/4J_{\text{CHav}}$ ensures efficient coherence transfer (where J_{CHav} is the median value of the ^1H - ^{13}C coupling constant). Simultaneous ^1H and ^{13}C 90° pulses transform the antiphase ^1H coherence into antiphase ^{13}C coherence. The 180° ^{13}C pulse during this second echo serves to refocus ^{13}C chemical shift evolution during τ_2 delays. τ_2 needs to be adjusted to an optimal value that depends on the nature of the resonance observed ($1/4J$ for CH, $1/8J$ for CH_2 and $1/10J$ for CH_3 spin systems, respectively). The optimized τ_2 (see ESI†, NMR spectrometry experiments) ensures efficient refocusing for CH, CH_2 and CH_3 and the glycerol signals (C1,3 and C2) are then in the same ratio as in the one pulse quantitative ^{13}C NMR experiment for a ^{13}C isotopic analysis. Biases remaining on the other signals are highly reproducible. For signals used in isotopic analysis, RSD values ≤ 1.2 per mil were observed. For other signals, only used for profiling, RSD values $\leq 3.1\%$ were observed.

The recovery delay in the one pulse sequence must be greater than 10 times the longest ^{13}C T_1 ;†† this condition is required to avoid the influence of the nuclear Overhauser effect (NOE). Here, it induces $\text{TR} = 42$ s (since the longest ^{13}C T_1 of the measured signals was 4.0 s). Furthermore, 72 scans are needed with this sequence to obtain a signal-to-noise ratio higher than 500 on the C2 of glycerol (Fig. 1), which is a prerequisite for a precision around 1 per mil. The measurement time was therefore 5 hours 45 min for six spectra. Since, in the INEPT

sequence, the proton magnetization is transferred to the bounded carbons, the intensity of the detected signal is only dependent on the ^1H longitudinal magnetization before the first 90° pulse and no longer depends on the carbon longitudinal relaxation. This avoids any influence of the NOE on the detected signal. The recovery delay is therefore only governed by the ^1H proton relaxation and must be greater than 7 times the longest ^1H T_1 in order to obtain an accuracy of 1 per mil. Here, the longest ^1H T_1 was measured for the methyl group of the linolenic acid (2.9 s). We have chosen to use a recovery delay equal to 8 times this T_1 , in order to overcome any variation in relaxation times from one sample to another; this condition induces $\text{TR} = 24$ s. The occurrence of the polarization transfer increases the sensitivity, thus only 16 scans were required to obtain a signal-to-noise ratio higher than 600 on the C2 of glycerol (Fig. 1). The measurement time was therefore 50 min 10 s for six spectra. The experiment time is therefore reduced by a factor of 7 with respect to the quantitative one pulse ^{13}C NMR.

For each olive oil sample, peak areas of the following signals were determined (see Fig. S2† in the ESI for signal designation): C2, C3, C ω 1, C ω 2 and C ω 3 of fatty acids; C2 and C1,3 of glycerol; the carbons 9 and 10 (oleic acid), 9, 10, and 12 (linoleic acid) at sn-1,3 and sn-2 positions of glycerol; allylic and di-allylic carbons, 24 signals related to the aliphatic methylene groups, other than those aforementioned, carbons 10, 12, 13, and 15 (linolenic acid) and carbons related to squalene were determined as well.

From the physical description of polarization transfer, it follows that the sensitivity gain in INEPT is only obtained for hydrogen-bearing carbons. In other words, only CH, CH_2 and CH_3 groups will be observed, while quaternary carbons will be missing. Carbonyl carbon signals are therefore lost, while they allow the determination of saturated, oleic, linoleic, and linolenic acids at sn-1,3 and sn-2 positions of the glycerol backbone. However, this information can still be obtained from other regions of the spectra such as from the olefinic region ranging from 132 to 127 ppm and from the region related to the C2 of fatty acids ranging from 34.2 to 33.7 ppm.¹²

As an example of the pertinence of these variables, we show in Fig. 2 that the plot of the peak area ratio of carbon 9 of linoleic acid at the sn-1,3 position of glycerol to that at the sn-2

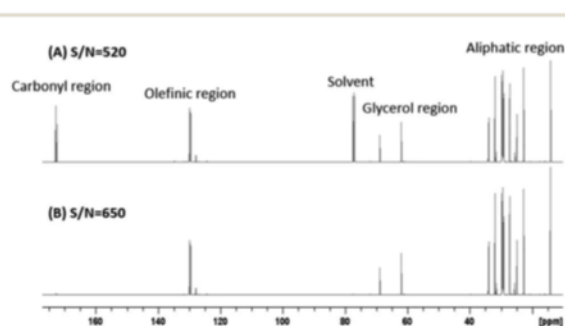


Fig. 1 ^{13}C NMR spectrum of an olive oil sample obtained by (A) one pulse ^{13}C acquisition in 54 min 30 s, and (B) adiabatic refocused INEPT in 8 min 21 s.

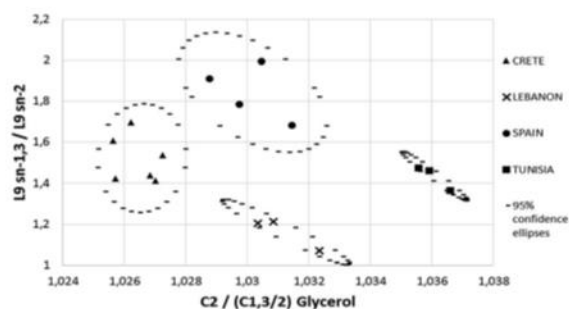


Fig. 2 Scatter plot of olive oil samples, from Crete, Lebanon, Spain, and Tunisia, based on the ratio $C2/(C1,3/2)$ of the glycerol backbone and the ratio of C9 linoleic acid-glycerol sn-1,3/C9 linoleic acid-glycerol sn-2.

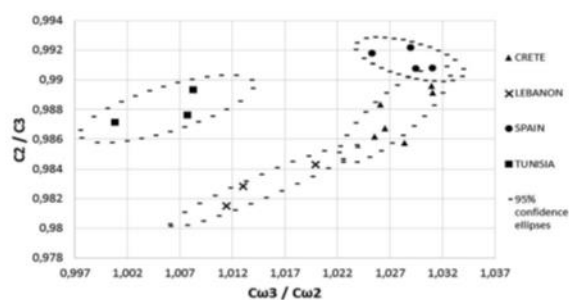


Fig. 3 Scatter plot of olive oil samples from 4 different countries (Crete, Lebanon, Spain, and Tunisia) based on the $C2/C3$ and $C\omega3/C\omega2$ of fatty acids.

position against the isotopic ratio of the glycerol carbons [$C2/(C1,3/2)$] permits an excellent separation between the different olive oil groups. It should be noted that the variables related to the position of fatty acids at sn-1,3 and sn-2 positions of glycerol could not be obtained by means of 1H NMR spectroscopy due to its lower resolution owing to its narrower range of resonance frequencies relative to ^{13}C NMR (see Fig. S3† in the ESI for a comparison).

As a result, the olive oils could be classified according to their geographical origin using solely the information related to the position of linoleic acid on the glycerol skeleton and the isotopic content of carbons in this moiety.

In the second example (Fig. 3), we show that it is also possible to distinguish olive samples from different countries using the ratios $C2/C3$ and $C\omega3/C\omega2$ of fatty acids.

In conclusion, adiabatic INEPT is a rapid and very precise method that allows the relative site specific ^{13}C content of triacylglycerols to be determined at the same time as the

metabolomic profiling with the information related to the position of fatty acids on the glycerol moiety being available.

Nevertheless, numerous elements are extracted from the corresponding ^{13}C INEPT spectra but only few of them were exploited in this paper. A global analysis of all the elements using a larger number of samples and statistical tools is currently underway in our laboratory, aiming to draw out their full characterization potential.

This method is a valuable tool for origin tracing and metabolic studies using triacylglycerols, extracted from vegetables, animal tissues or finished products, as probe molecules. Its characterization power is brought to a higher level since faint deviations between carbon signals in the spectra become rapidly quantifiable with high reproducibility and robustness, allowing the isotopic information inherited by the triacylglycerol molecules to be revealed.

Acknowledgements

N.M. acknowledges the financial support from the Lebanese National Council for Scientific Research and the Research Council of Saint-Joseph University. The CORSAIRE platform from Biogenouest is also acknowledged.

Notes and references

- 1 L. Mannina, A. P. Sobolev and S. Viel, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2012, **66**, 1–39.
- 2 E. Perri, C. Benincasa and I. Muzzalupo, in *Olive Germplasm - the Olive Cultivation, Table Olive and Olive Oil Industry in Italy*, ed. I. Muzzalupo, InTech, 2012, ch. 13, pp. 265–286.
- 3 J. M. McKenzie and K. R. Koch, *S. Afr. J. Sci.*, 2004, **100**, 349–354.
- 4 S. D. Kelly, in *Food Authenticity and Traceability*, ed. M. Lees, Woodhead Publishing, Cambridge, 2003, ch. 7, pp. 156–183.
- 5 Z. Muccio and G. P. Jackson, *Analyst*, 2009, **134**, 213–222.
- 6 E. Caytan, Y. Cherghaoui, C. Barril, C. Jouitteau, C. Rabiller and G. Remaud, *Tetrahedron: Asymmetry*, 2006, **17**, 1622–1624.
- 7 G. A. Morris and R. Freeman, *J. Am. Chem. Soc.*, 1979, **101**, 760–762.
- 8 D. M. Doddrell, D. T. Pegg and M. R. Bendall, *J. Magn. Reson.*, 1982, **48**, 323–327.
- 9 C. Thibaudeau, G. Remaud, V. Silvestre and S. Akoka, *Anal. Chem.*, 2010, **82**, 5582–5590.
- 10 U. Bussy, C. Thibaudeau, F. Thomas, J.-R. Desmurs, E. Jamin, G. S. Remaud, V. Silvestre and S. Akoka, *Talanta*, 2011, **85**, 1909–1914.
- 11 D. Canet, *J. Magn. Reson.*, 1976, **23**, 361–364.
- 12 F. J. Hidalgo and R. Zamora, *Trends Food Sci. Technol.*, 2003, **14**, 499–506.

III-3. Données complémentaires

III-3.1. NMR spectrometry experiments

For quantitative ^{13}C NMR, oils (403.2 mg) were dissolved in chloroform-d (630.0 mg) and the resulting solutions transferred into 5 mm NMR tubes. For each sample, 6 ^{13}C INEPT spectra were recorded using 11.7 T Bruker Avance-III spectrometer equipped with a 5 mm o.d. dual cryoprobe $^{13}\text{C}/^1\text{H}$ tuned at the recording frequency of 125.76 MHz for ^{13}C . The temperature of the probe was set at 293 K. The acquisition parameters for ^{13}C NMR spectra were as follows: ^{13}C and ^1H offsets were set at the middle of the frequency range (92.5 ppm for the ^{13}C and 3 ppm for the ^1H), pulse width 10 μs for the 90° ^1H and 11 μs for the 90° ^{13}C , 16 scans with a repetition delay of 24 s were recorded in order to have a signal-to-noise ratio higher than 600 on the C2 of glycerol. τ_1 was adjusted to 2.704 ms, and the refocusing period τ_2 was adjusted to 1.409 ms. Adiabatic full passage pulses were generated using Mathcad 8 (MathSoft, Inc.). They were designed with a cosine amplitude modulation of the RF field ($\omega_I^{\text{max}} = 157.1$ kHz or 93.89 kHz for ^{13}C or ^1H , respectively) and an offset independent adiabaticity (OIA) by optimizing the frequency sweep ΔF ($\Delta F = 39$ kHz or 17 kHz for ^{13}C or ^1H , respectively). For inversion pulses, adiabatic full passage pulses were used. For refocusing pulses, composite adiabatic pulses were used. ^1H decoupling was performed using adiabatic full passage RF pulses with cosine square amplitude modulation ($\nu_2^{\text{max}} = 17.6$ kHz) and offset independent adiabaticity with optimized frequency sweep ($\Delta F = 14$ kHz).

III-3.1. NMR data processing and analysis

FIDs were zero-filled to 128 K and submitted to an exponential multiplication inducing a line broadening of 1.5 Hz before Fourier transform. The ^{13}C NMR spectra were manually phased. An automatic polynomial baseline correction ($n = 5$) was applied to the resulting spectra. The curve fitting was carried out in accordance with a Lorentzian mathematical model using PERCH Software (PERCH NMR Software, University of Kuopio, Finland) and 97 peak areas were obtained for each sample.

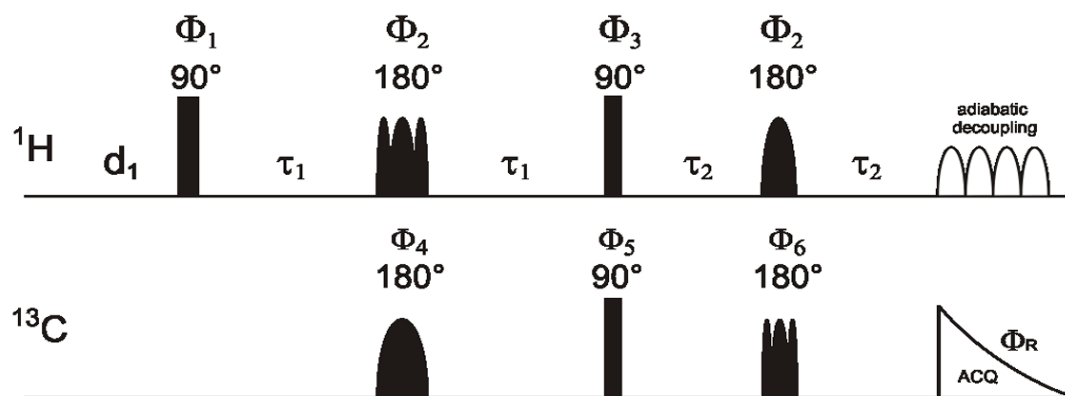
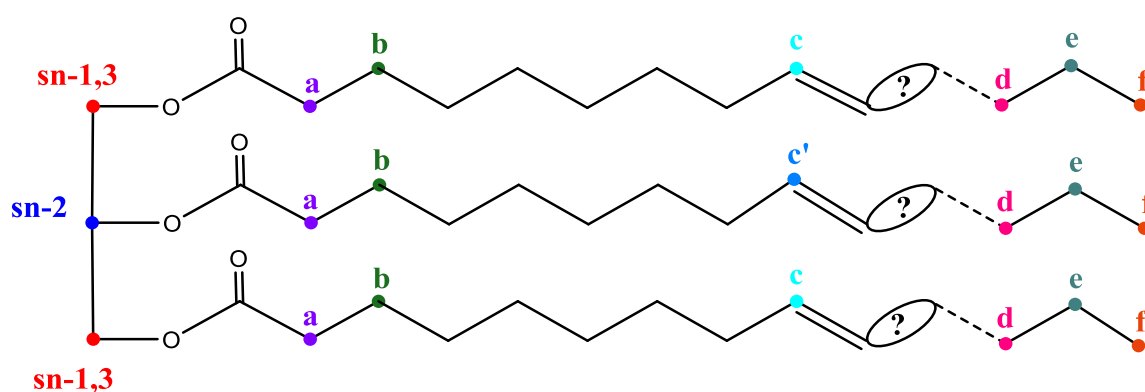


Figure S1: Adiabatic refocused INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) sequence with ^1H and ^{13}C 180° adiabatic composite refocusing pulses and adiabatic full passage inversion pulses.



sn-1,3: C1,3 of glycerol backbone

sn-2: C2 of glycerol backbone

a: C2 of fatty acids

b: C3 of fatty acids

c: C9 of linoleic acid at glycerol sn-1,3

c': C9 of linoleic acid at glycerol sn-2

d: C ω 3 of fatty acids

e: C ω 2 of fatty acids

f: C ω 1 of fatty acids

Figure S2: Designation of different carbons in a triacylglycerol molecule.

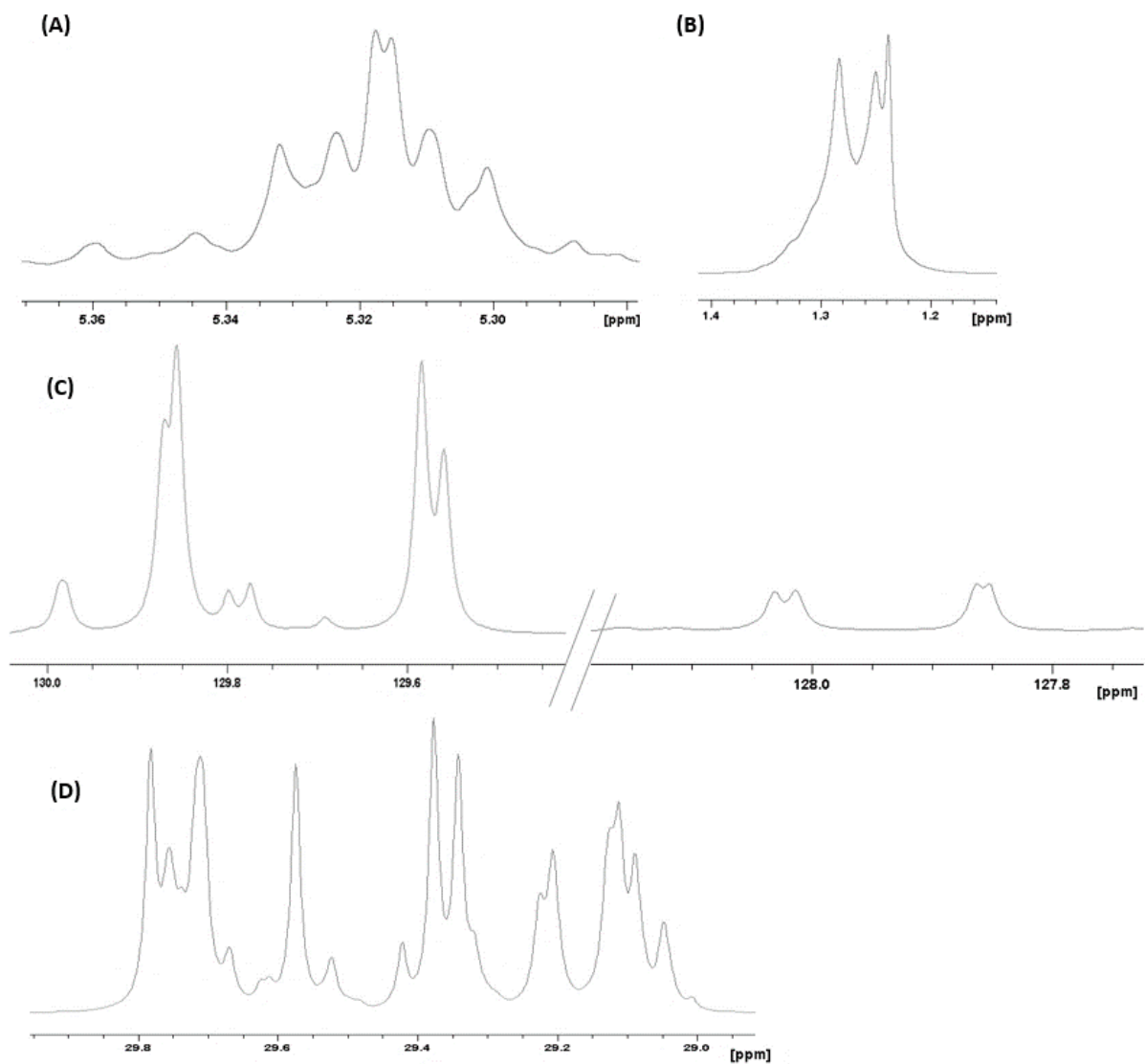


Figure S3: Comparison of different regions of the ^1H and ^{13}C NMR spectra of olive oil: olefinic (A) and aliphatic (B) regions of the ^1H NMR spectrum; olefinic (C) and aliphatic (D) regions of the ^{13}C NMR spectrum.

III-4. Conclusion

La durée de l'analyse isotopique des matrices riches en TAG par une séquence mono-impulsionnelle ^{13}C , étant relativement longue (54 min 30 s par spectre), une séquence INEPT optimisée dans une étude antérieure a été appliquée permettant d'augmenter la sensibilité et de réduire le temps d'analyse (8 min 21 s par spectre) ; et cela pour un même rapport signal sur bruit (S/B). Les délais des deux échos de spins ont été optimisés de telle manière à obtenir les mêmes compositions isotopiques ^{13}C ($\delta^{13}\text{C}$ en ‰) que celles obtenues par la séquence mono-impulsionnelle conventionnelle sur les 2 sites du glycérol des TAG de l'huile d'olive.

Les aires des autres signaux ne sont pas proportionnelles au nombre de noyaux ^{13}C . De ce fait, un facteur correctif moyen établi sur les résultats des échantillons analysés avec la séquence mono-impulsionnelle a été appliqué aux aires obtenues avec la séquence INEPT-refocalisée.

En conséquence, la séquence INEPT adiabatique permet d'atteindre la précision nécessaire (de l'ordre de quelques pour mille) pour analyser la teneur ^{13}C dans un temps d'expérience court.

En utilisant des informations obtenues de l'analyse INEPT- ^{13}C : (C2/C3 et C ω 3/C ω 2, ou bien C9 de l'acide linoléique en position sn-1,3 du glycérol/ sn-2 et le rapport C2/(C1,3/2) du glycérol) une bonne classification des huiles d'olive en fonction de leur origine géographique est observée. Il faut signaler que les informations positionnelles ne peuvent pas être obtenues à l'aide de la RMN ^1H en raison de sa faible résolution et sa gamme de déplacement chimique plus étroite que celle de la RMN ^{13}C .

La séquence INEPT adiabatique peut donc être considérée comme un outil très intéressant pour l'analyse isotopomique des TAG, molécules sondes des matrices animales et végétales.

*Partie IV - Développement d'une
nouvelle stratégie pour la
quantification rapide des acides gras
individuels dans les matrices
triacylglycérols par RMN ¹³C en
utilisant une technique de transfert de
polarisation*

IV-1. Introduction

Comme nous l'avons déjà évoqué dans la partie III, la méthode RMN ^{13}C offre des informations supplémentaires par rapport à la RMN ^1H , grâce à une meilleure discrimination des résonances. Nous avons également vu que la séquence INEPT adiabatique a été utilisée avec succès pour la caractérisation des triacylglycérols ; cette technique fournit des informations relatives à la distribution des acides gras sur la molécule du glycérol, elle permet d'étudier à la fois les profils métabolomique et isotopique des triacylglycérols.

Cependant, jusqu'à présent, la RMN ne permettait pas d'obtenir le profil complet des acides gras des TAG. C'est, la chromatographie gazeuse (GC) qui est considérée comme la méthode de référence pour la quantification des acides gras individuels.

Dans le cadre de cette étude, nous présentons une nouvelle stratégie basée sur l'utilisation de la séquence INEPT adiabatique et les techniques de régression linéaire multivariée dans le but de quantifier les acides gras individuels. Nous avons testé cette stratégie sur les huiles d'olive. À cette fin, 23 échantillons d'huile d'olive ont été analysés par INEPT- ^{13}C et GC ; parmi ces échantillons, 18 ont servi pour construire des modèles de prédiction et les 5 autres pour tester leur capacité prédictive. Il faut signaler que les échantillons ont été choisis de manière à couvrir toute la gamme des pourcentages des acides gras présents dans l'huile d'olive.

Les signaux résultant de la déconvolution des spectres INEPT- ^{13}C et les données obtenues par l'analyse chromatographique ont été respectivement utilisés comme variables indépendantes et dépendantes dans la régression des moindres carrés partiels (PLS) ou la régression linéaire multiple (MLR).

Ainsi, les meilleurs modèles permettant de prédire les pourcentages d'acides gras individuels et du squalène présents dans l'huile d'olive ont été sélectionnés, et ceci afin d'offrir en une seule expérience l'ensemble des informations isotopiques et métabolomiques, y compris les positions des acides gras sur le squelette du glycérol.

Les résultats obtenus ont fait l'objet d'un article intitulé « A novel quantification strategy of individual fatty acids in triacylglycerol matrices using optimized ^{13}C -NMR with a polarization transfer technique », soumis à la revue *Analytical and Bioanalytical Chemistry*.

IV-2. Article

A novel quantification strategy of individual fatty acids in triacylglycerol matrices using optimized ^{13}C -NMR with a polarization transfer technique

Noelle Merchak ^{a,b}, Virginie Silvestre ^b, Denis Loquet ^b, Toufic Rizk ^a, Serge Akoka ^b, and Joseph Bejjani ^{*a}

^a Research Unit: Technologies et Valorisation Alimentaire (TVA), Laboratory of Metrology and Isotopic Fractionation, Faculty of Science, Saint-Joseph University, P.O. Box 11-514 Riad el Solh, Beirut 1107 2050.

^b EBSI team, Interdisciplinary Chemistry: Synthesis, Analysis, Modelling (CEISAM), University of Nantes-CNRS UMR 6230, 2 rue de la Houssinière, BP 92208, F-44322 Nantes cedex 3, France

Abstract

Triacylglycerols, which are quasi-universal components of food matrices, consist of complex mixtures of molecules. Their fatty acids profile, position on the glycerol moiety, and position-specific ^{13}C content may significantly vary in function of the geographical, botanical or animal origin of the sample. Such variables could be valuable tracers for food authentication issues. The main objective of this work was to develop a new method based on a rapid and precise ^{13}C -NMR spectroscopy (using a polarization transfer technique) coupled with multivariate linear regression analyses in order to quantify the whole set of individual fatty acids within triacylglycerols. In this respect, olive oil samples were analyzed by means of an adiabatic ^{13}C -INEPT sequence and gas chromatography (GC). For each fatty acid within the studied matrix and for squalene, a multivariate prediction model was constructed using the deconvoluted peak areas of ^{13}C -INEPT spectra as predictors, and the data obtained by GC as response variables. This ^{13}C -NMR-based strategy, tested on olive oil, could serve as a rapid alternative to the gas chromatographic quantification of individual fatty acids in other matrices, while providing additional compositional and isotopic information.

Keywords

INEPT; Multivariate linear regression; Triacylglycerol; Olive oil; Fatty acid; Squalene

Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy, a powerful and rapid tool widely used in metabolomics,¹ advantageously provides qualitative and quantitative information about the composition of a multicomponent system in a single experiment and with minimal sample preparation. Thus, NMR technique was successfully employed for the characterization of triacylglycerols, which are quasi universal components of vegetable and animal food matrices.^{2,3} In this respect, ¹H-NMR permits to determine the amount of total saturated, unsaturated fatty acids, linoleic and linolenic acids by means of mathematical equations using appropriate peak areas. However, ¹H-NMR spectra of triacylglycerols mixtures are complicated by severe signal overlapping due to the small chemical shift range of protons (approximately 15 ppm).⁴ On the other hand, ¹³C-NMR affords useful and more valuable information than ¹H-NMR due to a better discrimination of resonances (spread over approximately 250 ppm); in addition to the metabolomic profile of triacylglycerols, it provides data related to the positional distribution of fatty acids on the glycerol backbone⁵⁻⁷ and to the position-specific ¹³C content.⁸ Furthermore, Scanoa et al. have quantified the total of 11-*cis* vaccenic and gondoic fatty acids in olive oil samples by means of corresponding overlapped ¹³C-NMR peaks;⁹ the individual determination of oleic, vaccenic, and gondoic acids present in oleifera seed oil was realized by Vlahov et al. using ¹³C-NMR; nevertheless, only vaccenic acid was quantified in one olive oil sample.¹⁰ Up to now, the quantification of the whole set of fatty acids in triacylglycerols requires the use of chromatographic techniques; gas chromatography (GC) is the reference method employed for this purpose.^{11,12} In this work, the complete profiling of triacylglycerol mixtures by ¹³C-NMR was considered. However, the conventional ¹³C-NMR is a lengthy experiment because of the low natural abundance of ¹³C isotope, its small gyromagnetic ratio, and its long longitudinal relaxation time. For this reason, an adiabatic INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) pulse sequence,^{13,14} that we have previously optimized for the isotopic analysis of triacylglycerol matrices,⁸ was used as an alternative allowing a considerable reduction in acquisition time. Besides, multivariate linear regression techniques were used aiming to quantify the complete set of individual fatty acids based on the available ¹³C-NMR variables; such a strategy, making use of spectroscopic data in combination with multivariate regression, was used either to predict concentrations of given compounds in a complex matrix or to determine physicochemical properties of a whole product.¹⁵⁻
¹⁷ In this respect, olive oil samples were analyzed by ¹³C-INEPT and GC. ¹³C-NMR

spectra were deconvoluted and the resulted peak areas were used as predictors in multiple linear regression (MLR) and partial least squares regression (PLSR) analyses along with GC data as response variables. Using this approach, we have developed models that permit to determine the percentages of individual fatty acids and squalene in olive oils while providing, in the same experiment, isotopic and positional information related to triacylglycerols.

Experimental

Samples and chemicals

Twenty-three olive samples (300 g each) from different varieties were collected from the olive-growing regions across Lebanon at different altitudes and latitudes. Authentic oils were obtained solely by mechanical methods in the laboratory. Olives were crushed, pitted and then ground to obtain a paste. The paste was centrifuged and the resulting oil layer was decanted. Oils were centrifuged to eliminate residual particles and then stored at -18 °C. Standard fatty acid methyl esters were purchased from sigma Aldrich and the deuterated chloroform solvent was purchased from Eurisotop (www.eurisotop.fr).

