UCLouvain

AGRO LOUVAIN

Faculté des bioingénieurs

Integrated approach to assess the freshness and authenticity of salmon fillets: a combination of Gas Chromatography and Raman Spectroscopy

Auteur : Authom Alexis
Promoteur(s) : Dr. Fernández Pierna Juan Antonio, Dr. Ir. Baeten Vincent
Lecteur(s) : Prof. Dr. Ir. Debier Cathy, Prof. Dr. Rees Jean-François
Année académique 2022-2023
Mémoire de fin d'études présenté en vue de l'obtention du diplôme de Bioingénieur : Sciences Agronomiques

Acknowledgements

A thousand thanks to Juan Antonio Fernandez Pierna, my supervisor, who introduced me to the world of research, shared his knowledge in research and statistics, and generously devoted some of his time to assisting me at every stage of this thesis. I am grateful for his guidance, rigour, and seriousness, which were invaluable throughout this journey.

A heartfelt thank you to Vincent Baeten, my thesis co-supervisor, for proposing the topic and for finding time to share his extensive knowledge and experience with me throughout the completion of this work. I am especially grateful for the opportunity to work within the CRA-W laboratories. It has been an immensely enriching experience, both academically and professionally, as well as on a personal level.

A special thanks to Quentin Arnould, who patiently introduced me to the fascinating world of spectroscopy, laboratory work, and data analysis. His expertise in these three domains was of great help, and his skilful guidance redirected me toward a better path more than once.

I also extend my gratitude to Delphine Delhotte for the time she dedicated to welcoming me and for her kindness and careful attention to my work. I would like to thank Nicaise Kayoka Mukendi as well for his unwavering wisdom and invaluable teachings on data analysis. In general, I am grateful to Quentin, Delphine, and Nicaise for warmly embracing me in their workspace and making me feel like a valued colleague.

Many thanks to Octave for providing significant assistance during the gas chromatography manipulations and for helping me refine the protocol. I also appreciate his guidance on chromatographic data processing.

I extend my thanks to Antoine Deryck for proofreading the statistical section of my thesis and providing accurate and precise feedback.

I am grateful to Cathy Debier and Jean-François Rees for accepting to be part of my master thesis jury, reading, and evaluating my work.

A huge thank you to all the wonderful staff at CRA-W for welcoming me into an unfamiliar environment. I admire their dedication and efficiency in carrying out their work, always striving for improvement, and creating a very human and caring atmosphere.

Table of contents

1 Introduction 1

1.1 Salmon 1

1.1.1 Farmed salmon 1

1.1.2 Organic and conventional farmed salmon 2

1.1.3 Wild salmon 2

1.2 Mislabelling 3

1.2.1 What motivates mislabelling? 3

1.3 Fraud impacts 4

1.4 How mislabelling is prevented? 5

1.4.1 Regulations 5

1.4.2 Analytical methods used against mislabelling 6

1.4.3 The wild and farmed salmon case 6

1.4.4 An alternative analytical method: Vibrational spectroscopy 10

1.5 Potential of Raman 12

1.5.1 Raman spectroscopy on salmon 12

1.5.2 Raman for discrimination between wild and farmed salmon 12

2 Objectives of the Study 12

3 Material and Methods 13

3.1 Raman Spectroscopy 13

3.1.1 Theory: Raman Vibrational Spectroscopy 13

3.1.2 Material 15

3.1.3 Sampling 17

3.1.4 Spectra acquisition 18

3.2 Gas Chromatography on Fatty Acids 20

3.2.1 Material 20

3.2.2 Method 20

3.3 Chemometrics 21

3.3.1 Data cleaning 21

3.3.2 Pre-processing 23

3.3.3 Principal Components Analysis (PCA) 29

3.3.4 Partial Least Square (PLS) 31

4 Results and Discussion 35
4.1 Gas Chromatography Results 35
4.1.1 The Fatty acids profile of Salmon 35
4.1.2 FAs in wild and farmed Salmon 37
4.1.3 Comparison of SFA, MUFA and PUFA in wild and farmed Salmon 38
4.1.4 Comparison of ω-6 and ω-3 FA in wild and farmed Salmon 39
4.1.5 Using PCA to detect FAs tendencies between wild and farmed Salmon. 40
4.2 Raman Spectra of Salmon 44
4.2.1 General salmon spectra, major bands and interpretation 44
4.3 Salmon Spectra by Classes 47
4.4 Challenges linked to the analysis of Salmon fillets by Raman spectroscopy 48
4.4.1 Reproducibility and Repeatability Studies 48
4.4.2 Heterogeneity in a Salmon Fillet 50
4.5 Fresh and thawed salmon samples Analysis 53
4.5.1 Thawing influence within a single sample 54
4.5.2 General thawing influence within 4 samples 56
4.5.3 Fresh and thawed samples PLS-DA 59
4.6 Wild and Farmed salmon analysis 62
4.6.1 PCA of wild, conventionally and organic farmed salmon spectra 62
4.6.2 PLS-DA model for wild, conventionally, and organic farmed salmon spectra. 66
4.6.3 A potential way to discriminate between wild and farmed salmon: Prediction of FAs concentration using Raman spectroscopy. 67
4.6.4 PLS model for FAs prediction 67
5 General Discussion 75
5.1 Feasibility of Raman measurement of salmon fillets. 75
5.2 Wild, organically, and conventionally farmed samples 76
5.3 Fresh and thawed samples 77
6 Conclusion 78
7 Perspectives 79
8 References 80

Summary (in EN and FR)

(EN) We are living in a world where food security is a global challenge, and where consumers' choices are continuously changing in terms of specific food quality. Routine methods for characterising and authenticating globally traded products like salmon are necessary. Still, today, controlling the fraud on the salmon market is difficult. In this context, Raman spectroscopy is rapidly advancing; it is among the fast, non-destructive, and cost-effective techniques highlighted in quality control. Coincidently, reliable tools for processing large databases become more accessible in open source.

In this exploratory study, several tests were conducted with Raman spectroscopy to assess the freshness, authenticity, and lipidic profiles associated with 28 salmon samples from different stores in Belgium. The acquisition parameters for spectral data and the feasibility of reliably analysing salmon flesh using Raman spectroscopy, have been discussed. The study also explored some data processing tools that can extract a wealth of interesting information from the salmon spectrum. A part of the study consisted of extracting lipids from the salmon fillets and analysing them using Gas Chromatography (GC). Both spectral and GC data were used to analyse the fatty acid profiles of the samples. The different types of salmon have been discussed, and predictive models targeting authenticity (wild, organically farmed, or conventionally farmed) and freshness (fresh or thawed) were developed and evaluated.

This end of study work shows the potential of Raman spectroscopy for the assessment of the freshness of a salmon fillet. The results also show the possibility to discriminate between wild, conventionally farmed, and organically farmed salmons according to their lipidic profiles, and the potential to evaluate some lipidic features of a salmon fillet with Raman spectroscopy.

This work was realised with the collaboration of the *Centre wallon de Recherches Agronomiques* (CRA-W), giving access to the spectroscopic and chromatographic instruments.

This work is open to perspectives for improvement and can serve as a foundation for further studies.

(FR) Nous vivons dans un monde où la sécurité alimentaire est un problème global, et dans lequel les choix des consommateurs changent continuellement en termes de qualité. Des méthodes de routine pour la caractérisation et l'authentification de produits vendus dans le monde entier, tel que le saumon, est plus que jamais nécessaire. Aujourd'hui encore, il est très difficile de contrôler la fraude sur le marché du saumon. Dans ce contexte, la spectroscopie Raman se développe à une vitesse fulgurante ; elle fait partie, dans le même temps, des méthodes rapides, non destructives et bon marché qui sont mises en avant en contrôle qualité, alors que des outils fiables de traitement de larges bases de données sont de plus en plus accessibles en open-source.

Dans cette étude exploratoire, quelques essais sur l'évaluation de la fraîcheur, de l'authenticité et des profils lipidiques avec la spectroscopie Raman ont été menés sur 28

échantillons de saumon acquis dans différents grands magasins en Belgique. Les paramètres d'acquisition des spectres ont été discutés, ainsi que la faisabilité d'analyser avec fiabilité de la chair saumon avec le Raman. L'étude s'est aussi portée sur l'essai de quelques outils de traitement de données qui permettent d'extraire un maximum d'informations intéressantes sur la chair de saumon. Une partie de l'étude a consisté en l'extraction de la matière grasse des filets de saumon et de son analyse en Chromatographie Gazeuse. A partir des données spectrales ainsi que de données de GC, les différents types de saumons ont été discutés, et des modèles de prédiction visant l'authenticité (saumons sauvages, d'aquaculture bio ou conventionnelle) et la fraicheur (frais ou décongelé) ont pu être construits et évalués.

Ce mémoire montre le potentiel de la spectroscopie Raman pour l'évaluation de la fraîcheur d'un filet de saumon. Les résultats montrent également la possibilité de distinguer les saumons sauvages et élevés en agriculture conventionnelle et biologique en fonction de leurs profils lipidiques, ainsi que le potentiel d'évaluation des caractéristiques lipidiques d'un filet de saumon avec la spectroscopie Raman.

Ce mémoire a été réalisé en collaboration avec le *Centre wallon de Recherches Agronomiques (CRA-W)*, qui a fourni l'accès aux instruments spectroscopiques et chromatographiques.

Ce travail est ouvert à des perspectives d'amélioration et peut servir de base à d'autres études.

List of Figures

Fig. 1. Simplified structure of a fatty acid. A fatty acid is a carboxyl group attached to a hydrocarbon chain (Source: Nagwa). 7
Fig. 2. The chemical structure of a saturated and an unsaturated fatty acid (Source: Nagwa) 7
Fig. 3. Raman microscope spectrometer Senterra II by Bruker used for this study. 15
Fig. 4. Rigaku Progeny 1064nm device used for this study. 16
Fig. 5. Shimadzu GC 2010 Plus used for the Gas Chromatography in this study 20
Fig. 6 : Removal of outliers on a salmon spectral dataset (acquired with 785nm laser and pre-processed with 1st derivative) with k-NN outlier detection algorithm. Figure A is the data before the removal and B after the removal of outliers. 23
Fig. 7. Illustration of the effect of SNV on a salmon spectral dataset. Figure A is a raw dataset, and Figure B is the same dataset after an SNV on each spectrum. 25
Fig. 8. Illustration of the effect of SG smoothing on a Raman spectral dataset of salmon. Figure A is a salmon spectral dataset from wavenumbers 1200 to 1500 cm^{-1}. Figure B is the smoothed dataset using SG smoothing (no derivative) with window size = 9. 27
Fig. 9. Illustration of the effect of SG derivatives on a Raman spectral dataset of salmon. Figure A is a salmon spectrum from wavenumbers 1200 to 1500 cm^{-1}. Figure D is its 1st derivative SG with window size = 21, and Figure E is its 2nd SG derivative with a window size = 21. 28
Fig. 10. Illustration of the effect of the window size of SG smoothing on the Raman spectral dataset of salmon. Figures B and F are both SG smoothed (no derivative) datasets of an original dataset (Image A from Fig.8) from wavenumbers 1200 to 1500 cm^{-1}. Figure A was made using a window size = 9 and Figure B a window size = 31. The Figure B appears to be smoother. ... 29
Fig. 11. Confusion matrix of a PLS-DA with 3 classes. 34
Fig. 12. Oleic acid (OA, C18:1n-9c) proportions in wild (blue), conventionally farmed (orange) and organically farmed (green) salmon samples fat. 38
Fig. 13. DHA (C22:6-n3) proportions in wild (blue), conventionally farmed (orange) and organically farmed (green) salmon samples fat. 39
Fig. 14. ω-3 proportions in wild (orange) and farmed (blue) salmon samples fat. 40
Fig. 15. 1st PCA scores on PC1 (44.94%) and PC3 (12.5%) of conventional farmed (blue), organic farmed (red) and wild (green) salmons (using 26FAs) 41
Fig. 16. PC1 and PC3 loadings of the Fig.15 PC scores graph. A FA with a longer bar has a higher influence on the PC (no matter the positive or negative direction). 41
Fig. 17. 2nd PCA scores on PC1 (84.94%) and PC3 (4.2%) of conventional farmed (blue), organic farmed (red) and wild (green) salmons (using 7 FAs). 42
Fig. 18. 3rd PCA scores on PC1 (84.94%) and PC3 (4.2%) of conventional farmed (blue), organic farmed (red) and wild (green) salmons (using 4 FAs). 44
Fig. 19. Salmon fillet average Raman spectra acquired with 532nm (green), 785nm (red) and 1064nm (blue) lasers. 44
Fig. 20. Average spectra of farmed (blue) and wild (red) salmons acquired with 785nm and pre-processed with SNV. 47
Fig. 21. Average spectra of farmed (blue) and wild (red) salmons acquired with 532nm and pre-processed with SNV. 48
Fig. 22. Orthogonal Distance (OD) vs. Score Distance (SD) of a 10 PC PCA of multiple 785nm Raman measurements of a sample, once an hour for 5 hours (4 spectra by hour so 4 points for each hour on the graph). 49