Gas chromatographic analysis

Fatty acids composition was determined by gas chromatography. Fatty acids methyl esters were prepared by shaking 200 mg of oil in 3mL of hexane with 0.4 mL of 2N methanolic potassium hydroxide^{11,12}. An Agilent Technologies chromatograph equipped with a flame ionization detector and a fused-silica capillary column (SGE-054616, 30m × 0.32mm i.d., BPX70 0.25µm) was used in this study with helium as carrier gas at flow of 1.0 mL/min. Oven temperature was programmed as follows: at 160 °C for 15 min, to 200 °C at a rate of 10 °C/min, at 200 °C for 10 min. A 1:10 dilution in hexane was performed. A volume of 1µL was injected with a split ratio of 50:1 and an injector temperature of 200 °C. The detector temperature was set at 270 °C. Two injections were performed for each sample. Retention times were determined by using standard fatty acid methyl esters. Molar percentages of palmitic, palmitoleic, margaric, margaroleic, hypogeic, stearic, oleic, vaccenic, linoleic, linolenic, arachidic, gondoic, behenic, and lignoceric acids (structures of fatty acids are given in **Table S1** in the ESI†) in olive oils were determined relatively to their total number of moles by

dividing each peak area by the molar weight of the corresponding fatty acid methyl ester. Squalene was also quantified in the same way.

¹³C-INEPT spectrometry experiments

Oil (403.2 mg) was dissolved in CDCl₃ (630.0 mg). A 500 MHz Bruker Avance-III spectrometer equipped with a 5 mm dual cryo-probe ¹³C/¹H carefully tuned at the recording frequency of 125.76 MHz for ¹³C was used. The temperature of the probe was set at 293 K. ¹³C-INEPT spectra were obtained by using an adiabatic INEPT pulse sequence (**Figure S1** in the ESI†).^{8,14} At the beginning of each measurement session, the pulse lengths were calibrated and the probe tuning and matching were adjusted. The acquisition parameters were as follows: ¹³C and ¹H offsets were set at the middle of the frequency range (92.5 ppm for the ¹³C and 3 ppm for the ¹H), pulse width 10 μs for the 90° ¹H and 11 μs for the 90° ¹³C, acquisition time of 1 s, repetition time of 24 s (the longest ¹H-T₁ (2.9 s) was measured, using the inversion-recovery pulse sequence, for the methyl group of the linolenic acid) and 16 scans were accumulated leading to a signal-to-noise ratio higher than 600 on the C2 of glycerol. Adiabatic full passage pulses were generated using Mathcad (MathSoft, Inc.). They were designed with a cosine amplitude modulation of the RF field ($\omega_1^{\max} = 157.1$ kHz or 93.89 kHz for ¹³C or ¹H, respectively) and an offset-independent adiabaticity with optimized frequency sweep ΔF ($\Delta F = 39$ kHz or 17 kHz for ¹³C or ¹H, respectively). For inversion pulses, adiabatic full passage pulses were used. For refocusing pulses, composite adiabatic pulses were used.¹³ ¹H decoupling was performed using adiabatic full passage RF pulses with cosine square amplitude modulation ($\omega_2^{\max} = 110.6$ kHz) and offset-independent adiabaticity with optimized frequency sweep ($\Delta F = 14$ kHz).¹⁸ For each sample, six spectra were recorded in 50 min 10 s.⁸

¹³C-INEPT data processing and analysis

The free induction decays (FIDs) were zero-filled to 128 K and exponentially apodized with $LB = 1.5$ Hz. The ¹³C-INEPT spectra were manually phased and an automatic polynomial baseline correction ($n = 5$) was applied to the resulting spectra. The curve fitting was carried out in accordance with a Lorentzian mathematical model using PERCH Software (PERCH NMR Software, University of Kuopio, Finland). Ninety-seven peak areas were obtained for each sample.

Statistical analysis

Statistical analyses were achieved with TANAGRA software.¹⁹ Data obtained from ¹³C-INEPT (deconvoluted peak areas) and GC (molar percentages of fatty acids and squalene) analyses of olive oil samples were used as independent and dependent variables, respectively, in the multivariate regression analyses. Both MLR and PLSR procedures were applied in order to find suitable models for fatty acids and squalene quantification. The MLR models were developed by reducing the number of variables based on their relevance (*p*-value) by aid of the backward elimination technique; the robustness of models were assessed by means of the leave-one-out cross-validation test with Q^2 used as indicator. Q^2 is given by:

$$Q^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{(-i)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

With *n*: number of samples in the training set; y_i : a fatty acid amount determined by GC; $\hat{y}_{(-i)}$: prediction of y_i with a model constructed after omitting the corresponding sample from the training set and using the initial descriptors; \bar{y} represents the mean of y_i values for all the training set samples.

The PLSR models were constructed by eliminating non-relevant variables according to their standardized regression coefficient while adjusting the number of components based on the cross-validation test (by randomly leaving out 10% of the training samples); maximum cumulative Q^2 (Q^2_{cum}) parameter was used as robustness indicator.^{20,21} Q^2_{cum} is given by:

$$Q^2_{\text{cum}} = 1 - \prod_{j=1}^h \frac{PRESS_j}{RSS_{j-1}} \quad (2)$$

where *h* is the number of components in a PLSR model; $PRESS = \sum_{i=1}^n (y_i - \hat{y}_{(-i)})^2$; $RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$; *n*, y_i , and $\hat{y}_{(-i)}$ have the same definition as in equation 1, \hat{y}_i is the prediction of y_i while the corresponding sample is included in the training set when the model is constructed.

Adjusted- R^2 was used to compare between models with different number of predictors. This parameter is given by:

$$\text{Adjusted-}R^2 = 1 - \frac{(n-1)}{(n-m-1)} (1-R^2) \quad (3)$$

with *n*: sample size, *m*: number of predictors taking into account the constant term, and R^2 : coefficient of multiple determination.

Furthermore, since Q^2 or Q^2_{cum} are not sufficient to evaluate the predictive power of models,²² external validation samples were used for this end. The closer the predictive

R^2 (Pred- R^2) is to 1, the higher is the predictive ability of the model.²³ Pred- R^2 is given by:

$$\text{Pred-}R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

In this case, y_i and \hat{y}_i correspond to the validation samples; \hat{y}_i values are predicted by the initial regression model; \bar{y} has the same definition as in equation 1 (it is only for the training set samples).

However, Pred- R^2 does not guarantee a true expression of external predictability since it is dependent on the mean value of the training set samples and hence affected by the choice of test samples. In this respect, Roy et al. recommend the use of Modified R^2 (Modif- R^2) to better understand the predictability of models.²³ This parameter is given by:

$$\text{Modif-}R^2 = r^2 \left(1 - \sqrt{r^2 - r_0^2} \right) \quad (5)$$

where r^2 and r_0^2 are the squared correlation coefficients between the observed and the predicted values of the external validation samples with and without intercept, respectively. A Modif- R^2 greater than 0.5 is an indicator of good predictability.²³

Results and discussion

To the best of our knowledge, there are no previous studies reporting the quantification of the whole set of individual fatty acids in triacylglycerols matrices by means of NMR spectroscopy.

In this work, a strategy based on ^{13}C -INEPT and multivariate linear regression (MLR and PLSR) analyses was developed, compared to GC, and tested on the olive oil matrix in order to fill this gap.

Thus, olive oil samples were analysed by ^{13}C -INEPT and GC. INEPT spectra were deconvoluted and detected resonances were assigned (**Figure 1**). Signals can be divided into three sets: unsaturated carbon from 127 to 132 ppm, carbons of glycerol backbone from 62 to 69 ppm, and carbons of methylene and methyl groups from 10 to 35 ppm. The deconvoluted peak areas (see labeling in **Figure 1**) were normalized relatively to the mean value of alpha (C2) and beta (C3) carbons signals and used as inputs in the multivariate analyses, while molar percentages of individual fatty acids obtained by GC were used as targets. For each fatty acid and for squalene, samples were split into training ($n = 18$) and validation ($n = 5$) sets in order to construct a prediction model and to check its reliability; It must be noted that fatty acids ranges of

samples used in this study are comparable to those reported by international sources (Table S1 in the ESI†).

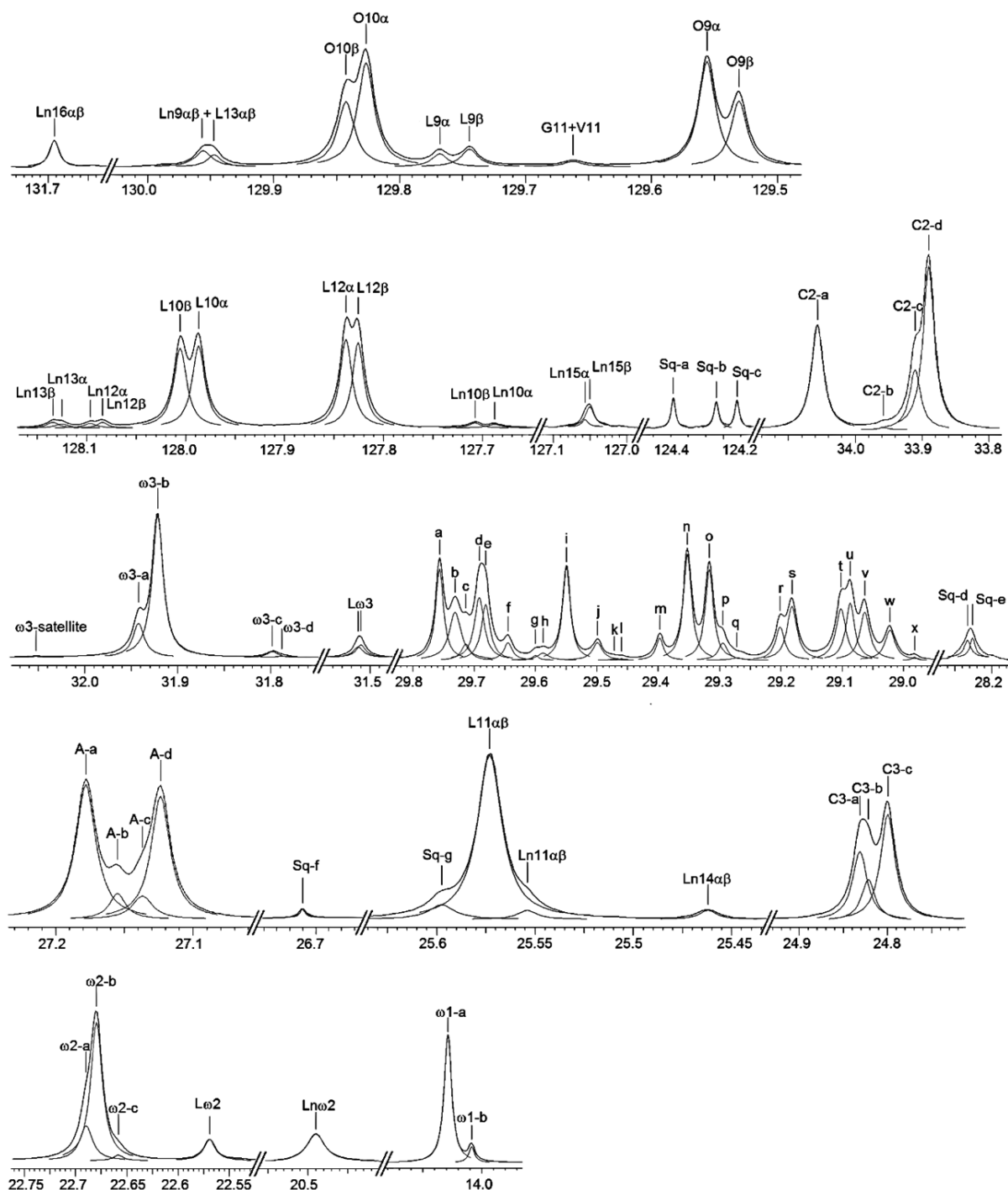


Figure 1: Deconvoluted lines of olive oil ^{13}C -INEPT spectra and signals labelling used in the multiple linear regression. Signals labelling: linolenic (Ln); linoleic (L); oleic (O); gondoic (G); vaccenic (V); squalene (Sq); glycerol sn-1,3 position (α); glycerol sn-2 position (β); signals of the aliphatic carbons region ranging from 29.82 to 28.85 ppm are labeled from a to x; carbons 2 and 3 of fatty acids chains (C2-a to d) and (C3-a to c), respectively; allylic carbons (A-a to d), starting from the methyl group of the fatty acids chains, carbons 1, 2, and 3 are denoted (ω 1-a and b), (ω 2-a to c), and (ω 3-a to d), respectively.

The best fatty acids quantification models obtained are summarized in **Table 1** with p -values of predictors and intercepts to show their relevance in the equations. They were constructed using the variables reduction procedures described in the experimental section. The following parameters were used in the assessment of models: R^2 and adjusted- R^2 (eq. 3) for the training set to describe how well regression models fit GC data; $F(df1,df2)$ - and p -values (F-test) for the overall significance of models; Q^2 (eq. 1) or Q^2_{cum} (eq. 2) for models' robustness; Pred- R^2 (eq. 4) and Modif- R^2 (eq. 5) for the external predictability evaluation. These parameters are reported in **Figure 2** where correlations between GC and predicted data are presented for the training and the validation sets of samples separately.

In the cases of squalene, oleic, linoleic, and linolenic acids, only one or two variables specific to these compounds have contributed to the best model; for this reason, the validation issue was not a limitation and thus, the whole set of samples (23) was used to construct the final model.

Table 1: Models for individual fatty acids quantification olive oil samples.

Predictor ^a	Intercept and regression coefficients corresponding to the different fatty acids regression models ^b (<i>p-value</i>) ^c									
	Palmitic	Margaric	Stearic	Arachidic	Behenic	Palmitoleic	Oleic ^d	Vaccenic	Linoleic ^d	Linolenic ^d
Intercept	8.6618 (0.0043)	9.9846 (0)	-32.4824 (0.00033)	6.7513 (0.00059)	-0.5560 (0.0044)	0.4392 (0.00047)	3.6492 (0.0032)	-2.6003 (0.00018)	0.4924 (0)	0.0447 (0.075)
Ln16 $\alpha\beta$										0.2159 (0.0042)
L9 α									0.3752 (0)	
L12 β									0.3828 (0)	
ω 3-a			-0.5901 (0)							
ω 3-b							0.3227 (0)			
ω 3-c			-0.1636 (0.000089)							
ω 3-d										
B	-0.0830 (0.00026)		0.1134 (0)	0.0090 (0)	0.0028 (0.0040)					
C				0.0140 (0)						
D					-0.0008 (0.012)					
I		-0.0101 (0)						0.0969 (0.00060)		
J		-0.0618 (0)						0.0916 (0.00024)		
K		-0.0916 (0.000033)						0.5472 (0)		
L								0.4832 (0)		
M	0.4690 (0)		0.4863 (0.000011)	-0.0554 (0.00013)	-0.0326 (0.00028)					
N			0.1111 (0.00030)	-0.0242 (0.00037)						
O	-0.0296 (0.0027)			0.0016 (0.0079)	0.0029 (0.0013)			-0.0889 (0.00057)		
P					0.0316 (0.00055)			-0.0723 (0.00092)		
Q					-0.0189 (0.000088)	-0.1009 (0.0022)				
R		-0.0668 (0)								
S		-0.0484 (0)								
T		0.0282 (0)								
U		0.0312 (0)								
X				-0.0501 (0)		0.2653 (0)				
Ln14 $\alpha\beta$										0.1402 (0.036)
ω 2-a			-0.0348 (0.0030)							
ω 2-c	0.1966 (0)									

^a See Figure 1 for variables labelling. ^b Each model is based on 18 olive oil samples; 5 additional samples were used to test the regression equation. ^c *p*-values reflect the relevance of variables in the regression models, *p* = 0 means that it is less than 0.00001. ^d The whole set of samples (23) was considered in the construction of the final model since only one or two variables assigned to the fatty acid in question were found sufficient to obtain the best result in a preliminar evaluation based on 18 samples for training and 5 samples for validation.

For the quantification of squalene, a minor component of olive oil, an excellent regression model based on signals Sq-b, and Sq-f (**Figure 1**) has allowed a better prediction than using each squalene signal apart [$Sq = (-0.0053 \times Sq-b) + (0.0072 \times Sq-f) + 0.0073$; $n = 23$, $R^2 = 0.9283$, adjusted- $R^2 = 0.9211$, $F(2,20) = 129.43$, $p = 0$, and $Q^2 = 0.9076$; p -values corresponding to regression coefficients and intercept are: 0.012, 0.00093, and 0.00067, respectively].

As shown in **Figure 2**, excellent prediction models were obtained for palmitic, stearic, palmitoleic, oleic, vaccenic, linoleic, and linolenic acids, which are present in high or at least noticeable amounts in olive oil. Besides, excellent linear correlations found between several ^{13}C -INEPT variables and these fatty acids, except stearic one, are reported in **Table 2**; they allow assigning several signals to these specific fatty acids and offer alternatives to their quantification models but with slightly lower accuracies. On the other hand, very good prediction models were constructed for Margaric, Arachidic, and Behenic acids, but weaker models were obtained for Lignoceric and gondoic acids present in trace amounts ($< 0.4\%$, see **Table S1** in the ESI†) in olive oil (see **Figure 2** for models assessment).²⁴ Besides, Lignoceric acid model, reached by means of the PLSR procedure while using the predicted values of Palmitic (Pred-P), Oleic (Pred-O), and Arachidic (Pred-A) acids as inputs variables [$Lg = (0.00099 \times \text{Pred-P}) + (-0.000058 \times \text{Pred-O}) + (0.070 \times \text{Pred-A}) + 0.0081$], and Gondoic acid model, obtained by using the PLSR method with ^{13}C -INEPT signals as inputs [$G = (-0.0053 \times \omega 3-c) + (-0.00096 \times f) + (-0.011 \times l) + (-0.00083 \times p) + (-0.0046 \times r) + 0.80$], are due to correlations existing between fatty acids within the olive oil matrix and not to Lignoceric or Gondoic acid signals in the ^{13}C -INEPT spectrum.

Nevertheless, no suitable models were reached for hypogeic and margaroleic acids which, in the analysed samples, range in the intervals 0.086 – 0.18% and 0.067 – 0.32%, respectively.

Furthermore, detailed examination of models (**Table 1**) and correlations (**Table 2**) provides additional insights into the relations existing between or within the dependent and independent variables. Thus, it is likely that signal x (**Figure 1**) is related to both palmitoleic and vaccenic acids (**Table 2**), which are highly correlated compounds, since in the prediction model for palmitoleic acid (**Table 1**) variable x figures with a positive coefficient while variable q (specific to vaccenic acid, **Table 2**) has a negative one (i.e., subtracting vaccenic part from signal x provides a better correlation with palmitoleic acid). Moreover, the presence of two signals (predictors) in linoleic and linolenic acids models is due to the occurrence of these acids at glycerol sn-1,3 (α) and sn-2 (β) positions (**Figure 1**).

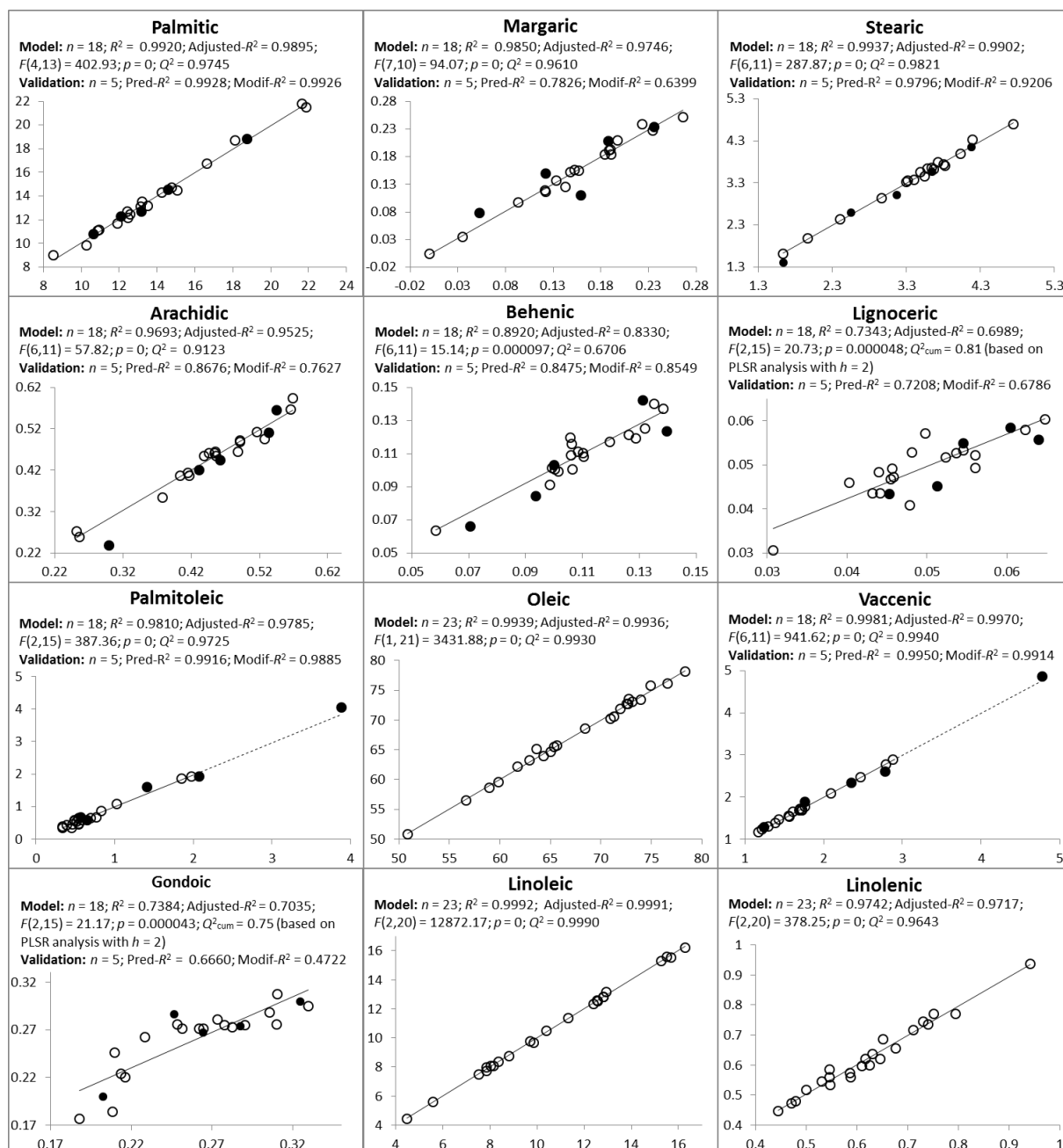


Figure 2: Correlations between fatty acids percentages obtained by GC (x-axis) and predicted using ^{13}C -INEPT method (y-axis). \circ Model samples, \bullet Validation samples, $p = 0$ means that it is less than 0.00001.

Besides, in some regression models (**Table 1**), signals mainly belonging to the same fatty acid may appear with positive and negative coefficients. It is likely that this phenomenon permits to extract information contained in overlapped signals; e.g., in the vaccenic acid model, an improved correlation was obtained when oleic acid signals i (positive sign) and o (negative sign), palmitic acid signals j (positive sign) and p (negative sign), and signal k were incorporated. They probably compensate some overlapping effect with signal l , responsible for the correlation with vaccenic acid (**Table 2**).

Table 2: Linear correlations between specific peak areas from ^{13}C -INEPT spectra and fatty acids determined by GC.

Fatty acid	Variable (ppm) ^a	R^2
Palmitic	j (29.50)	0.9804
	C2-c (33.90)	0.9765
	p (29.30)	0.9703
Linoleic	L ω 2 (22.57)	0.9989
	L11 $\alpha\beta$ (25.57)	0.9988
	ω 1-b (14.10)	0.9961
Linolenic	Ln11 $\alpha\beta$ (25.60)	0.9768
	Ln16 $\alpha\beta$ (131.60)	0.9677
	Ln14 $\alpha\beta$ (24.46)	0.9608
Oleic	ω 3-b (31.92)	0.9939
	i (29.55)	0.9919
	a (29.76)	0.9913
	o (29.32)	0.9848
Palmitoleic	x (28.98)	0.9769
	ω 2-c (22.66)	0.9497
	ω 3-d (31.79)	0.9061
	ω 3-c, ω 3-d, ω 3-satellite ^b	0.9809
Vaccenic	G11+V11 (129.66)	0.9909
	q (29.27)	0.9858
	l (29.46)	0.9734
	x (28.98)	0.9721
	ω 2-c (22.66)	0.9369
	ω 3-c (31.80)	0.9241 ^c
	ω 3-c, ω 3-d, ω 3-satellite ^b	0.9856

^a See variables labelling in Figure 1. ^b Combination of ω 3-c, ω 3-d, and ω 3-satellite using MLR analysis. It is likely that ω 3-c and ω 3-d signals correspond to both palmitoleic and vaccenic fatty acids at glycerol sn-1,3 (α) and glycerol sn-2 (β) positions; the ω 3-satellite peak at 32.06 ppm permits to correct the overlapping between ω 3-c and the ^{13}C satellite corresponding to ω 3-b signal (Figure 1). ^c A better correlation ($R^2 = 0.9436$) was reached by means of MLR when including the ω 3-satellite peak.

Finally, it is remarkable that for most of the models, except those of oleic, linoleic, and linolenic acids, the predictors are predominantly among the methylene signals confined within an interval of less than 1 ppm between 28.9 and 29.8 ppm (signals *a* to *x*).

Conclusions

In summary, ^{13}C -INEPT combined to deconvolution fitting and multivariate linear regression techniques offers a valid strategy that permits the quantification of individual fatty acids. The experimental time is short – 8 min 22 s per spectrum – and no chemical treatment of sample is required. These findings make ^{13}C -INEPT spectroscopy a promising, rapid and valuable method to determine, within a single experiment, individual fatty acids amounts, information regarding their position on the glycerol moiety, and their relative site specific ^{13}C content. Further works are now underway to test this strategy on a larger set of olive oil samples and to adapt it for other triacylglycerol matrices in order to fully characterize their fatty acids profile.

Acknowledgements

N.M. acknowledges the financial support of the Lebanese National Council for Scientific Research and the Research Council of Saint-Joseph University. The CORSAIRE platform from Biogenouest is also acknowledged.

References

- 1 D. Wishart, *Trends Analyt. Chem.*, 2008, **27**, 228-237.
- 2 K. F. Wollenberg, *J. Am. Oil Chem. Soc.*, 1990, **67**, 487-494.
- 3 R. Sacchi, F. Addeo and L. Paolillo, *Magn. Reson. Chem.*, 1997, **35**, S133-S145.
- 4 G. Vlahova, *Prog. Nucl. Mag. Res. Sp.* 1999, **35**, 341-357.
- 5 G. Vlahova, A. D. Shawb and D. B. Kell, *J. Am. Oil Chem. Soc.*, 1999, **76** 1223-1231.
- 6 R. Zamora, V. Alba and F. J. Hidalgo, *J. Am. Oil Chem. Soc.*, 2001, **78**, 89-94.
- 7 G. Vlahova, *Anal. Chim. Acta*, 2006, **577**, 281-287.
- 8 N. Merchak, J. Bejjani, T. Rizk, V. Silvestre, G. Remaud and S. Akoka, *Anal. Methods*, 2015, **7**, 4889-4891.
- 9 P. Scanoa, M. Casua, A. Laia, G. Sabaa, M.A. Dessib, M. Deianab, F. P. Corongiub and G. Bandinoc, *Lipids*, 1999, **34**, 757-759.
- 10 G. Vlahov, P. K. Chepkwony and P. K. Ndalut, *J. Agric. Food Chem.*, 2002, **50**, 970-975.
- 11 European Economic Community (1977). European Community regulation no. 72/77. Amending Regulation (EEC) No 1470/68 on the drawing and reduction of samples and the determination of oil content, impurities and moisture in oil seeds.
- 12 International Olive Oil Council (2001). Preparation of the fatty acid methyl esters from olive oil and olive-pomace oil. <http://www.internationaloliveoil.org/estaticos/view/222-standards>. (Accessed 02.06.14).
- 13 C. Thibaudeau, G. Remaud, V. Silvestre, S. Akoka, *Anal. Chem*, 2010, **82**, 5582-5590.
- 14 U. Bussy, C. Thibaudeau, F. Thomas, J.R. Desmurs, E. Jamin, G. S. Remaud, V. Silvestre and S. Akoka, *Talanta*, 2011, **85**, 1909-1914.
- 15 O. Galtier, N. Dupuy, Y. Le Dréau, D. Ollivier, C. Pinatel, J. Kister, and J. Artaud, *Anal. Chim. acta*, 2007, **595**, 136-144.
- 16 M. J. Lerma-García, E. F. Simó-Alfonso, A. Bendini and L. Cerretani, *Food chem.*, 2011, **124**, 679-684.
- 17 G. S. Kapur, A. Ecker and R. Meusinger, *Energy & fuels*, 2001, **15**, 943-948.
- 18 E. Tenailleau and S. Akoka, *J. Magn. Reson.*, 2007, **1**, 50-58.
- 19 R. Rakotomalala, In: Proceedings of EGC'2005, RNTI-E-3, vol. 2, 697-702.
- 20 D. M. Hawkins, S. C. Basak and D. Mills, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 579-586.
- 21 J. P. Gauchi and P. Chagnon, *Chemometr. Intell. Lab.*, 2001, **58**, 171-193.
- 22 A. Golbraikh and A. Tropsha, *J. Mol. Graphics Model.*, 2002, **20**, 269-276.
- 23 P. P. Roy and K. Roy, *QSAR Comb. Sci.*, 2008, **3**, 302-313.
- 24 D. Boskou, G. Blekas and M. Tsimidou, In *Olive oil: Chemistry and Technology*, 2nd ed.; Boskou, D.; AOCS Publishing, 2006; pp 41-72.