Fig. 23. Orthogonal Distance (OD) vs. Score Distance (SD) of a 10 PC PCA of multiple 532nm Raman measurement of a sample, analysed once a day for 4 days (36 spectra for a day = 36 points for a day on the graph)..... 50
Fig. 24. Salmon piece used for the heterogeneity study. The referenced zones are the area where the four slices were cut..... 51
Fig. 25. Graph of PC3 and PC4 scores of a PCA on all spectra, coloured by zones of the sampling on the salmon piece of Fig.24. 52
Fig. 26. Loadings of the PC4 from PCA from Fig.25 and its highest loading band annotated (around 1655 cm^{-1})..... 52
Fig. 27. Average Raman spectra of zone 1 (blue), zone 2 (red), zone 3 (green) and zone 4 (yellow) acquired with 532nm laser, from wavenumber 1800 cm^{-1} to 1400 cm^{-1}. The zone between 1670 cm^{-1} to 1645 cm^{-1} is annotated..... 53
Fig. 28. Mean spectra of fresh (blue) and thawed (in red) salmon samples from 1800 cm^{-1} to 700 cm^{-1}. Both spectra were acquired with 532nm laser and pre-processed with SG 1st derivative and SNV. 54
Fig. 29. Scores on PC1 and PC3 of 72 spectra of the same sample, analysed fresh (red) and thawed (blue). 55
Fig. 30. Loadings of PC 1 and 3 of the PCA in Fig.29. 55
Fig. 31. Scores of 4 samples, fresh (red) and thawed (blue) on PC1 and PC2 of the PCA with the 1st pre-processing : SNV and SG 1st derivative (pol = 2, d = 1, window = 7). 56
Fig. 32. Loading of PC1 and PC2 of the PCA with the 1st pre-processing : SNV and SG 1st derivative (pol = 2, d = 1, window = 7). 57
Fig. 33. Scores of 4 samples, fresh (red) and thawed (blue) on PC1 and PC2 of the PCA with the 2nd pre-processing: SNV and SG (pol = 2, d = 2, window = 21). 57
Fig. 34. Loadings of PC1 and PC2 of the PCA with the 2nd pre-processing: SNV and SG 2nd derivative (pol = 2, d = 2, window = 21). 58
Fig. 35. Scores of train (red) and test (blue) sets on PC1 and PC2 of a PCA applied to the whole dataset. Each point is an individual spectrum. 60
Fig. 36. Performance of the fresh/thawed discrimination model based on the Error (1-Accuracy/100) of models composed of 1 to 24 latent variables. The lowest error is given with 3 LV. 60
Fig. 37. Confusion matrix of the prediction of the test set by the fresh/thawed PLS-DA model. An occurrence is an individual sample. 61
Fig. 38. Loadings of the LV1 (red), LV2 (blue) and LV3 (green) of the PLS-DA model to discriminate between fresh and thawed spectra, built from 3 samples. 62
Fig. 39. Scores on PC2 and PC4 of conventionally farmed (red), organically farmed (green) and wild (blue) salmons from PCA on 785 nm Raman spectra from ground samples aliquots. 63
Fig. 40. Loadings of the PC2 (red) and PC4 (blue) from Fig.39. 63
Fig. 41. PCA biplot of spectra acquired on salmon slices with 785 nm laser. Points are coloured by classes: wild (blue), conventionally farmed (red) and organically farmed (green) salmon. 64
Fig. 42. Loadings of PC1 (red) and PC8 (blue) from Fig. 41. 65
Fig. 43. Confusion matrix of classification of the test set by the PLS-DA model for wild, conventionally farmed, and organically farmed salmon. Each occurrence is an individual spectrum..... 67

Fig. 44. Histograms of the density of the FAs data for train (blue) and test (yellow) sets after the Kennard-Stone algorithm splitting for SFA (A), ALA (B), C22:0 (C), ω-3 FAs (D), OA (E), EPA (F) and MUFA (G). 70
Fig. 45. RMSEP on 10 validation datasets vs. the number of latent variables (LVs). The model with 3 LVs has the lowest mean RMSEP. 70
Fig. 46. Loadings of the LV1 and LV2 of the SFA prediction model with 2LV. 71
Fig. 47. Illustrations of the PLS-R model observed values (Y axe) vs. predicted values (X axe) of FA concentration for ALA (A), OA (B), EPA (C), MUFA (D) and SFA (E). The blue line is the "ideal regression line", passing by the origin with a slope of 1. 73
Fig. 48. Loading plot of LV1 (red) and LV2 (blue) of SFA PLS-regression. 74
Fig. 49. Loading plot of LV1 (red), LV2 (blue) and LV3 (green) of MUFA PLS-Regression..... 75

List of Tables

Tab. 1. Table of all salmon samples and their principal characteristics. 17
Tab. 2. Synthetic table of the different measurements made in the study. 19
Tab. 3. Mean fatty acids composition (%) of the wild, conventionally farmed and organic farmed salmon samples. 36
Tab. 4. Correlation matrix of the 7 fatty acids chosen for 2nd PCA. The highest correlation values between the two FAs are in bold. 43
Tab. 5. Major bands found in Fig. 19, their vibration types, and their most probable associated component of the salmon muscle matrix, towards literature. The relative intensities of the bands for each laser are given by vw (very weak), w (weak), s (strong) and vs (very strong). 45
Tab. 6. Distribution of the data to build the fresh/thawed PLS discrimination model. 59
Tab. 7. PLS-R models: Their performance indicators (RMSEP and R^2) and the mean of the values to predict. 71

List of Acronyms

°C	Celsius degree
ALA	Alpha-Linolenic Acid (C18:3n-3c)
C	Carbon (element)
cm	Centimeter
DHA	Docosahexaenoic Acid (C22:6n-3c)
EPA	Eicosapentaenoic Acid (C20:5n-3c)
FA(s)	Fatty Acid(s)
FID	Flame Ionisation Detector
FT	Fourier transformation
GC	Gas Chromatography
H	Hydrogen (element)

H-NMR	Proton Nuclear Magnetic Resonance
kPa	Kilopascal
LA	Linoleic Acid (C18:2n-6c)
LV(s)	Latent Variable(s)
Min	Minute
MIR	Mid-Infrared
mL	Milliliter
mm	Millimeter (10^{-3}m)
ms	Millisecond
MUFA(s)	Mono-Unsaturated Fatty Acid(s)
mW	Milliwatt
NIR	Near-Infrared
nm	Nanometer (10^{-6}m)
OA	Oleic Acid (C18:1n-9c)
OD	Orthogonal Distance
PC(s)	Principal Component(s)
PCA	Principal Components Analysis
PLS	Partial Least Squares
PLS-DA	Partial Least Squares – Discriminant Analysis
PLS-R	Partial Least Squares – Regression
PUFA(s)	Poly-Unsaturated Fatty Acid(s)
Rcf	Relative Centrifugal Force
Rpm	Rotation per Minute
SD	Score Distance
SG	Savitsky-Golay
SNV	Standard Normal Variate
SVD	Singular Values Decomposition
t	Time

1 Introduction

One of the main challenges for the food industry today and in the future is the control of the increasingly complex and interlinked food supply chains. Also, the current context of climate change, the ageing population and other parameters lead to a continuous change in consumer choice trends in terms of specific food quality (Beddington et al., 2005; Carcea et al., 2009; Galvez et al., 2018; Hansen and Onozaka, 2011). Fish is one of the most internationally traded food products, making it highly prone to mislabelling and fraud (Rašković et al., 2016). Moreover, since fish meats, in particular processed fish meats, are often indistinguishable in appearance, texture, and taste, it is relatively easy to substitute products with high marketing value with products with a lower trade value. Cases of this kind of substitution, for example, farmed salmons sold as wild ones, are well reflected in the literature and are frequently reported by industry and academics (Hellberg and Morrissey, 2011; Hu et al., 2018; Jacquet and Pauly, 2008; Ward et al., 2009). The ability of the market to discriminate fish meat is therefore of great interest to protect consumers (Rašković et al., 2016).

1.1 Salmon

Salmon is a general name applied to different species of fish of the family Salmonidae native to the North Atlantic (genus *Salmo*) and the North Pacific (genus *Oncorhynchus*) oceans. The Pacific salmon's native areas traditionally extend from Alaska to Mexico in the eastern Pacific, and from northern Siberia to Taiwan in the western Pacific. Wild Atlantic salmon are found in the northern Atlantic Ocean and in the rivers that flow into it. They can be found in at least 17 countries in Europe and North America ("Atlantic salmon," 2022; Warner et al., 2013).

Salmons are sensitive animals. Environmental changes, pollution, overfishing, climate change, overexploitation of water resources, habitat destruction, or ocean acidification can negatively affect salmon populations. In the United States, for example, many populations of salmons are protected under the Endangered Species Act (ESA), including all wild salmon in the Atlantic and some in the Pacific Northwest (Warner et al., 2013).

Salmon is a valuable source of polyunsaturated fatty acids and other compounds of interest to human health such as carotenoids, although the actual amounts and integrity of the latter in its tissues depend largely on the living conditions of the salmon.

1.1.1 Farmed salmon

Atlantic salmon (*Salmo salar*) is the most commercially farmed and harvested salmon worldwide. It constitutes more than 90% of the farmed salmon market and more than 50% of the total global world salmon market (FAO, 2004). Its production has increased considerably in the two last decades, exceeding 2,4 million tonnes in 2018 and providing 50% of the salmon

consumed worldwide (*World Aquaculture 2020 – A brief overview*, 2022). The production takes place in 17 countries, with Norway and Chile leading. Pacific salmon are also farmed, but to a lesser extent (Warner et al., 2013). Farmed salmon are available year-round and their production is not subject to year-to-year fluctuations (Capuano et al., 2012). Aquaculture can occur for commercial purposes or recreational and subsistence fishing.

Most of the salmon feed is owned and produced by a few large companies. Traditionally, farmed salmon diet is based on fishmeal and fish oils largely produced by huge industrial fisheries in South America. Increasing pressure on these products is leading to more research into replacing fish products with vegetable sources of protein and oils (FAO, 2004).

1.1.2 Organic and conventional farmed salmon

If a salmon purchased in the EU is certified "organic", it guarantees that the farming conditions have respected a set of specifications. The first principles for an organic production method are good animal health and environmental conditions adapted to the species. The fish must have enough space to respect their physical integrity and species-specific behaviour. For example, the maximum density for organic farming is 20 kg of salmon/m^3. Other requirements are sufficient oxygen concentration, natural light, and optimal temperature in accordance with the species. The installations and techniques must avoid contaminants/pollutants. In addition, their feed must be adapted to their physiological needs and produced and then processed according to specifications for each species. To respect the organic label, feed must come entirely from organic farming and/or from products that have not undergone any chemical treatment during storage and processing, and/or from authorized complementary products. Furthermore, the use of GMOs, products derived from GMOs, or produced using GMOs, is prohibited at all stages of production from the fish to the final product (DIRECTION GENERALE DES POLITIQUES ECONOMIQUES, 2007). In the case of *Salmo Salar*, the proportion of plant-derived ingredients must be 100% from organic production, and represent a minimum of 30% of the supplementary (or provided) feed.

1.1.3 Wild salmon

The life cycle of wild salmons is full of remarkable journeys and transformations. They hatch in freshwater streams, where they spend a year or two before heading out to the oceans. Then, they remain in the ocean for most of their lives before returning to their native waters to spawn after two to seven years. This journey causes dramatic morphological changes. The body may change colour, turning bright red in some cases, as they compete for mates (Warner et al., 2013). Also, this ascent is exhausting. Low feeding during the upstream swim, their bodies provide sustenance. Upon reaching their birth waters, they spawn, and eventually, they die (Warner et al., 2013). Salmon caught before the journey to their native streams also undergo many transformations; they land on a boat that travels thousands of kilometres and takes many names and forms before the plate.

The wild salmon species that are sold as fresh are for example the *Oncorhynchus Tshawytscha*, *Oncorynchus Keta* or *Oncorynchus Nerka* species, and their meat is seen as a luxury product. They are rare and very good quality fishes. The only wild salmon species that was found in frozen pieces on the market in Belgium was *Oncorhynchus Gorbuscha,* the

smallest salmon species (less than 50 cm for less than 2.5 kg) and sold much cheaper than the other *Oncorynchus* due to the lower taste qualities of its meat.

Wild salmon is known to be richer in the more valuable omega-3 fatty acids than farmed salmon (Blanchet et al., 2005; Lundebye et al., 2017) and is therefore perceived by consumers as superior in terms of nutritional properties and quality (Muñoz-Colmenero et al., 2017). This results in a significant increase in the price of wild salmon compared to farmed salmon and, consequently, in serious concerns about fraud in the sector (Cline, 2012; Kappel and Schröder, 2016; Thomas et al., 2008), but we will see that price is not the only reason that leads to fraud.

1.2 Mislabelling

Mislabelling can relate to freshness, species, origin, or whether the fish is wild or farmed. In addition, mislabelling appears to be higher in highly processed products (Filonzi et al., 2010; Muñoz-Colmenero et al., 2017; Rasmussen and Morrissey, 2008) as their production and selling are more permissive by law, compared to whole and fresh sold products (Muñoz-Colmenero et al., 2017).

Mislabelling is reported on all continents (Miller et al., 2012) and the proportion of mislabelled products appears to be heterogeneous between areas of the world (Muñoz-Colmenero et al., 2017). Even in some markets with developed quality control mechanisms, such as the US, one-third of the fish consumed is mislabelled (Jacquet and Pauly, 2008). This is due to the high complexity of sources (aquaculture, wild, freshwater) and identification (populations and species) of fish and seafood products in general (Hellberg and Morrissey, 2011; Hu et al., 2018; Jacquet and Pauly, 2008; Leal et al., 2015). Despite the implementation of strict regulations and labelling protocols in many countries, these issues remain a global concern (Griffiths et al., 2014).

1.2.1 What motivates mislabelling?

There are several reasons why mislabelling in the fish market is global and constitutes a real issue for years. It is often assumed that the main reason is the desire to label a product as a higher value. In the case of salmon, strong demand coupled with inter-species substitutability provides a strong incentive for mislabelling, as it is the case for example with the substitution of wild Pacific salmon with cheaper farmed Atlantic salmon (Cline, 2012; Muñoz-Colmenero et al., 2017).

The degree of mislabelling of salmon would be highly correlated with the season of the year in which the salmon is purchased. In a nationwide study in the USA, results identified the highest level of mislabelling of salmon in restaurants when salmon was out of season (Warner et al., 2013). Studies mentioned that the substitution of Oncorhynchus species by Salmo species, which is known to be the most common mislabelling detected in other works (Pardo et al., 2016), would be higher in winter when wild species are not available (Muñoz-Colmenero et al., 2017). Also, consumers would be eight times more likely to be misled in small markets than in supermarket chains, which are required to provide information such as freshness, country of origin, species, and origin of farmed or wild salmon (Warner et al., 2013). Other factors, such as regulatory avoidance or market access, are ultimately connected to profit. But

it may also be unrelated to profit, such as confusion due to mixed fisheries, similar species, informal supply chains or poor regulations of product labels (Crego-Prieto et al., 2012).

1.3 Fraud impacts

Mislabelling of seafood products can result in a wide range of negative consequences for consumers, industry stakeholders and the environment.

There is an impact at the economic level. When a less valuable product, such as farmed Atlantic salmon, is sold as a more valuable species such as Chinook salmon, consumers do not get the value they pay for (Warner et al., 2013). This leads to economic losses for consumers, as well as for responsible fishermen who compete with fraudulent products and may receive less money for their catch (Warner et al., 2013). As an example, in 2006, 15% of farmed salmon was mislabelled. About 210,000 tonnes of farmed salmon were sold as wild salmon. Just by considering this, retailers and distributors won at least 2 million US $ in illicit profits (Jacquet and Pauly, 2008).

At the societal level, fraud undermines the efforts of sustainable fisheries and aquaculture, as well as the environmental awareness of consumers. It is the case for instance with South African hake (Marine Stewardship Council: certified sustainable seafood 2004; as https://www.msc.org/). They are sustainably caught but sold abroad with the wrong label (Garcia-Vazquez et al., 2011; Muñoz-Colmenero et al., 2015). The enormous effort of stakeholders to achieve sustainability may not be recognized by the unaware foreign consumer.

At the ecological level, mislabelling may hide cases of illegal fishing (Jacquet and Pauly, 2008). In addition, the species being substituted may reflect an erroneous conclusion that it is more available than it is, leading to errors in catch estimates. It may destabilize the calculation of stock sizes and leads to confusion among aware consumers (Muñoz-Colmenero et al., 2015). For example, with the extensive mislabelling of farmed salmon as wild, it is more difficult to make informed decisions about the true sustainability of their seafood choices.