IV-3. Données complémentaires

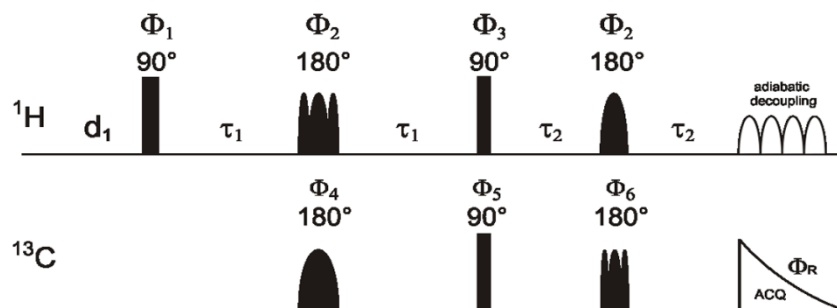


Figure S1: Adiabatic refocused INEPT sequence with ^1H and ^{13}C 180° adiabatic composite refocusing pulses and adiabatic full passage inversion pulses. The duration τ_1 was adjusted to 2.704 ms, and the refocusing period τ_2 was adjusted to 1.409 ms.

Table S1: Fatty acids ranges in olive oils samples of this study versus that reported by international sources

Fatty acid		Fatty acid range (%)	
		Analyzed samples ^a	International sources ^b (average)
Saturated	Palmitic (P, C16:0)	8.5 – 21.9	7.5 – 20 (11.9)
	Margaric (M, C17:0)	0.035 – 0.27	0.03 – 0.2 (0,08)
	Stearic (S, C18:0)	1.6 – 4.7	0.5 – 5 (2.26)
	Arachidic (A, C20:0)	0.25 – 0.56	< 0.6 (0.38)
	Behenic (B, C22:0)	0.058 – 0.14	< 0.3 (0.12)
	Lignoceric (Lg, C24:0)	0.030 – 0.064	< 0.2 (0.05)
Monounsaturated	Palmitoleic (Po, <i>c</i> -9-C16:1)	0.34 – 3.9	0.3 – 3.5 (0,8)
	Hypogeic (H, <i>c</i> -7-C16:1)	0.086 – 0.18	0.09 – 0.20 (0,12)
	Margaroleic (Mo, <i>c</i> -8-C17:1)	0.067 – 0.32	0.06 – 0.36 (0.15)
	Oleic (O, <i>c</i> -9-C18:1)	50.7 – 78.0	55 – 83 (72.2)
	Vaccenic (V, <i>c</i> -11-C18:1)	1.2 – 4.8	1.2 – 3.9 (2,3)
	Gondoic (G, <i>c</i> -11-C20:1)	0.19 – 0.33	0.2 – 0.4 (0,28)
Polyunsaturated	Linoleic (L, <i>c</i> -9,12-C18:2)	4.5 – 16.2	3.5 – 21 (8.35)
	Linolenic (Ln, <i>c</i> -9,12,15-C18:3)	0.44 – 0.94	< 1.5 (0.65)

^a Determined by gas chromatography. ^b Boskou, D.; Blekas, G.; Tsimidou, M. In *Olive oil: Chemistry and Technology*, 2nd ed.; Boskou, D.; AOCS Publishing, 2006; pp 41-72. Ollivier D., Pinatel C., Artaud J., *Le Nouvel Olivier*, N°44 mars-avril 2005.

PLS regression results for lignoceric and gondoic acid

➤ Proportion of Variance Explained:

It describes the proportion of variance explained by the latent factors, for the predictors (X) and the dependent variables (Y).

Latent Factor	Lignoceric			
	Input variables (X)		Target Variables (Y)	
	Current X (%)	Cumulative X (%)	Current Y (%)	Cumulative Y (%)
1	77.831	77.831	65.622	65.622
2	19.017	96.848	7.813	73.434

Latent Factor	Gondoic			
	Input variables (X)		Target Variables (Y)	
	Current X (%)	Cumulative X (%)	Current Y (%)	Cumulative Y (%)
1	70.417	70.417	63.153	63.153
2	19.262	89.68	10.687	73.84
3	8.972	98.652	0.34	74.179

➤ R-squared:

It shows the correlation of each variable with the latent factor (i.e., the proportion of variance explained for each variable). This is an important tool for the interpretation and the comprehension of the factors.

Input	Lignoceric			
	Input(s) vs. X-Scores			
	R-squared		Cumulative R-squared	
	Factor 1	Factor 2	Factor 1	Factor 2
Pred-P	0.8296	0.1117	0.8296	0.9413
Pred-A	0.6774	0.3167	0.6774	0.9941
Pred-O	0.8279	0.1421	0.8279	0.97
Total Exp.	0.7783	0.1902	0.7783	0.9685

Input	Gondoic					
	Input(s) vs. X-Scores					
	R-squared			Cumulative R-squared		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
ω3-c	0.8607	0.0517	0.0476	0.8607	0.9124	0.96
F	0.7761	0.0299	0.1846	0.7761	0.806	0.9906
L	0.8467	0.0059	0.1399	0.8467	0.8526	0.9925
P	0.9022	0.0377	0.0548	0.9022	0.9398	0.9946
R	0.1352	0.8379	0.0218	0.1352	0.9731	0.9949
Total Exp.	0.7042	0.1926	0.0897	0.7042	0.8968	0.9865

Target	Target(s) vs. X-Scores			
	R-squared		Cumulative R-squared	
	Factor 1	Factor 2	Factor 1	Factor 2
Lignoceric	0.6562	0.0781	0.6562	0.7343
Total Exp.	0.6562	0.0781	0.6562	

Target	Target(s) vs. X-Scores					
	R-squared			Cumulative R-squared		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Gondoic	0.6315	0.1069	0.0034	0.6315	0.7384	0.7418
Total Exp.	0.6315	0.1069	0.0034	0.6315	0.7384	0.7418

➤ X-loadings (Model Effect Loadings - Vector Ph):

The loadings represent the correlation between factors and predictors. They are important for the interpretation of models; they do not exactly correspond to correlation coefficients.

However, they allow to position the variables in the same way with regard to the factors.

Lignoceric		
Input	Factor1	Factor2
Pred-P	0.6044	-0.4573
Pred-A	0.5461	0.7702
Pred-O	-0.6037	0.516

Gondoic			
Input	Factor1	Factor2	Factor3
$\omega 3-c$	-0.5061	-0.2331	0.367
F	-0.4806	0.1773	-0.7228
L	-0.502	-0.0789	0.6292
P	-0.5182	0.1989	-0.3938
R	0.2006	-0.9382	-0.2484

➤ **X-Weights (Model Effect Weights - Vector Wh and Wh*):**

They indicate how much the predictors are correlated with the PLS Responses. The vectors Wh*, contrary to Wh, are directly related to the predictors. They can interpret them easily.

Input	Lignoceric			
	Wh		Wh*	
	Factor 1	Factor 2	Factor 1	Factor 2
Pred-P	0.5624	-0.2509	0.5624	-0.1568
Pred-A	0.664	0.7049	0.664	0.8159
Pred-O	-0.4927	0.6635	-0.4927	0.5811

Input	Gondoic					
	Wh			Wh*		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
$\omega 3-c$	-0.5573	-0.2335	-0.0038	-0.5573	-0.3555	-0.0424
f	-0.4519	0.131	-0.4264	-0.4519	0.032	-0.423
l	-0.5004	0.0074	0.7942	-0.5004	-0.1022	0.7831
p	-0.4845	0.1537	-0.4162	-0.4845	0.0476	-0.411
r	-0.0076	-0.9511	-0.1189	-0.0076	-0.9528	-0.2224

➤ **Y-Weights (Dependent Variable Weights - Vector Ch):**

They indicate how much the dependent variables are correlated to the PLS responses. It enables to determine which are the variables that are well explained by the PLS responses. These are not really the correlation coefficient, but the interpretation is the same.

Target	Factor 1	Factor 2
Lignoceric	0.5375	0.3826

Target	Factor 1	Factor 2	Factor 3
Gondoic	0.4336	0.3351	0.0981

➤ **VIP (Variable Importance in Projection):**

They reflect the relative importance of the predictors, through the factors, in the prediction model.

Lignoceric		
Input	Factor 1	Factor 2
Pred-P	0.9741	0.9317
Pred-A	1.1501	1.1578
Pred-O	0.8535	0.8896

Gondoic			
Input	Factor 1	Factor 2	Factor 3
$\omega 3$ -c	1.2461	1.1694	1.1667
f	1.0106	0.9412	0.9413
L	1.1189	1.0348	1.0394
P	1.0835	1.0105	1.0101
R	0.0171	0.8093	0.8076

➤ **Unstandardized Regression Parameters (by target variable):**

The unstandardized regression coefficients can be used for the prediction on new instances.

	Target Variable(s)
Input	Lignoceric
Pred-P	0.000992
Pred-A	0.069775
Pred-O	-0.000058
constant	0.00808

	Target Variable(s)
Input	Gondoic
$\omega 3$ -c	-0.005356
f	-0.001179
l	-0.007436
p	-0.001005
r	-0.004862
Constant	0.846188

➤ **Standardized Regression Parameters (by target variable):**

Because, the variables are not in the same unit, we can use the standardized regression coefficients to compare the relative importance of the variables for the prediction.

	Target Variable(s)		
	Average	Std.Deviation	
			0.0491
			0.0081
Input	Average	Std.Deviation	Lignoceric
Pred-P	12.8238	1.9824	0.242308
Pred-A	0.4634	0.0778	0.669041
Pred-O	69.0434	5.9733	-0.042554

	Target Variable(s)		
	Average	Std.Deviation	
			0.2595
			0.0417
Input	Average	Std.Deviation	Gondoic
$\omega 3$ -c	7.4678	2.8397	-0.364876
f	49.6371	8.014	-0.226681
l	1.5824	0.9774	-0.174372
p	51.2092	9.7261	-0.234434
r	87.4153	2.9524	-0.344369

➤ **Regression assessment (using external validation samples):**

Lignoceric		Error Sum of Squares		
		Default (Mean)	Model	Pred-R ²
Observed	Predicted	0.0004	0.0001	0.7208

Gondoic		Error Sum of Squares		
		Default (Mean)	Model	Pred-R ²
Observed	Predicted	0.0085	0.0024	0.7125

Lignoceric		
h	Q ²	Q ² cum
1	0.772	0.772
2	0.482	0.882
3	0.233	0.909

Gondoic		
h	Q ²	Q ² cum
1	0.649	0.649
2	0.299	0.754
3	0.106	0.78
4	-0.122	0.753

➤ **Standardized Regression Parameters for all individual fatty acid models obtained by MLR:**

Target Variable(s)			
Average			14.0129
Std.Deviation			3.6115
Input	Average	Std.Deviation	Palmitic
b	183.8865	33.5689	-0.771832
m	53.6413	10.3542	1.344647
o	203.3063	23.6027	-0.193233
ω2-c	7.4965	5.3947	0.293601

Target Variable(s)			
Average			0.1544
Std.Deviation			0.0663
Input	Average	Std.Deviation	Margaric
i	199.8012	22.4955	-3.412733
j	56.2271	13.6518	-12.724236
k	1.9037	0.31	-0.428316
r	87.9918	2.9574	-2.978709
s	144.0547	11.2265	-8.191213
t	150.4294	7.5586	3.220329
u	141.6862	8.0598	3.793505

Target Variable(s)			
Average			3.4032
Std.Deviation			0.7617
Input	Average	Std.Deviation	Stearic
ω 3-a	55.3433	8.3434	-6.463574
ω 3-c	6.8304	2.0689	-0.44438
b	175.5139	26.6575	3.96834
m	50.8439	7.748	4.946303
n	242.6082	8.5858	1.252618
ω 2-a	55.355	10.6077	-0.484325

Target Variable(s)			
Average			0.4455
Std.Deviation			0.0872
Input	Average	Std.Deviation	Arachidic
b	181.5559	33.0734	3.414693
c	47.7349	11.7854	1.887807
m	52.9728	10.4493	-6.639207
n	239.5616	12.881	-3.569843
o	203.4072	20.5843	0.370549
x	4.0447	2.8536	-1.641294

Target Variable(s)			
Average			0.1107
Std.Deviation			0.0184
Input	Average	Std.Deviation	Behenic
b	182.1545	24.244	3.647394
d	166.1071	13.6449	-0.607258
m	52.7417	8.3313	-14.857829
o	202.8678	18.9306	2.967784
p	49.4333	9.3483	16.020073
q	6.1196	2.0683	-2.123512

Target Variable(s)			
Average			0.7157
Std.Deviation			0.4708
Input	Average	Std.Deviation	Palmitoleic
q	5.6016	1.691	-0.362392
x	3.173	2.3419	1.319429

Target Variable(s)			
Average			67.5023
Std.Deviation			6.9886
Input	Average	Std.Deviation	Oleic
ω 3-b	197.8949	21.5931	0.996955

Target Variable(s)			
Average			1.7657
Std.Deviation			0.4932
Input	Average	Std.Deviation	Vaccenic
i	201.7117	15.562	3.058251
j	54.7322	8.884	1.649544
k	1.8514	0.2068	0.229447
l	1.3318	0.5777	0.566063
o	206.0833	16.1579	-2.913173
p	48.7746	7.6873	-1.127484

Target Variable(s)			
Average			10.7247
Std.Deviation			3.3092
Input	Average	Std.Deviation	Linoleic
L9 α	13.465	4.6923	0.53202
L12 β	13.5333	4.1144	0.475897

Target Variable(s)			
Average			0.6244
Std.Deviation			0.1176
Input	Average	Std.Deviation	Linolenic
Ln16 $\alpha\beta$	1.6494	0.3187	0.58478
Ln14 $\alpha\beta$	1.5949	0.3414	0.407033

IV-4. Conclusion

Des modèles multivariés linéaires basés sur des prédicteurs fournis par les spectres INEPT- ^{13}C déconvolués des huiles d'olive ont permis de quantifier leurs contenus en squalène et en acides palmitique, margarique, stéarique, arachidique, béhénique, lignocérique, palmitoléique, oléique, vaccénique, gondoïque, linoléique et linoléinique.

L'approche utilisée se caractérise par :

- un temps expérimental court (8 min 22 s par spectre),
- une préparation des échantillons très rapide et simple, aucun traitement chimique de l'échantillon n'est nécessaire,
- l'obtention, en une seule expérience, des pourcentages en acides gras individuels, des teneurs isotopiques relatives en ^{13}C sur plusieurs sites clés des triacylglycérols, ainsi que la distribution des acides oléique, linoléique et linoléinique sur le squelette du glycérol.

Cette stratégie a été testée ultérieurement sur un plus grand nombre d'échantillons d'huile d'olive pour confirmer la robustesse des modèles construits (les résultats sont présentés dans la partie V). De plus, elle pourrait être adaptée à d'autres matrices de triacylglycérols afin de construire des modèles adéquats pour déterminer leur profil complet d'acides gras.

*Partie V - Caractérisation et
classification des huiles d'olive en
utilisant une technique de transfert de
polarisation en RMN ¹³C :
comparaison avec la GC et la RMN ¹H*

V-1. Introduction

Dans la quatrième partie de ce manuscrit, nous avons montré que la séquence INEPT-¹³C permet de déterminer rapidement les concentrations en acides gras individuels et en squalène dans l'huile d'olive et ceci sans avoir besoin des réactions préliminaires de trans-méthylesterification.

Une nouvelle série d'échantillons d'olives qui n'a pas été introduite dans l'étude précédente (partie IV), a été analysée par les deux techniques INEPT-¹³C et GC ; les corrélations linéaires entre les valeurs prédites des acides gras et celles obtenues par l'analyse chromatographique sont présentées pour valider la robustesse des modèles construits dans l'étude précédente.

Par ailleurs, un échantillon d'huile d'olive a été analysé plusieurs fois par RMN (¹³C et ¹H) et GC afin de comparer les répétabilités à long terme de ces méthodes.

Finalement, les variables obtenues par chacune de ces trois techniques (INEPT-¹³C, RMN ¹H et GC) ont été évaluées en tant que descripteurs pour la classification des huiles d'olive en fonction de l'altitude du terroir d'origine et de la couleur des olives. Une étude comparative a été effectuée à ce niveau en traitant les données correspondantes à chaque technique par les analyses discriminantes linéaire et canonique (LDA et CDA).

Les résultats obtenus ont fait l'objet d'un article intitulé « Olive oil characterization and classification by ¹³C-NMR with a polarization transfer technique: A comparison with gas chromatography and ¹H-NMR », qui va être soumis à la revue Lipid Research.

V-2. Article

Olive oil characterization and classification by ^{13}C -NMR with a polarization transfer technique: A comparison with gas chromatography and ^1H -NMR

Noelle Merchak,^{a,b} Toufic Rizk,^b Virginie Silvestre,^a Gerald S. Remaud,^a Joseph Bejjani,^{b,*} and Serge Akoka^a

^a EBSI team, Interdisciplinary Chemistry: Synthesis, Analysis, Modelling (CEISAM), University of Nantes-CNRS UMR 6230, 2 rue de la Houssinière, BP 92208, F-44322 Nantes cedex 3, France

^b Research Unit: Technologies et Valorisation Agroalimentaire (TVA), Laboratory of Metrology and Isotopic Fractionation, Faculty of Science, Saint-Joseph University, P.O. Box 11-514 Riad el Solh, Beirut 1107 2050.

Abstract

Multivariate linear regression models using variables from ^{13}C -INEPT spectra of olive oil samples were constructed in a previous work aiming to predict fatty acids and squalene amounts within the studied matrix. In this study, a new set of olive oil samples was analyzed using the same NMR method and by gas chromatography (GC) to prove the reliability of the constructed models and to assess their predictability. Furthermore, ^1H -NMR measurements were achieved on the same samples in order to compare the long term repeatability of the three techniques. Finally, a comparative evaluation of the classification power of these techniques, according to a morphological and a geographical factor affecting olive oils, was performed by means of the linear discriminant analysis.

Keywords

INEPT; RMN ^1H ; GC; Olive oil; Classification; long term repeatability; Fatty acids

1. Introduction

In the past few years, the authenticity control of edible oils has gained increasing attention and the search for the origin of olive oils has been the object of several studies.¹⁻⁵ In this respect, the fatty acid compositional data, obtained by means of the gas chromatography (GC) method, have been used as analytical variables in order to classify oils and to ensure their botanical and geographical traceability.⁶⁻¹⁰ However, the GC analysis used to this end is time-consuming and requires prior chemical manipulations such as trans-esterification. An alternative can be found with Nuclear Magnetic Resonance (NMR) spectroscopy, which delivers direct information about the molecular composition of fatty acids in triacylglycerols matrices only by dissolving oils in deuterated solvent without any previous chemical preparations. For instance, ¹H-NMR method provides the percentages of saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), polyunsaturated fatty acids (PUFA), linoleic and linolenic acids (the only fatty acids that can be individually quantified by this method), and allows the identification of minor components in oils. On the other hand, ¹³C-NMR, owing to its broad range of resonance frequencies, affords additional information regarding the individual fatty acids amounts,^{11,12} the positional distribution of fatty acids on the glycerol backbone,¹³⁻¹⁵ and the position-specific isotopic ¹³C content of the studied molecules.¹⁶ Nevertheless, it is a very lengthy experiment due to the low natural abundance of ¹³C, its small gyromagnetic ratio and its long longitudinal relaxation time. A dramatic reduction in experimental time can be reached by applying polarization transfer methods from ¹H to ¹³C nuclei such as INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) pulse sequence. However, the regular version of this sequence does not satisfy the isotopic analysis requirement since it suffers from a low repeatability (about 5%) due to off-resonance effects and radiofrequency inhomogeneity in the 180° pulses applied to ¹H and ¹³C nucleus.¹⁷ For this reason, appropriate adiabatic pulses must be used in the INEPT sequence to reach a high precision (repeatability and reproducibility of per mil).^{17,18} In a previous work, we have investigated this issue for the metabolomic and isotopic characterization of triacylglycerols matrices.¹⁶ A precision of about 1 per mil was obtained for carbons of the glycerol moiety (C1,3 and C2) and for carbons 2 and 3 of fatty acids (C2 and C3), allowing their use for isotopic purposes. The other peaks observed in the spectrum were used as compositional profile elements.¹⁶ Moreover, we have successfully used the precise ¹³C-INEPT data thus obtained to construct multivariate linear regression models permitting the individual quantification of fatty acids in triacylglycerols matrices.¹² In the present work, a new set of olive oil samples was analyzed by this approach, by GC, and by ¹H-NMR to prove the

reliability of the constructed models, to compare the long term repeatability of these three techniques and to evaluate their classification power, i.e., classification according to the altitude of the olive field and to the colour of olives.

2. Materials and methods

2.1. Olive oil samples

Thirty-six authentic green and black Lebanese olive samples were collected from different olive-growing regions at different altitudes designated as low (Lw, 0-350 m) and high (Hh, above 700 m). Authentic oils were obtained solely by mechanical methods in the laboratory. Olives were crushed, pitted and then ground to obtain a paste. The paste was centrifuged and the resulting oil layer was decanted. Oils were centrifuged to eliminate residual particles and then stored at -18 °C.

2.2. NMR spectroscopy

2.2.1. ¹H-NMR analysis

A 400.13 MHz Bruker Avance II spectrometer was used. Oil samples were prepared in 5 mm NMR tubes; a concentration of 25% (v/v) of olive oil (0.18 mL) in chloroform-d (0.56 mL, purchased from Deutero GmbH) was used in order to obtain the best compromise between sensitivity and resolution.¹⁹ For each sample, 7 ¹H-NMR spectra were acquired at 298 K using the following conditions: time domain size 64 K; pulse angle 30°; pulse width 27.4 μs; spectral width 9 ppm; acquisition time 9.1 s; relaxation delay 1 s; 4 dummy scans and 32 transients. The longest longitudinal relaxation times T_1 (3.19 s) was observed for the methyl group of the α -linolenic acid by using the inversion-recovery method. The recording time of the seven spectra was 43 min. Free induction decays (FIDs) were zero-filled to 128 K and exponentially apodized with $LB = 0.3$ Hz. The spectra were manually phased. The baseline correction was performed using the Cubic Spline technique in Bruker TOPSPIN 1.3 software. All the spectra were calibrated by setting the peak with the highest chemical shift, corresponding to the CH₂ signal of the glycerol moiety at 4.299 ppm. The spectral resolution, expressed as intensity of the minimum between the signals at 4.288 and 4.299 ppm relative to intensity of the signal at 4.299 ppm (**Figure S1** in SM), was systematically determined; it never exceeds 25%. Peak areas determination was achieved using fixed chemical shift intervals and the corresponding values were normalised relative to the signal of protons at position 2 of acyl chains. The same protons were taken as reference when measuring intensities of signals (**Figure S2** in SM). For each olive oil sample, one discrete and sixteen

continuous variables related to fatty acids and minor components were obtained. These variables, listed in our previous study²⁰ (see **Table S1** in SM), were used as inputs in the statistical analysis.

2.2.2. ¹³C-NMR analyses

Samples were prepared by dissolving oils (403.2 mg) in CDCl₃ (630.0 mg, purchased from Eurisotop). ¹³C-NMR spectra were recorded on a 500 MHz Bruker Avance-III spectrometer equipped with a 5 mm dual cryoprobe ¹³C/¹H carefully tuned at the recording frequency of 125.76 MHz for ¹³C. The temperature of the probe was set at 293 K. The pulse lengths were calibrated and the probe tuning and matching were adjusted at the beginning of each measurement session. ¹³C and ¹H offsets were set at the middle of the frequency range (92.5 ppm for the ¹³C and 3 ppm for the ¹H).

2.2.2.1. ¹³C-NMR one pulse sequence

Quantitative ¹³C-NMR spectra were acquired by using: pulse width 11 μs (90°), acquisition time of 1 s, a repetition delay of 42 s, which is greater than 10 times the longest ¹³C T₁, since the longest ¹³C T₁ using the inversion-recovery pulse sequence was 4.0 s for the measured signals. Moreover, 4 dummy scans and 72 scans were recorded leading to a signal-to-noise ratio (SNR) > 500 on the C2 of glycerol, which is a requirement for a precision around 1 per mil. Inverse-gated decoupling techniques were applied in order to avoid the influence of the nuclear Overhauser effect (NOE). ¹H decoupling was performed using adiabatic full passage RF pulses with cosine square amplitude modulation ($\omega_2^{\max} = 110.6$ kHz) and offset independent adiabaticity with optimized frequency sweep ($\Delta F = 14$ kHz) and appropriate phase cycles.²¹ The recording time of six spectra was 5 hours 45 min.

2.2.2.2. ¹³C-INEPT adiabatic sequence

¹³C-INEPT spectra were obtained by using the adiabatic INEPT pulse sequence^{16,18} (**Figure S3** in SM). The acquisition parameters were as follows: pulse width 10 μs for the 90° ¹H and 11 μs for the 90° ¹³C, acquisition time of 1 s, repetition time of 24 s (The longest ¹H-T₁ (2.9 s) was measured (using the inversion-recovery pulse sequence) for the methyl group of the linolenic acid), 4 dummy scans and 16 scans were accumulated leading to a SNR higher than 600 on the C2 of glycerol. Adiabatic full passage pulses were generated using Mathcad 8 (MathSoft, Inc.). They were designed with a cosine amplitude modulation of the RF field ($\omega_1^{\max} = 157.1$ kHz or 93.89 kHz for ¹³C or ¹H, respectively) and an offset independent adiabaticity by optimizing the frequency sweep ΔF ($\Delta F = 39$ kHz or 17 kHz for ¹³C or ¹H,

respectively).²¹ For inversion pulses, adiabatic full passage pulses were used. For refocusing pulses, composite adiabatic pulses were used.¹⁷ ¹H decoupling was performed as for ¹³C-NMR one pulse sequence. Six spectra were recorded for each sample. The experiment time of each sample (6 spectra) was reduced to 50 min 10 s.

2.2.3. Data Processing of ¹³C-NMR experiments

Free induction decays (FIDs) were zero-filled to 128 K and submitted to an exponential multiplication inducing a line broadening of 1.5 Hz. ¹³C-NMR spectra were manually phased. An automatic polynomial baseline correction ($n = 5$) was applied to the resulting spectra. The curve fitting was carried out in accordance with a Lorentzian mathematical model using Perch Software (Perch NMR Software, University of Kuopio, Finland).

2.3. Gas Chromatographic analysis

The complete fatty acids composition was quantified by gas chromatography. Fatty acids methyl esters were prepared by shaking 200 mg of oil in 3mL of hexane with 0.4 mL of 2N methanolic potassium hydroxide.^{22,23} An Agilent Technologies chromatograph equipped with a flame ionization detector and a fused-silica capillary column (SGE-054616, 30m × 0.32mm i.d., BPX70 0.25 μ m) was used in this study with helium as carrier gas at flow of 1.0 mL/min. Oven temperature was programmed as follows: at 160 °C for 15 min, to 200 °C at a rate of 10 °C/min, at 200 °C for 10 min. A 1:10 dilution in hexane was performed. A volume of 1 μ L was injected with a split ratio of 50:1 and an injector temperature of 200 °C. The detector temperature was set at 270 °C. Two injections were performed for each sample with a total elution time of 58 min. Standard fatty acid methyl esters, purchased from sigma Aldrich, were used to determine the retention time of each fatty acid. Molar percentages of palmitic, palmitoleic, margaric, margaroleic, hypogeic, stearic, oleic, vaccenic, linoleic, linolenic, arachidic, gondoic, behenic, and lignoceric acids in olive oils were determined relatively to their total number of moles by dividing each peak area by the molar weight of the corresponding fatty acid methyl ester. Squalene was also quantified in the same way.

2.4. Statistical analysis

Multivariate statistical analysis of the resulting data was accomplished by means of TANAGRA software.²⁴ External validation samples were used in this study in order to evaluate the predictive power of fatty acids and squalene prediction models developed in our previous work.¹² Predictive R^2 (Pred- R^2) was used to assess the predictability of the models.²⁵ Pred- R^2 is given by:

$$\text{Pred-}R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_{train})^2} \quad (1)$$

where n is the number of validation samples; y_i is a fatty acid amount determined by GC; \hat{y}_i is the prediction of a fatty acid amount given by the appropriate regression model; and \bar{y}_{train} represents the mean of y_i values for all the training set samples used to construct the prediction model reported in our previous study.¹²

However, $\text{Pred-}R^2$ does not guarantee a true expression of external predictability since it is dependent on the mean value of the training set samples and hence affected by the choice of test samples. In this respect, the parameter Modified R^2 ($\text{Modif-}R^2$), reported by Roy et al., permits to better understand the predictability of models.²⁵ This parameter is given by:

$$\text{Modif-}R^2 = r^2 \left(1 - \sqrt{r^2 - r_0^2} \right) \quad (2)$$

with r^2 and r_0^2 the squared correlation coefficients between GC and predicted values of the external validation samples with and without intercept, respectively. A $\text{Modif-}R^2$ greater than 0.5 is an indicator of good predictability.²⁵

Besides, deviations observed between values obtained by GC analysis and those predicted by models are expressed as relative sample standard deviation of difference (RSDD) which is given by:

$$\text{RSDD} (\%) = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \times \frac{100}{\bar{y}} \quad (3)$$

where n , y_i , \hat{y}_i have the same definition as in equation 1, and \bar{y} represents the mean of y_i values for the validation samples used in this study.

On the other hand, for the classification issues, canonical discriminant analysis (CDA) was carried out while eliminating irrelevant input variables using the backward stepwise discriminant analysis and based on the contribution of each variable to *wilks' λ* criterion. The linear discriminant analysis (LDA) was used as classifier and the performance of the classification model was assessed using *wilks' λ* and the classification error rate (*Er*) as indicators. The overall significance of models was evaluated based on Rao's $F(\text{df1}, \text{df2})$ - and p -values. The robustness of the selected models was assessed by means of the leave-one-out (LOO) cross-validation test. The result of the test is expressed as classification error rate (LOO-*Er*).

3. Results and discussion

3.1. Predictability of INEPT based regression models for fatty acids and squalene

In a previous work,¹² multivariate regression models based on ¹³C-INEPT variables were constructed for the individual quantification of fatty acids and squalene within the olive oil matrix. In order to further assess the predictability of these models, eleven olive oil samples, not included in the preceding study, were used for an additional validation. They were analyzed by GC and ¹³C-INEPT methods. The correlations between the two sets of data thus obtained are depicted in **Figure 1** along with the prediction ability indicators r^2 , r_0^2 , $\text{Pred-}R^2$ (eq. 1), $\text{Modif-}R^2$ (eq. 2), and RSDD (eq. 3) described in Materials and Methods section.

As shown in **Figure 1**, the prediction values for squalene, palmitic, stearic, arachidic, palmitoleic, oleic, vaccenic, linoleic and linolenic acids using ¹³C-INEPT fits perfectly with the values obtained by GC analysis where $\text{Pred-}R^2 > 0.78$, $r^2 > 0.92$, $r_0^2 > 0.88$ and $\text{Modif-}R^2 > 0.70$. However, for margaric and behenic acids present in trace amounts in olive oil, it is likely that the divergence between R^2 and $\text{Pred-}R^2$ or $\text{Modif-}R^2$ is due to a deviation on the GC calibration. Nevertheless, for gondoic and especially lignoceric acid (the correlation for this latter is not presented in **Figure 1**, the corresponding parameters are as follows: $\text{Pred-}R^2 = 0.2414$, $r^2 = 0.4030$, $r_0^2 = 0.2873$ and $\text{Modif-}R^2 = 0.2659$), present in olive oil at less than 0.4% and 0.2%, respectively,²⁶ weak predictions were obtained. Concerning the RSDD, it is clear that this parameter is affected by deviations due to both GC and INEPT methods.

3.2. Variables obtained by GC, ¹H-NMR, one-pulse ¹³C-NMR, and ¹³C-INEPT

As shown in **Table 1**, GC method affords individual quantification of 14 fatty acids and squalene present in olive oils; ¹H-NMR technique provides, in one single experiment, amounts of linoleic and linolenic acids, SFA, MUFA, other minor components such as phenolic compounds (Ph), a terpene (Trp-1), β -sitosterol (Ss), diglycerides (1,2-dig) and an unknown compound (X); and ¹³C-NMR method, due to its wide chemical shift range (approximately 250 ppm), allows to quantify oleic, linoleic, and linolenic acids on the 1,3- and 2-positions of the glycerol backbone, and it is the only technique that permits a position-specific isotopic analysis (PSIA), i.e., site-by-site relative isotopic ratio determination.

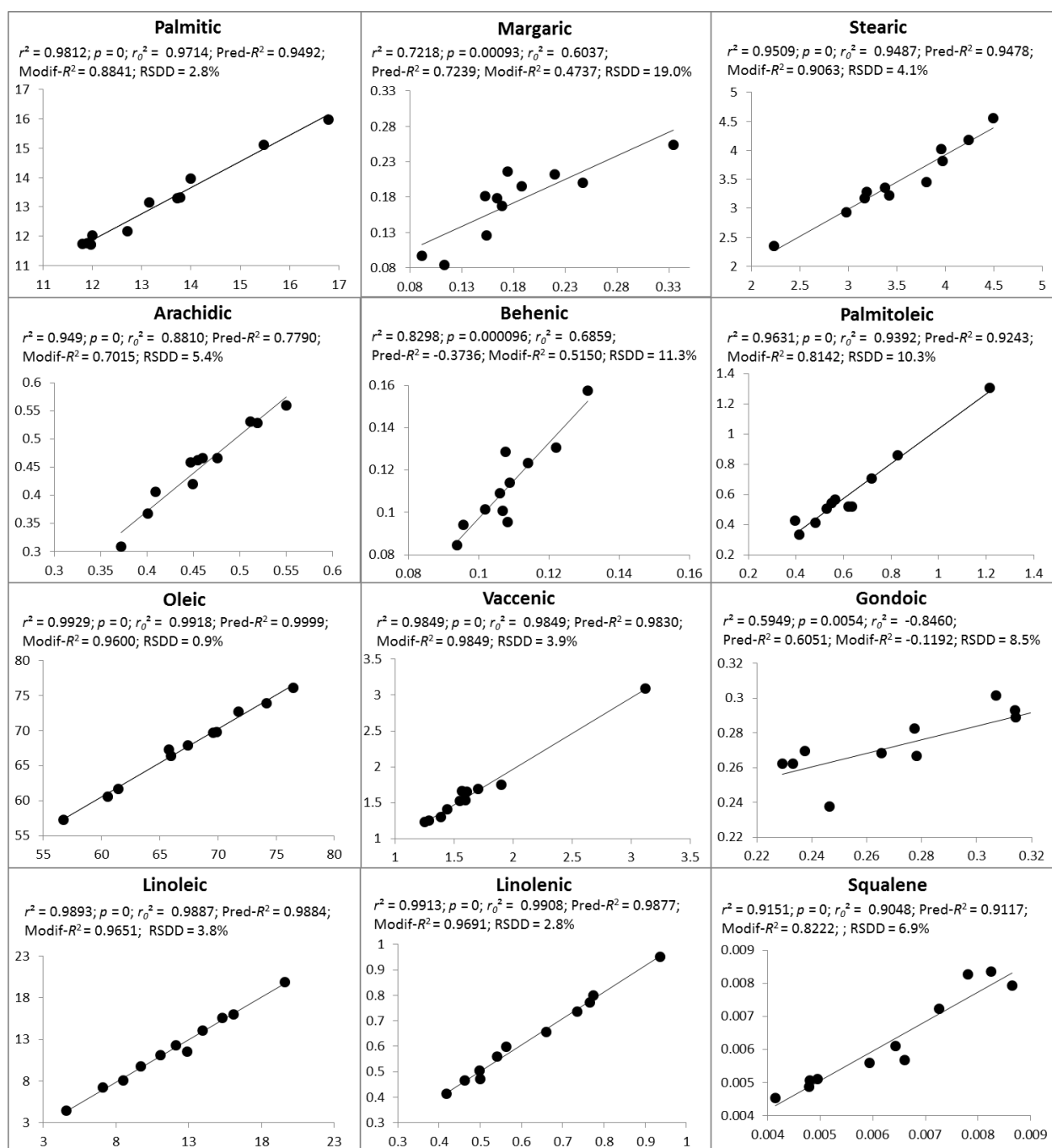


Figure 1: Linear correlations between fatty acids and squalene percentages determined by GC (x-axis) and predicted ^{13}C -INEPT models (y-axis) for a new set of olive oil samples ($n = 11$).

However, since the classical ^{13}C -NMR experiment is time-consuming, an adiabatic INEPT sequence was optimized in such a way that the glycerol signals (C1,3 and C2) are still obtained with the correct ratio and with high sensitivity satisfying the required precision for ^{13}C isotopic analysis.¹⁶ For the other signals, correction factors can be determined and the information regarding the positional distribution of fatty acids on the glycerol skeleton lost with the carboxylic carbons signals can be retrieved from other regions of the spectra such as from the olefinic region ranging from 132 to 127 ppm and from the region related to the C2 of

fatty acids ranging from 34.2 to 33.7 ppm.^{16,27} Thus, relatively to GC analysis of triacylglycerols, INEPT advantageously provides, without a previous derivatization step, ¹³C isotopic information related to different sites in the molecules¹⁶ and positional distribution ratios of oleic, linoleic, and linolenic acids,¹⁶ while allowing the prediction of most fatty acids amounts¹² with excellent to acceptable accuracy (**Figure 1**).

Table 1: GC, ¹H-NMR, ¹³C-INEPT, and one-pulse ¹³C-NMR variables with their long-term repeatability.

Variables	RSD % ^a (Robust-RSD %) ^b of variables afforded by each technique			
	GC (<i>n</i> = 11)	¹ H-NMR (<i>n</i> = 12)	Adiabatic INEPT sequence (<i>n</i> = 11)	One pulse ¹³ C-NMR sequence ^c (<i>n</i> = 6)
Saturated fatty acids (SFA)	5.2 (1.2)	2.1 (1.7)	1.9 (1.2) ^c	+
Palmitic (P)	6.6 (2.3)		1.9 (1.0) ^c	+
Margaric (M)	3.1 (4.1)		8.2 (7.9) ^c	
Stearic (S)	1.0 (0.9)		5.3 (4.5) ^c	+
Arachidic (A)	3.5 (4.8)		3.5 (3.5) ^c	+
Behenic (B)	5.1 (6.7)		10.2 (6.1) ^c	+
Lignoceric (Lg)	4.2 (3.5)		nd ^d	
Monounsaturated fatty acids (MUFA)	0.9 (0.4)	0.4 (0.3)	0.2 (0.2) ^c	+
Margaroleic (Ma)	2.9 (2.5)			
Hypogeic (H)	2.9 (2.9)			
Palmitoleic (Po)	2.7 (2.0)		2.4 (1.2) ^c	+
Oleic (O)	0.9 (0.6)		0.2 (0.1) ^c	+
Vaccenic (V)	1.8 (1.8)		2.4 (3.1) ^c	+
Gondoic (G)	4.8 (5.2)		nd ^d	+
Polyunsaturated fatty acids (PUFA)	2.3 (1.5)	0.5 (0.7)	0.7 (0.4) ^c	+
Linoleic (L)	2.3 (1.6)	0.6 (0.8)	0.9 (0.4) ^c	+
Linolenic (Ln)	2.8 (2.7)	3.5 (3.9)	1.2 (1.4) ^c	+
Minor components				
Squalene (Sq)	7.8 (7.8) ^f	1.7 (2.6)	5.9 (5.5) ^c	4.7 (2.3)
Phenolic compounds (Ph)		12.6 (7.9)		
Terpene (Trp-1)		18.3 (22.7)		
β-Sitosterol (Ss)		7.5 (6.3)		
Diglyceride (1,2 dig)		3.2 (5.1)	3.2 (3.1)	7.0 (8.8)
Unknown compound (X)		11.0 (9.4)		
Positional distribution ratios of fatty acids				
O9α/O9β ^g			0.8 (0.4)	0.9 (0.5)
O10α/O10β ^h			0.9 (1.0)	0.8 (0.6)
Oα/Oβ ⁱ			0.4 (0.4)	0.8 (0.6)
Lα/Lβ ^j			0.9 (0.9)	0.4 (0.5)
Lnα/Lnβ ^k			9.7 (14.6)	40.3 (23.3)
Isotopomic information				
Cα/Cβ of glycerol			0.1 (0.1)	0.1 (0.1)
C2/C3			0.1 (0.1)	0.1 (0.1)
C1/C2				0.4 (0.1)
Cω3/Cω2 ^l			0.1 (0.1)	0.2 (0.2)
Cω1/Cω2 ^l			0.1 (0.1)	0.1 (0.1)

^a Within-laboratory long term repeatabilities determined from control charts over a period of one year and expressed as relative standard deviations (RSD %), *n* is the number of measurements. ^b Robust-RSD = 1.4826 × MAD × 100 / Median, (MAD: Median Absolute Deviation = Median (|x_i - Median(x_i)|)). ^c amounts of fatty acids and squalene were predicted by the models reported in our previous study.¹² ^d nd: Repeatability not determined for this fatty acid since the correlation existing between GC data and predicted values is weak. ^e +: information could be obtained by one pulse ¹³C-NMR sequence but the model was developed using the INEPT sequence (reference 12). ^f *n* = 9 after excluding outliers deviating more than 5 × SD (standard deviation). ^g The ratio of C-9 of oleic acid at α (1,3) and β (2) positions on the glycerol moiety; ^h The ratio of C-10 of oleic acid at α (1,3) and β (2) positions on the glycerol moiety; ⁱ Average of unsaturated carbons of oleic acid (C-9 and C-10) at α (1,3) and β (2) positions on the glycerol moiety. ^j Average of unsaturated carbons of linoleic acid (C-9, C-10, C-12) at the α and β positions of the glycerol moiety. ^k Average of unsaturated carbons of linolenic acid (C-12 and C-13) at the α and β positions of the glycerol moiety. ^l this ratio could be affected by the composition of the oil samples. For variables labeling see Figure S4 in SM.

3.3. Comparison of the long-term repeatability of GC and NMR methods

The long-term repeatabilities of measurements performed by GC, $^1\text{H-NMR}$, $^{13}\text{C-INEPT}$, and one-pulse $^{13}\text{C-NMR}$, expressed as relative standard deviations (RSD), were determined by analyzing the same oil several times over a period of one year (**Table 1**). Moreover, the robust-RSDs, based on the median absolute deviation (MAD) statistics and thus are less affected by outliers than RSDs, were determined. The results show that, in general, comparable precisions were obtained with GC and $^{13}\text{C-INEPT}$ techniques. For INEPT specific signals, i.e., positional and isotopic variables, excellent precisions were reached except for $\text{Ln}\alpha/\text{Ln}\beta$ where RSD is 9.7%. However, this precision is still beneficial for the classification issues since, in the analyzed samples, the ratio $\text{Ln}\alpha/\text{Ln}\beta$ varies from 0.75 to 2.23 with an average value of 1.26 and an RSD of 26.2%.

3.4. Comparison of the classification power of GC and NMR methods

In a previous work based on $^1\text{H-NMR}$ analysis of olive oils, we showed that obvious discriminations are possible according to the colour of olives and in function of the olive field altitude.²⁰ Thus, in order to compare the classification power of the analytical methods used in this study (GC, $^1\text{H-NMR}$, and $^{13}\text{C-INEPT}$), we used these factors (i.e., colour and altitude) as target discrete variables in the multivariate analysis. In this respect, olive oil samples were divided into black (Bk), green (Gn), low (Lw, for altitudes from 0 to 350 m), and high (Hh, for altitudes above 700 m) groups. Therefore, classification models were constructed and their robustness tested according to the procedures described in material and methods (by means of CDA, LDA, and LOO techniques). The corresponding unidimensional scatterplots based on the canonical axes, along with the appropriate statistical indicators (cf. section 2.4.) are reported in **Figures 2** and **3**.

3.4.1. Classification of oils according to the colour of olives

Significant discriminations between oils from black and green olives were observed (**Figure 2**). Small *Wilks' λ* and *Er* values were obtained indicating a very good classification power of the models; values obtained for the parameters $F(df1,df2)$, *p*, and LOO-*Er* confirm the significance and robustness of models. Comparable classifications were noticed while using INEPT and GC methods (**Figure 2-A and B**). On the other hand, less discrimination was obtained using the $^1\text{H-NMR}$ variables (**Figure 2-C**). As a result, GC and INEPT tools were

found to be alternative to each other in separating olive oil samples according to the drupes colour.

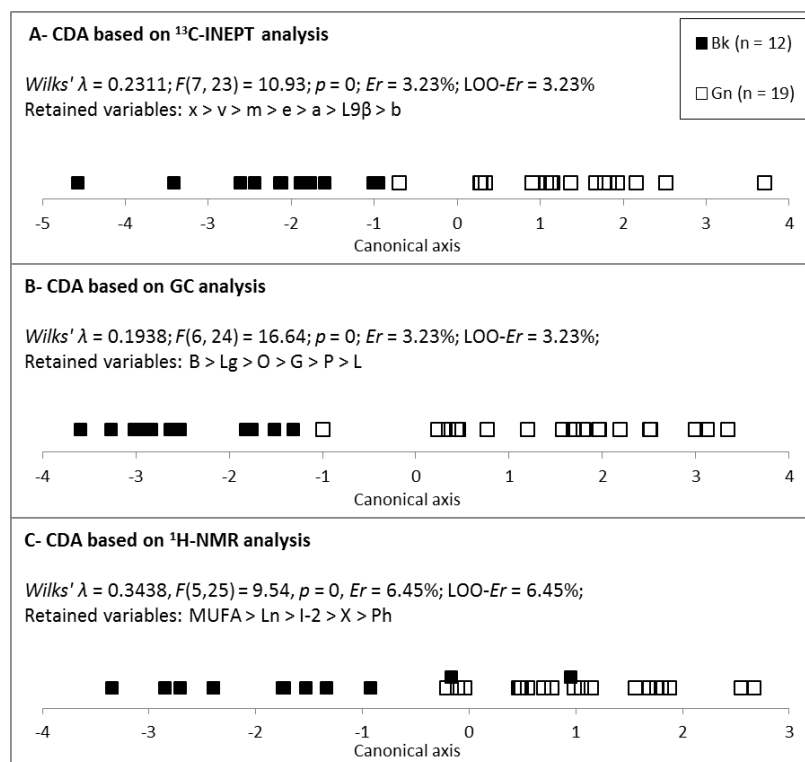


Figure 2: Classification of oils according to the colour of olives (Bk: black olives and Gn: green olives). Retained variables used in the classification are reported in decreasing order of importance based on their contribution to *Wilks' λ*. (see ¹H-NMR and INEPT spectra labelling in Figures S2 and S4 in SM respectively).

3.4.2. Classification of oils according to the altitude of the terroir

As shown in **Figure 3**, INEPT variables provided the best separation of oil samples, as Lw and Hh, with zero percent *Er* and LOO-*Er* (**Figure 3-A** and **B**), the ratio Cω1/Cω2 plays a significant role in discriminating the olive oils from different altitude. However, the *Er* increases to 3.45% and 10.34 % while using variables obtained from GC and ¹H-NMR analyses, respectively (**Figure 3-D** and **E**). Thus, INEPT method has shown a better classification power in separating olive samples according to the altitude of the terroir than GC and ¹H-NMR methods.

Moreover, **Figure 3-C** shows that it is also possible to differentiate olive oil samples from different altitudes using simply the two ratios O10α/O10β and Lnα/Lnβ related to the positional distribution of oleic and linolenic acids on the glycerol backbone. It was found that olives from high altitude (> 700 m) have higher values of O10α/O10β and lower values of Lnα/Lnβ than olives from low altitude (< 350 m). Here it should be noted that oleic and linolenic contents of oils vary significantly with the altitude of the olive field; oleic acid

increases while linolenic acid decreases with elevation, similar results were reported in previous studies.^{20,28} However, the use of oleic and linolenic acids instead of the ratios $O10\alpha/O10\beta$ and $Ln\alpha/Ln\beta$ clearly decreases the classification power of the model.

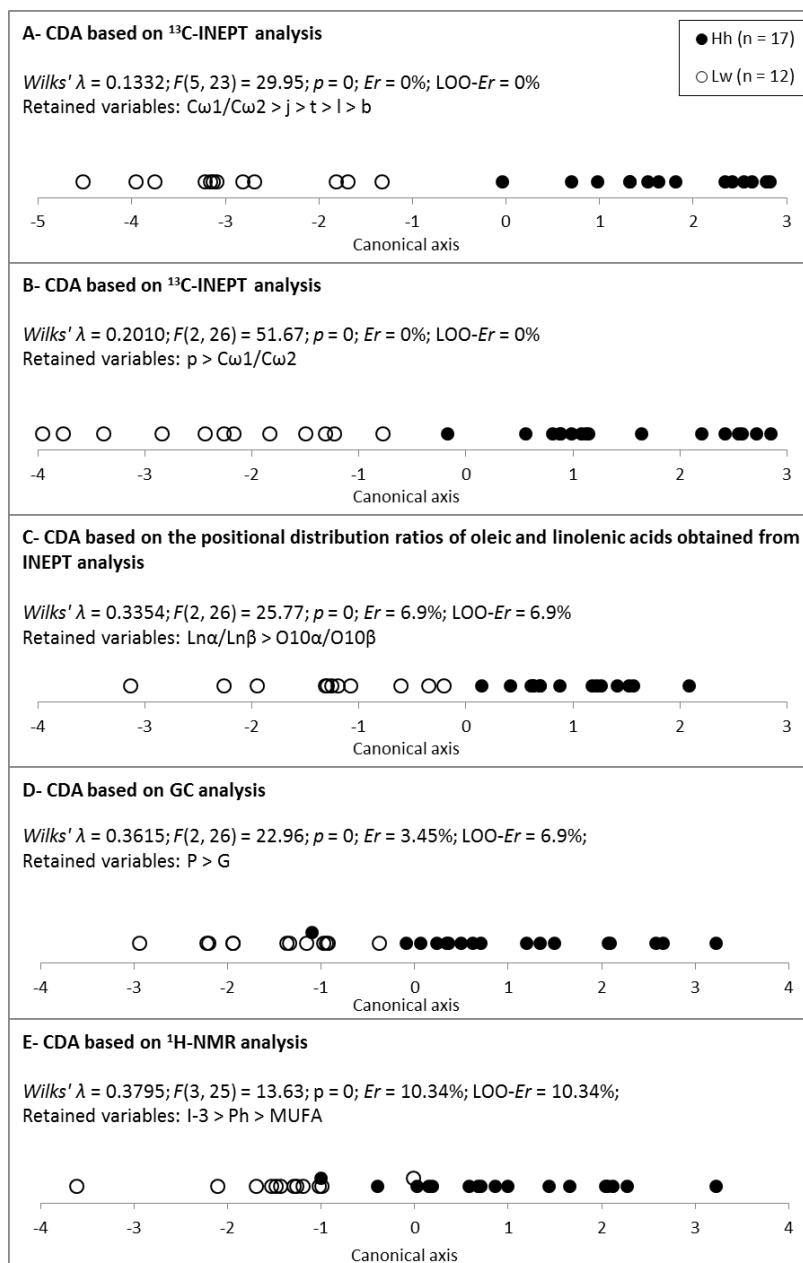


Figure 3: Classification of olive oils according to the altitude of the terroir (Hh: >700 m and Lw: <350 m). Retained variables used in the classification are reported in decreasing order of importance based on their contribution to Wilks' λ . (see INEPT and ^1H -NMR spectra labelling in Figures S4 and S2 in SM).

4. Conclusion

The findings of this study constitute an additional validation of the ^{13}C -INEPT prediction models previously proposed for the quantification of individual fatty acids in the olive oil

matrix. Besides, comparable precisions were obtained by GC and ^{13}C -INEPT techniques. Thus, comparatively to GC, ^{13}C -INEPT advantageously affords, in one single experiment and without prior chemical manipulations, the individual quantification fatty acids, the positional distribution of oleic, linoleic, and linolenic acids on the glycerol backbone, and the relative position-specific isotopic ^{13}C content at different sites of the triacylglycerol molecules. Moreover, ^{13}C -INEPT combined with multivariate statistics shows higher potential for the classification of olive oils according to morphological and geographical factors than ^1H -NMR and GC. Finally, this analytical methodology, validated for olive oils, could be used for the characterization of other triacylglycerol matrices.

Acknowledgements

N.M. acknowledges the financial support of the Lebanese National Council for Scientific Research and the Research Council of Saint-Joseph University. The CORSAIRE platform from Biogenouest is also acknowledged.

References

1. Sacco, A., Brescia, M. A., Liuzzi, V., Reniero, F., Guillou, G., Ghelli, S., & van der Meer, P. (2000). Characterization of Italian olive oils based on analytical and nuclear magnetic resonance determinations. *Journal of the American Oil Chemists' Society*, 77(6), 619-625.
2. Mannina, L., Marini, F., Gobino, M., Sobolev, A. P., & Capitani, D. (2010). NMR and chemometrics in tracing European olive oils: The case study of Ligurian samples. *Talanta*, 80(5), 2141-2148.
3. D'Imperio, M., Mannina, L., Capitani, D., Bidet, O., Rossi, E., Bucarelli, F. M., ... & Segre, A. (2007). NMR and statistical study of olive oils from Lazio: a geographical, ecological and agronomic characterization. *Food chemistry*, 105(3), 1256-1267.
4. Brescia, M. A., Alviti, G., Liuzzi, V., & Sacco, A. (2003). Chemometric classification of olive cultivars based on compositional data of oils. *Journal of the American Oil Chemists' Society*, 80(10), 945-950.
5. Alonso Garc a, M. V., & Aparicio Lopez, R. (1993). Characterization of European virgin olive oils using fatty acids. *Grasas y aceites*, 44(1), 18-24.
6. Eddib, O., & Nickless, G. (1987). Elucidation of olive oil classification by chemometrics. *Analyst*, 112(4), 391-395.
7. Alonso Garc a, M. V., & Aparicio Lopez, R. (1993). Characterization of European virgin olive oils using fatty acids. *Grasas y aceites*, 44(1), 18-24.
8. Ollivier, D., Artaud, J., Pinatel, C., Durbec, J. P., & Gu r re, M. (2003). Triacylglycerol and fatty acid compositions of French virgin olive oils. Characterization by chemometrics. *Journal of agricultural and food chemistry*, 51(19), 5723-5731
9. D'Imperio, M., Dugo, G., Alfa, M., Mannina, L., & Segre, A. L. (2007). Statistical analysis on Sicilian olive oils. *Food Chemistry*, 102(3), 956-965
10. Ollivier, D., Artaud, J., Pinatel, C., Durbec, J. P., & Gu r re, M. (2006). Differentiation of French virgin olive oil RDOs by sensory characteristics, fatty acid and triacylglycerol compositions and chemometrics. *Food Chemistry*, 97(3), 382-393
11. Scano, P., Casu, M., Lai, A., Saba, G., Dessi, M. A., Deiana, M., ... & Bandino, G. (1999). Recognition and quantitation of cis-vaccenic and eicosenoic fatty acids in olive oils by ¹³C nuclear magnetic resonance spectroscopy. *Lipids*, 34(7), 757-759.
12. Merchak, N., Silvestre, V., Loquet, D., Rizk, T., Akoka, S. & Bejjani, J. (2015). A novel quantification strategy of individual fatty acids in olive oil using optimized ¹³C-NMR with a polarization transfer technique. Submitted to *Analyst* "not published yet".
13. Vlahov, G., Shaw, A. D., & Kell, D. B. (1999). Use of ¹³C nuclear magnetic resonance distortionless enhancement by polarization transfer pulse sequence and multivariate analysis to discriminate olive oil cultivars. *Journal of the American Oil Chemists' Society*, 76(10), 1223-1231.
14. Vlahov, G. (2006). ¹³C nuclear magnetic resonance spectroscopy to determine olive oil grades. *Analytica chimica acta*, 577(2), 281-287.
15. Zamora, R., Alba, V., & Hidalgo, F. J. (2001). Use of high-resolution ¹³C nuclear magnetic resonance spectroscopy for the screening of virgin olive oils. *Journal of the American Oil Chemists' Society*, 78(1), 89-94.
16. Merchak, N., Bejjani, J., Rizk, T., Silvestre, V., REMAUD, G., & Akoka, S. (2015). ¹³C isotopomics of triacylglycerols using NMR with polarization transfer techniques. *Analytical Methods*.

17. Thibaudeau, C., Remaud, G., Silvestre, V., & Akoka, S. (2010). Performance evaluation of quantitative adiabatic ^{13}C NMR pulse sequences for site-specific isotopic measurements. *Analytical chemistry*, 82(13), 5582-5590.
18. Bussy, U., Thibaudeau, C., Thomas, F., Desmurs, J. R., Jamin, E., Remaud, G. S., ... & Akoka, S. (2011). Isotopic finger-printing of active pharmaceutical ingredients by ^{13}C NMR and polarization transfer techniques as a tool to fight against counterfeiting. *Talanta*, 85(4), 1909-1914.
19. Faulh, C., Reniero, F., & Guillou, C. (2000). ^1H NMR as a tool for the analysis of mixtures of virgin olive oil with oils of different botanical origin. *Magnetic Resonance in Chemistry*, 38(6), 436-443.
20. Merchak, N., El Bacha, E., Bou Khouzam, R., Rizk, T., Akoka, S., & Bejjani, J., (2015). Geoclimatic, morphological, and temporal effects on Lebanese olive oils composition and classification: A ^1H NMR metabolomic study. Submitted to *Food Chemistry* "not published yet".
21. Tenailleau, E., & Akoka, S. (2007). Adiabatic ^1H decoupling scheme for very accurate intensity measurements in ^{13}C NMR. *Journal of Magnetic Resonance*, 185(1), 50-58.
22. European Economic Community (1977). European Community regulation no. 72/77. Amending Regulation (EEC) No 1470/68 on the drawing and reduction of samples and the determination of oil content, impurities and moisture in oil seeds.
23. International Olive Oil Council (2001). Preparation of the fatty acid methyl esters from olive oil and olive-pomace oil.
<http://www.internationaloliveoil.org/estaticos/view/222-standards>.
(Accessed 02.06.14).
24. R. Rakotomalala, In: Proceedings of EGC'2005, RNTI-E-3, vol. 2, 697-702.
25. Roy, P. P., & Roy, K. (2008). On some aspects of variable selection for partial least squares regression models. *QSAR & Combinatorial Science*, 27(3), 302-313.
26. Boskou, D.; Blekas, G.; Tsimidou, M. In *Olive oil : Chemistry and Technology*, 2nd ed.; Boskou, D.; AOCS Publishing, 2006; pp 41-72. Ollivier D., Pinatel C., Artaud J., *Le Nouvel Olivier*, N°44 mars-avril 2005.
27. Hidalgo, F. J., & Zamora, R. (2003). Edible oil analysis by high-resolution nuclear magnetic resonance spectroscopy: recent advances and future perspectives. *Trends in Food Science & Technology*, 14(12), 499-506.
28. Aguilera, M. P., Beltrán, G., Ortega, D., Fernández, A., Jiménez, A., & Uceda, M. (2005). Characterisation of virgin olive oil of Italian olive cultivars: Frantoio' and Leccino', grown in Andalusia. *Food chemistry*, 89(3), 387-391.