Accurate labelling of seafood products is necessary not only for economic and environmental reasons but also to ensure public health. For example, studies have shown that in the United States, seafood is responsible for 18-20% of foodborne illnesses that affect 76 million Americans annually (Alert, 2007). Mislabelling contributes to illness in the US and beyond. A study of more than 700 salmon samples from around the world also confirmed the need for proper labelling. According to a study by Hites et al. (2004), farmed salmon contains higher levels of contaminants than wild salmon. Furthermore, the study found that farmed salmon from European countries have significantly higher levels of contamination than salmon farmed in North and South America. The contaminants present in farmed Atlantic and wild Pacific salmon raise important questions about the competing health benefits and risks of fish consumption. Wild salmon is preferred at least in part due to the general belief that farmed fish contain more mercury than wild fish (an average of 68.5% of European consumers; Pieniak et al., 2013) (Muñoz-Colmenero et al., 2017). However, farmed salmon would have higher concentrations of fatty acids than wild salmon, which may offset the risks associated with non-carcinogenic contaminants (Foran et al., 2005).

This highlights the importance of informative labelling, such as indicating whether the fish is farmed or wild, fresh or thawed, and the country of origin, to help consumers make choices. Differences in the benefit-risk ratio of farmed vs. wild salmon, and differences in farmed salmon by regions support this recommendation.

In conclusion, seafood fraud, including mislabelling, can have serious social implications and impacts on public health, economy, and ecology. Proper labelling is crucial to enable consumers to make informed choices, support sustainable fisheries and aquaculture and reduce the health risks associated with seafood consumption.

1.4 How mislabelling is prevented?

1.4.1 Regulations

The need for more regulation and oversight in the seafood industry is evident. This is especially true in regions where salmon is a resource of high cultural value and the potential ecological consequences of mislabelling are significant (Muñoz-Colmenero et al., 2017).

Over the past years, numerous regulations and protocols have been enacted to prevent fraud and adulteration in seafood products (D'Amico et al., 2016; Hofherr et al., 2016). For example, Council Regulation (EC-No. 104/2000) requires seafood products to bear the commercial and scientific name of the species, the geographical origin and the production method (farmed or wild caught) of all fishery and aquaculture products (Moretti et al., 2003), and at every step of the chain between fishery and marketing (Regulation, 2013).

In the United States for example, the Bioterrorism Act of 2005 required traceability of all links in the food and feed supply chains, but the FDA (U.S. Food and Drug Administration) has limited resources to enforce this regulation (Barrionuevo, 2007). In 2006, the FDA had the resources to check only 1% of the 8.9 million imported food shipments and tested only 0.59% of seafood imports (Alert, 2007). The fishing industry often argues that labelling legislation is too costly, but the costs associated with mislabelling are also high, including consumer and resource losses, ineffective ecolabelling and health problems (Jacquet and Pauly, 2008).

Another issue is that many labelling regulations only apply to wholesalers, not to restaurants (Jacquet and Pauly, 2008).

Despite these challenges, efforts have been made in recent years to improve labelling requirements and regulations. For example, in 2005, Country of Origin Label (COOL) Legislation for seafood products came into force in the United States but was subsequently weakened by industry lobbying efforts (Corbo et al., 2005). To effectively combat fish fraud and mislabelling, governments and industry associations must implement stricter regulations and labelling requirements, including global mandates for species, country of origin and production method labelling, as well as verifiability of ecolabels. The FAO Codex Alimentarius should require the actual geographic place of origin, not just the place of processing. In addition, third-party validation from 'cradle to plate', such as the MSC Chain of Custody certification, could be adopted at national and/or global levels to improve traceability and maintain public confidence in labelling claims. Penalties for mislabelling and illegal fish catches could also be increased. Currently, fines for illegally caught and mislabelling fish are minimal, which is not a strong deterrent against these practices. Governments could increase resources

and funding for organizations such as the FDA to enforce these regulations and conduct regular inspections of seafood products. In addition, consumers need to play an active role in the fight against fish fraud and mislabelling by being aware of the different labelling requirements and regulations, and by supporting sustainable and traceable seafood products.

1.4.2 Analytical methods used against mislabelling

Reliable analytical methods are necessary to authenticate fish and seafood products (Capuano et al., 2012). In the case of salmon, many sensory and physical-based techniques, as well as analytical methods for various compounds, have been developed in recent years to assess the freshness of the fish, its geographic origin, the species and whether it is wild or farmed (Cheng et al., 2015; Fiorino et al., 2019).

There are methods such as biochemical techniques, DNA barcoding and assays that can effectively check the validity of seafood at the species level. However, these methods require complex technologies, human resources, and political will, and may not be realistic in the developing world (Jacquet and Pauly, 2008).

One of the main methods used to detect fraud is DNA barcoding, which has been used in numerous studies to identify mislabelling in salmonids. Accurate species identification by DNA barcoding is based on a suitable DNA barcode, which refers to a standardized sequence (usually less than 1,000 bp) of the genome. In the case of animal identification, the most commonly used barcode marker is the mitochondrial cytochrome c oxidase subunit I (COI) (Hebert et al., 2003). For example, studies by Pardo et al. (2016), Cawthorn et al. (2012), Cline (2012), Warner et al. (2013) and Wong and Hanner (2008) have used DNA barcoding to identify mislabelling in salmon products. However, the use of DNA barcoding is not without its difficulties. The method can be costly, time-consuming and difficult to implement on a large scale.

1.4.3 The wild and farmed salmon case

1.4.3.1 What is a fatty acid?

Fatty acids are found in lipids in animal tissues. Lipids are a group of nutrients primarily made up of carbon, hydrogen, and oxygen atoms; these nutrients that provide energy are hydrophobic in nature. From a nutritional perspective, they serve as a significant energy source (providing 9 kcal per gram) and transport fat-soluble vitamins (A, D, E, and K) throughout the body. They are also crucial suppliers of essential fatty acids that the body is unable to synthesize (Moghadasian and Shahidi, 2017)

The fatty acids constitute the basic chemical structure of lipids. Around 95 mg of fatty acids can be obtained from roughly 100 milligrams of fat or oil. The physical and chemical properties of fats and oils are significantly affected by the types and proportions of the individual fatty acids and their arrangement on the glycerol molecule. Fatty acids consist of both saturated and unsaturated carbon chains, typically with an even number of carbon atoms and a solitary carboxyl group (Moghadasian and Shahidi, 2017).

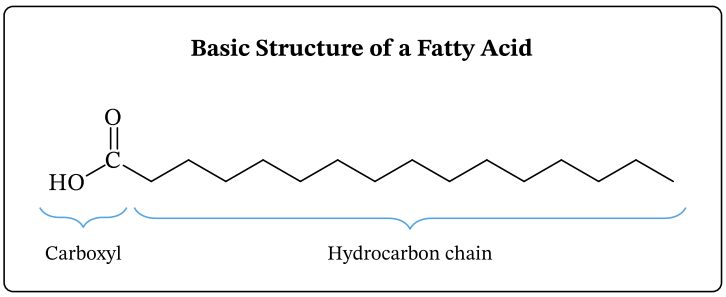


Fig. 1. Simplified structure of a fatty acid. A fatty acid is a carboxyl group attached to a hydrocarbon chain (Source: Nagwa).

Fatty acids are classified as saturated (SFA) and unsaturated fatty acids. Unsaturated FAs are subdivided into monounsaturated (MUFA) and polyunsaturated fatty acids (PUFAs).

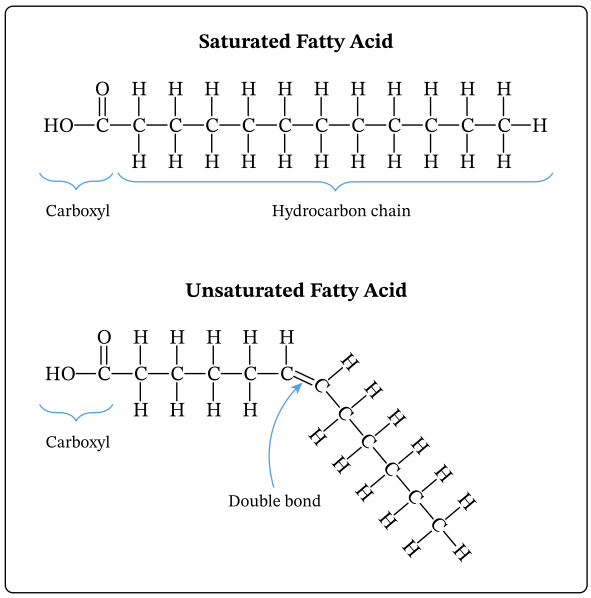


Fig. 2. The chemical structure of a saturated and an unsaturated fatty acid (Source: Nagwa)

1.4.3.2 What is a carotenoid?

Carotenoids, numbering more than 750 types, rank as the second most prevalent naturally occurring pigments worldwide. These pigments, primarily categorized as C_{40} lipophilic isoprenoids, are produced within photosynthetic life forms like bacteria, algae, and plants, and in certain non-photosynthetic bacteria and fungi. The palette of carotenoid colours spans from colourless to shades of yellow, orange, and red, influencing the appearance of numerous fruits, flowers, and vegetables. This diverse colouring also contributes to their economic value (Nisar et al., 2015).

While carotenoids play a crucial role in maintaining the normal well-being and behaviour of animals, most animals cannot synthesise these pigments. Instead, they depend on their diet to acquire these essential compounds. Consumption of carotenoids, such as astaxanthin, through the diet, can imbue specific marine animals (like salmon, trout, shrimp, and lobster) and avian species (such as flamingos and quails) with tissue pigmentation. This process can enhance their immune systems and often leads to improved health outcomes (Nisar et al., 2015).

1.4.3.3 Potential general differences between wild and farmed salmons

Some potential factors exist to determine whether a salmon is farmed or wild. One of them is the carotenoid content. Salmon cannot synthesise carotenoids, so it only absorbs them through feed (Tolasa et al., 2005).

The metabolism and retention of the carotenoid in salmon are then determined by the structure of the molecule (Elmadfa and Majchrzak, 1998), and by other parameters such as the sex, diet, size, age or species of the fish (Torrissen, 1986). The two main carotenoids in salmon are astaxanthin and canthaxanthin (Megdal et al., 2009; Molkentin et al., 2015; Tolasa et al., 2005; Torrissen, 1986). Since the measure of redness of a salmon filet is highly correlated with a high concentration of carotenoids, these molecules have been studied extensively in salmonids because the colour represents an important quality parameter for the salmon industry (Hatlen et al., 1998). The consumer normally expects the flesh colour of farmed salmon to be similar to that of free-living salmon (West European Fish Technologist's Association, 2004). Therefore, farmed salmons fed are enriched with astaxanthin or canthaxanthin to enhance flesh pigmentation (Christiansen et al., 1995; Torrissen et al., 1989), while wild salmon assimilate carotenoids by consuming krill. However, the difference in colour between wild and farmed salmon is often indistinguishable by visual inspection (Megdal et al., 2009).

In the marine habitat, astaxanthin is produced by microalgae and phytoplankton and then consumed by zooplankton and crustaceans, which constitute the main diet of salmon. Thus, the carotenoid accumulates in the food chain (Yuan et al., 2011). While organic farmed salmon are fed with shrimp shells or natural astaxanthin from Phaffia yeast (Molkentin et al., 2015), the conventional farmed salmon industry uses some synthetic astaxanthin. Synthetic pigments account for 15% of feed costs (Prendergast et al., 1994) and less than 20% of the ingested carotenoid is retained in the flesh (Storebakken and No, 1992).

Many methods of synthetic astaxanthin have been developed in recent years, as astaxanthin is a valuable molecule for human health. Many evidences suggests that astaxanthin has a strong antioxidant capacity, so its consumption could have many health benefits (Ambati et al., 2014; Davinelli et al., 2018; Guerin et al., 2003; Yuan et al., 2011). This leads to a change of astaxanthin sources in aquaculture feeds over the years. In aquaculture, as synthetic astaxanthin is cheaper than natural astaxanthin, it is currently more used than natural astaxanthin and astaxanthin contained in microalgae (Villaró et al., 2021). However, there could be some changes in the future. A study predicted a method to produce natural astaxanthin from microalgae *Haematococcus pluvialis* at a 610€.kg^{-1} cost, compared to the actual production cost of 880€.kg^{-1} for synthetic astaxanthin (Li et al., 2011).

Synthetic and natural astaxanthin differ in their optical isomers ratios composing it (Molkentin et al., 2015; Villaró et al., 2021; West European Fish Technologist's Association,

2004). This leads to different bioactivity. Some studies showed that ingested astaxanthin isomers are deposited in the flesh without changes in isomer distribution (Storebakken et al., 1985; Whyte et al., 1998). Therefore, the configurational ratio in the flesh reflects the isomer distribution in the diet. The different ratios of astaxanthin isomers found in salmon flesh could represent a good parameter to differentiate farmed from wild salmon, as wild salmon are unlikely to consume synthetic astaxanthin. In extension, for example, the meso-astaxanthin content of a salmon fed with synthetic astaxanthin represents 50% of the total isomers. As the meso-form level is very low in wild fish, the detection of meso-form means that the salmon is more likely to be farmed (Bjerkeng, 1997; Lura and Saegrov, 1991; Ostermeyer, 2004). Another interesting difference is the fatty acid profiles between the farmed and wild salmon. This will be explained in the next sections.

1.4.3.4 Analytical methods in the wild and farmed salmon case

So far, some interesting methods have been developed during the last years in various works using mass spectrometry analysis (Fiorino et al., 2019), gas chromatography (Megdal et al., 2009), chemical and stable isotopic analysis (Molkentin et al., 2015) or H-NMR spectroscopy (Capuano et al., 2012), depending on the target molecules used to discriminate. Currently, the method approved by the US Food and Drug Administration (FDA) to assess the authenticity of a wild or farmed salmon, uses liquid chromatography to analyse the content of carotenoid optical isomers (Turujman et al., 1997). Mass spectroscopy and gas chromatography methods were based on a difference in fatty acid composition between wild and farmed fish. The study of Fiorino et al. (2019) corroborated the hypothesis that aquaculture may have induced relevant oxidative stress in fish, explaining this difference. The study also reports a higher abundance of polyunsaturated FA with 20 or 22 carbon atoms in wild salmons, and a more relevant presence of unsaturated FA (18:1, 18:2 and 18:3) in the latter of farmed salmons (Fiorino et al., 2019).

In a study by Megdal et al. (2009), discrimination between farmed and wild salmon was performed using a method based on gas chromatography (GC) of fatty acid. The results of the study showed that there was 100% agreement between the GC method and the method used by the FDA, with statistical validation showing elevated levels of 18:2n-6 in farmed samples compared to the wild samples. The GC method was found to be suitable for wide adaptation, as it is a well-established procedure in laboratories performing analysis of lipid composition and food constituents. The study concludes that the fatty acid 18:2n-6 could be used as a key indicator to determine the origin of North American salmon and that it could be applied to other seafood (Megdal et al., 2009).

H-NMR (proton nuclear magnetic resonance) spectroscopy has also been used to analyse FAs content and has provided a correct classification tool (Capuano et al., 2012). Studies using stable isotopic analysis are considered an excellent tool for assessing the origin of fish (Martinez et al., 2003). Some are based on ^{13}C (Dempson and Power, 2004), while others focus on isotopic ratios such as oxygen ($^{16}O/^{18}O$) and hydrogen (2H/1H), which are assumed to be good indicators of environmental conditions (Ziegler et al., 1976). Isotopic and chemical analysis of fatty acids and carotenoids could be used to discriminate organic farmed salmon from conventionally farmed salmon (Megdal et al., 2009; Molkentin et al., 2015).

1.4.4 An alternative analytical method: Vibrational spectroscopy

Although the methods listed in the previous paragraph have already proven to be effective, they are generally time-consuming, laborious and require a considerable amount of solvents and chemical reagents to prepare the samples due to their complex organic composition (Ren et al., 2023). It is in this general context that vibrational spectroscopy (infrared, Raman) has been one of the most attractive and commonly used methods for seafood analysis as it allows fast, accurate, non-invasive and environmental-friendly determination of many major components (Power and Cozzolino, 2020).

Mid-infrared (MIR), near-infrared (NIR), and Raman spectroscopy are molecular/vibrational spectroscopy techniques used to study the interactions between electromagnetic waves and a sample. The electromagnetic spectrum encompasses various types of radiation, from radio waves to gamma rays. Molecules absorb infrared (IR) radiation due to the vibrational movements of their chemical bonds, such as bending, stretching, rocking, wagging, or scissoring. This absorption occurs at specific energy levels, resulting in chemical bonds absorbing IR radiation at distinct wavenumbers or wavelengths. Infrared spectroscopy measures this absorption, producing a spectrum with peaks that represent the chemical bonds present in the analysed sample.

The NIR region of the spectrum is located between 700–2500 nm. Overall, the NIR spectra are the result of overtone and combination bands of fundamental vibrations of C—H, O—H, and N—H bonds (Cozzolino and Murray, 2012; Liu et al., 2013; Weeranantanaphan et al., 2011).

The mid-infrared (MIR) region is between 4000 and 400 cm^{-1}. MIR is of great significance in analytical studies due to its inclusion of the biochemical and chemical fingerprint region (1800–900cm^{-1}). This region contains distinct peaks representing various common biochemical and chemical compounds found in complex food matrices, such as Amide I, II, and III, carbohydrates, glycogen, lipids, ester groups, asymmetric phosphate stretching vibrations, symmetric phosphate stretching vibrations, and protein phosphorylation (Power and Cozzolino, 2020).

Raman spectroscopy involves the inelastic scattering of monochromatic light when interacting with molecules. This technique provides rich information about molecular vibrational modes and unique structures. Despite its inherently weak effect due to minimal inelastically scattered light, Raman scattering can be harnessed to analyse the chemical composition of samples. Recent advancements in high-resolution and portable instruments have facilitated the in-situ measurement of Raman spectra in complex samples (Ellis et al., 2015; Zheng and He, 2014).

Vibrational spectroscopy methods are combined with chemometrics to create new applications. Chemometrics encompasses the integration of algebra, mathematics, and statistical analysis techniques to analyse and process data, especially in the context of large datasets derived from food analysis. These methods facilitate multivariate analysis, enabling the extraction of crucial insights from intricate datasets, visualizing underlying patterns, and constructing calibration and validation models (Amigo et al., 2013; Kamal and Karoui, 2015; Workman et al., 1996). In recent years, the use of chemometric methods has surged due to advancements in computing power and the availability of specialized software packages. In applications related to authenticity, traceability, and provenance, chemometric techniques

can be broadly categorized into unsupervised and supervised methods. Unsupervised methods allow for sample clustering without prior knowledge, aiding in identifying internal groupings within unlabelled data. On the other hand, supervised methods require prior knowledge about the samples or dataset, allowing for the prediction of categories for unknown objects (Amigo et al., 2013; Kamal and Karoui, 2015; Workman et al., 1996).

1.4.4.1 Authentication, Provenance and Traceability

Many studies exist about applications of vibrational spectroscopy on seafood samples. Brodersen and Bremner (2001) used NIR and chemometrics to classify shrimps that undergo several treatments and could differentiate between frozen and thawed samples, salt content, pH in the flesh, cooking period or processing (whole or minced shrimps). Ottavian and co-workers demonstrated that NIR could be reliably applied to discriminate between wild and farmed sea bass, with the same performance as methods that use chemical properties and morphometric traits (Ottavian et al., 2012). A study by Lv et al. (2017), also utilised NIR spectroscopy to discriminate several species of freshwater fish.

1.4.4.2 Freshness, spoilage and storage

Dai and co-researchers (2015) explored the applicability of near-infrared (NIR) hyperspectral imaging as a swift and non-intrusive approach to determine prawn freshness. The team assessed a total of 280 prawns, including both unfrozen and frozen samples, and achieved rapid and accurate classification of unfrozen and frozen prawns, with average correct prediction rates of 98.33% and 95%, respectively. Wu et al. (2018) assessed the potential of Fourier transform mid-infrared (FT-MIR) spectroscopy in examining the adulteration of Norwegian salmon with Heilongjiang salmon at nearby fish markets.

In their study, Ding et al. (2014) demonstrated the fast and non-invasive evaluation of silver chub (*Macrhybopsis storeriana*) freshness using FT-NIR spectroscopy. They found that the most optimal location to position the NIR optic fibre for spectrum collection was the eyeball of the fish (Ding et al., 2014). In another study, Saraiva et al. (2017) reported the use of FT-MIR to identify and monitor the spoilage of salmon samples stored in three different environments, in real-time. Identification of the MIR region that was associated with the spoilage process was made by PCA and the researcher also used a combination of MIR spectral data and sensory to quantify better the sample's freshness (Saraiva et al., 2017). Zhu and co-workers (2013) assessed the potential of combining visible (VIS) and near-infrared (NIR) hyperspectral imaging to distinguish between frozen and thawed samples of halibut fish (*Psetta maxima*) and got above 97% of classification rates (Zhu et al., 2013), concluding concluded that VIS and NIR hyperspectral imaging was able to differentiate between fresh and frozen/thawed samples.

Raman spectroscopy also has been used by Murat et al. (2015) as a means to identify and classify the species of fish as well as to determine their freshness (Velioğlu et al., 2015).

1.5 Potential of Raman

1.5.1 Raman spectroscopy on salmon

Salmon has been extensively studied by Raman since 2005 (Esmonde-White and Lewis, 2021). Its applications in this area include, for example, discrimination of fish fillets (Rašković et al., 2016), direct in-situ detection of astaxanthin in salmon fillets (Hikima et al., 2017), or assessment of important biochemical details on fish tissues through the skin (Landry et al., 2020). Recent studies show that Raman spectroscopy can be more robust than NIR spectroscopy for measuring FA in salmon muscle tissues (Afseth et al., 2022). Raman has also been used to identify salmon processing, storage, breeding and adulteration (Thawornchinsombut et al., 2006; Zhong et al., 2021; Zumelzu et al., 2018).

1.5.2 Raman for discrimination between wild and farmed salmon

As mentioned above, Raman may be able to detect and quantify carotenoids and fatty acids in salmon samples, and these parameters would be more relevant to discriminate between wild and farmed salmon. The study by Ermakov et al. (2006) found that Raman spectroscopy is a potential method for determining carotenoid concentrations in salmonids muscle tissue (Ermakov et al., 2006). The researchers conducted calibration experiments and found that the signal intensity of the Raman peaks corresponds linearly to the concentration of carotenoids in the tissue if the tissue is not too dense. They found that the C=C carotenoid peak at 1524 cm^{-1} produced the strongest signal and could be used for data analysis. The Raman signals of carotenoids in fish tissue samples were directly correlated to the absorption levels of the corresponding acetone extracts, indicating the potential of Raman spectroscopy as a fast, accurate, and non-destructive method for monitoring astaxanthin and canthaxanthin content in fish muscle tissue (Ermakov et al., 2006).

Hikima and al. demonstrated that the use of Raman spectroscopy for the quantification of astaxanthin in salmon muscles is significantly easier and more accurate than measurements using a visual method of determination (Hikima et al., 2017). Thus, the Raman spectra appear to be sensitive to carotenoid variations. Raman technology could also detect the different optical isomers of astaxanthin (Balaji et al., 2010; Yao et al., 2021). However, these studies have been conducted on pure astaxanthin extract and not on salmon samples. Raman spectroscopy has also been successfully used to quantify fat content, predict fat unsaturation or identify individual fatty acids in salmon muscles in several studies (Afseth et al., 2022, 2014; Wold et al., 2004).

2 Objectives of the Study

This study aims to demonstrate the potential of Raman spectroscopy for the rapid and non-destructive characterization of salmon fillets found in the market. Characterization is a broad term that encompasses any characteristic that a salmon fillet may have, ranging from colour to origin to fatty acid composition, for example. The idea is to provide information about the quality of the salmon based on spectral data. The quality indicators on which this

study focuses are fillet freshness, authenticity of wild or farmed salmon and fatty acid composition. This study attempts to evaluate these three quality parameters based on Raman spectra obtained from fillets.

The freshness and authenticity parameters were explored using chemometric tools such as PCA and PLS methods. For the prediction of fatty acid composition, reference values were obtained by gas chromatography (GC). These values will be used to create prediction models for fatty acid composition from the Raman spectrum using the PLS method. This objective is aligned with the assessment of salmon fillet authenticity. Indeed, differences in fatty acids could represent a means to assess the authenticity of the fillets. Throughout these various objectives, numerous tools (pre-processing, regression, discrimination, and exploration) were employed to process the Raman data. The general principles behind these tools are explained in this work. This study is not derived from any other and is intended to be exploratory rather than exhaustive. The main ideas and results presented in this paper could serve as starting points for further research on Raman spectroscopy in salmon.

3 Material and Methods

3.1 Raman Spectroscopy

3.1.1 Theory: Raman Vibrational Spectroscopy

When light interacts with matter, photons may be absorbed or scattered, or may not interact with material and pass straight through it. It is when the energy of a photon corresponds to the energy gap between the excited and ground state of a molecule, that the photon may be absorbed, and that the molecule reaches a higher energetic excited state.

However, a photon could also interact with the molecule and scatter from it. The scattered protons can be observed by collecting the light after the interaction. Scattering is a commonly used technique, and Raman is the main one for molecular identification (Smith and Dent, 2019).

There are two types of scattering. When the energy of an incident photon is unaltered after hitting a molecule, the scattered photon should have the same frequency as the incident photon, defining Rayleigh or elastic scattering (Movasaghi et al., 2007). The other type is known as inelastic Raman scattering. It is a way rarer event as it involves only one in 10^6 - 10^8 of the scattered photons (Smith and Dent, 2019).

Inelastic Raman scattering encompasses Stokes Raman scattering and anti-Stokes Raman scattering. The scattered light can be a lower-frequency photon than the original one, in this case, it is known as Stokes Raman scattering. In the case of anti-Stokes Raman scattering, the outcoming light is a higher-frequency photon (Rostron and Gerber, 2016). The photon gets this energy from an initially excited bond (Kudelski, 2008).

Raman spectroscopy is specialized in measuring the frequency shift of inelastically scattered light (Das and Agrawal, 2011; Efremov et al., 2008; Li-Chan, 1996; Smith and Dent, 2019). Raman scattering is given as a shift in energy from the energy of the laser beam. The

shift, expressed in wavelength (nm) or wavenumber (cm^{-1}) of the scattered light varies with the chemical composition of the molecules responsible for scattering. The intensity of the response is proportional to the change in molecule polarization.

Comparing infrared and Raman spectroscopy, the way radiation is employed is different. For infrared spectroscopy, a range of frequencies is directed onto the sample, and the absorption occurs where the frequency of radiation matches the vibration of a molecule, promoting it to a vibrational excited state. This is the loss of this radiation frequency that is then detected. In the case of Raman, one single frequency of radiation irradiates the sample, and this is the radiation scattered from the molecule, one vibrational unit of energy different from the incident beam, that is detected. The Raman scattering happens when the light interacts with the molecule and polarizes the cloud of electrons around the nuclei into a short-lived state, and does not requires matching the incident radiation to the energy difference between the ground and the excited state (Smith and Dent, 2019).

3.1.2 Material

3.1.2.1 Senterra II

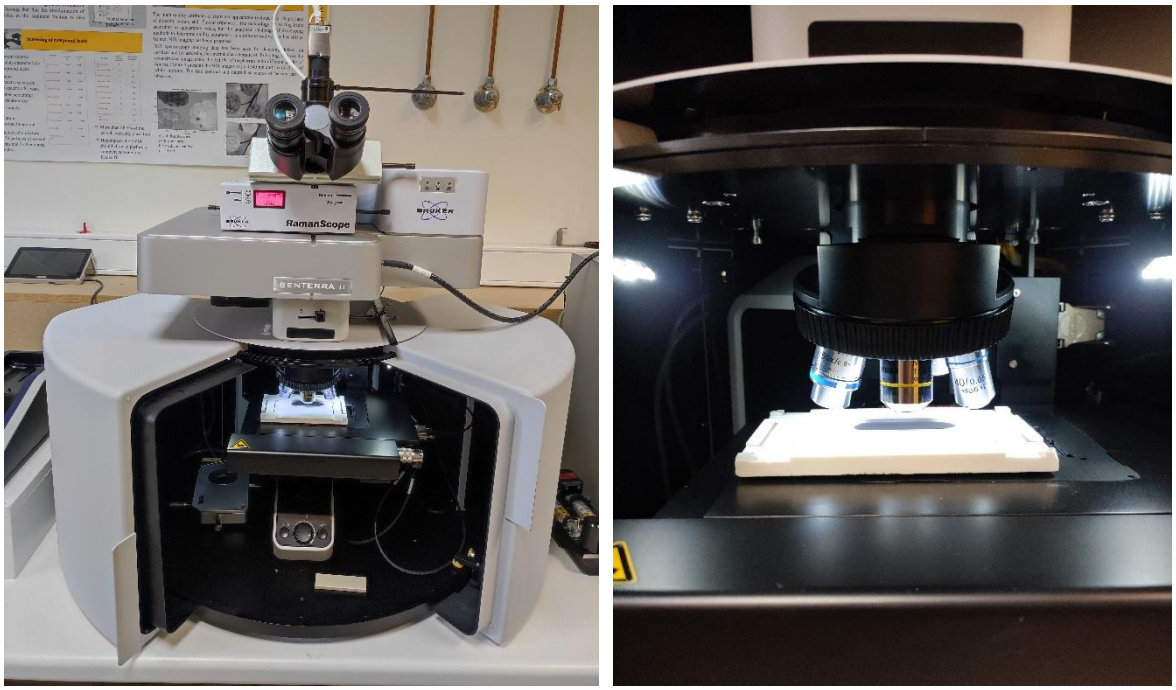


Fig. 3. Raman microscope spectrometer Senterra II by Bruker used for this study.