V-3. Données complémentaires

Table S1 : Olive oils variables obtained by ¹H NMR spectroscopy.

Variable ^a (abbreviation)	δ ppm	S _{Rw} ^b %	σ^c %	Mean	Range	
					Min	Max
Unknown compound (X)	0.512-0.528	11.0	62.9	0.12	0.02	0.63
β -Sitosterol (Ss)	0.655-0.672	7.5	27.1	0.74	0.34	1.57
Linolenic acid CH ₃ (Ln) ^d	0.929-0.979	3.5	24.1	0.52	0.32	1.17
CH ₂ intensity (I-1) ^e	1.238	0.8	8.8	4.37	3.13	5.71
CH ₂ intensity (I-2) ^e	1.249	0.5	3.1	3.32	2.97	3.56
CH ₂ intensity (I-3) ^e	1.283	0.5	2.6	3.49	3.17	3.74
Squalene (Sq)	1.640-1.666	1.7	25.5	8.0	2.5	14.6
Allylic CH ₂ ^f	1.891-2.099	0.3	2.8	1022	925	1096
CH ₂ position 2 (α) ^g	2.182-2.409					
Diallylic CH ₂ ^h	2.653-2.866	0.5	22.0	74.0	34.0	128.1
1,2-Diglycerides (1,2-dig)	3.678-3.718	3.2	24.5	4.17	2.26	7.97
Terpenes (Trp-1)	4.563-4.579	18.3	54.4	0.07	0.00	0.29
Terpenes (Trp-2) ⁱ	4.6412; 4.6919				Absent ⁱ	High ⁱ
Phenolic compounds (Ph)	6.740-6.764	12.6	68.2	0.12	0.00	0.71
Linoleic acid (L) ^j		0.6	23.1	11.30	4.83	20.16
Monounsaturated fatty acids (MUFA) ^k		0.4	5.7	71.1	58.5	81.2
Saturated fatty acids (SFA) ^l		2.1	13.5	17.0	10.1	25.0

^a All variables are determined relative to the signal of methylene at position 2 of fatty acids and expressed as peak areas, unless specified otherwise. ^b Relative standard deviation of measurement corresponding to within-laboratory reproducibility (based on control charts over a one-year period). ^c Relative standard deviation of the population. ^d Expressed in molar percentage (Table S1 in SM). ^e Intensity of the signal at the mentioned chemical shift relative to the highest peak (2.285 ppm) of methylene signal at position 2 (Figure S2-B in SM). ^f Only used to calculate molar percentages of monounsaturated and saturated fatty acids. ^g Only used as a reference signal for peak areas and representing the total number of moles of fatty acids. ^h Only used to calculate molar percentages of monounsaturated and linoleic fatty acids. ⁱ Semiquantitated relative to the ¹³C satellite signal of the glycerol moiety (4.435-4.489 ppm) and reported as: absent, faint, moderate or high (Figure 2). ^j Expressed in molar percentage; calculated from diallylic CH₂ and linolenic acid CH₃ signals (Table S1 in SM). ^k Expressed in molar percentage; calculated from allylic CH₂, diallylic CH₃, and squalene (1.653 ppm, 6H) signals; the last one is used in order to correct the fatty acids allylic CH₂ signal which overlaps with that of the 20 allylic protons of squalene (Table S1 in SM). ^l Expressed in molar percentage; calculated from allylic CH₂ and squalene (1.653 ppm) signals (cf. the precedent note and Table S1 in SM).

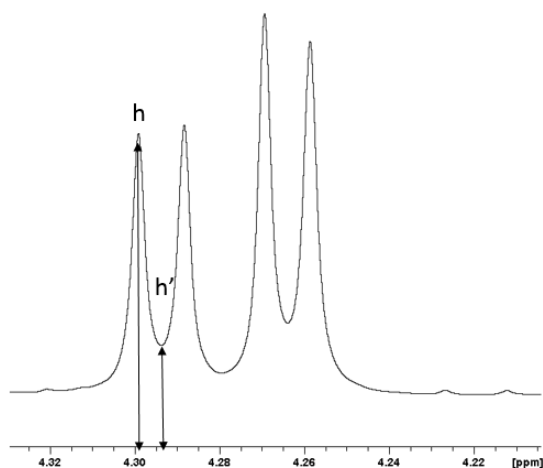


Figure S1: Intensities used to determine the spectral resolution ($100 \cdot h'/h$).

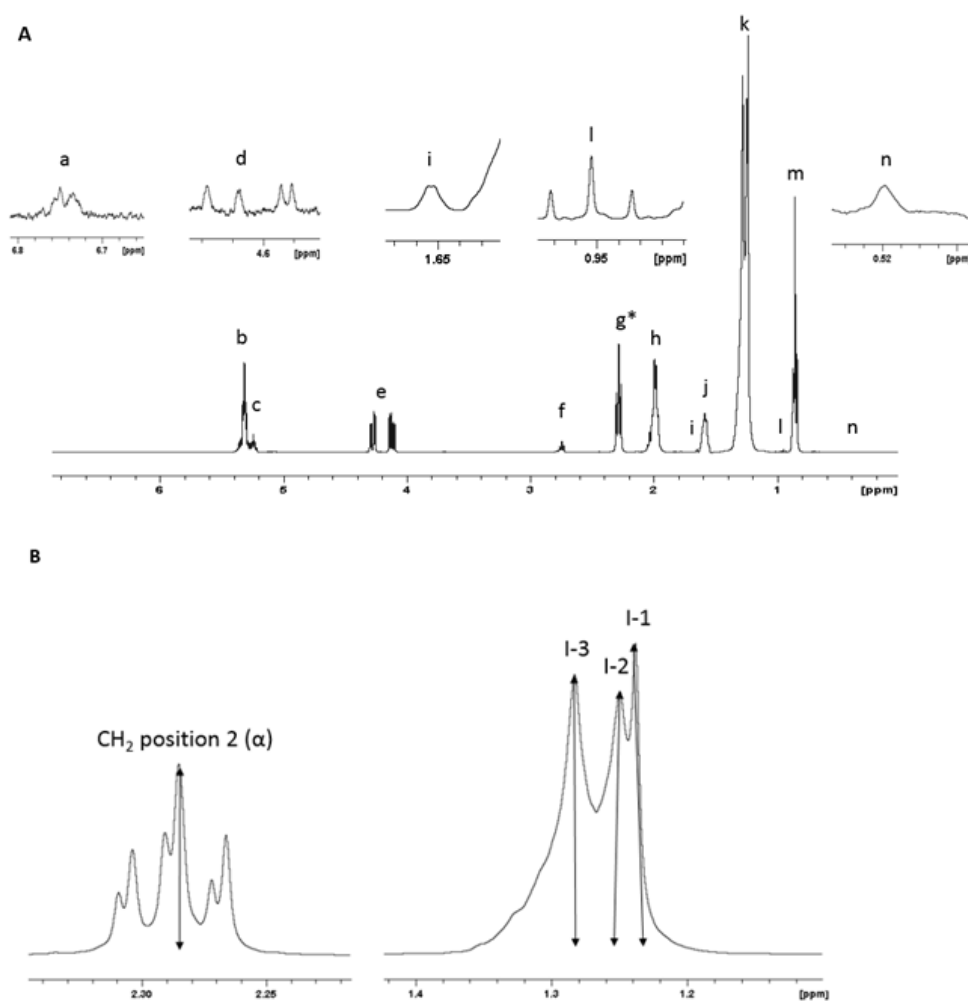


Figure S2: A) ^1H NMR spectrum of olive oil: phenolic compounds (a); olefinic protons (b); sn-2 glycerol protons (c); terpenes (d); sn-1,3 glycerol protons (e); diallylic protons (f); CH_2 at position 2 (g^*); allylic protons (h); squalene (i); CH_2 at position 3 (j); aliphatic CH_2 protons (k); methylenic protons of linolenic acid (l); methylenic protons of fatty acids chains (m); unknown compound (n). B) Peaks used to calculate intensities I-1, I-2 and I-3

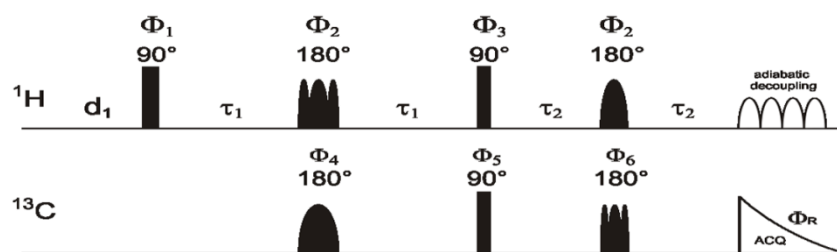


Figure S3: Adiabatic refocused INEPT sequence with ^1H and ^{13}C 180° adiabatic composite refocusing pulses and adiabatic full passage inversion pulses. The duration τ_1 was adjusted to 2.704 ms, and the refocusing period τ_2 was adjusted to 1.409 ms.

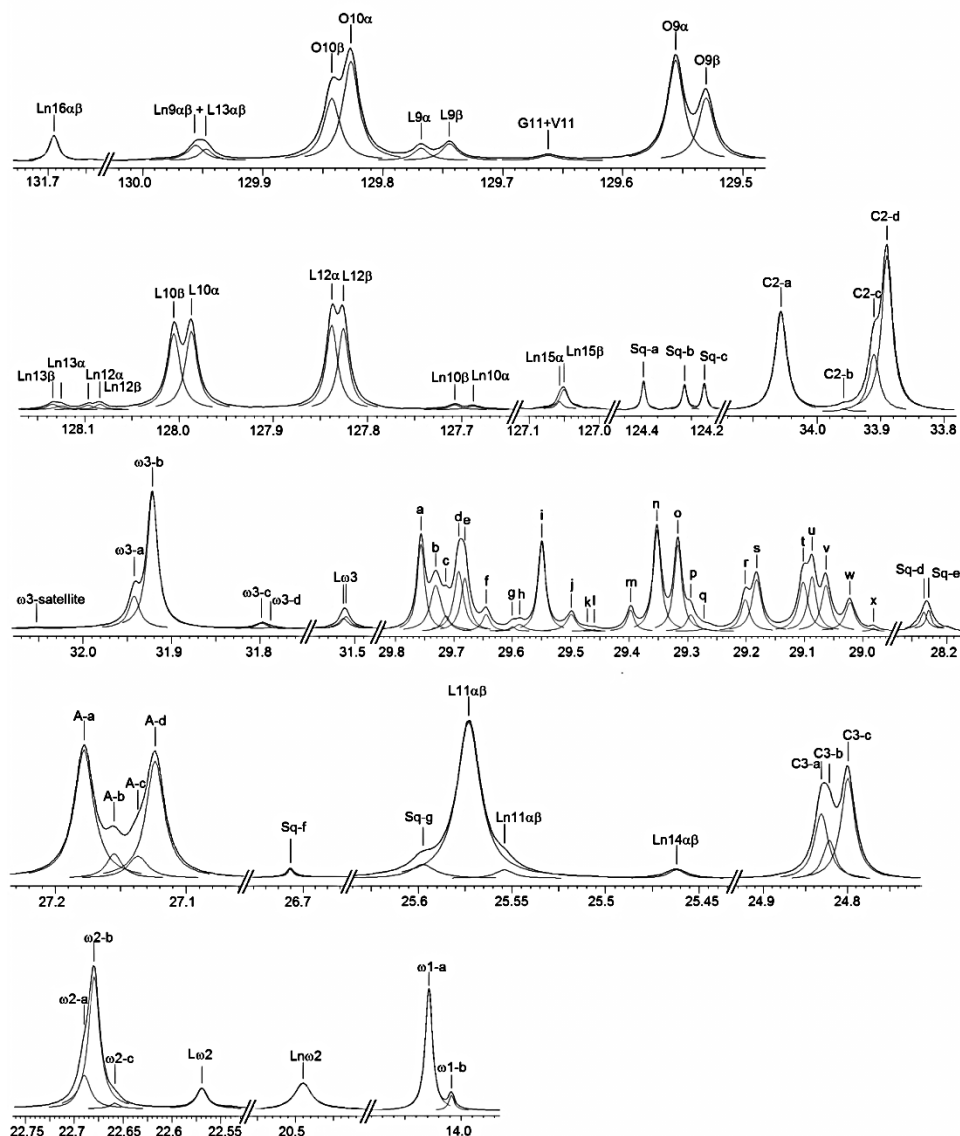


Figure S4: Deconvoluted lines of olive oil ^{13}C -INEPT spectra and signals labelling used in the multiple linear regression. Signals labelling: linolenic (Ln); linoleic (L); oleic (O); gondoic (G); vaccenic (V); squalene (Sq); glycerol sn-1,3 position (α); glycerol sn-2 position (β); signals of the aliphatic carbons region ranging from 29.82 to 28.85 ppm are labeled from a to x; carbons 2 and 3 of fatty acids chains (C2-a to d) and (C3-a to c), respectively; allylic carbons (A-a to d); starting from the methyl group of the fatty acids chains, carbons 1, 2, and 3 are denoted (ω 1-a and b), (ω 2-a to c), and (ω 3-a to d), respectively.

V-4. Conclusion

Les résultats obtenus dans cette étude montrent que la stratégie proposée précédemment (partie IV), pour la quantification individuelle des acides gras dans les matrices riches en triacylglycérols, notamment les huiles d'olive, a été vérifiée par application des modèles déjà construits sur une nouvelle série d'échantillons d'huile d'olive.

De très fortes corrélations entre les pourcentages des acides gras prédits et ceux obtenus par la GC ont été de nouveau observées dans le cas des acides palmitique, stéarique, arachidique, palmitoléique, oléique, vaccénique, linoléique et linoléinique, ainsi que pour le squalène. Par contre, les prédictions des acides béhénique et margarique sont moins robustes, cela est vraisemblablement dû à une déviation au niveau de l'étalonnage de la GC. La prédiction des acides gondoïque et lignocérique est aussi faible, ces acides gras sont présents sous forme de traces dans l'huile d'olive (< 0,4% et < 0,2%, respectivement). D'autre part, les répétabilités à long terme des mesures réalisées par INEPT-¹³C et GC sont comparables.

De plus, la méthode INEPT-¹³C permet d'obtenir dans un temps d'expérience raisonnable des informations isotopiques et positionnelles supplémentaires par rapport à la GC et la RMN ¹H.

Le traitement statistique, des données obtenues a montré que les deux techniques INEPT et GC sont plus avantageuses que la RMN ¹H pour la classification des huiles en fonction de la couleur des olives. Tandis que l'INEPT a fourni la meilleure séparation des huiles en fonction de l'altitude du terroir d'origine.

Finalement, la technique l'INEPT adiabatique montre un potentiel considérable pour la caractérisation des huiles d'olive et leur classification en fonction de leurs origines.

De plus, la stratégie basée sur l'INEPT-¹³C et la régression linéaire multivariée qui a été testée sur l'huile d'olive pourrait être adaptée à d'autres matrices riches en TAG afin de déterminer le profil complet de leurs acides gras.

***Partie VI - Analyse isotopomique
rapide et précise en utilisant une
séquence ^1H - ^{13}C HSQC : Application
sur les triacylglycérols***

VI-1. Introduction

Les techniques d'analyse par RMN 1D ne permettent pas de résoudre complètement le problème de chevauchement des signaux. Cette limitation peut cependant être surmontée en utilisant la RMN 2D. Toutefois, les expériences 2D sont longues et donc plus sensibles aux instabilités du spectromètre ; leur utilisation pour l'analyse quantitative reste donc limitée.

Des travaux antérieurs, réalisés au sein du laboratoire CEISAM, ont montré que la symétrie intrinsèque de la séquence HSQC permet de compenser les imperfections dues aux inhomogénéités du champ et à l'effet d'« off-resonance ». Ainsi, une séquence HSQC modifiée a permis d'atteindre une bonne répétabilité, mais des améliorations supplémentaires doivent être apportées afin que cette méthode puisse répondre aux exigences de l'analyse isotopique.

Dans le cadre de cette partie, nous avons développé une nouvelle version de la séquence HSQC, en travaillant sur l'organisation des impulsions RF pour la rendre parfaitement symétrique, et cela afin de répondre aux exigences de l'analyse isotopique.

D'autre part, nous avons utilisé deux approches, le repliement spectral et l'échantillonnage non-uniforme, pour réduire le temps d'expérience tout en obtenant une précision élevée (de quelques pour mille). Cette précision a été évaluée par l'estimation de la répétabilité des mesures à long terme tout en tenant compte du nombre de spectres enregistré, le but étant de déterminer le nombre de spectres minimal qui préserve la précision nécessaire pour l'analyse isotopique.

En outre, le potentiel de cette méthode en tant qu'outil d'authentification des matrices riches en triacylglycérols a été testé. Des échantillons d'huiles végétales commerciales de différentes origines botaniques (arachide, noix, noisette, tournesol, colza, olive) et géographiques (Tunisie, Espagne, Liban, Crète) ont été analysés selon la procédure développée afin de les classer en fonction de leurs origines. L'analyse discriminante linéaire (LDA) et l'analyse canonique discriminante (CDA) ont été utilisées pour le traitement des données obtenues.

Les résultats rapportés dans cette étude ont fait l'objet d'un article intitulé « Precise and rapid isotopic analysis by ^1H - ^{13}C HSQC: Application to triacylglycerol matrices », soumis à la revue Talanta.

VI-2. Article

Precise and rapid isotopomic analysis by ^1H - ^{13}C HSQC: Application to triacylglycerol matrices

Noelle Merchak ^{a, b}, Virginie Silvestre ^a, Laetitia Rouger ^a, Patrick Giraudeau ^{a,c}, Toufic Rizk ^b, Joseph Bejjani ^{b, *}, Serge Akoka ^a

^a EBSI team, Interdisciplinary Chemistry: Synthesis, Analysis, Modelling (CEISAM), University of Nantes-CNRS UMR 6230, 2 rue de la Houssinière, BP 92208, F-44322 Nantes cedex 3, France

^b Research Unit: Technologies et Valorisation Alimentaire (TVA), Laboratory of Metrology and Isotopic Fractionation, Faculty of Science, Saint-Joseph University, P.O. Box 11-514 Riad el Solh, Beirut 1107 2050.

^c Institut Universitaire de France, 1 rue Descartes, 75005 Paris, France

Abstract

An optimized 2D HSQC sequence has been tested and applied to triacylglycerol matrices to determine their isotopic and metabolomic profiles. Spectral aliasing and non-uniform sampling approaches were used to decrease the experimental duration and to improve the resolution, respectively. Long-term repeatability was evaluated and different commercial vegetable oils were analyzed by this methodology. The results show that this method can be used to classify oil samples according to their geographical and botanical origins.

Keywords

HSQC; Isotopic profile; Metabolomic profile; Triacylglycerol; Fast 2D NMR

1. Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy, combined with statistical analysis, shows considerable potential in many fields including metabolomics [1-3] and products authentication [4-6]. ^1H -NMR allows quantitative compositional analysis of complex mixtures. ^{13}C -NMR offers an interesting alternative as it permits a much better discrimination of resonances and is the only technique providing site-specific isotopic information [7]. However, a high precision (a few per mil) is required for isotopic analysis at natural

abundance since the deviations of ^{13}C contents in natural or synthetic compounds is very limited (e.g., $\Delta\delta^{13}\text{C}$ between C1, C2, and C3 of glycerol are in the range from 1.6 to 40.9 per mil) [8,9]. This limitation was recently encompassed by using multipulse ^{13}C NMR relying on polarization transfer methods such as INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) in order to enhance spectral sensitivity and reduce experiment time [10,11]. A high precision was reached after a careful optimization of the INEPT sequence by incorporating adiabatic pulses. This approach was integrated in an isotopomic workflow to simultaneously determine the position-specific ^{13}C content and the metabolomic profile of triacylglycerols [12]. However, peak overlaps are still observed and reduce the precision of the isotopic content determination. It is therefore necessary to extend this methodology to a higher dimensionality in order to overcome this limitation [13,14].

Numerous studies have reported the application of two dimensional (2D) NMR in various fields of metabolomic studies [15-18]. An HSQC (Heteronuclear Single Quantum Coherence spectroscopy) pulse sequence was previously optimized aiming to achieve, in a reasonable time, a high precision (of per mil) required for ^{13}C isotopic analysis. A high precision, was obtained although the pulse sequence used did not include adiabatic pulses. This can be probably explained by the intrinsic symmetry of this sequence, which efficiently compensates several pulse imperfections. However, further precision improvement is needed for isotopic applications [19].

In the present work, we propose an HSQC pulse sequence with perfect symmetry in order to improve measurements precision in the complex triacylglycerol matrix. Non-uniform sampling (NUS) and aliasing approaches were used to reduce the experiment duration. Moreover, the minimum number of replicate spectra necessary to obtain the required precision for isotopic analysis was determined by considering the long-term repeatability of the method. At the end of this work, different commercial vegetable oils were analyzed using this methodology. The 2D heteronuclear (^1H - ^{13}C) NMR data thus obtained were treated by means of linear and canonical discriminant analyses in order to classify the oil samples according to their geographical and botanical origins.

2. Materials and methods

2.1. Oil samples

Thirty-two commercial oil samples from 6 different botanical origins were used in this study. Oil samples were as follows (the number of samples analyzed is reported in brackets): Peanut

(3), colza (4), hazelnut (3), nut (3), sunflower (4), and olive (15). Olive oils were purchased from different origins: Tunisia (3), Crete (5), Spain (4) and Lebanon (3).

2.2. Sample preparation

For quantitative analysis, a concentration of 25% (v/v) of oil (0.21 mL) in chloroform-d (0.63 mL, purchased from Eurisotop) was used in order to obtain a high resolution. The solution was carefully filtered into a 5 mm NMR tube.

2.3. NMR acquisition parameters

NMR experiments were performed on a Bruker Avance III HD 700 spectrometer equipped with an inverse $^1\text{H}/^{13}\text{C}/^{15}\text{N}/^2\text{H}$ cryogenically cooled probe carefully tuned to the recording frequency of 176.09 MHz for the ^{13}C and to 700.36 MHz for the ^1H nuclei. The temperature of the probe was set at 303 K. For each experiment, pulse widths were calibrated to 10.3 μs for the 90° ^1H and 10.4 μs for the 90° ^{13}C . An automatic shimming was performed to preserve the B_0 field homogeneity. The HSQC spectra were recorded using the pulse sequence presented in **Figure 1** using the following acquisition parameters: acquisition time of 0.5 s, recovery delay of 17.5 s, 64 acquired points in the indirect dimension (^{13}C), 2 scans and 8 dummy scans, with an aliased ^1H spectral width of 7.5 ppm and ^{13}C spectral width of 24.5 ppm [20]. The ^1H and ^{13}C offsets were set at 2.75 and 29.75 ppm, respectively. The ^{13}C decoupling was performed with a garp-4 composite decoupling pulse sequence (pw90 = 65.5 μs at 0.596 W). The INEPT delay Δ was optimized to $\frac{1}{4} \cdot ^1J_{\text{CH}}$, with $^1J_{\text{CH}} = 140$ Hz (the average of the ^1H - ^{13}C one bond coupling constants). Purge gradients were set to $g1 = 80\%$ and $g3 = 40\%$ of the maximum gradient strength (50 G/cm), while coherence order selection was achieved by $g2 = 80\%$ and $g4 = 20.1\%$ of the maximum gradient strength, the duration of all gradient pulses was 1 ms followed by a recovery delay of 300 μs . A NUS scheme was applied with a compression rate of 50% and an exponentially weighted sparse sampling scheme ($T_2 = 0.025$ s), resulting in an experiment time divided by two compared to the classical experimentation. For each sample, 5 spectra were recorded (22 min per spectrum).

2.4. NMR processing and integration parameters

The HSQC spectra were processed with Bruker Topspin version 3.1. Before Fourier Transform, the data were weighted with a TRAP function ($TM_1 = 0.01$ and $TM_2 = 0.5$) and a lorentzian-gaussian apodization ($GB = 0.001$, $LB = -0.01$ Hz) in the F_1 and F_2 dimensions, respectively. The data matrices were zero-filled to 8192 points in F_2 and 512 points in F_1 . All the spectra were processed in magnitude mode, because phasing 2D spectra with an adequate

repeatability is time-consuming and it is impossible to reach the precision needed for isotopic analysis. Moreover, an automatic polynomial baseline correction ($n = 3$) was achieved. Non-uniform sampling spectra were reconstructed with the compressed sensing (CS) algorithm available with the Bruker software Topspin 3.1. All the spectra were calibrated by setting the peak with the highest chemical shift corresponding to the glycerol moiety at 4.29 and 21.88 ppm in the F_2 and F_1 dimensions, respectively. The peak volumes were obtained by using the integration routine in the Bruker software and the integration box widths were carefully adjusted in such a way to obtain same level of truncation for all peaks and they were used for each series of analyzed samples. Peaks volumes were normalized by the sum of all the integrated peaks.

2.5. Statistical analysis

A chemometrics analysis of the resulting data was carried out by means of TANAGRA software. Canonical discriminant analysis (CDA) was carried out while eliminating irrelevant input variables using the backward stepwise discriminant analysis and based on the contribution of each variable to *wilks' λ* criterion. The linear discriminant analysis (LDA) was used as classifier and the performance of the classification model was assessed using *wilks' λ* and the classification error rate as indicators. The overall significance of models was evaluated based on Rao's $F(df1,df2)$ and p -values. The robustness of the selected models was assessed by means of the leave-one-out cross-validation test.

3. Results and discussion

3.1. HSQC pulse sequence optimization

The HSQC pulse sequence optimized in this study is presented in **Figure 1** (See the classical HSQC sequence in **Figure S1 in SI**). It is the standard HSQC sequence in which RF pulses were added in a way to have a perfect symmetry in terms of RF scheme and coherence pathway selection by gradient pulses. The expected result of this high symmetry is an improved repeatability through an efficient compensation of pulse imperfections arising from RF field inhomogeneities and off-resonance effects.

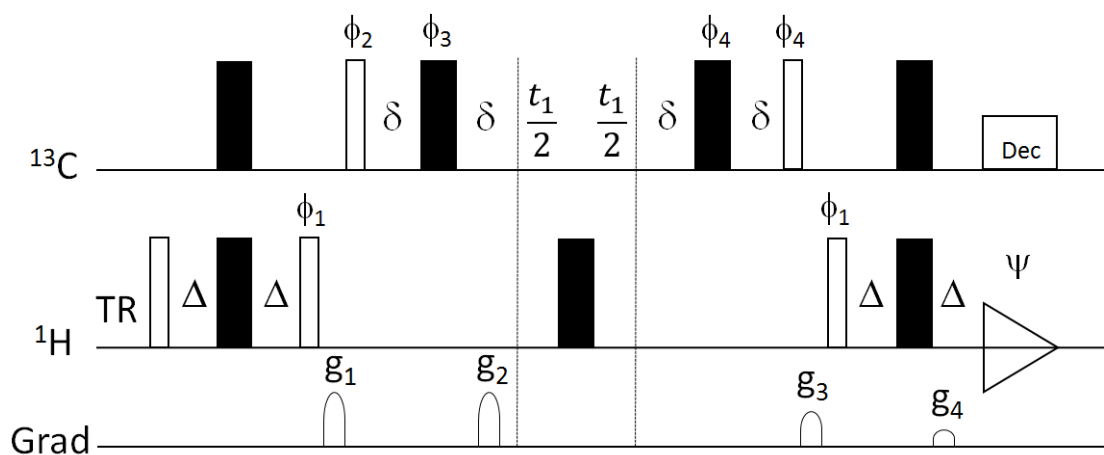


Figure 1: Symmetrized HSQC pulse sequence. Narrow (open) and wide (filled) rectangles represent 90° and 180° pulses, respectively. Pulse phases are x unless indicated otherwise. Phase cycling is given by $\phi_1 = y$; $\phi_2 = y, y$; $\phi_3 = x$; $\phi_4 = x, -x$ and $\Psi = x, -x$. Quadrature detection is obtained in F_1 dimension by incrementing ϕ_2 and ϕ_3 , according to States-TPPI method. TR is the recovery delay (17.5 s), Δ was set to $^1J_{\text{CH}}/4$, with $^1J_{\text{CH}}$ corresponding to the average of the ^1H - ^{13}C one-bond coupling constants (140 Hz), δ was set to 1.3 ms (minimum value needed to include a gradient pulse and its recovery delay).

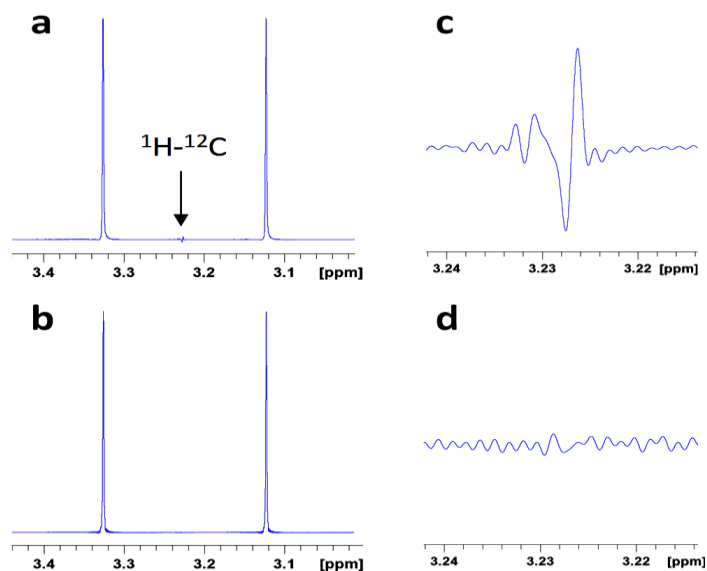


Figure 2: 1D HSQC spectra corresponding to the first t_1 increment of the conventional (a,c) and symmetrized (b,d) HSQC pulse sequence without ^{13}C decoupling, recorded on a methanol/ D_2O 1:10 sample, showing how the symmetrized pulse sequence improves the suppression of the ^1H - ^{12}C signal.