The Raman spectrometer utilized is the Senterra II model manufactured by Bruker. Its spectral range extends from 4000 to 50 cm^{-1}. The laser employed in this instrument is generated by diodes and has a wavelength of whether 785 nm or 532 nm, with adjustable power ranging from 0 to 100 mW for 785 nm, and 0 to 20 mW for 532 nm. For all the analysed samples, the highest power setting was utilized. The instrument is of the dispersive type.

The laser beam is split into different wavelengths using a grid, and the resulting Raman scattering is detected by a CCD detector. To maintain cooling, the spectrometer employs a Peltier cooling system, which keeps the instrument at a temperature of -60°C. The Senterra II functions as a confocal microscope, with the laser penetrating the sample to a depth of 5 to 10 microns depending on the wavelength used (greater wavelength results in deeper penetration). Two laser wavelengths are available: 785 nm and 532 nm. The spectrometer provides a resolution of 4 cm^{-1}. Multiple spectra can be automatically measured at different locations, it is called a mapping.

The Senterra II operates with the OPUS 7.8 software. It allows the consultation of the data and enables parametrical adjustments such as integration time and the number of co-additions. The integration time is the time of exposition of the scattered light to the detector. Longer integration time means that the detector will collect scattered light longer, and

therefore that more signals will be collected. The integration times used for the 532 nm and the 785 nm in this study were respectively 2 000 ms and 10 000 ms.

Co-addition is the process of combining a certain number of spectra to form a single spectrum. It is commonly employed to reduce measurement noise. A co-addition value of 5 was used for all the samples.

The sample can be observed either through the binocular eyepiece or on a video camera. A joystick is used to navigate the sample, adjust focus, and control brightness. The optical microscope lens can be selected with magnifications of 4, 10, 20, 50, or 100x (in this study only the 10x lens was used).

3.1.2.2 Rigaku Progeny 1064 nm handheld Raman analyzer

Fig. 4. Rigaku Progeny 1064nm device used for this study.

The Rigaku Progeny 1064 nm handheld Raman analyser is a 29.9 cm long portable spectroscopic instrument that uses a 1064 nm laser beam and a 512 pixels TE Cooled InGaAs detector. The incident laser power can be adjusted from 30 mW to 490 mW. Its spectral range extends from 200 to 2500 cm^{-1} (1087-1450 nm).

The spectral resolution goes from 8 to 11 cm^{-1}. The instrument is the dispersive type. The detector is a 512-pixel TE Cooled InGaAs. The integration time can be set from 5ms to 30s.

The device is equipped with an integrated camera for sample documentation, a 5.7-inch touchscreen display that is used for data visualisation and a rechargeable battery for extended field use. For real-time data analysis and management, it is equipped with the RAD Progeny Software.

3.1.3 Sampling

A total of 28 different samples of fresh or frozen raw salmon were used in this study. These samples consisted of pre-cut and packaged pieces of salmon, which were purchased from various supermarkets in Belgium. Due to the challenge of obtaining wild salmon, the sample set was made of 24 raised salmon samples and 4 wild salmon samples. Of the raised salmon samples, 3 were farmed in Chile, 2 in the Faroe Islands, 15 in Norway, and 4 in Scotland. Of these, 8 were labelled as "organic" All the farm-raised salmon were from the *Salmo Salar* species, while all the wild-caught salmon were *Oncorhinchus Gorbuscha*. The sample set included 14 frozen and 14 fresh samples, with prices ranging from 10 to 70 euros per kilogram (average price of 35 euros per kilogram).

Tab. 1. Table of all salmon samples and their principal characteristics.

SAMPLE SET						
Sample name	**Species**	**Type**	**Farm place**	**Production mode**	**Freshness**	**Price/kg (€)**
DQ210884-003	*salmo salar*	farmed	Chile	conventional	frozen	25,32
DQ210884-004	*salmo salar*	farmed	Norway	organic	frozen	36,9
DQ210884-005	*salmo salar*	farmed	Chile	conventionnal	frozen	27,31
DQ230046-001	*oncorhinchus gorbuscha*	wild	/	/	frozen	12
DQ230046-002	*oncorhinchus gorbuscha*	wild	/	/	frozen	10,5
DQ230046-003	*salmo salar*	farmed	Norway	conventionnal	fresh	19,9
DQ230046-004	*salmo salar*	farmed	Norway	conventionnal	fresh	24,99
DQ230046-005	*salmo salar*	farmed	Norway	organic	fresh	36,99
DQ230046-006	*salmo salar*	farmed	Scotland	conventionnal	fresh	36,99
DQ230046-007	*oncorhinchus gorbuscha*	wild	/	/	frozen	27
DQ230046-008	*salmo salar*	farmed	Faroe Islands	conventionnal	frozen	45,56
DQ230046-009	*salmo salar*	farmed	Norway	organic	frozen	40,2
DQ230046-010	*salmo salar*	farmed	Norway	organic	fresh	37,99
DQ230046-011	*salmo salar*	farmed	Scotland	conventionnal	fresh	42,99
DQ230046-012	*salmo salar*	farmed	Norway	organic	fresh	36,99
DQ230046-013	*salmo salar*	farmed	Scotland	conventionnal	fresh	38,99
DQ230046-014	*salmo salar*	farmed	Norway	quality carrefour	fresh	30
DQ230046-015	*salmo salar*	farmed	Norway	organic	fresh	37,99
DQ230046-016	*salmo salar*	farmed	Scotland	conventionnal	fresh	56,99
DQ230046-017	*salmo salar*	farmed	Norway	conventionnal	fresh	33,99
DQ230046-018	*oncorhinchus gorbuscha*	wild	/	/	frozen	25
DQ230046-019	*salmo salar*	farmed	Faroe Islands	conventionnal	frozen	45,56
DQ230046-020	*salmo salar*	farmed	Norway	organic	frozen	40,2
DQ230046-021	*salmo salar*	farmed	Norway	conventionnal	frozen	31
DQ230046-022	*salmo salar*	farmed	Norway	organic	fresh	68
DQ230046-023	*salmo salar*	farmed	Norway	conventionnal	fresh	39
DQ230046-025	*salmo salar*	farmed	Chile	conventionnal	frozen	31,85
DQ230046-026	*salmo salar*	farmed	Norway	conventional	frozen	25

3.1.3.1 Sample preparation

The frozen samples were purchased and stocked in a freezer at -20°C. Before analysis, the frozen pieces of salmon were removed from the freezer and thawed at room temperature (approximately 20°C) for 20 minutes.

The fresh samples were purchased and stored in a refrigerator at a temperature of 5°C while awaiting analysis. Once taken out of the refrigerator, they were analysed within a maximum of 10 minutes at room temperature.

To prepare each sample for analysis, a 3 to 5 mm slice of salmon sample was cut and placed onto an aluminium plate.

3.1.4 Spectra acquisition

The slices were then analysed using a Senterra II with 532nm and 785 nm lasers, as well as a 1064nm laser on the Progeny Portable Raman device. Spectra were measured on the pink muscle between the white stripes of the fillet. In the case of the 532nm laser, measurements were taken using a 3x3 grid of spectra forming a square of 210 µm on each side at four different areas of the salmon, totalling 36 spectra per measure. For the 785nm laser, a 2x2 grid of spectra forming a square of 210 µm on each side was taken at four different areas, totalling 12 spectra per measure. Several experiments have been made in the case of this study, as shown in Tab.2. The measurement parameters were the same for each experiment. For the 532 nm analysis, the integration time used was 2000 ms with 5 co-additions, the lens used was x10, the power of the laser was 20 mW and the resolution was 4 cm^{-1}. For the 785 nm analysis, the integration time was 10 000 ms with 5 co-additions, the lens was x10, the power of the laser was 100 mW and the resolution was 4cm^{-1}.

The total measurement time for a sample was approximately 10 and 20 minutes for the 532 nm and 785 nm lasers, respectively. To analyse the slice with the handheld Raman device, a glass plate was placed between the Rigaku laser and the sample slice. Three spectra were then taken for each sample.

Each sample is at least analysed once with the three lasers. Due to configuration issues, the 1064 nm laser of the Rigaku handheld device was only used during the construction of the global database. It primarily allowed for the comparison of average responses from the three lasers. As a result, most of the research data for this study comes from a non-portable device.

Tab. 2. Synthetic table of the different measurements made in the study.

EXPERIMENTS			
Experiments	**Method**	**n samples**	**Lasers**
Successive freezings and thawings	The samples are completely frozen and thawed 5 times, the thawed sample is analysed once a day.	3	532nm, 785nm
One weekthe in fridge (5°C)	The sample stays at 5°C degrees for a week and is analysed once a day.	2	532nm, 785nm
Spatial test	Four different areas of the same salmon filet are analysed.	1	532nm, 785nm
Fresh and thawed	Each sample is analysed once, fresh and thawed.	5	532nm, 785nm
One day at room temperature	A salmon sample is left for 1 day a room temperature and is analysed every hour.	1	532nm, 785nm
28 samples spectra database	Global database (at least one measure/sample)	28	532nm, 785nm, 1064nm
24 ground samples analysis	Homogenised samples analysis	24	532nm, 785nm
Repeatability test	A sample is analysed every hour for 6 hours	1	532nm, 785nm
Reproducibility test	A sample is analysed once a day for 4 days	1	532nm, 785nm

3.2 Gas Chromatography on Fatty Acids

3.2.1 Material

Fig. 5. Shimadzu GC 2010 Plus used for the Gas Chromatography in this study

The material used for the gas chromatography is the Shimadzu GC 2010 Plus gas chromatograph, equipped with a flame ionisation detector (FID) and a split injector. The gas chromatography (GC) method used in this thesis involved specific parameters to ensure accurate and reliable analysis. The GC system employed a 100-meter-long fused silica capillary column with an internal diameter of 0.25 mm and a film thickness of 0.2 µm. The stationary phase of the column was cyanopropyl-polysiloxane, which is a commonly used stationary phase for separating and analysing compounds.

The carrier gas used in this method was helium, and the column head pressure was maintained at 225 kPa. The sample injection was performed using a split injection technique, with a split flow of 25.5 mL/min. The split ratio was set at 10:1.

The injector temperature was set to 250°C, and the detector temperature was also maintained at 250°C throughout the analysis. The GC analysis started with an initial temperature of 60°C, which was held for 5 minutes to allow for equilibration. The temperature was then raised at a rate of 15°C per minute until reaching 165°C, which was held for 1 minute. Subsequently, the temperature was further increased at a rate of 2°C per minute until reaching 225°C and held for 20 minutes.

3.2.2 Method

3.2.2.1 Fat extraction

To extract their fatty acids, a Folch method has been used on the salmon muscles. The protocol used for that fat extraction has been described by Molversmyr et al. (2022). To

prepare the samples, a piece of salmon (around 5 to 10 grams) was frozen in liquid nitrogen and then ground with a portable grinder (IKA®-A11 basic) to obtain a homogenous frozen salmon powder. An aliquot of 0.5 grams of the homogenate was then added to a 50ml Falcon tube. To this, 10ml of a 2:1 chloroform:methanol (volume/volume) solution was added and shaken at 390 rpm (rotation per minute) for 20 minutes in an orbital shaker (IKA®-Rocker 2D). Then, 2ml of a 0.9% NaCl in Milli-Q water solution was added a shaken using a vortex mixer (RSLAB®-6). The organic and aqueous phases were then separated by centrifugation (in an Eppendorf® Centrifuge 5804-R) at 720 rcf (relative centrifugal force) for 5 minutes. The upper, aqueous phase was then discarded, and the organic phase was transferred into 15ml Falcon tubes. The samples were dried using a CentriVap® (LabConco™ Acid Resistant) concentrator at 40°C under vacuum. Some samples were sufficiently dried after 2 hours, but others required 3 to 4 hours. Each resulting fat sample was weighted.

3.2.2.2 Derivatisation

In a 250 ml volumetric flask, a standard solution was then prepared adding 500 mg of tritridecanoin and 500 mg of methyl undecanoate. The whole was dissolved up to the mark with MTBE. Then, 5 ml of this internal standard solution was added to each fat sample, followed by 5 ml of 5 % (weight/vol) methanolic sodium methoxide solution, and at this step a timer can be started. The tube was then shaken for 10 s using a vortex mixer (RSLAB®-6). At t = 180s, after the first drop of methanolic sodium methoxide solution was added, 2ml of hexane was added. At t = 210 s, 10 ml of disodium hydrogen citrate and sodium chloride aqueous solution were added to stop transesterification. Transesterification time should not exceed t = 240 s. The tubes were then centrifuged at 1750 min^{-1} for 5 min (Eppendorf® Centrifuge 5804-R). After that, 200 µl of the supernatant was collected and diluted with 10 ml (if the fat sample weight after extraction was > 15 mg) or 5 ml (if < 15 mg) of *n*-hexane.

A total of 24 samples, a standard solution containing 37 fatty acids, a blank and some chloroform: methanol (2:1) solution used for the extraction, were then analysed in triplicate within the gas chromatograph.

3.3 Chemometrics

Chemometrics tools were all from a package called Jchemo.jl. Jchemo.jl runs on JULIA v1.8.5, a free open-source programming language. JULIA was run via the open-source code editor Visual Studio Code (VSCode) developed by Microsoft. The Jchemo.jl package contains utility functions for spectral analysis and chemometrics to transform and explore data, and to build and tune models of regression and discrimination (Lesnoff, 2023).

3.3.1 Data cleaning

It happens in Raman spectroscopy that, for some reason, a spectrum is considered abnormally different from the other. These types of spectra are called outliers. The hazard with outliers is that they can heavily affect the data analysis because they represent a very high variation in the dataset. These spectra are perceived as non-representative objects due to instrumental error. For example, if too much information is sent to a detector, it can be saturated. This results in incomplete spectra. Another problem could be due to a particle that

is not from the sample which gets between the laser and the sample. The information of this particle is then on the spectrum, and it becomes unusable. Some analytical tools exist to discard them. In this work, a “distance-based outlier spectra detector” was used.