Another potential source of irreproducibility is the residual signal arising from ^1H bound to ^{12}C . While this signal should be suppressed by gradient-based coherence selection, an incomplete and non-repeatable suppression may lead to a loss of repeatability. This effect was studied through preliminary investigations on a model methanol sample. **Figure 2** shows non-

decoupled 1D HSQC spectra (corresponding to the first increment of the 2D experiment) obtained with a standard versus symmetrized experiment. The results demonstrate that the symmetrized pulse sequence completely eliminates the residual ^1H - ^{12}C signal, contrary to the standard experiment.

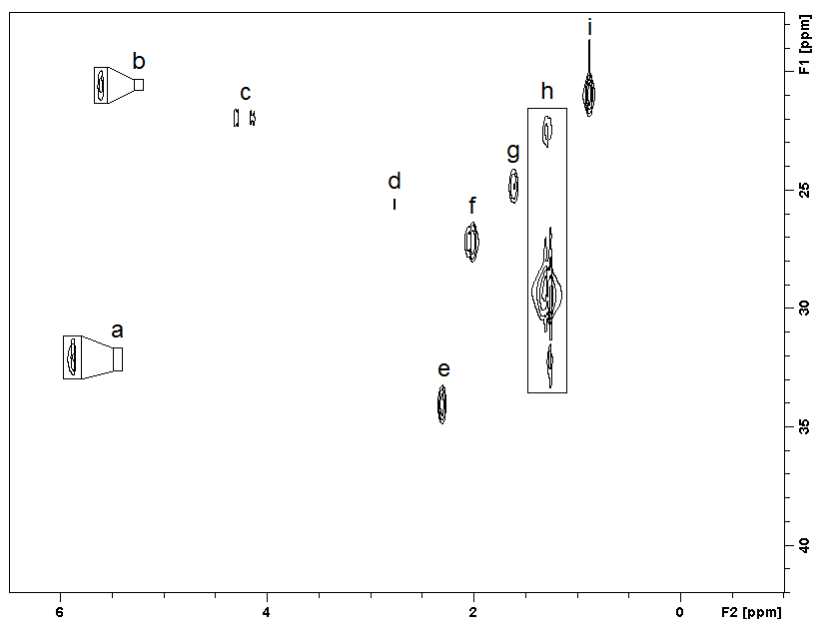


Figure 3: HSQC spectrum of olive oil recorded at 700 MHz with a cryoprobe in 22 min, using a spectral aliasing strategy which consists on selecting a smaller spectral width (24.5 ppm) in the F_1 dimension (^{13}C) and with 50% NUS and $N_I = 64 t_I$. The cross-signals are the following: a- olefinic, b- C2 of glycerol, c- C1,3 of glycerol, d- diallylic, e- alpha (C2 of fatty acids), f- allylic, g- beta (C3 of fatty acids), h- aliphatic, i- methyl. A zoom on the a and b regions was performed.

The 2D NMR experiments are limited by the long acquisition time and they are more likely to be affected by spectrometer instabilities over time [21]. Our aim is to reduce the experiment duration and reach in a reasonable time a precision of a few per mil required for isotopic analysis. In 2D NMR, the experiment time is given by $Ta = NS \cdot N_I \cdot TR$, where NS is the number of scans, N_I the number of t_I increments and TR the repetition time between two experiments [22,23]. Here, NS was set to a minimum value of 2 to preserve the coherence-selection scheme. N_I was reduced to 64 to decrease the experiment duration as much as possible while preserving a sufficient resolution to avoid peak overlap. This low value of N_I was made possible by the use of spectral aliasing which consists on selecting a smaller spectral width (24.5 ppm) in the F_1 dimension (^{13}C) [19]. Finally, the recovery delay between two experiments (TR) was set to a value of 17.5 s permitting the complete ^1H longitudinal relaxation in the case of the triacylglycerol matrix to preserve high precision for the quantitative and isotopic analyses. Note that as a consequence of aliasing, chemical shifts

along F_1 are no longer correct but it is not a problem in the case of a well-known matrix. Methods are available to recover true chemical shifts from aliased HSQC spectra [24,25]. The (^1H - ^{13}C) 2D HSQC spectrum of olive oil with the assignment of the different cross-signals is shown in **Figure 3**. Furthermore, NUS was applied to increase the resolution of the 2D HSQC spectra of triacylglycerol matrices with no increase in conventional experimental time by acquiring only a part of the time increments in the indirect dimension [26-29]. In this respect, NUS rate of 25 and 50% were tested and HSQC spectra were recorded in the same experiment time (22 min) as an HSQC spectrum recorded using 32 t_1 increments and regular sampling. The spectrum recorded with 25% NUS and $N_1 = 128$ t_1 shows better resolution but more artifact than the one acquired with 50% NUS and $N_1 = 64$ t_1 . Therefore, the NUS rate of 50% was used in this study to achieve a good resolution with fewer artifacts (The corresponding spectrum is shown in **Figure S2** in SI).

3.2. Long-term repeatability

Table 1: Long-term repeatabilities of HSQC depending on the number of spectra recorded.

Number of recorded spectra Experimental time (min)	Relative Standard Deviation RSD (%) ^a		
	5	3	1
Olefinic	0.51	0.57	0.60
C2 of glycerol	0.17	0.18	0.18
C1,3 of glycerol	0.25	0.25	0.24
Di-allylic	0.31	0.33	0.36
C2 of fatty acids	0.22	0.23	0.24
Allylic	0.11	0.12	0.09
C3 of fatty acids	0.25	0.24	0.26
Aliphatic	0.11	0.10	0.11
Methyl	0.46	0.44	0.44

^aRSD (%) was calculated from ten analyses of the same oil over a period of three months.

In this study, we aimed to determine the optimal acquisition and processing conditions yielding an excellent precision required for isotopic analysis in the shortest time. In this respect, the long-term repeatability of this technique was determined by analyzing the same oil 10 times over a period of three months, five spectra were recorded each time. The corresponding Relative Standard Deviations (RSD) were determined as a function of the number of spectra recorded each time and are reported in **Table 1**. We observed that increasing the number of recorded spectra does not significantly improve the precision. As a result, one spectrum recorded in only 22 min is sufficient and allows to obtain a high precision required for isotopic analysis (**Table 1**).

3.3. Authentication applications

The optimized HSQC pulse sequence was used to classify commercial oils from different origins (cf. Materials and methods section). The classification results are presented in **Figure 4**. Fifteen olive oil samples from different countries were analyzed using the 2D HSQC method, and a significant separation with an error rate of 0% was observed. The retained discriminating variables are reported in **Figure 4-a** in decreasing order of importance based on Wilks' λ parameter. As it is shown, the olive oil samples from Tunisia and those from Lebanon were considerably different and well distinguished from the other countries on the score plots (**Figure 4-a**). On the other hand, six varieties of oils (Peanut, colza, hazelnut, nut, sunflower and olive) were analyzed as well, and a significant separation of these samples with an error rate of 0% was obtained using the retained variables reported in **Figure 4-b**. Moreover, Despite the olive oil samples used in this classification are from different geographical origin, they still very well grouped and separated from the oils of other varieties.

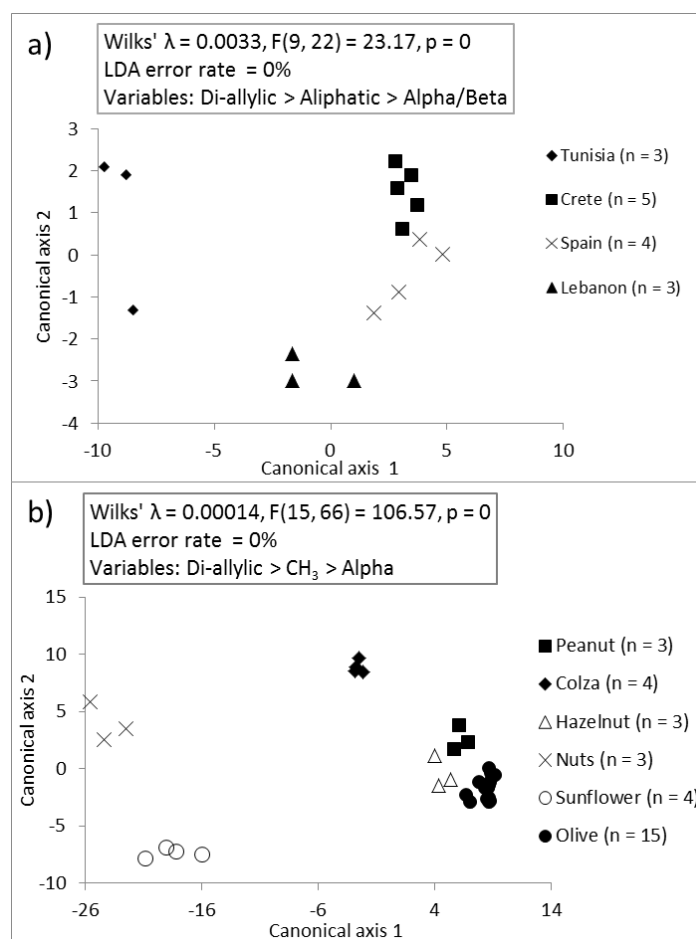


Figure 4: Classification of vegetable oils using 2D (¹H-¹³C) HSQC sequence. a) According to their geographical origin. b) According to their botanical origin.

4. Conclusion

An excellent precision (few per mil) required for isotopic analysis was attained using an optimized ^1H - ^{13}C HSQC sequence with an experiment time of 22 min only. The high precision achieved is due to the symmetry of the HSQC sequence which permitted to compensate pulse imperfections and completely eliminate of signals coming from protons attached to ^{12}C . Moreover, the combination of spectral aliasing and non-uniform sampling techniques has allowed to decrease the experimental duration without losing in terms of resolution and precision. These findings could be useful to verify the authenticity of food and pharmaceutical products in a reasonable time.

Acknowledgements

N.M. acknowledges the financial support of the Lebanese National Council for Scientific Research and the Research Council of Saint-Joseph University. The CORSAIRE platform from Biogenouest is also acknowledged.

References

- [1] J.C. Lindon, J.K. Nicholson, E. Holmes, J.R. Everett, *Concepts Magn. Reson.* 12 (2000) 289-320.
- [2] D.S. Wishart, *TrAC, Trends Anal. Chem.* 27 (2008) 228-237.
- [3] S. Zhang, G.A. Nagana Gowda, V. Asiago, N. Shanaiah, C. Barbas, D. Raftery, *Anal. Biochem.* 383 (2008) 76-84.
- [4] E. Tenailleau, P. Lancelin, R.J. Robins, S. Akoka, *J. Agric. Food Chem.* 52 (2004) 7782-7787.
- [5] F. Le Grand, G. George, S. Akoka, *J. Agric. Food. Chem.* 53 (2005) 5125-5129.
- [6] G.F. Pauli, *Phytochem. Anal.* 12 (2001) 28-42.
- [7] G.J. Martin, M.L. Martin, *Annu. Rep. NMR Spectrosc.* 31 (1995) 81-104.
- [8] E. Caytan, Y. Cherghaoui, C. Barril, C. Jouitteau, C. Rabiller, G. S. Remaud, *Tetrahedron: Asymmetry.* 17 (2006) 1622-1624.
- [9] E. Caytan, E. P. Botosoa, V. Silvestre, R.J. Robins, S. Akoka, G.S. Remaud. *Anal. Chem.* 79 (2007) 8266-8269.
- [10] C. Thibaudeau, G. Remaud, V. Silvestre, S. Akoka, *Anal. Chem.* 82 (2010) 5582-5590.
- [11] U. Bussy, C. Thibaudeau, F. Thomas, J.-R. Desmurs, E. Jamin, G. S. Remaud, V. Silvestre, S. Akoka, *Talanta.* 85 (2011) 1909-1914.
- [12] N. Merchak, J. Bejjani, T. Rizk, V. Silvestre, G. Remaud, S. Akoka, *Anal. Methods.* 7 (2015) 4889-4891.
- [13] S. Heikkinen, M.M. Toikka, P.T. Karhunen, A. Kilpeläinen, *J. Am. Chem. Soc.* 125 (2003) 4362-4367.
- [14] P. Giraudeau, *Magn. Reson. Chem.* 52 (2014) 259-272.
- [15] E. Martineau, I. Tea, S. Akoka, P. Giraudeau, *NMR Biomed.* 25 (2012) 985-992.
- [16] W. Gronwald, M.S. Klein, H. Kaspar, S.R. Fagerer, N. Nurnberger, K. Dettmer, T. Bertsch, P.J. Oefner, *Anal. Chem.* 20 (2008) 9288-9297.
- [17] S. Massou, C. Nicolas, F. Letisse, J.C Portais, *Phytochemistry.* 68 (2007) 2330-2340.
- [18] I.A. Lewis, S.C. Schommer, B. Hodis, K.A. Robb, M. Tonelli, W.M. Westler, M.R. Sussman, J.L. Markley, *Anal. Chem.* 79 (2007) 9385-9390.
- [19] E. Martineau, S. Akoka, R. Boisseau, B. Delanoue, P. Giraudeau, *Anal. Chem.* 85 (2013) 4777-4783.
- [20] D. Jeannerat, *Magn. Reson. Chem.* 41 (2003) 3-17.
- [21] G.A. Morris, *J. Magn. Reson.* 100 (1992) 316-328.

- [22] P. Giraudeau, N. Guignard, E. Hillion, E. Baguet, S. Akoka, *J. Pharm. Biomed. Anal.* 43 (2007) 1243-1248.
- [23] E. Martineau, P. Giraudeau, I. Tea, S. Akoka, *J. Pharm. Biomed. Anal.* 54 (2011) 252-257.
- [24] B. Vitorge, S. Bieri, M. Humam, P. Christen, K. Hostettmann, O. Muñoz, O ... & D. Jeannerat, *Chem. Comm.* 8 (2009) 950-952.
- [25] K. Ramirez-Gualito, D. Jeannerat, *Magn. Reson. Chem.* 53 (2015) 901-907.
- [26] J.C.J. Barna, E.D. Laue, M.R. Mayger, J. Skilling, S.J.P Worrall, *J. Magn. Reson.* 73 (1987) 69-77.
- [27] D.J. Holland, M.J. Bostock, L.F. Gladden, D. Nietlispach, *Angew. Chem., Int. Ed.* 50 (2011) 6548-6551.
- [28] C.M. Thiele, W. Bermel, *J. Magn. Reson.* 216 (2012) 134-143.
- [29] A. Le Guennec, J.N. Dumez1, P. Giraudeau, S. Caldarelli, *Magn. Reson. Chem.* 53 (2015) 913-920.

VI-3. Données complémentaires

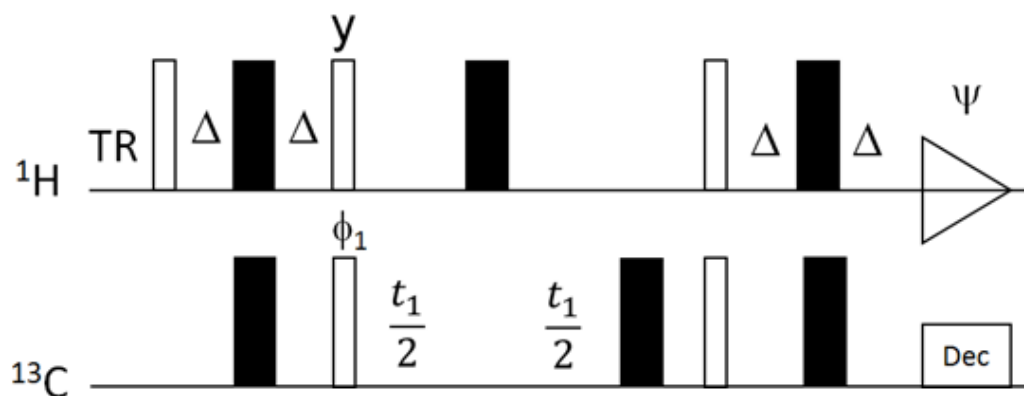


Figure S1: Previously used HSQC pulse sequence. Narrow (open) and wide (filled) rectangles represent 90° and 180° pulses, respectively.

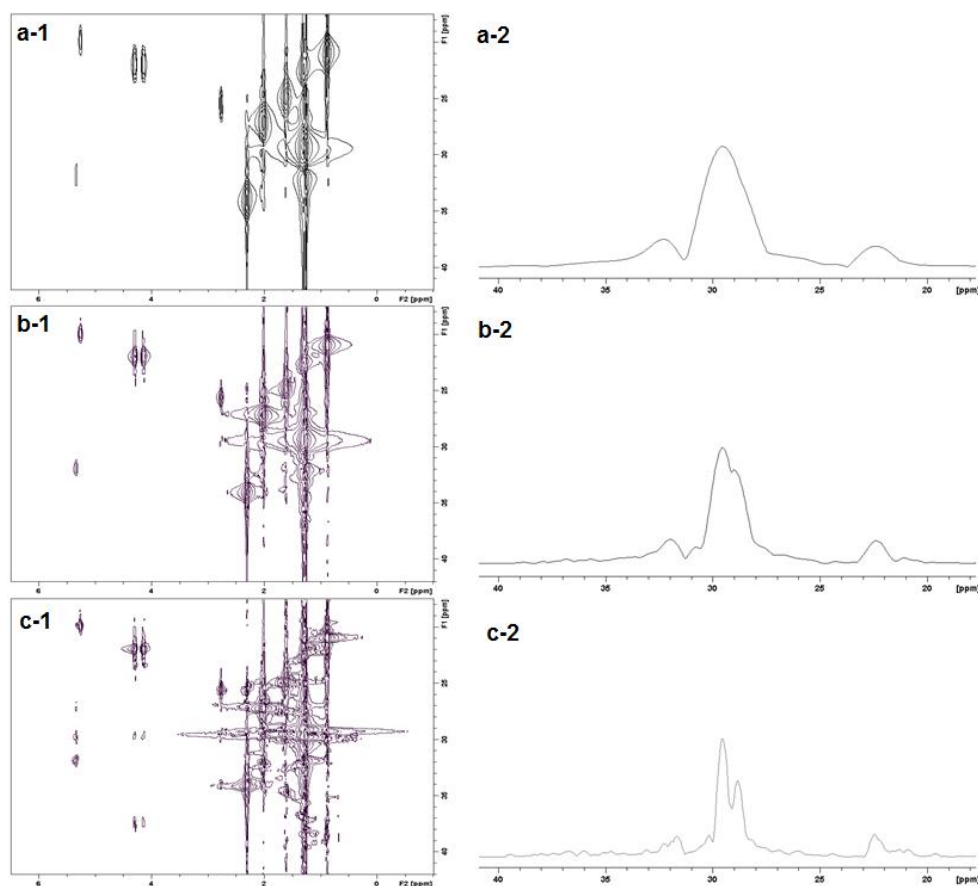


Figure S2: 2D HSQC spectra and 1D projection of the aliphatic region of olive oil in CDCl_3 recorded in 22 min on 700 MHz Bruker spectrometer equipped with an inverse $^1\text{H}/^{13}\text{C}$ cryoprobe, using spectral aliasing strategy, $\text{SW} = 24.5$ ppm in F_1 dimension. a-1, a-2) $N_1 = 32$ t_1 increments. b-1, b-2) $N_1 = 64$ t_1 increments and a NUS rate of 50%. c-1, c-2) $N_1 = 128$ t_1 increments and a NUS rate of 25%.

VI-4. Conclusion

La séquence HSQC classique a été optimisée en ajoutant des impulsions RF de telle manière à avoir une symétrie parfaite permettant de compenser au mieux les imperfections dues aux inhomogénéités du champ et à l'effet d'« off-resonance » ; des impulsions de gradient sont aussi été ajoutées pour sélectionner le chemin de cohérence souhaité et supprimer le signal résiduel résultant des ^1H liés à ^{12}C , car une suppression incomplète de ces signaux peut conduire à une dégradation de la précision. Les résultats obtenus démontrent que la séquence HSQC symétrisée élimine complètement le signal ^1H - ^{12}C résiduel, contrairement à la séquence classique.

D'un autre côté, les expériences RMN 2D sont limitées par une longue durée d'acquisition et peuvent ainsi être affectées par les instabilités du spectromètre au cours du temps [179]. Notre approche qui consiste aussi à réduire la durée de l'expérience a permis d'atteindre, dans un délai raisonnable, une précision de quelques pour mille nécessaire pour l'analyse isotopique.

Le nombre d'incrément (N_1) dans la dimension F_1 a été réduit à 64 pour diminuer la durée de l'expérience autant que possible tout en conservant une résolution suffisante pour éviter le chevauchement des pics. Cette faible valeur de N_1 a été rendue possible par l'utilisation du « repliement spectral » qui consiste à sélectionner une largeur spectrale plus faible (24,5 ppm) dans la dimension de F_1 (^{13}C). Il faut signaler que les déplacements chimiques le long de la dimension F_1 ne sont plus corrects mais ce n'est pas un problème dans le cas d'une matrice bien connue.

En outre, l'échantillonnage non uniforme « non-uniform sampling » (NUS) a été appliquée pour améliorer la résolution des spectres 2D HSQC des matrices de TAG sans augmenter le temps expérimental, cette approche consiste en l'acquisition d'une partie des incréments dans la dimension indirecte (F_1). À cet égard, les taux 25 et 50% de NUS ont été testés, et les spectres HSQC ont été enregistrés dans le même temps d'expérience (22 min) que celui enregistré avec $N_1 = 32$ incréments sans NUS. Le spectre enregistré avec un NUS de 25% et $N_1 = 128$ montre une meilleure résolution, mais nous avons observé des artefacts plus importants que dans le cas des acquisitions réalisées avec un NUS de 50% et $N_1 = 64$. Suite à ce résultat, le NUS de 50% a été appliqué lors de l'enregistrement des spectres.

Ainsi, nous avons optimisé les paramètres d'acquisition et de traitement des spectres HSQC pour atteindre, dans un temps relativement court, une précision suffisante pour l'analyse isotopique. La répétabilité à long terme de cette technique a été évaluée sur une période de trois mois. Nous n'avons pas observé une amélioration avec l'augmentation du nombre de spectres enregistrés à chaque fois. Cela est vraisemblablement dû à l'absence d'écarts

significatifs entre les mesures effectuées sur le même tube RMN durant la même session d'analyse.

En conséquence, un seul spectre enregistré en 22 min est suffisant pour obtenir la précision requise pour l'analyse isotopique.

La séquence HSQC optimisée a été utilisée pour classer des huiles végétales commerciales de différentes origines géographiques et botaniques. Les résultats montrent que cette méthode, appliquée sur les TAG, fournit dans un délai compétitif des informations isotopiques et métabolomiques à potentiel d'authentification élevé.

Conclusion générale et perspectives

L'objectif de cette thèse était de disposer de méthodes RMN, métabolomiques et isotopiques, documentées, fiables, rapides et efficaces pour assurer la traçabilité des origines géographique et botanique des produits alimentaires riches en TAG. Et cela dans le but de détecter les fraudes, de protéger les consommateurs et de leur garantir des produits authentiques.

Les résultats de l'étude métabolomique par RMN ^1H , première étude, qui a porté sur 242 échantillons d'huile d'olive de différentes régions libanaises montrent que leur composition chimique varie en fonction de leurs coordonnées géographiques et de la morphologie des olives correspondantes. Une stratégie séquentielle, basée sur l'ensemble des variables obtenues, a été développée et des modèles statistiques ont été construits permettant la projection d'un nouvel échantillon et son attribution au groupe qui lui ressemble le plus.

Les olives vertes sont plus riches en squalène que les olives noires et cette différence est significative. De plus les échantillons d'olives noires se caractérisent par la présence d'une variable discrète associée à un composé terpénique (Trp-2) non identifié. Une fois la couleur des olives reconnue, la détermination de l'origine géographique (altitude et latitude du terroir) devient plus précise et ceci en choisissant le modèle d'analyse canonique discriminante (CDA) approprié. Par ailleurs, les résultats de l'analyse multivariée révèlent que la date de récolte dans un intervalle de trois semaines n'affecte pas la classification géographique et variétale des huiles d'olive au niveau sous-régional. Enfin, une corrélation inverse est observée entre la quantité des composés phénoliques et le niveau de précipitation au cours des mois d'avril et de mai indépendamment du taux de précipitations annuel.

Cependant, les spectres RMN ^1H des mélanges de TAG sont de loin moins résolus que les spectres ^{13}C correspondants ; cela est dû à la gamme restreinte de déplacement chimique des protons. Une meilleure discrimination des signaux est donc observée en RMN ^{13}C . De plus, les spectres ^{13}C sont plus propres suite à un découplage des protons et permettent d'obtenir des informations supplémentaires telles que la position des acides gras sur le squelette du glycérol et les rapports isotopiques de différents sites de la molécule. Pour toutes ces raisons, la RMN ^{13}C semble être plus intéressante pour les applications d'authenticité et de classification. Toutefois, la longue durée d'expérience constitue un sérieux problème. Nous avons donc optimisé, dans la deuxième étude, une séquence de transfert de polarisation, 7 fois plus rapide que la séquence mono impulsionnelle conventionnelle. Cela nous a permis de mieux caractériser les matrices riches en TAG, et d'obtenir la précision élevée, de l'ordre de quelques pour mille, indispensable à l'analyse isotopique.

La RMN ^{13}C ne permet cependant pas d'obtenir directement le profil complet des acides gras. C'est la GC qui est actuellement considérée comme la méthode de référence pour ce type de profilage. Dans la troisième étude, des modèles linéaires multivariés pour la quantification des acides gras individuels et du squalène dans les huiles d'olive ont été construits en se basant sur des variables obtenues à partir des spectres INEPT- ^{13}C après déconvolution (variables indépendantes) et des analyses GC (variables dépendantes). Des modèles pour les acides gras suivants ont pu être obtenus : palmitique, margarique, stéarique, arachidique, béhénique, lignocérique, palmitoléique, oléique, vaccénique, gondoïque, linoléique et linoléinique.

La séquence INEPT utilisée s'est révélée être une technique rapide et efficace, elle ne nécessite aucune préparation chimique préalable des échantillons comme la trans-méthylesterification. Cette technique permet d'obtenir à la fois : des pourcentages en acides gras individuels, des teneurs isotopiques relatives en ^{13}C site par site pour les métabolites invariants, ainsi que la distribution de certains acides gras sur le squelette du glycérol.

Dans une quatrième étude, une nouvelle série d'échantillons d'huile d'olive a été analysée par INEPT- ^{13}C et GC. Les fortes corrélations obtenues ont confirmé la robustesse des modèles construits pour la détermination des acides gras. De plus, les répétabilités à long terme des mesures effectuées par INEPT- ^{13}C et GC sont comparables. Dans une deuxième partie de cette étude, il a été remarqué que les techniques INEPT et GC possèdent des pouvoirs discriminants comparables dans la classification des huiles selon la couleur des olives. Au niveau de la classification en fonction de l'altitude du terroir de provenance, l'INEPT a fourni le meilleur résultat. Dans les deux cas, la RMN ^1H était moins efficace.

La séquence INEPT- ^{13}C adiabatique présente donc un fort potentiel de caractérisation qui pourra être mis à profit sur d'autres matrices riches en TAG par la construction de modèles linéaires multivariés pour le calcul des pourcentages des acides gras individuels.

Dans une cinquième étude, la RMN 2D, plus particulièrement la séquence HSQC, a été utilisée pour l'analyse des huiles végétales. La symétrie parfaite de la séquence utilisée a permis de compenser les imperfections RF et d'obtenir la précision requise pour l'analyse isotopique. Cette technique ne permet pas de donner des informations concernant la position des acides gras ni les concentrations des composés minoritaires. En revanche, elle s'est révélée très robuste, et ne nécessite pas l'enregistrement de plusieurs spectres par expérience, un seul (22 min) est suffisant pour obtenir la précision élevée nécessaire à l'analyse isotopique. De ce fait, cette séquence semble être avantageuse pour gagner du temps. De plus, elle a été utilisée pour classer des huiles végétales commerciales de différentes origines géographiques et botaniques. Les résultats des CDA ont montré que cette approche

pourrait être utilisée pour vérifier l'authenticité des produits alimentaires dans un délai raisonnable, et pour faire de l'analyse isotopique.

Ce travail devrait constituer une étape importante dans le domaine de la sécurité alimentaire et le respect des normes internationales. Il nous a permis de mettre au point des méthodes fiables et efficaces pour vérifier les origines des denrées alimentaires et assurer leur traçabilité, répondant ainsi à la demande croissante des consommateurs pour des produits authentiques. Cette thèse devrait donc contribuer à l'instauration, au Liban, d'Appellations d'Origine Contrôlée (AOC) ou d'autres labels garantissant l'authenticité des huiles d'olive.