3.3.1.1 Distance-based outlier spectra detector (Angiulli and Pizzuti, 2002; Ramaswamy et al., 2000)

The kNN-based outlier spectra detector is the algorithm that was used on the dataset to clean the data of spectra that have a significantly different profile compared to other spectra. These outliers could potentially cause errors during statistical processes such as PCA. This algorithm was found in a Julia package called "OutlierDetection.jl". The package includes a collection of algorithms based on relevant research in the field, which can be used for outlier detection (OutlierDetection.jl, 2023). The basic principles of the *k*-NN detector will be explained in this section, but for a more in-depth understanding, the source works are referenced too (Angiulli and Pizzuti, 2002; Knorr et al., 2000; Ramaswamy et al., 2000).

3.3.1.2 Basic Definitions of the k-NN OD Method

The k-NN Outlier Distance method is an algorithm that identifies outliers based on their distance from other data points (Sikder and Batarseh, 2023). It is a popular method for detecting outliers in high-dimensional data sets because it does not rely on assumptions about the underlying distribution of the data. Instead, it uses the distance between data points to identify those that are far away from the rest of the data. The distance metric used can be Euclidean distance, Manhattan distance, or any other distance metric that is appropriate for the data set.

Given a positive integer k, k-NN looks at the k points closest to a point $p(x, y)$. For a point p, the distance between a point p, and its k^{th} nearest neighbour $p_k(x_k, y_k)$ is the k-distance of p and can be denoted as:

$$D^k(p) = \sqrt{(y - y_k)^2 + (x - x_k)^2} \qquad (1)$$

The maximum D^k value is considered as the threshold to set if a point p is an outlier or not. Given k and m, a data point p is considered an outlier if less than m – 1 other data points possess a larger magnitude of D^k than p. The top m points with the largest D^k magnitude are classified as outliers. The m and k parameters have been set experimentally. The main goal was to discard outlying samples that could badly influence analysing tools such as PCA and PLS, while keeping a maximal variance in the spectral data set.

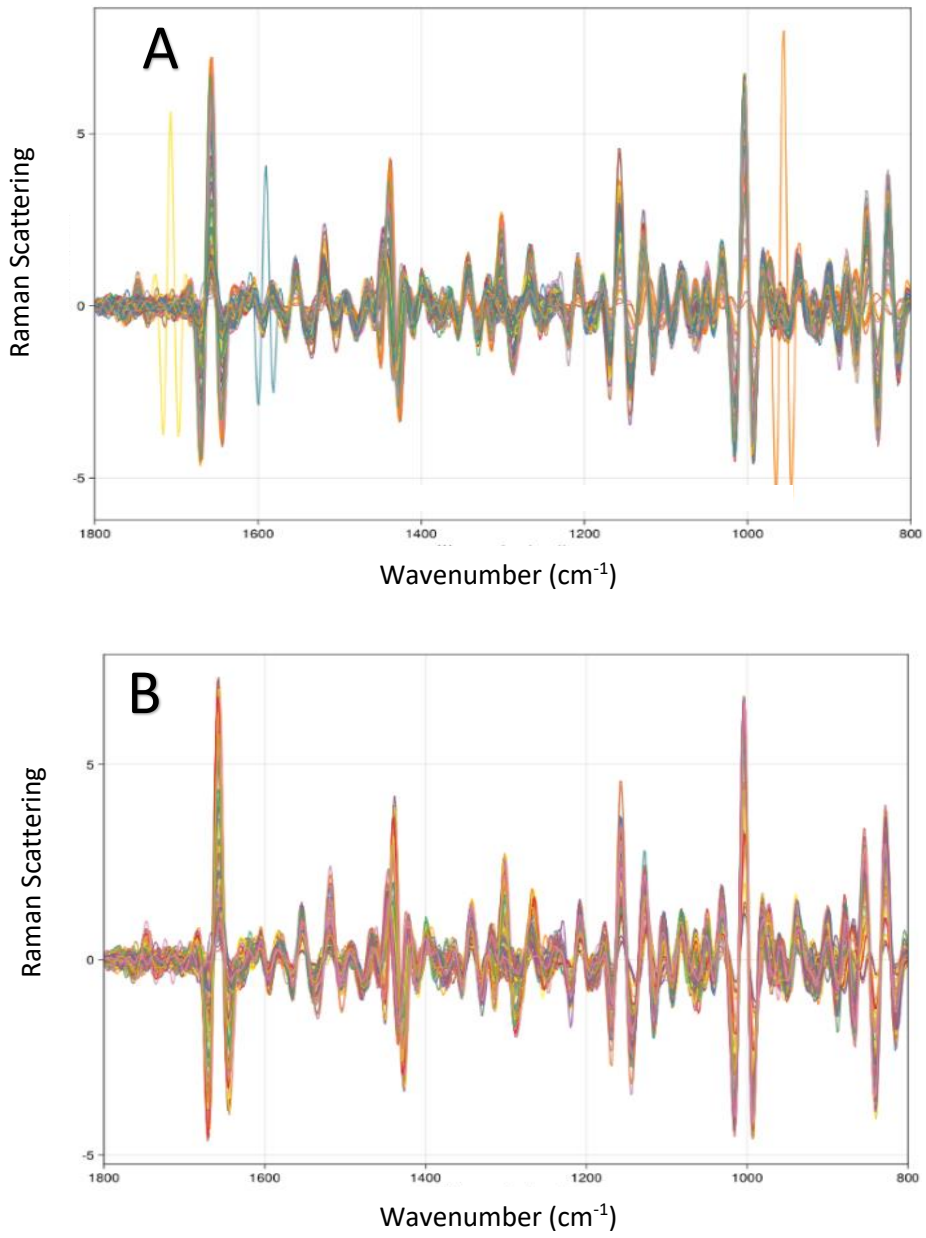


Fig. 6 : Removal of outliers on a salmon spectral dataset (acquired with 785nm laser and pre-processed with 1st derivative) with k-NN outlier detection algorithm. Figure A is the data before the removal and B after the removal of outliers.

3.3.2 Pre-processing

The information obtained from Raman signals can be influenced by various experimental factors and by the physical and chemical properties of the sample matrix. Several pre-processing steps are performed on the spectra dataset before applying exploratory or regression-based methods. Their main objective is to reduce sources of variability such as noise and physical effects (Rinnan et al., 2009). The removal of baseline features and the extraction of relevant information from the data are crucial pre-processing steps. Pre-processing methods are mathematical correction methods that can be applied to

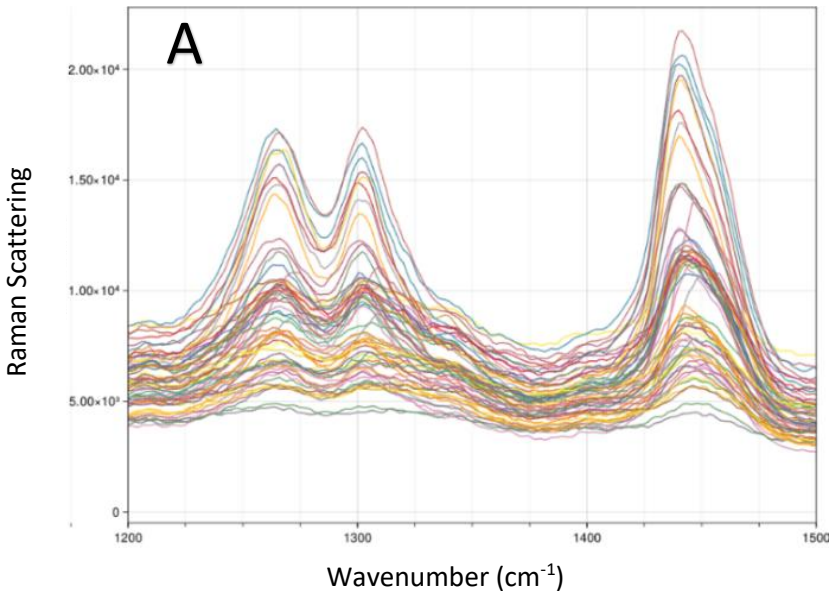
the Raman spectra. In this study, two pre-processing methods are applied to the salmon spectra, which are the SNV (standard normal variate) and the Savitsky-Golay (SG) derivative.

3.3.2.1 Standard Normal Variate

The SNV transformation purpose is to standardize the spectra on an individual sample basis. This transformation involves performing the same calculation on each spectrum variable.

$$X_{SNV} = \frac{X_{original} - \mu}{\sigma} \quad (2)$$

Where μ is the average value of the sample spectrum to be corrected and σ is the standard deviation of the sample's spectrum (Rinnan et al., 2009). After SNV, a spectrum will have a mean equals to 0 and a standard deviation equal to 1. The SNV is applied individually on each spectrum.



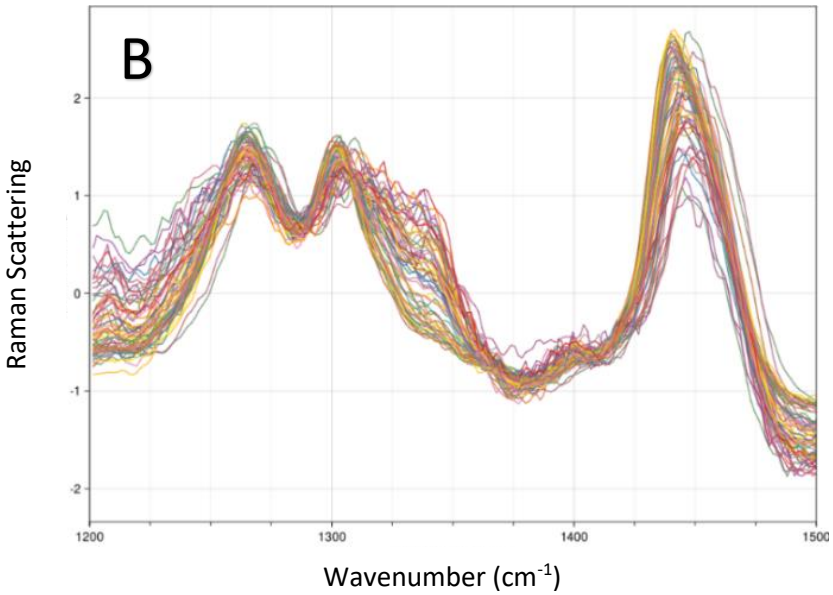


Fig. 7. Illustration of the effect of SNV on a salmon spectral dataset. Figure A is a raw dataset, and Figure B is the same dataset after an SNV on each spectrum.

3.3.2.2 Savitsky-Golay smoothing and derivatives

Savitzky and Golay popularized a method for numerical derivation of a vector, that includes a smoothing step (Savitzky and Golay, 2002).

The SG smoothing and derivative method first proceed with a polynomial fitting according to the least square criterion. Consider a dataset of n data points, a spectrum. Each point (x_j, y_j) $(j = 1, \dots, n)$ is a Raman scattering intensity x_j at a certain wavenumber y_j. Suppose a subset of *m* adjacent data points, each separated by an interval *h*. This is called a window and the size of the window is *m* (an odd number). Firstly, a change of variable is made by:

$$z = \frac{x - \bar{x}}{h} \qquad (3)$$

Where $\bar{x}$ is the value of the central point of the window. This set of points can then be approximated by a polynomial equation. The a 's are selected for the fitting equation to be optimal towards the least square criterion. This polynomial P, of degree k, is defined by:

$$Y = P(z) = a_0 + a_1 z + a_2 z^2 \cdots + a_k z^k. \qquad (4)$$

The Y values, noted y, are given by the fitting polynomial for each point of the window. Savitsky-Golay filter then calculates "convolution coefficients". These can be applied to all data subsets, to give estimates of the smoothed signal, (or derivatives of the smoothed signal) at the central point of each sub-set. They are calculated by:

$$C = (J^T J)^{-1} J^T \qquad (5)$$

Where *J* is the Jacobian matrix of *P* :

$$J = \left(\frac{\partial P}{\partial a_j}(z_i)\right) = \begin{pmatrix} 1 & z_1 & z_1^2 & \cdots & z_1^d \\ 1 & z_2 & z_2^2 & \cdots & z_2^d \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & z_{2l+1} & z_{2l+1}^2 & \cdots & z_{2l+1}^d \end{pmatrix} \qquad (6)$$

The value of the central point Y_j given by the Savitsky-Golay filter is calculated by :

$$Y_j = \sum_{i=-\frac{m-1}{2}}^{\frac{m-1}{2}} C_i \, y_{j+i}, \qquad \frac{m+1}{2} \leq j \leq n - \frac{m-1}{2} \qquad (7)$$

Where y are the values given by the polynomial P. The process is done for each window of *m* points, going point by point. The window moves across the spectrum to calculate the smoothed value of each central point. The numerical derivatives are obtained by differentiating the polynomial P. The SG transformation is used for smoothing (when P is not differentiated) as well as for differentiation. The parameters of the SG are thus : the size of *m* (the window size), the polynomial order of P (*k*), and the derivative order.

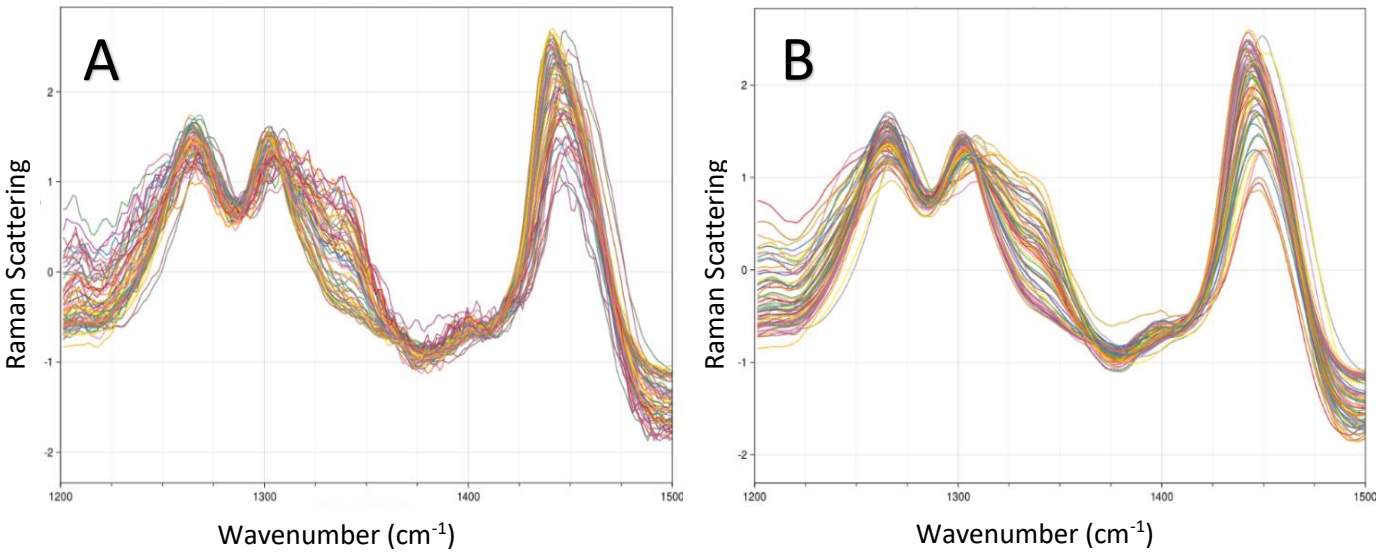


Fig. 8. Illustration of the effect of SG smoothing on a Raman spectral dataset of salmon. Figure A is a salmon spectral dataset from wavenumbers 1200 to 1500 cm^{-1}. Figure B is the smoothed dataset using SG smoothing (no derivative) with window size = 9.

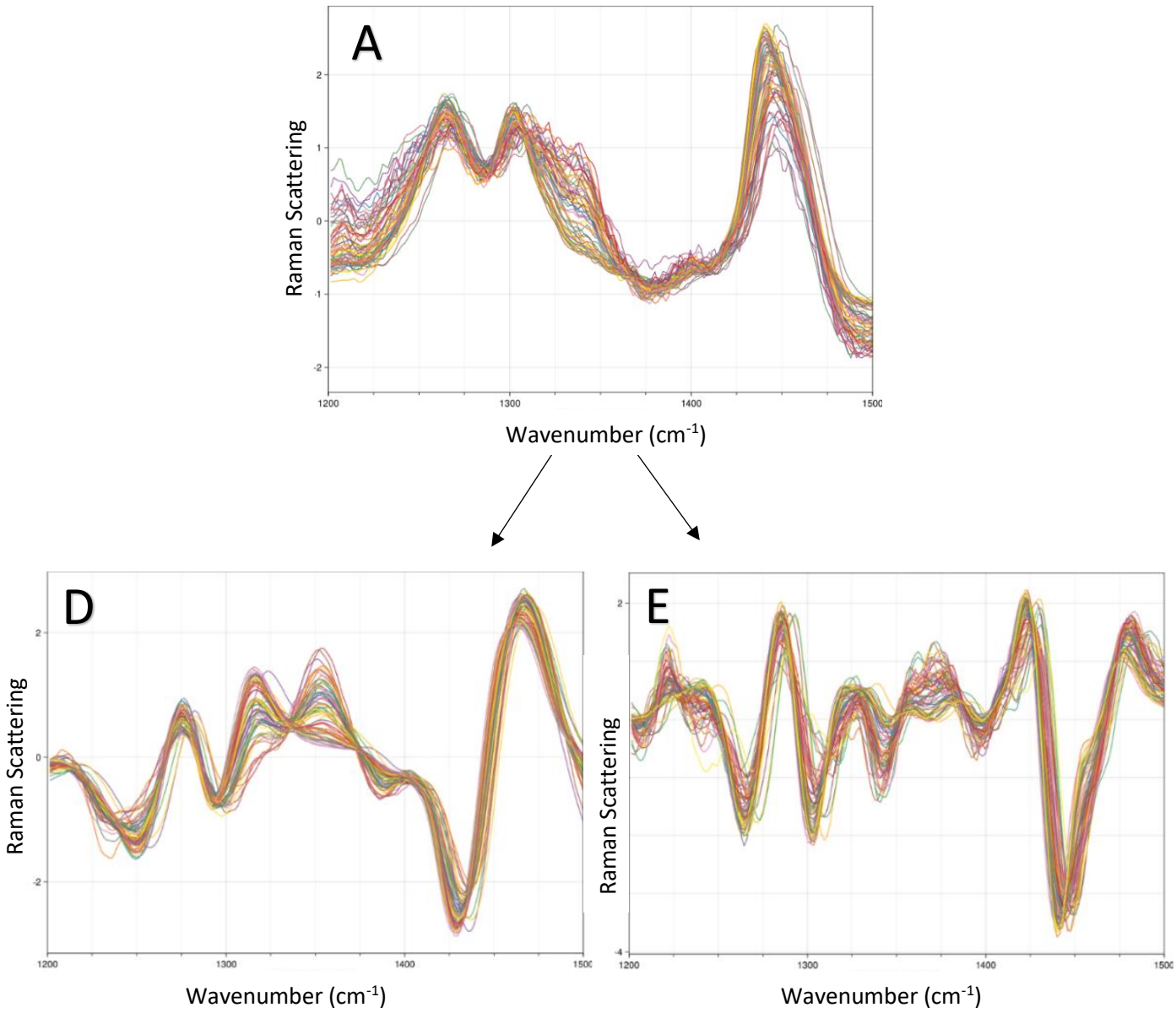


Fig. 9. Illustration of the effect of SG derivatives on a Raman spectral dataset of salmon. Figure A is a salmon spectrum from wavenumbers 1200 to 1500 cm^{-1}. Figure D is its 1st derivative SG with window size = 21, and Figure E is its 2nd SG derivative with a window size = 21.

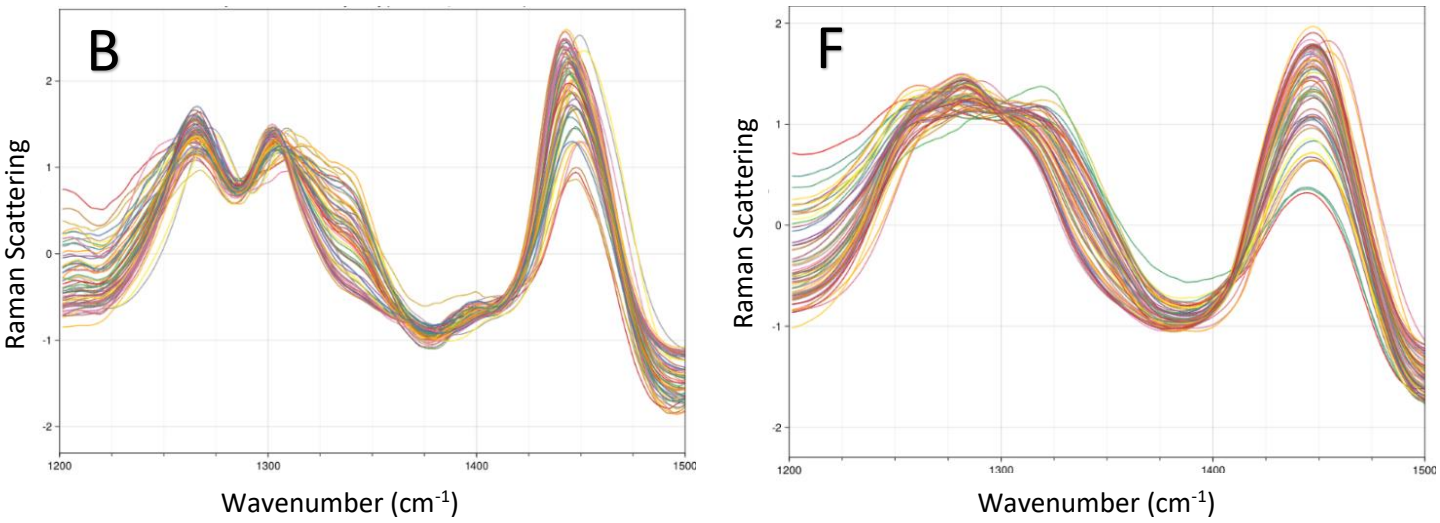


Fig. 10. Illustration of the effect of the window size of SG smoothing on the Raman spectral dataset of salmon. Figures B and F are both SG smoothed (no derivative) datasets of an original dataset (Image A from Fig.8) from wavenumbers 1200 to 1500 cm^{-1}. Figure A was made using a window size = 9 and Figure B a window size = 31. The Figure B appears to be smoother.

3.3.3 Principal Components Analysis (PCA)

PCA is an exploratory method for the visualisation of multifactorial data. As a Raman spectrum is defined by thousands of wavenumbers, it is difficult to visualise the data in a few dimensions. PCA aims to offer a simpler visualisation of the samples by reducing the dimensionality of the data. To do so, PCA creates new variables, that are some linear combinations of the initial variables (in this case, the wavenumbers). These new variables are called principal components (PCs), and they are uncorrelated. Each PC could therefore be seen as a "building block" of the dataset. A PCA can be applied using the eigenvectors method or the SVD (singular value decomposition) methods. Both eigenvectors and SVD aim to decompose a X matrix into new components and identify patterns that explain better the covariance in the data than the original variables. The PCA used in this work is SVD-PCA, so the following text uses the SVD terms for the explanation.

The PCA identifies new orthogonal directions (or PCs), by linear combinations of the initial variables, which explain the maximum cumulative variance in the data. Geometrically, the way to find the first PC is to maximise the square distance of the projected samples to the origin on a new axe in the real data space. The vector that gives the maximal variance can be called the singular value, and it is the direction of the first component (PC1). Each principal component is characterized by its singular value, and by its direction, called the singular vectors. There are two singular vectors for each PC, called the left and the right singular vectors. Each PC is related to the original variables via the singular vectors. A sample has a

value for each principal component, obtained by its projection on these singular vectors. These new coordinates are called the scores of a sample for each PC.

The second principal component is a perpendicular axe to the first component that passes through the origin point and maximises the singular value of the samples on the axe. The process is repeated to obtain the other principal components. The principal components are ranked in such a way that the earlier ones explain the most variability across the dataset. The variability that is explained by each PC in the PCA can be calculated by dividing the singular value of a PC by the sum of the singular values of all the other PCs. The maximal number of PCs that a PCA can create is equal to the number of samples, or to the number of original variables, whichever is the smallest.

Let's set *X* as the real data matrix. The *X* matrix is *n* x *p* size, where *n* is the number of samples (rows) and *p* is the number of variables (columns). The first step of PCA is the centring of the data matrix. The mean value of each variable must be subtracted from each observed value in the data matrix. The mean values for all variables (all columns) are now equal to 0. This centred matrix is called *B* and is calculated by:

$$B = X - \bar{X} \quad (8)$$

Where $\bar{X}$ is the matrix of the means for each column. The next step is to compute the singular vectors, also called loadings. In this work, the PCA uses the singular value decomposition method to decompose the *B* matrix into new variables. The centred *B* matrix is therefore written as:

$$B = USV^T \quad (9)$$

Where U is a unitary matrix (which columns are called left singular vectors), S is the matrix of singular values s_i, and V^T columns are called right singular vectors. Singular vectors can be seen as a coefficient of orthogonal transformation. The singular values are the variances explained by each PC. The columns of *US* are the principal components "scores", that we will denote as T:

$$BV = US = T \quad (10)$$

The scores matrix T is the multiplication of the B matrix with the "loadings" matrix V. A loading plot shows variables as points in the space of principal components. They are weights/coefficients that link the original variables to the principal components. The loadings plot shows the pattern of these weights across the principal components (convenient to identify which variables contribute most to each PC).

A PCA model is made of a certain number of principal components and does not necessarily explain the total variation of the data. A correct way to write a data matrix from the PCA point of view would be:

$$X = TV^T + E \quad (11)$$

Where X is a data matrix, T is the scores matrix, V^T is the loading matrix and E is the error matrix. E contains residual values; the differences between the position of projected samples in the PCs space, and their locations in the original space of the data matrix X. This difference is a part of data variation which a PCA model does not capture. These differences can be described using two distances: *orthogonal distance*, OD, and *score distance*, SD.

The orthogonal distance OD can be calculated by taking sums of squares values from $E = \{e_{ij}\}$ along every row:

$$q_i = \sum_{j=1}^{J} e_{ij}^2 \quad (12)$$

This OD distance is denoted as q.

The score distance SD is a squared Mahalanobis distance between the projection of the objects and the origin. It calculates how far a projection of an object to the PC space is from the origin. It is a measure of the extremeness of an object and is usually denoted as h or T^2.

These two distances are important statistics to assess how well samples are described by the PCA model, and if some observations are outliers or extreme, they will have high OD and/or SD compared to other samples. In this study, a distance plot (scatter plot where OD are plotted against SD for each sample) was used to determine how different were some samples from the others.

3.3.4 Partial Least Square (PLS)

In opposition to PCA, which operates dimensionality direction based solely on information contained within the X matrix (independent variables), PLS considers the information within a X and a Y matrix (dependent variables).

PLS and PCA both apply a dimensional reduction. PLS also creates new variables, that are called components or latent variables, as PCA does. But these new variables are not created with the same criterion as PCA. In PCA, the variance of the scores of the data matrix X must be maximised, and in PLS, this is the covariance of the scores between the data of the matrices X and Y that is maximised. This method allows the new latent variables to maximise the linear relation between the scores of X and Y. The latent variables are therefore more relevant for prediction than the principal components that PCA would create.

Let there be two datasets: one spectral dataset of 4 samples, and one dataset of their compositions in fatty acid. The goal is to create a model that predicts fatty acids concentration based on the spectral data set, and to evaluate its performance. To create the model, set that we use 3 samples to build the model, and 1 sample to evaluate it. These are called the train set and the test set, respectively. Now take the train set, composed of 3 samples, and write its spectral data matrix X, and its fatty acids compositions matrix Y. The Y and X matrices can each be decomposed into 3 terms:

$$Y = UQ' + F \quad (13)$$

$$X = TP' + E \quad (14)$$

With U and T, respectively the scores matrices of Y and X, and Q and P respectively the loadings matrices of Y and X, with F and E being the error terms. The principle is that in prediction, the loading matrices are part of the prediction model, and the only data that is needed is the T, the score matrix of X, to make the relation between the X and the Y matrices. Indeed, if we can estimate U, we can calculate Y. For a latent variable i, the score of Y, noted u_i, can be estimated by the score of X, noted t_i via the relation :

$$u_i = t_i r_i \quad (15)$$

With r_i, the coefficient of regression for a latent variable i. That is what defines PLS regression method. The PLS finds components such as u_i has maximum covariance with t_i.

To build a PLS model, the number of latent variables is important because we don't want a model that is too complex, as it could lead to overfit the relation between X and Y. So, practically, the train set is split into two parts: the calibration set, and the validation set. The calibration set is the data that is used to build the model in the first place. Then, the validation set is necessary to tune the model and determine the optimal number of latent variables before testing it on the new samples in the test set. Generally, it is wise to use 20% of the train set for the construction of the validation set. The distribution of the data into calibration and validation sets can be chosen randomly or using a predefined criterion as the Kennard-Stone algorithm. This algorithm ensures a maximal variability in both calibration and validation sets when splitting the train set.