De nombreuses perspectives sont ouvertes et les futurs travaux pourraient notamment s'orienter vers :

- l'utilisation de plus hauts champs pour déterminer l'impact du gain en termes de sensibilité et de résolution ;
- l'application de la séquence INEPT adiabatique à bas champs compte tenu du coût des spectromètres correspondants ;
- la caractérisation de matrices animales riches en TAG en fonction de leur origine géographique et de l'alimentation des animaux ;
- l'identification des biomarqueurs liés à certaines pathologies (notamment les maladies cardiovasculaires) ;
- l'application de la séquence HSQC optimisée pour analyser des molécules autres que les TAG ;
- la réalisation d'études multi-centriques qui permettront de vérifier la robustesse de ces méthodes et de préciser leurs exigences en termes de configuration des appareils et de procédures de calibrage.

Les résultats de cette thèse ont fait l'objet de cinq articles.

Production scientifique

Articles

- 1- N. Merchak, J. Bejjani, T. Rizk, V. Silvestre, G.S. Remaud, S. Akoka, ^{13}C isotopomics of triacylglycerols using NMR with polarization transfer techniques. *Analytical Methods*, 7(12), 4889-4891 **2015**.
- 2- N. Merchak, E. El Bacha, R. Bou Khouzam, T. Rizk, V. Silvestre, G.S. Remaud, S. Akoka, J. Bejjani, Geoclimatic, morphological, and temporal effects on Lebanese olive oils composition and classification : A ^1H NMR metabolomic study. Article en cours d'évaluation par la revue Food Chemistry.
- 3- N. Merchak, V. Silvestre, D. Loquet, T. Rizk, S. Akoka, J. Bejjani, A novel quantification strategy of individual fatty acids in olive oil using optimized ^{13}C -NMR with a polarization transfer technique, Article en cours d'évaluation par la revue Analytical and Bioanalytical Chemistry.
- 4- N. Merchak, T. Rizk, V. Silvestre, G.S. Remaud, J. Bejjani, S. Akoka, Olive oil characterization and classification by ^{13}C -NMR with a polarization transfer technique: A comparison with gas chromatography and ^1H -NMR», qui va être soumis à la revue Lipid Research.
- 5- N. Merchak, T. Rizk, V. Silvestre, L. Rouger, P. Giraudeau, J. Bejjani, S. Akoka, Precise and rapid isotopomic analysis by ^1H - ^{13}C HSQC: Application to triacylglycerol matrices», soumis à la revue Talanta.

Communications orales dans des congrès

- 1- N. Merchak, V. Silvestre, J. Lalande, S. Akoka, J. Bejjani, 8^{ème} journées scientifiques du Réseau Francophone de Métabolomique et Fluxomique (RFMF) ; Lyon, le 20/05/2014.
- 2- N. Merchak, V. Silvestre, G.S. Remaud, S. Akoka, T. Rizk, J. Bejjani, 21^{ème} conférence scientifique internationale de la LAAS ; Beyrouth, le 17/04/2015.
- 3- N. Merchak, V. Silvestre, G.S. Remaud, S. Akoka, T. Rizk, J. Bejjani, 9^{ème} journées scientifiques du Réseau Francophone de Métabolomique et Fluxomique (RFMF) ; Lille, le 10/06/2015.

Références bibliographiques

-
- [1] IUPAC-IUB (1976). Nomenclature of lipids, World Wide Web version, prepared by G.P. Moss. <http://www.chem.qmul.ac.uk/iupac/lipid/>.
- [2] Raynal-Ljutovac, K., Bouvier, J., Gayet, C., Simon, N., Joffre, F., Fine, F., ... & Genot, C. (2011). Organisation structurale et moléculaire des lipides dans les aliments: impacts possibles sur leur digestion et leur assimilation par l'Homme. *Oléagineux, Corps gras, Lipides*, 18(6), 324-351.
- [3] Gunstone, Frank D. "Lipid Glossary 2".
- [4] Morin, O., & Pagès-Xatart-Parès, X. (2012). Huiles et corps gras végétaux: ressources fonctionnelles et intérêt nutritionnel. *Oléagineux, Corps gras, Lipides*, 19(2), 63-75.
- [5] Rahman, S. A., Huah, T. S., Nassan, O., & Daud, N. M. (1995). Fatty acid composition of some Malaysian freshwater fish. *Food chemistry*, 54(1), 45-49.
- [6] Pieroni, G., & Coste, T. C. (2010). Composition en acides gras des œufs. Intérêt nutritionnel et valeur santé. *Cahiers de Nutrition et de Diététique*, 45(5), 261-266.
- [7] Kochhar, S. P. (1993). Deterioration of edible oils, fats and foodstuffs. *Atmospheric oxidation and antioxidants*, 2, 71-131.
- [8] Armanino, C., Leardi, R., Lanteri, S., & Modi, G. (1989). Chemometric analysis of Tuscan olive oils. *Chemometrics and Intelligent Laboratory Systems*, 5(4), 343-354.
- [9] Lotti, G., Riu, R., & Izzo, R. (1982). Influenza del clima sulla composizione acidica e sterolica degli olii di oliva. *Rivista della Societa Italiana di Scienza dell'Alimentazione*.
- [10] French, P., Stanton, C., Lawless, F., O'riordan, E. G., Monahan, F. J., Caffrey, P. J., & Moloney, A. P. (2000). Fatty acid composition, including conjugated linoleic acid, of intramuscular fat from steers offered grazed grass, grass silage, or concentrate-based diets. *Journal of Animal Science*, 78(11), 2849-2855.
- [11] Brockerhoff, H., & Yurkowski, M. (1966). Stereospecific analyses of several vegetable fats. *Journal of lipid research*, 7(1), 62-64.
- [12] Christie, W. W., & Moore, J. H. (1970). A comparison of the structures of triglycerides from various pig tissues. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, 210(1), 46-56.
- [13] Christie, W. W., & Moore, J. H. (1972). The structures of adipose tissue and heart muscle triglycerides in the domestic chicken (*Gallus gallus*). *Journal of the science of food and agriculture*, 23(1), 73-77.
- [14] O'Leary, M. H. (1981). Carbon isotope fractionation in plants. *Phytochemistry*, 20(4), 553-567.
- [15] Ziegler, H., Osmond, C. B., Stichler, W., & Trimborn, P. (1976). Hydrogen isotope discrimination in higher plants: correlations with photosynthetic pathway and environment. *Planta*, 128(1), 85-92.
- [16] Chilliard, Y., Ferlay, A., & Doreau, M. (2001). Effect of different types of forages, animal fat or marine oils in cow's diet on milk fat secretion and composition, especially conjugated linoleic acid (CLA) and polyunsaturated fatty acids. *Livestock Production Science*, 70(1), 31-48.
- [17] Aursand, M., Mabon, F., & Martin, G. J. (2000). Characterization of farmed and wild salmon (*Salmo salar*) by a combined use of compositional and isotopic analyses. *Journal of the American Oil Chemists' Society*, 77(6), 659-666.

-
- [18] Molquentin, J., Meisel, H., Lehmann, I., & Rehbein, H. (2007). Identification of organically farmed Atlantic salmon by analysis of stable isotopes and fatty acids. *European Food Research and Technology*, 224(5), 535-543.
- [19] Renou, J.P., Deponge, C., Gachon, P., Bonnefoy, J.C., Coulon, J.B., Garel, J.P., Vérité, R., and Ritz, P. (2004). Characterization of animal products according to geographic origin and feeding diet using nuclear magnetic resonance and isotope ratio mass spectrometry: cow milk, *Food Chemistry* 85(1):63–66.
- [20] Renou, J. P., Bielicki, G., Deponge, C., Gachon, P., Micol, D., & Ritz, P. (2004). Characterization of animal products according to geographic origin and feeding diet using nuclear magnetic resonance and isotope ratio mass spectrometry. Part II: Beef meat. *Food Chemistry*, 86(2), 251-256.
- [21] Molquentin, J., & Gieseemann, A. (2007). Differentiation of organically and conventionally produced milk by stable isotope and fatty acid analysis. *Analytical and Bioanalytical Chemistry*, 388(1), 297-305.
- [22] Guardiola, F., Codony, R., Rafecas, M., Boatella, J., & López, A. (1994). Fatty acid composition and nutritional value of fresh eggs, from large-and small-scale farms. *Journal of Food Composition and Analysis*, 7(3), 171-188.
- [23] Vigli, G., Philippidis, A., Spyros, A., & Dais, P. (2003). Classification of edible oils by employing ^{31}P and ^1H NMR spectroscopy in combination with multivariate statistical analysis. A proposal for the detection of seed oil adulteration in virgin olive oils. *Journal of Agricultural and Food Chemistry*, 51(19), 5715-5722.
- [24] Longobardi, F., Sacco, D., Casiello, G., Ventrella, A., & Sacco, A. (2015). Characterization of the Geographical and Varietal Origin of Wheat and Bread by Means of Nuclear Magnetic Resonance (NMR), Isotope Ratio Mass Spectrometry (IRMS) Methods and Chemometrics: A Review. *Agricultural Sciences*, 6(01), 126.
- [25] Hu, F. B., Stampfer, M. J., Manson, J. E., Rimm, E., Colditz, G. A., Rosner, B. A., ... & Willett, W. C. (1997). Dietary fat intake and the risk of coronary heart disease in women. *New England Journal of Medicine*, 337(21), 1491-1499.
- [26] Parkinson, L., & Keast, R. (2014). Oleocanthal, a phenolic derived from virgin olive oil: a review of the beneficial effects on inflammatory disease. *International journal of molecular sciences*, 15(7), 12323-12334.
- [27] Owen, R. W., Mier, W., Giacosa, A., Hull, W. E., Spiegelhalter, B., & Bartsch, H. (2000). Phenolic compounds and squalene in olive oils: the concentration and antioxidant potential of total phenols, simple phenols, secoiridoids, lignans and squalene. *Food and Chemical Toxicology*, 38(8), 647-659.
- [28] Martínez-González, M. A., Fernández-Jarne, E., Serrano-Martínez, M., Martí, A., Martínez, J. A., & Martín-Moreno, J. M. (2002). Mediterranean diet and reduction in the risk of a first acute myocardial infarction: an operational healthy dietary score. *European journal of nutrition*, 41(4), 153-160
- [29] Fernández-Jarne, E., Martínez-Losa, E., Prado-Santamaria, M., Brugarolas-Brufau, C., Serrano-Martínez, M., & Martínez-González, M. A. (2002). Risk of first non-fatal myocardial infarction negatively associated with olive oil consumption: a case-control study in Spain. *International Journal of Epidemiology*, 31(2), 474-480.
- [30] Abbey, M., Belling, G. B., Noakes, M., Hirata, F., & Nestel, P. J. (1993). Oxidation of low-density lipoproteins: intraindividual variability and the effect of dietary linoleate supplementation. *The American journal of clinical nutrition*, 57(3), 391-398.

- [31] Bonanome, A., Pagnan, A., Biffanti, S., Opportuno, A., Sorgato, F., Dorella, M., ... & Ursini, F. (1992). Effect of dietary monounsaturated and polyunsaturated fatty acids on the susceptibility of plasma low density lipoproteins to oxidative modification. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 12(4), 529-533.
- [32] Reaven, P., Parthasarathy, S., Grasse, B. J., Miller, E., Almazan, F., Mattson, F. H., ... & Witztum, J. L. (1991). Feasibility of using an oleate-rich diet to reduce the susceptibility of low-density lipoprotein to oxidative modification in humans. *The American journal of clinical nutrition*, 54(4), 701-706.
- [33] Reaven, P., Parthasarathy, S. A. M. P. A. T. H., Grasse, B. J., Miller, E., Steinberg, D., & Witztum, J. L. (1993). Effects of oleate-rich and linoleate-rich diets on the susceptibility of low density lipoprotein to oxidative modification in mildly hypercholesterolemic subjects. *Journal of Clinical Investigation*, 91(2), 668.
- [34] Barzi, F., Woodward, M., Marfisi, R. M., Tavazzi, L., Valagussa, F., & Marchioli, R. (2003). Mediterranean diet and all-causes mortality after myocardial infarction: results from the GISSI-Prevenzione trial. *European journal of clinical nutrition*, 57(4), 604-611.
- [35] Aparicio, R., & Harwood, J. (2013). *Handbook of olive oil* (pp. 163-395). Springer: New York, NY, USA.
- [36] Niklis, D., Baourakis, G., Thabet, B., & Manthoulis, G. (2014). Chapitre 11. Commerce et logistique: le cas de la filière huile d'olive.
- [37] Lebanese Ministry of Agriculture (2003). L'olivier: Contraintes et Potentialités.
- [38] Investment Development Authority of Lebanon (2014). Olive oil industry in Lebanon.
- [39] León-Camacho, M., Morales, M. T., & Aparicio, R. (2013). Chromatographic Methodologies: Compounds for Olive Oil Traceability Issues. In *Handbook of Olive Oil* (pp. 163-217). Springer US.
- [40] Ollivier, D., Artaud, J., Pinatel, C., Durbec, J. P., & Guérère, M. (2003). Triacylglycerol and fatty acid compositions of French virgin olive oils. Characterization by chemometrics. *Journal of agricultural and food chemistry*, 51(19), 5723-5731
- [41] Ollivier D., Pinatel C., Artaud J., Le Nouvel Olivier, N°44 mars-avril 2005.
- [42] Ostlund Jr, R. E., Racette, S. B., & Stenson, W. F. (2002). Effects of trace components of dietary fat on cholesterol metabolism: phytosterols, oxysterols, and squalene. *Nutrition reviews*, 60(11), 349-359.
- [43] Newmark, H. L. (1997). Squalene, olive oil, and cancer risk: a review and hypothesis. *Cancer Epidemiology Biomarkers & Prevention*, 6(12), 1101-1103.
- [44] Owen, R. W., Haubner, R., Würtele, G., Hull, W. E., Spiegelhalder, B., & Bartsch, H. (2004). Olives and olive oil in cancer prevention. *European Journal of Cancer Prevention*, 13(4), 319-326.
- [45] Bonanome, A., Colomer, R., Corella-Piquer, D., Covas, I., Chamorro-Quiros, J., Escrich, E., ... & Villar-Ortiz, J. R. (2005). virgin olive oil. *European journal of clinical investigation*, 35(7), 421-424.
- [46] Tripoli, E., Giammanco, M., Tabacchi, G., Di Majò, D., Giammanco, S., & La Guardia, M. (2005). The phenolic compounds of olive oil: structure, biological activity and beneficial effects on human health. *Nutrition Research Reviews*, 18(01), 98-112.
- [47] Visioli, F., Poli, A., & Gall, C. (2002). Antioxidant and other biological activities of phenols from olives and olive oil. *Medicinal research reviews*, 22(1), 65-75.

- [48] Franconi, F., Coinu, R., Carta, S., Urgeghe, P. P., Ieri, F., Mulinacci, N., & Romani, A. (2006). Antioxidant effect of two virgin olive oils depends on the concentration and composition of minor polar compounds. *Journal of agricultural and food chemistry*, 54(8), 3121-3125.
- [49] Galli, C., & Visioli, F. (1999). Antioxidant and other activities of phenolics in olives/olive oil, typical components of the Mediterranean diet. *Lipids*, 34, S23-S26.
- [50] Marrugat, J., Covas, M. I., Fitó, M., Schröder, H., Miró-Casas, E., Gimeno, E., ... & Farré, M. (2004). Effects of differing phenolic content in dietary olive oils on lipids and LDL oxidation. *European journal of nutrition*, 43(3), 140-147.
- [51] Salvini, S., Sera, F., Caruso, D., Giovannelli, L., Visioli, F., Saieva, C., ... & Palli, D. (2006). Daily consumption of a high-phenol extra-virgin olive oil reduces oxidative DNA damage in postmenopausal women. *British Journal of Nutrition*, 95(04), 742-751.
- [52] Covas, M. I., de la Torre, K., Farré-Albaladejo, M., Kaikkonen, J., Fitó, M., López-Sabater, C., ... & de la Torre, R. (2006). Postprandial LDL phenolic content and LDL oxidation are modulated by olive oil phenolic compounds in humans. *Free Radical Biology and Medicine*, 40(4), 608-616.
- [53] Visioli, F., Bellomo, G., Montedoro, G., & Galli, C. (1995). Low density lipoprotein oxidation is inhibited in vitro by olive oil constituents. *Atherosclerosis*, 117(1), 25-32.
- [54] Schuler, P. (1990). Natural antioxidants exploited commercially. In *Food antioxidants* (pp. 99-170). Springer Netherlands.
- [55] Quilez, J., Garcia-Lorda, P., & Salas-Salvado, J. (2003). Potential uses and benefits of phytosterols in diet: present situation and future directions. *Clinical Nutrition*, 22(4), 343-351.
- [56] Awad, A. B., Von Holtz, R. L., Cone, J. P., Fink, C. S., & Chen, Y. C. (1997). beta-Sitosterol inhibits growth of HT-29 human colon cancer cells by activating the sphingomyelin cycle. *Anticancer research*, 18(1A), 471-473.
- [57] Awad, A. B., Downie, A. C., & Fink, C. S. (2000). Inhibition of growth and stimulation of apoptosis by beta-sitosterol treatment of MDA-MB-231 human breast cancer cells in culture. *International journal of molecular medicine*, 5(5), 541-546.
- [58] Covas, M. I., Ruiz-Gutiérrez, V., De La Torre, R., Kafatos, A., Lamuela-Raventós, R. M., Osada, J., ... & Visioli, F. (2006). Minor components of olive oil: evidence to date of health benefits in humans. *Nutrition reviews*, 64(10), S20.
- [59] Raicht, R. F., Cohen, B. I., Fazzini, E. P., Sarwal, A. N., & Takahashi, M. (1980). Protective effect of plant sterols against chemically induced colon tumors in rats. *Cancer Research*, 40(2), 403-405.
- [60] Fedeli, E., Lanzani, A., Capella, P., & Jacini, G. (1966). Triterpene alcohols and sterols of vegetable oils. *Journal of the American Oil Chemists' Society*, 43(4), 254-256.
- [61] Luna, G. (2002). Characterisation of monovarietal virgin olive oils. *Eur. J. Lipid Sci. Technol*, 104, 614-627.
- [62] Ferreiro, L., & Aparicio, R. (1992). Influence of altitude on chemical composition of Andalusian virgin olive oil. Mathematical classification equations. *Grasas y Aceites*, 43(3), 149-156.
- [63] Mohamed Mousa, Y., Gerasopoulos, D., Metzidakis, I., & Kiritsakis, A. (1996). Effect of altitude on fruit and oil quality characteristics of 'Mastoides' olives. *Journal of the Science of Food and Agriculture*, 71(3), 345-350.

-
- [64] Aguilera, M. P., Beltrán, G., Ortega, D., Fernández, A., Jiménez, A., & Uceda, M. (2005). Characterisation of virgin olive oil of Italian olive cultivars: Frantoio and Leccino, grown in Andalusia. *Food chemistry*, 89(3), 387-391.
- [65] Houlali, I., Rabi, S., Elbir, M., Ider, A. A., Amhoud, A., Moubarik, A., ... & Mbarki, M. Caractérisation chimique de l'huile d'olive dans la région Marocaine de Tadla Azilal (Chemical characterization of the virgin olive oil in Tadla Azilal Moroccan area).
- [66] Ranalli, A., De Mattia, G., Ferrante, M. L., & Giansante, L. (1997). Incidence of olive cultivation area on the analytical characteristics of the oil. Note 1. *Rivista Italiana delle Sostanze Grasse*, 74(11), 501-508.
- [67] Osman, M., Metzidakis, I., Gerasopoulos, D., & Kiritsakis, A. (1994). Qualitative changes in olive oil of fruits collected from trees grown at two altitudes. *Rivista Italiana delle Sostanze Grasse*, 71(4), 187-190.
- [68] Aparicio, R., Ferreiro, L., & Alonso, V. (1994). Effect of climate on the chemical composition of virgin olive oil. *Analytica Chimica Acta*, 292(3), 235-241.
- [69] Stefanoudaki, E., Kotsifaki, F., & Koutsaftakis, A. (1999). Classification of virgin olive oils of the two major Cretan cultivars based on their fatty acid composition. *Journal of the American Oil Chemists' Society*, 76(5), 623-626.
- [70] Esmaceli, A., Shaykhoradi, F., & Naseri, R. (2012). Comparison of oil content and fatty acid composition of native olive genotypes in different region of Liam, Iran. *Intl J Agric Crop Sci*, 4(8), 434-438.
- [71] Temime, S. B., Wael, T., Bechir, B., Leila, A., Douja, D., & Mokhtar, Z. (2006). Changes in olive oil quality of Chétoui variety according to origin of plantation. *Journal of Food Lipids*, 13(1), 88-99.
- [72] Ferreiro, L., & Aparicio, R. (1992). Influencia de la altitud en la composición química de los aceites de oliva vírgenes de Andalucía. Ecuaciones matemáticas de clasificación. *Grasas y aceites*, 43(3), 149-156.
- [73] García-González, D. L., Aparicio-Ruiz, R., Aparicio, R., Moreau, R. A., & Kamal-Eldin, A. (2009). Olive oil. *Gourmet and health-promoting speciality oils.*, 33-72.
- [74] D'Imperio, M., Mannina, L., Capitani, D., Bidet, O., Rossi, E., Bucarelli, F. M., ... & Segre, A. (2007). NMR and statistical study of olive oils from Lazio: a geographical, ecological and agronomic characterization. *Food chemistry*, 105(3), 1256-1267.
- [75] Issaoui, M., Flamini, G., Brahmi, F., Dabbou, S., Hassine, K. B., Taamali, A., ... & Hammami, M. (2010). Effect of the growing area conditions on differentiation between Chemlali and Chétoui olive oils. *Food Chemistry*, 119(1), 220-225.
- [76] Vossen, P. (2013). Growing olives for oil. In *Handbook of Olive Oil* (pp. 19-56). Springer US.
- [77] Aparicio, R., Alonso, V., & Morales, M. T. (1994). Detailed and exhaustive study of the authentication of European virgin olive oils by SEXIA expert system. *Grasas y aceites*, 45(4), 241-252.
- [78] Grattan, S. R., Berenguer, M. J., Connell, J. H., Polito, V. S., & Vossen, P. M. (2006). Olive oil production as influenced by different quantities of applied water. *Agricultural water management*, 85(1), 133-140.
- [79] Gómez-Rico, A., Salvador, M. D., Moriana, A., Pérez, D., Olmedilla, N., Ribas, F., & Fregapane, G. (2007). Influence of different irrigation strategies in a traditional Cornicabra cv. olive orchard on virgin olive oil composition and quality. *Food chemistry*, 100(2), 568-578.

- [80] Servili, M., Esposto, S., Lodolini, E., Selvaggini, R., Taticchi, A., Urbani, S., ... & Gucci, R. (2007). Irrigation effects on quality, phenolic composition, and selected volatiles of virgin olive oils cv. Leccino. *Journal of Agricultural and Food Chemistry*, 55(16), 6609-6618.
- [81] Patumi, M., d'Andria, R., Fontanazza, G., Morelli, G., Giorgio, P., & Sorrentino, G. (1999). Yield and oil quality of intensively trained trees of three cultivars of olive (*Olea europaea* L.) under different irrigation regimes. *Journal of Horticultural Science and Biotechnology*, 74(6), 729-737.
- [82] Parr, A. J., & Bolwell, G. P. (2000). Phenols in the plant and in man. The potential for possible nutritional enhancement of the diet by modifying the phenols content or profile. *Journal of the Science of Food and Agriculture*, 80(7), 985-1012.
- [83] Morelló, J. R., Romero, M. P., Ramo, T., & Motilva, M. J. (2005). Evaluation of L-phenylalanine ammonia-lyase activity and phenolic profile in olive drupe (*Olea europaea* L.) from fruit setting period to harvesting time. *Plant Science*, 168(1), 65-72.
- [84] Dag, A., Ben-Gal, A., Yermiyahu, U., Basheer, L., Nir, Y., & Kerem, Z. (2008). The effect of irrigation level and harvest mechanization on virgin olive oil quality in a traditional rain-fed 'Souri' olive orchard converted to irrigation. *Journal of the Science of Food and Agriculture*, 88(9), 1524-1528.
- [85] Ramos, A. F., & Santos, F. L. (2010). Yield and olive oil characteristics of a low-density orchard (cv. Cordovil) subjected to different irrigation regimes. *Agricultural Water Management*, 97(2), 363-373.
- [86] Berenguer, M. J., Vossen, P. M., Grattan, S. R., Connell, J. H., & Polito, V. S. (2006). Tree irrigation levels for optimum chemical and sensory properties of olive oil. *HortScience*, 41(2), 427-432.
- [87] Lazzez, A., Perri, E., Caravita, M. A., Khlif, M., & Cossentini, M. (2008). Influence of olive maturity stage and geographical origin on some minor components in virgin olive oil of the Chemlali variety. *Journal of agricultural and food chemistry*, 56(3), 982-988.
- [88] D'Imperio, M., Gobbino, M., Picanza, A., Costanzo, S., Della Corte, A., & Mannina, L. (2010). Influence of harvest method and period on olive oil composition: an NMR and statistical study. *Journal of agricultural and food chemistry*, 58(20), 11043-11051.
- [89] Baccouri, O., Guerfel, M., Baccouri, B., Cerretani, L., Bendini, A., Lercker, G., ... & Miled, D. D. B. (2008). Chemical composition and oxidative stability of Tunisian monovarietal virgin olive oils with regard to fruit ripening. *Food Chemistry*, 109(4), 743-754.
- [90] Gimeno, E., Castellote, A. I., Lamuela-Raventós, R. M., De la Torre, M. C., & Lopez-Sabater, M. C. (2002). The effects of harvest and extraction methods on the antioxidant content (phenolics, α -tocopherol, and β -carotene) in virgin olive oil. *Food Chemistry*, 78(2), 207-211.
- [91] Lazzez, A., Perri, E., Caravita, M. A., Khlif, M., & Cossentini, M. (2008). Influence of olive maturity stage and geographical origin on some minor components in virgin olive oil of the Chemlali variety. *Journal of agricultural and food chemistry*, 56(3), 982-988.
- [92] Gutiérrez, F., Jimenez, B., Ruiz, A., & Albi, M. A. (1999). Effect of olive ripeness on the oxidative stability of virgin olive oil extracted from the varieties Picual and Hojiblanca and on the different components involved. *Journal of Agricultural and Food Chemistry*, 47(1), 121-127.
- [93] Matos, L. C., Cunha, S. C., Amaral, J. S., Pereira, J. A., Andrade, P. B., Seabra, R. M., & Oliveira, B. P. (2007). Chemometric characterization of three varietal olive oils (Cvs. Cobrançosa, Madural and Verdeal Transmontana) extracted from olives with different maturation indices. *Food chemistry*, 102(1), 406-414.

- [94] Beltrán, G., del Rio, C., Sánchez, S., & Martínez, L. (2004). Influence of harvest date and crop yield on the fatty acid composition of virgin olive oils from cv. Picual. *Journal of agricultural and food chemistry*, 52(11), 3434-3440.
- [95] Mailer, R. J., Ayton, J., & Graham, K. (2010). The influence of growing region, cultivar and harvest timing on the diversity of Australian olive oil. *Journal of the American Oil Chemists' Society*, 87(8), 877-884.
- [96] Anastasopoulos, E., Kalogeropoulos, N., Kaliora, A. C., Kountouri, A., & Andrikopoulos, N. K. (2011). The influence of ripening and crop year on quality indices, polyphenols, terpenic acids, squalene, fatty acid profile, and sterols in virgin olive oil (Koroneiki cv.) produced by organic versus non-organic cultivation method. *International Journal of Food Science & Technology*, 46(1), 170-178.
- [97] Freihat, N. M., Al-Shannag, A. K., & El Assi, N. (2008). Qualitative responses of “Nabali” olive oil to harvesting time and altitudes at sub-humid Mediterranean. *International Journal of Food Properties*, 11(3), 561-570.
- [98] Famiani, F., Proietti, P., Farinelli, D., & Tombesi, A. (2000, September). Oil quality in relation to olive ripening. In *IV International Symposium on Olive Growing 586* (pp. 671-674).
- [99] Casado, F. H., NAVAS FERNANDEZ, M. A., Guinda Garín, A., Ruiz Gómez, A., Camacho, M. L., Lanzón Rey, A., ... & GARCIANI CONSTANTE, E. (1993). La calidad del aceite de oliva virgen: posibles nuevos criterios para su evaluación. *Grasas y aceites*, 44(1), 10-17.
- [100] Dag, A., Kerem, Z., Yogev, N., Zipori, I., Lavee, S., & Ben-David, E. (2011). Influence of time of harvest and maturity index on olive oil yield and quality. *Scientia Horticulturae*, 127(3), 358-366.
- [101] Oteros, J., García-Mozo, H., Vázquez, L., Mestre, A., Domínguez-Vilches, E., & Galán, C. (2013). Modelling olive phenological response to weather and topography. *Agriculture, ecosystems & environment*, 179, 62-68.
- [102] Abaza, L., Msallem, M., Daoud, D., & Zarrouk, M. (2002). Caractérisation des huiles de sept variétés d'olivier tunisiennes. *Oléagineux, Corps Gras, Lipides*, 9(2), 174-179.
- [103] D'Imperio, M., Dugo, G., Alfa, M., Mannina, L., & Segre, A. L. (2007). Statistical analysis on Sicilian olive oils. *Food Chemistry*, 102(3), 956-965.
- [104] Nicholson, J. K., Lindon, J. C., & Holmes, E. (1999). Metabonomics. *understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data*. *Xenobiotica*, 29(11), 1181-1189.
- [105] Oliver, S. G., Winson, M. K., Kell, D. B., & Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends in biotechnology*, 16(9), 373-378.
- [106] Fiehn, O. (2002). Metabolomics—the link between genotypes and phenotypes. *Plant molecular biology*, 48(1-2), 155-171.
- [107] Kind, T., Tolstikov, V., Fiehn, O., & Weiss, R. H. (2007). A comprehensive urinary metabolomic approach for identifying kidney cancer. *Analytical biochemistry*, 363(2), 185-195.
- [108] Sreekumar, A., Poisson, L. M., Rajendiran, T. M., Khan, A. P., Cao, Q., Yu, J., ... & Chinnaiyan, A. M. (2009). Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature*, 457(7231), 910-914.
- [109] German, J. B., Roberts, M. A., Fay, L., & Watkins, S. M. (2002). Metabolomics and individual metabolic assessment: the next great challenge for nutrition. *The Journal of nutrition*, 132(9), 2486-2487.

- [110] Jones, O. A., Sdepanian, S., Lofts, S., Svendsen, C., Spurgeon, D. J., Maguire, M. L., & Griffin, J. L. (2014). Metabolomic analysis of soil communities can be used for pollution assessment. *Environmental toxicology and chemistry*, 33(1), 61-64.
- [111] López-Gresa, M. P., Lisón, P., Kim, H. K., Choi, Y. H., Verpoorte, R., Rodrigo, I., ... & Bellés, J. M. (2012). Metabolic fingerprinting of tomato mosaic virus infected *Solanum lycopersicum*. *Journal of plant physiology*, 169(16), 1586-1596.
- [112] Verpoorte, R., Choi, Y. H., Mustafa, N. R., & Kim, H. K. (2008). Metabolomics: back to basics. *Phytochemistry Reviews*, 7(3), 525-537.
- [113] REGULATION, H. A. T. 15. Official Journal of the European Communities No L 12/11.
- [114] Armougom, R., Grondin, I., & Smadja, J. (1998). Composition en acides gras des extraits lipidiques de quelques graines de cucurbitacées tropicales. *OCL. Oléagineux, corps gras, lipides*, 5(4), 323-327.
- [115] Hussein, L., Ali, M., Abouelhassan, A., Grzeskiewicz, S., & Cantellops, D. (2001). Assessment of the fatty acid patterns in vegetable oils, fats and fat-rich foods commonly consumed in Egypt. *Grasas y Aceites*, 52(3-4), 163-170.
- [116] Sato, T. (1994). Application of principal-component analysis on near-infrared spectroscopic data of vegetable oils for their classification. *Journal of the American Oil Chemists' Society*, 71(3), 293-298.
- [117] Bertran, E., Blanco, M., Coello, J., Iturriaga, H., MasPOCH, S., & Montoliu, I. (2000). Near infrared spectrometry and pattern recognition as screening methods for the authentication of virgin olive oils of very close geographical origins. *Journal of Near Infrared Spectroscopy*, 8(1), 45-52.
- [118] Beckonert, O., Keun, H. C., Ebbels, T. M., Bundy, J., Holmes, E., Lindon, J. C., & Nicholson, J. K. (2007). Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols*, 2(11), 2692-2703.
- [119] Luykx, D. M., & Van Ruth, S. M. (2008). An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry*, 107(2), 897-911.
- [120] Kim, H. K., Choi, Y. H., & Verpoorte, R. (2011). NMR-based plant metabolomics: where do we stand, where do we go?. *Trends in biotechnology*, 29(6), 267-275.
- [121] Seger, C., & Sturm, S. (2007). Analytical aspects of plant metabolite profiling platforms: current standings and future aims. *Journal of proteome research*, 6(2), 480-497.
- [122] Hatzakis, E., Koidis, A., Boskou, D., & Dais, P. (2008). Determination of phospholipids in olive oil by ³¹P NMR spectroscopy. *Journal of agricultural and food chemistry*, 56(15), 6232-6240.
- [123] Spyros, A., & Dais, P. (2000). Application of ³¹P NMR spectroscopy in food analysis. 1. Quantitative determination of the mono-and diglyceride composition of olive oils. *Journal of agricultural and food chemistry*, 48(3), 802-805.
- [124] Petrakis, P. V., Agiomyrgianaki, A., Christophoridou, S., Spyros, A., & Dais, P. (2008). Geographical characterization of Greek virgin olive oils (cv. Koroneiki) using ¹H and ³¹P NMR fingerprinting with canonical discriminant analysis and classification binary trees. *Journal of agricultural and food chemistry*, 56(9), 3200-3207.
- [125] Siciliano, C., Belsito, E., De Marco, R., Di Gioia, M. L., Leggio, A., & Liguori, A. (2013). Quantitative determination of fatty acid chain composition in pork meat products by high resolution ¹H NMR spectroscopy. *Food chemistry*, 136(2), 546-554.

- [126] Miyake, Y., Yokomizo, K., & Matsuzaki, N. (1998). Determination of unsaturated fatty acid composition by high-resolution nuclear magnetic resonance spectroscopy. *Journal of the American Oil Chemists' Society*, 75(9), 1091-1094.
- [127] Vlahov, G., Shaw, A. D., & Kell, D. B. (1999). Use of ^{13}C nuclear magnetic resonance distortionless enhancement by polarization transfer pulse sequence and multivariate analysis to discriminate olive oil cultivars. *Journal of the American Oil Chemists' Society*, 76(10), 1223-1231.
- [128] Mavromoustakos, T., Zervou, M., Theodoropoulou, E., Panagiotopoulos, D., Bonas, G., Day, M., & Helmis, A. (1997). ^{13}C NMR analysis of the triacylglycerol composition of Greek virgin olive oils. *Magnetic resonance in chemistry*, 35(13), S3-S7.
- [129] Mannina, L., Luchinat, C., Emanuele, M. C., & Segre, A. (1999). Acyl positional distribution of glycerol tri-esters in vegetable oils: a ^{13}C NMR study. *Chemistry and physics of lipids*, 103(1), 47-55.
- [130] Giraudeau, P., Tea, I., Remaud, G. S., & Akoka, S. (2014). Reference and normalization methods: Essential tools for the intercomparison of NMR spectra. *Journal of pharmaceutical and biomedical analysis*, 93, 3-16.
- [131] Gillet, S., & Delpuech, J. J. (1980). Optimum conditions for nondestructive quantitative analysis by carbon-13 NMR. *Journal of Magnetic Resonance (1969)*, 38(3), 433-445.
- [132] Rodrigues, C. I., Maia, R., Miranda, M., Ribeirinho, M., Nogueira, J. M. F., & Máguas, C. (2009). Stable isotope analysis for green coffee bean: A possible method for geographic origin discrimination. *Journal of Food Composition and Analysis*, 22(5), 463-471.
- [133] Wang, J., & Li, Q. X. (2011). Chemical composition, characterization, and differentiation of honey botanical and geographical origins. *Advances in food and nutrition research*, 62, 89-137.
- [134] Longobardi, F., Casiello, G., Sacco, D., Tedone, L., & Sacco, A. (2011). Characterisation of the geographical origin of Italian potatoes, based on stable isotope and volatile compound analyses. *Food Chemistry*, 124(4), 1708-1713.
- [135] Guyon, F., Auberger, P., Gaillard, L., Loublanches, C., Viateau, M., Sabathié, N., ... & Médina, B. (2014). $^{13}\text{C}/^{12}\text{C}$ isotope ratios of organic acids, glucose and fructose determined by HPLC-co-IRMS for lemon juices authenticity. *Food chemistry*, 146, 36-40.
- [136] Košir, I. J., Kocjančič, M., Ogrinc, N., & Kidrič, J. (2001). Use of SNIF-NMR and IRMS in combination with chemometric methods for the determination of chaptalisation and geographical origin of wines (the example of Slovenian wines). *Analytica Chimica Acta*, 429(2), 195-206.
- [137] Drivelos, S. A., & Georgiou, C. A. (2012). Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. *TrAC Trends in Analytical Chemistry*, 40, 38-51.
- [138] Martin, G. J. (1988). Les applications de la mesure par resonance magnetique nucleaire du fractionnement isotopique naturel specifique (RMN-FINS) en viticulture et oenologie. *Revue Francaise d'Oenologie*.
- [139] Angerosa, F., Breas, O., Contento, S., Guillou, C., Reniero, F., & Sada, E. (1999). Application of stable isotope ratio analysis to the characterization of the geographical origin of olive oils. *Journal of Agricultural and Food Chemistry*, 47(3), 1013-1017.

- [140] Camin, F., Larcher, R., Perini, M., Bontempo, L., Bertoldi, D., Gagliano, G., ... & Versini, G. (2010). Characterisation of authentic Italian extra-virgin olive oils by stable isotope ratios of C, O and H and mineral composition. *Food chemistry*, 118(4), 901-909.
- [141] Alonso-Salces, R. M., Moreno-Rojas, J. M., Holland, M. V., Reniero, F., Guillou, C., & Héberger, K. (2010). Virgin olive oil authentication by multivariate analyses of ^1H NMR fingerprints and $\delta^{13}\text{C}$ and $\delta^2\text{H}$ data. *Journal of agricultural and food chemistry*, 58(9), 5586-5596.
- [142] Iacumin, P., Bernini, L., & Boschetti, T. (2009). Climatic factors influencing the isotope composition of Italian olive oils and geographic characterisation. *Rapid Communications in Mass Spectrometry*, 23(3), 448-454.
- [143] Remaud, G. S & Akoka, S. (2004). L'Isotope : Traceur d'Origine. Distribution isotopique dans les composés naturels.
- [144] Coplen, T. B. (2011). Guidelines and recommended terms for expression of stable-isotope-ratio and gas-ratio measurement results. *Rapid communications in mass spectrometry*, 25(17), 2538-2560.
- [145] Caytan, E., Cherghaoui, Y., Barril, C., Jouitteau, C., Rabiller, C., & Remaud, G. S. (2006). Strategy for specific isotope ratio determination by quantitative NMR on symmetrical molecules: application to glycerol. *Tetrahedron: Asymmetry*, 17(11), 1622-1624.
- [146] Tenailleau, E., & Akoka, S. (2007). Adiabatic ^1H decoupling scheme for very accurate intensity measurements in ^{13}C NMR. *Journal of Magnetic Resonance*, 185(1), 50-58.
- [147] Botosoa, E. P. (2008). *Protocoles analytiques pour l'étude de la vanilline par RMN ^{13}C isotopique en abondance naturelle: reproductibilité méthodologique, purification, origines du fractionnement isotopique* (Doctoral dissertation, Nantes).
- [148] Caytan, E., Botosoa, E. P., Silvestre, V., Robins, R. J., Akoka, S., & Remaud, G. S. (2007). Accurate quantitative ^{13}C NMR spectroscopy: repeatability over time of site-specific ^{13}C isotope ratio determination. *Analytical chemistry*, 79(21), 8266-8269.
- [149] Tenailleau, E., Lancelin, P., Robins, R. J., & Akoka, S. (2004). NMR approach to the quantification of nonstatistical ^{13}C distribution in natural products: Vanillin. *Analytical chemistry*, 76(13), 3818-3825.
- [150] Tenailleau, E. J., Lancelin, P., Robins, R. J., & Akoka, S. (2004). Authentication of the origin of vanillin using quantitative natural abundance ^{13}C NMR. *Journal of agricultural and food chemistry*, 52(26), 7782-7787.
- [151] Silvestre, V., Mboula, V. M., Jouitteau, C., Akoka, S., Robins, R. J., & Remaud, G. S. (2009). Isotopic ^{13}C NMR spectrometry to assess counterfeiting of active pharmaceutical ingredients: site-specific ^{13}C content of aspirin and paracetamol. *Journal of pharmaceutical and biomedical analysis*, 50(3), 336-341.
- [152] Morris, G. A., & Freeman, R. (1979). Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society*, 101(3), 760-762.
- [153] Bardet, M., Foray, M. F., & Robert, D. (1985). Use of the DEPT pulse sequence to facilitate the ^{13}C NMR structural analysis of lignins. *Die Makromolekulare Chemie*, 186(7), 1495-1504.
- [154] Vlahov, G., Schiavone, C., & Simone, N. (2001). Quantitative ^{13}C NMR method using the DEPT pulse sequence for the determination of the geographical origin (DOP) of olive oils. *Magnetic Resonance in Chemistry*, 39(11), 689-695.

- [155] Vlahov, G., Schiavone, C., & Simone, N. (2001). Quantitative ^{13}C NMR method using the DEPT pulse sequence for the determination of the geographical origin (DOP) of olive oils. *Magnetic Resonance in Chemistry*, 39(11), 689-695.
- [156] Vlahov, G., Del Re, P., & Simone, N. (2003). 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. *Journal of agricultural and food chemistry*, 51(19), 5612-5615.
- [157] Vlahov, G. (1997). Quantitative ^{13}C NMR method using the DEPT pulse sequence for the detection of olive oil adulteration with soybean oil. *Magnetic resonance in chemistry*, 35(13), S8-S12.
- [158] Karabulut, N., Baguet, E., Trierweiler, M., & Akoka, S. (2002). Improvement in quantitative accuracy of ^{13}C DEPT integrals by parameter-optimization. *Analytical letters*, 35(15), 2549-2563.
- [159] Thibaudeau, C., Remaud, G., Silvestre, V., & Akoka, S. (2010). Performance evaluation of quantitative adiabatic ^{13}C NMR pulse sequences for site-specific isotopic measurements. *Analytical chemistry*, 82(13), 5582-5590.
- [160] Bussy, U., Thibaudeau, C., Thomas, F., Desmurs, J. R., Jamin, E., Remaud, G. S., ... & Akoka, S. (2011). Isotopic finger-printing of active pharmaceutical ingredients by ^{13}C NMR and polarization transfer techniques as a tool to fight against counterfeiting. *Talanta*, 85(4), 1909-1914.
- [161] Wishart, D. S. (2008). Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry*, 27(3), 228-237.
- [162] Lindon, J. C., Nicholson, J. K., Holmes, E., & Everett, J. R. (2000). Metabonomics: metabolic processes studied by NMR spectroscopy of biofluids. *Concepts in Magnetic Resonance*, 12(5), 289-320.
- [163] Zhang, S., Gowda, G. N., Asiago, V., Shanaiah, N., Barbas, C., & Raftery, D. (2008). Correlative and quantitative ^1H NMR-based metabolomics reveals specific metabolic pathway disturbances in diabetic rats. *Analytical biochemistry*, 383(1), 76-84.
- [164] Le Grand, F., George, G., & Akoka, S. (2005). Natural abundance ^2H -ERETIC-NMR authentication of the origin of methyl salicylate. *Journal of agricultural and food chemistry*, 53(13), 5125-5129.
- [165] Pauli, G. F. (2001). qNMR a versatile concept for the validation of natural product reference compounds. *Phytochemical Analysis*, 12(1), 28-42.
- [166] Martineau, E., Akoka, S., Boisseau, R., Delanoue, B., & Giraudeau, P. (2013). Fast Quantitative ^1H - ^{13}C Two-Dimensional NMR with Very High Precision. *Analytical chemistry*, 85(9), 4777-4783.
- [167] Koskela, H., Kilpeläinen, I., & Heikkinen, S. (2005). Some aspects of quantitative 2D NMR. *Journal of Magnetic Resonance*, 174(2), 237-244.
- [168] Morris, G. A. (1992). Systematic sources of signal irreproducibility and t1 noise in high-field NMR spectrometers. *Journal of Magnetic Resonance (1969)*, 100(2), 316-328.
- [169] Giraudeau, P. (2008). *Développement de la RMN quantitative multi-dimensionnelle* (Doctoral dissertation, Nantes).

- [170] Jeannerat, D. (2003). High resolution in heteronuclear ^1H - ^{13}C NMR experiments by optimizing spectral aliasing with one-dimensional carbon data. *Magnetic Resonance in Chemistry*, 41(1), 3-17.
- [171] Jeannerat, D. (2007). Computer optimized spectral aliasing in the indirect dimension of ^1H - ^{13}C heteronuclear 2D NMR experiments. A new algorithm and examples of applications to small molecules. *Journal of Magnetic Resonance*, 186(1), 112-122.
- [172] Vitorge, B., Bieri, S., Humam, M., Christen, P., Hostettmann, K., Muñoz, O., ... & Jeannerat, D. (2009). High-precision heteronuclear 2D NMR experiments using 10-ppm spectral window to resolve carbon overlap. *Chemical communications*, (8), 950-952.
- [173] Ramírez-Gualito, K., & Jeannerat, D. (2015). Exploiting the phase of NMR signals to carry useful information. Application to the measurement of chemical shifts in aliased 2D spectra. *Magnetic Resonance in Chemistry*, 53(11), 901-907.
- [174] Hyberts, S. G., Arthanari, H., Robson, S. A., & Wagner, G. (2014). Perspectives in magnetic resonance: NMR in the post-FFT era. *Journal of Magnetic Resonance*, 241, 60-73.
- [175] Hyberts, S. G., Heffron, G. J., Tarragona, N. G., Solanky, K., Edmonds, K. A., Luithardt, H., ... & Falchuk, K. H. (2007). Ultrahigh-resolution ^1H - ^{13}C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction. *Journal of the American Chemical Society*, 129(16), 5108-5116.
- [176] Rai, R. K., & Sinha, N. (2012). Fast and Accurate Quantitative Metabolic Profiling of Body Fluids by Nonlinear Sampling of ^1H - ^{13}C Two-Dimensional Nuclear Magnetic Resonance Spectroscopy. *Analytical chemistry*, 84(22), 10005-10011.
- [177] Barna, J. C. J., Laue, E. D., Mayger, M. R., Skilling, J., & Worrall, S. J. P. (1987). Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *Journal of Magnetic Resonance (1969)*, 73(1), 69-77.
- [178] Holland, D. J., Bostock, M. J., Gladden, L. F., & Nietlispach, D. (2011). Fast multidimensional NMR spectroscopy using compressed sensing. *Angewandte Chemie*, 123(29), 6678-6681.
- [179] Thiele, C. M., & Bermel, W. (2012). Speeding up the measurement of one-bond scalar (1 J) and residual dipolar couplings (1 D) by using non-uniform sampling (NUS). *Journal of Magnetic Resonance*, 216, 134-143.
- [180] Le Guennec, A., Dumez, J. N., Giraudeau, P., & Caldarelli, S. (2015). Resolution-enhanced 2D NMR of complex mixtures by non-uniform sampling. *Magnetic Resonance in Chemistry*, 53(11), 913-920.
- [181] Barkhuijsen, H., De Beer, R., Bovee, W. M. M. J., & Van Ormondt, D. (1985). Retrieval of frequencies, amplitudes, damping factors, and phases from time-domain signals using a linear least-squares procedure. *Journal of Magnetic Resonance (1969)*, 61(3), 465-481.
- [182] Stern, A. S., Li, K. B., & Hoch, J. C. (2002). Modern spectrum analysis in multidimensional NMR spectroscopy: comparison of linear-prediction extrapolation and maximum-entropy reconstruction. *Journal of the American Chemical Society*, 124(9), 1982-1993.
- [183] Eddib, O., & Nickless, G. (1987). Elucidation of olive oil classification by chemometrics. *Analyst*, 112(4), 391-395.

- [184] Alonso Garcça, M. V., & Aparicio Lopez, R. (1993). Characterization of European virgin olive oils using fatty acids. *Grasas y aceites*, 44(1), 18-24.
- [185] Tsimidou, M., & Karakostas, K. (1993). Geographical classification of Greek virgin olive oil by non-parametric multivariate evaluation of fatty acid composition. *Journal of the Science of Food and Agriculture*, 62(3), 253-257.
- [186] Vlahov, G., & Angelo, C. S. (1996). The Structure of Triglycerides of Monovarietal Olive Oils: A ^{13}C -NMR Comparative Study. *Lipid/Fett*, 98(6), 203-205.
- [187] Shaw, A. D., di Camillo, A., Vlahov, G., Jones, A., Bianchi, G., Rowland, J., & Kell, D. B. (1997). Discrimination of the variety and region of origin of extra virgin olive oils using ^{13}C NMR and multivariate calibration with variable reduction. *Analytica Chimica Acta*, 348(1), 357-374.
- [188] Sacco, A., Brescia, M. A., Liuzzi, V., Reniero, F., Guillou, G., Ghelli, S., & van der Meer, P. (2000). Characterization of Italian olive oils based on analytical and nuclear magnetic resonance determinations. *Journal of the American Oil Chemists' Society*, 77(6), 619-625.
- [189] Mannina, L., Fontanazza, G., Patumi, M., Ansanelli, G., & Segre, A. (2001a). Italian and Argentine olive oils: a NMR and gas chromatographic study. *Grasas y Aceites*, 52(6), 380-388.
- [190] Mannina, L., Patumi, M., Proietti, N., & Segre, A. L. (2001b). PDO (protected designation of origin): geographical characterization of Tuscan extra virgin olive oils using high-field ^1H NMR spectroscopy. *Italian journal of food science*, 13(1), 53-63.
- [191] Tsimidou, M., Macrae, R., & Wilson, I. (1987). Authentication of virgin olive oils using principal component analysis of triglyceride and fatty acid profiles: Part 1-Classification of Greek olive oils. *Food Chemistry*, 25(3), 227-239.
- [192] Brescia, M. A., Alviti, G., Liuzzi, V., & Sacco, A. (2003). Chemometric classification of olive cultivars based on compositional data of oils. *Journal of the American Oil Chemists' Society*, 80(10), 945-950.
- [193] Downey, G., McIntyre, P., & Davies, A. N. (2003). Geographic classification of extra virgin olive oils from the eastern Mediterranean by chemometric analysis of visible and near-infrared spectroscopic data. *Applied spectroscopy*, 57(2), 158-163.
- [194] Aranda, F., Gomez-Alonso, S., Del Álamo, R. R., Salvador, M. D., & Fregapane, G. (2004). Triglyceride, total and 2-position fatty acid composition of Cornicabra virgin olive oil: Comparison with other Spanish cultivars. *Food Chemistry*, 86(4), 485-492.
- [195] Ollivier, D., Pinatel, C., Dupuy, N., Guérère, M., & Artaud, J. (2007). Caractérisations sensorielles et chimiques d'huiles d'olive vierges de six AOC françaises. *Oléagineux, Corps gras, Lipides*, 14(2), 116-129.
- [196] Longobardi, F., Ventrella, A., Casiello, G., Sacco, D., Tasioula-Margari, M., Kiritsakis, A. K., & Kontominas, M. G. (2012). Characterisation of the geographical origin of Western Greek virgin olive oils based on instrumental and multivariate statistical analysis. *Food chemistry*, 133(1), 169-175.
- [197] Rezzi, S., Axelson, D. E., Héberger, K., Reniero, F., Mariani, C., & Guillou, C. (2005). Classification of olive oils using high throughput flow ^1H NMR fingerprinting with principal component analysis, linear discriminant analysis and probabilistic neural networks. *Analytica Chimica Acta*, 552(1), 13-24.
- [198] Mannina, L., D'IMPERIO, M., Lava, R., Schievano, E., & Mammi, S. (2005). Caratterizzazione NMR e analisi statistica di oli di oliva DOP veneti. *Rivista Italiana delle Sostanze Grasse*, 82(2), 59-63.

-
- [199] Schievano, E., Arosio, I., Lava, R., Simionato, V., Mammi, S., & Consonni, R. (2006). Olio di oliva DOP del lago di Garda: uno studio NMR e analisi statistica multivariata. *Rivista Italiana delle Sostanze Grasse*, 83(1), 14-17.
- [200] Ollivier, D., Artaud, J., Pinatel, C., Durbec, J. P., & Guérére, M. (2006). Differentiation of French virgin olive oil RDOs by sensory characteristics, fatty acid and triacylglycerol compositions and chemometrics. *Food Chemistry*, 97(3), 382-393.
- [201] Galtier, O., Dupuy, N., Le Dréau, Y., Ollivier, D., Pinatel, C., Kister, J., & Artaud, J. (2007). Geographic origins and compositions of virgin olive oils determined by chemometric analysis of NIR spectra. *Analytica chimica acta*, 595(1), 136-144.
- [202] Mannina, L., Marini, F., Gobino, M., Sobolev, A. P., & Capitani, D. (2010). NMR and chemometrics in tracing European olive oils: The case study of Ligurian samples. *Talanta*, 80(5), 2141-2148.
- [203] García-González, D. L., Tena, N., & Aparicio, R. (2012). Describing the chemical singularity of the Spanish protected designations of origin for virgin olive oils in relation to oils from neighbouring areas. *grasas y aceites*, 63(1), 26-34.
- [204] Abdallah, M., Vergara-Barberán, M., Lerma-García, M. J., Herrero-Martínez, J. M., Simó-Alfonso, E. F., & Guerfel, M. (2015). Classification of Tunisian extra virgin olive oils according to their genetic variety and maturity index using fatty acid profiles established by direct infusion mass spectrometry. *European Journal of Lipid Science and Technology*.
- [205] Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics.
- [206] Ramousse, R., Berre, M. L., & Guelte, L. L. (1996). Introduction aux statistiques. *disponible sur le lien <http://www.cons-dev.org/elearning/stat/index.html>*. 02-01-2016
- [207] Trygg, J., Holmes, E., & Lundstedt, T. (2007). Chemometrics in metabonomics. *Journal of proteome research*, 6(2), 469-479.
- [208] Klecka, W. R. (1980). *Discriminant analysis* (No. 19). Sage.
- [209] Vancolen, S. (2004). *La régression PLS* (Doctoral dissertation, Université de Neuchâtel).
- [210] Bélanger, M., El-Jabi, N., Caissie, D., Ashkar, F., & Ribí, J. M. (2005). Estimation de la température de l'eau de rivière en utilisant les réseaux de neurones et la régression linéaire multiple. *Revue des sciences de l'eau/Journal of Water Science*, 18(3), 403-421.
- [211] Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2), 109-130.
- [212] Desbois, D. (1999). Introduction à la régression des moindres carrés partiels avec la procédure PLS de SAS. *Modulad*, 24, 41-97.
- [213] Roy, P. P., & Roy, K. (2008). On some aspects of variable selection for partial least squares regression models. *QSAR & Combinatorial Science*, 27(3), 302-313.
- [214] Anderssen, E., Dyrstad, K., Westad, F., & Martens, H. (2006). Reducing over-optimism in variable selection by cross-model validation. *Chemometrics and intelligent laboratory systems*, 84(1), 69-74.

[215] Hawkins, D. M., Basak, S. C., & Mills, D. (2003). Assessing model fit by cross-validation. *Journal of chemical information and computer sciences*, 43(2), 579-586.

Thèse de Doctorat

Noelle MERCHAK

Caractérisation et traçabilité des matrices riches en triacylglycérols par des techniques isotopiques et RMN métabolomique

Characterisation and traceability of mat

Résumé

De nos jours, les fraudes et les adultérations n'ont pas de frontières ; pour cette raison, la vérification de l'origine des aliments à risque doit être assurée par des méthodes fiables et standardisées. Les triacylglycérols (TAG) sont considérés comme des composants quasi universels des matrices alimentaires; leurs profils métabolomique et isotopique fournissent des marqueurs d'origine efficaces pour l'authentification. À cet égard, une étude métabolomique par RMN ^1H a été réalisée sur 242 échantillons d'olives collectés de différentes régions libanaises. L'analyse statistique des données ainsi obtenues met en évidence la variation de la composition des huiles en fonction de leurs coordonnées géographiques, morphologiques et temporelles. Dans une seconde étape, nous avons exploité les potentialités de la RMN ^{13}C . Une séquence INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) adiabatique a été optimisée pour les TAG, permettant simultanément les analyses isotopique et métabolomique (isotopomie) de ces mélanges. Par la suite, une stratégie pour la quantification des acides gras individuels a été développée et testée sur les huiles d'olive en utilisant l'INEPT. En comparaison avec la chromatographie en phase gazeuse et la RMN ^1H , la séquence INEPT optimisée a permis une meilleure classification des huiles d'olive au niveau sous-régional. Enfin, une séquence à deux dimensions HSQC (Heteronuclear Single Quantum Coherence spectroscopy) a été développée et optimisée pour les huiles végétales, des informations isotopomiques peuvent être obtenues rapidement ce qui a permis de classer les huiles selon leurs origines géographiques et botaniques.

Mots clés

Triacylglycérols – Huile d'olive – Authentification – Isotopomique – RMN ^1H – INEPT – HSQC – Analyse statistique

Abstract

Nowadays, fraud and adulteration have no borders; for this reason, verifying the origin of food products at risk must be ensured by reliable and standardized methods. Triacylglycerols (TAG) are considered as quasi universal components of food matrices; metabolomic and ^{13}C isotopic profiling of these molecules afford very efficient origin tracers used for authentication issues. In this context, a metabolomic study by ^1H NMR was carried out on 242 olive samples collected from different Lebanese regions. The statistical analysis of the data thus obtained has revealed significant variations of different parameters within the oil matrix in function of the geographical origin of the oil, the olive morphology, the harvest date, and the production year. In a second step, we have explored potentialities of ^{13}C NMR. An adiabatic INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) sequence was optimized for the TAG matrices providing simultaneously their position-specific ^{13}C content and metabolomic profile (isotopomics). Based on this method, a strategy for the quantification of individual fatty acids was developed and tested on olive oils. In comparison with gas chromatography and ^1H NMR, the optimized INEPT sequence has provided a better classification of olive oils at subregional level. Finally, an optimized 2D HSQC (Heteronuclear Single Quantum Coherence spectroscopy) sequence was developed for vegetable oils. Using this analytic tool, valuable isotopic information can be rapidly reached allowing the classification of oils according to their geographical and botanical origins.

Key Words

Triacylglycerols – Olive oil – Authentication – Isotopic – ^1H NMR – INEPT – HSQC – Statistical analysis