PLS aims to predict accurately the Y, that in our case are the concentration of several fatty acids, from a new matrix of X data, that is the Raman spectral data. The model is built with the calibration set. The Y and the X of the calibration set are known, so their loadings matrices Q and P, and the regression coefficients R can be found by PLS algorithm. Once the model is built, new samples can be predicted. The process that is used to predict with PLS model can be noted as:

$$\hat{T} = X_{new}P \rightarrow \hat{U} = \hat{T}R \rightarrow \hat{Y} = \hat{U}Q' \quad (16)$$

Where R *is* the matrix of r_i coefficients that links the T matrix to the U matrix as shown in equation 15. The new samples must be unknown to the model. The spectral data from new samples is noted X_{new} (the spectral data matrix of the test set) and it is the only input that the model needs to predict the new Y, noted $\hat{Y}$ (an estimator of the fatty acids compositions of the test set). The model is made of the P, R and Q matrices calculated from the calibration set. The score matrix $\hat{T}$ of the new samples spectral data matrix (X_{new}) is calculated by multiplying the X_{new} matrix by the loadings matrix P of the model. These scores are multiplied by the R matrix found in equation 15 to get the score matrix of the new samples noted $\hat{U}$, which allows to predict the $\hat{Y}$ data, the concentration in fatty acids of the new samples.

The main performance indicator of PLS-Regression models is the RMSEP, that will be used in this work. The RMSEP is the root mean square error of prediction of *Y* and it gives an error in the same scale as the original Y values. It is given by :

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (17)$$

Where y_i are the observed values and $\hat{y}_i$ are the predicted values. RMSEP is a relative value. There is no general RMSEP threshold that defines a good or bad model. However, a lower RMSEP is associated to a lower error, thus it is a good mean to evaluate a model. An eventual threshold must be set towards the type of data and the objectives of the model, to determinate if the model is accurate or poor. Another indicator is the R^2. Let there be $\bar{y}$, the mean of the observed values of Y:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (18)$$

Then R^2 is measured by these formula :

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (19)$$

$$with\ SS_{res} = \sum (y_i - \hat{y}_i)^2$$

$$and\ SS_{tot} = \sum (y_i - \bar{y}_i)^2$$

R^2 represents the percentage of the total variance in the target variable Y that can be explained by the predictive variables X. The closer R^2 is to 1, the more effectively the PLS-R model explains the variance in the *Y* data. However, a very high R^2 does not necessarily mean

that the model is good. The high complexity of a model increases the R^2. It should be avoided as overfitting would be a risk.

In the case of PLS-R, PLS is used for quantification. But PLS can also be used for classification. In this case, this is called PLS-Discriminant Analysis (PLS-DA). The PLS-DA works as a PLS-R, as it uses a PLS method to reduce dimensionality. The aim is to predict which sample belongs to which class. The Y matrix is made such as each class has a Y value called "dummy value". For example, wild salmon has a value equal to -1 and farmed salmon has a value equal to 1. The PLS-DA seeks the components that represent the highest covariance between the instrumental data (the Raman spectra) and the dummy values. When a new Raman spectrum is given to the model, it predicts a value between -1 and 1. In PLS-DA, a threshold must be set such as, based on the predicted dummy value, it classifies the object in a class or another. In our example, a threshold of 0, between -1 and 1 would be logical. An object with a dummy value between 0 and -1 would be classified in A, and an object with a dummy value between more than 0 and 1 would be classified in B, for example. The final output is the classification of the object. A useful way to have a good understanding of the model performance is to look at the confusion matrix. Set a confusion matrix of a model that discriminates between 3 classes, where the rows are the observed classes and the columns are the predicted classes:

		Class 1	**Class 2**	**Class 3**
Observed classes	**Class 1**	TP1	FP2	FP3
	Class 2	FP1	TP2	FP3
	Class 3	FP1	FP2	TP3

Predicted classes

Fig. 11. Confusion matrix of a PLS-DA with 3 classes.

The TP1, TP2 and TP3 cases are the "true positives", the objects that have been correctly classified. The other cases are the "false positive", and they are the misclassified objects. For example, if a class 2 object has been classified as a class 3 object, it will be placed in the FP3 case at row 2 and column 3 of the confusion matrix.

A numerical indicator of the performance of a PLS-DA model that is widely used is the accuracy. Based on the confusion matrix of the Fig.11, it would calculated by:

$$Accuracy = \frac{TP1+TP2+TP3}{TP1+TP2+TP3+FP1+FP2+FP3} * 100\% \quad (20)$$

4 Results and Discussion

4.1 Gas Chromatography Results

4.1.1 The Fatty acids profile of Salmon

Gas chromatography was used to estimate the fatty acid profile of the salmon fat. The total lipid quantitation of the salmons has not been evaluated. Each fatty acid composition is evaluated as a percentage of the total peak area given by gas chromatography. The GC consisted of the analysis of 24 samples, each measured in triplicate and averaged. The 24 samples set was composed of 3 wild, 6 organic farmed and 15 conventionally farmed samples.

Supposing that the fat extraction method gave good results and that every component detected in the GC are FAs (except for the solvents), then the arithmetic means of each FAs proportions in wild, conventionally farmed, and organic labelled farmed salmons were calculated as the peak area for each FA, on the sum of all the detected FAs peak areas. The salmon fat composition of a sample in any fatty acid is therefore calculated by:

$$\%Composition(FattyAcid_i) = \frac{FattyAcid_i\ peak\ area}{\sum_j^{i=1} FattyAcid_i\ peak\ area} \quad (21)$$

With $\sum_j^{i=1} FattyAcid_i\ peak\ area$, the sum of all peaks detected by the GC for a sample.

For each FA, a comparison between their proportion in the three groups was assessed using an ANOVA test. Statistical significance was set at $P < 0.05$. If the p-value < 0.05, the difference between the groups is significant, and if the p-value > 0.05, the difference is not considered significant. Therefore, the low quantities of samples in two of the three groups (3 wild and 6 organic labelled) can cause issues regarding the accuracy of the test.

The Tab.3 below only includes the fatty acids present in the standard solution injected before the samples. This is why some peaks are not identified. The proportion of unidentified fatty acids is listed in the table in the "Unidentified Fatty Acids" row.

Tab. 3. Mean fatty acids composition (%) of the wild, conventionally farmed and organic farmed salmon samples.

FAs	Wild	Conventionally Farmed	Organic farmed	
	Composition (%)	Composition (%)	Composition (%)	p-value
C12:0	0.0 ± 0.0	0.03 ± 0.02	0.01 ± 0.01	p < 0.05
C14:0	3.35 ± 1.03	2.27 ± 0.4	3.51 ± 1.31	p < 0.05
C15:0	0.33 ± 0.29	0.2 ± 0.02	0.24 ± 0.07	p > 0.05
C16:0 (Palmitic acid)	12.87 ± 4.86	9.37 ± 0.96	10.74 ± 2.07	p < 0.05
C17:0	0.19 ± 0.16	0.15 ± 0.02	0.14 ± 0.01	p > 0.05
C18:0	2.82 ± 1.57	2.7 ± 0.51	2.75 ± 0.4	p > 0.05
C20:0	0.11 ± 0.1	0.15 ± 0.13	0.18 ± 0.09	p > 0.05
C22:0	0.0 ± 0.0	0.15 ± 0.03	0.11 ± 0.03	p < 0.05
C24:0	0.0 ± 0.0	0.06 ± 0.02	0.04 ± 0.02	p < 0.05
SFA	19.77 ± 6.04	15.12 ± 1.35	17.72 ± 3.8	p < 0.05
C14:1n-5c	0.0 ± 0.0	0.01 ± 0.01	0.03 ± 0.03	p < 0.05
C16:1n-7c	3.43 ± 0.54	2.47 ± 0.28	4.23 ± 1.81	p < 0.05
C17:1n-7c	0.23 ± 0.25	0.09 ± 0.04	0.09 ± 0.04	p < 0.05
C18:1n-9t	0.05 ± 0.09	0.05 ± 0.03	0.07 ± 0.01	p > 0.05
C18:1n-9c (OA)	10.13 ± 2.67	36.07 ± 2.96	24.73 ± 13.03	p < 0.05
C20:1n-9c	3.58 ± 1.85	4.53 ± 1.89	7.19 ± 2.31	p < 0.05
C22:1n-9c	1.87 ± 0.7	0.55 ± 0.22	0.88 ± 0.29	p < 0.05
C24:1n-9c	1.12 ± 0.21	0.38 ± 0.14	0.45 ± 0.11	p < 0.05
MUFA	23.07 ± 2.86	47.32 ± 2.1	40.35 ± 8.09	p < 0.05
C18:2n-6c (LA)	1.78 ± 0.26	12.95 ± 3.05	13.67 ± 0.36	p < 0.05
C18:3n-6c	0.0 ± 0.0	0.1 ± 0.05	0.08 ± 0.01	p < 0.05
C18:3n-3c (ALA)	0.89 ± 0.13	4.79 ± 1.42	3.74 ± 1.88	p < 0.05
C20:2n-6c	0.63 ± 0.56	1.02 ± 0.14	0.95 ± 0.19	p < 0.05
C20:3 c8,11,14 (n-6)	0.07 ± 0.12	0.27 ± 0.12	0.19 ± 0.03	p < 0.05
C20:3 c11,14,17 (n-6)	0.17 ± 0.2	0.42 ± 0.13	0.33 ± 0.21	p < 0.05
C20:4n-6c	0.44 ± 0.19	0.26 ± 0.1	0.22 ± 0.02	p < 0.05
C22:2n-6c	0.81 ± 0.3	0.1 ± 0.02	0.08 ± 0.02	p < 0.05
C20:5n-3c (EPA)	7.33 ± 0.92	2.65 ± 0.3	2.97 ± 0.76	p < 0.05
C22:6n-3c (DHA)	19.37 ± 9.43	4.6 ± 1.08	4.94 ± 0.84	p < 0.05
PUFA	31.4 ± 11.01	26.89 ± 3.42	26.97 ± 1.25	p > 0.05
Unidentified FAs	25.76 ± 6.34	10.67 ± 5.82	14.96 ± 4.88	p < 0.05

∑ ω-9	16.7 ± 0.95	41.53 ± 2.01	33.27 ± 10.39	$p < 0.05$
∑ ω-6	3.73 ± 1.1	14.71 ± 3.37	15.17 ± 0.36	$p < 0.05$
∑ ω-3	27.77 ± 9.95	12.48 ± 1.41	11.98 ± 1.06	$p < 0.05$
n-6/n-3	0.13	1.18	1.27	$p < 0.05$
C18:1 cis	13.63 ± 3.5	39.77 ± 2.76	28.32 ± 12.12	$p < 0.05$
C18:1 TFA	1.27 ± 0.61	0.68 ± 0.4	0.96 ± 0.41	$p < 0.05$
C18:2 TFA	0.03 ± 0.06	0.09 ± 0.03	0.09 ± 0.03	$p < 0.05$

4.1.2 FAs in wild and farmed Salmon

Total lipid content in wild and farmed salmon have not been evaluated. Therefore, experimentally, the wild samples were the only three to give a lipid recuperation percentage from the muscle under 3% with the Folch extraction, compared to a mean of 15% (+- 7%) of lipid recuperation for the farm-raised salmons. This could show that the wild salmons used in the study are less fat than the farm-raised ones.

To identify the components, retention times were given using a standard free fatty acid solution composed of 37 fatty acids injected in GC just before the samples. A total of 31 FAs have been identified amongst these 37, C12:0 being the shortest and C24:1 being the longest. Compared to wild salmon, three FAs particularly show up in the farmed salmon profile: oleic acid (OA, C18:1n-9c), linoleic acid (LA, C18:2n-6c) and alpha-linolenic acid (ALA, C18:3n-3c). There are present in high concentration and account for more than 50% of the total composition of conventionally farmed, and 40% for organic farmed salmons. In the wild counterpart, they only account for about 12%.

In the wild salmons, the docosahexaenoic acid (DHA, C22:6n-3c), the eicosapentaenoic acid (EPA, C20:5n-3c), oleic acid (OA, C18:1n-9c) and palmitic acid (C16:0) were the major FAs in concentration. The dominant FA in farmed salmon is OA and it represents 36% of the total lipid composition. This is already discussed in the literature and this difference could be due to the increased amount of plant-based ingredients in farmed salmon feed (Friesen et al., 2015; Sprague et al., 2016). OA, LA and ALA are the most commonly found FA in plants made feed, and particularly in rapeseed oil (Sharafi et al.,2015), which is one of the main ingredients in salmon feed in Norway (Aas et al., 2019).

As farmed salmon would have a higher lipid content (Molversmyr et al., 2022), the composition do not reflect the concentration of FAs in salmon muscle. However, it gives some qualitatives features of the fat. Both EPA and DHA have been found in farmed and wild salmon, but their proportions are respectively 3 and 4 times higher in wild salmon (7% and 19% respectively) than in farmed salmon (2,5% and 4,5% respectively).

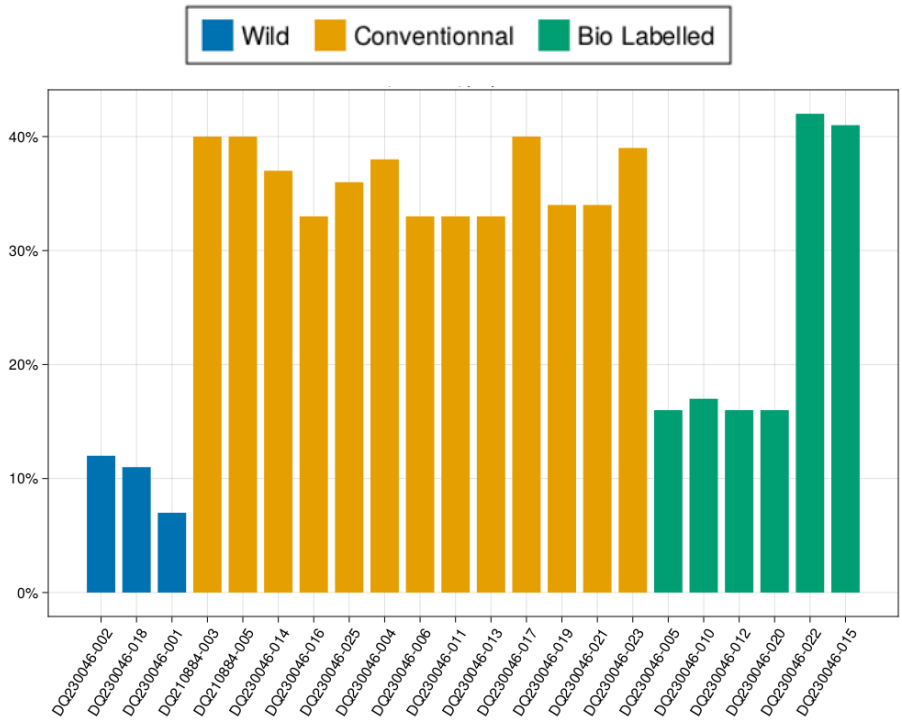


Fig. 12. Oleic acid (OA, C18:1n-9c) proportions in wild (blue), conventionally farmed (orange) and organically farmed (green) salmon samples fat.

4.1.3 Comparison of SFA, MUFA and PUFA in wild and farmed Salmon

It is well established now that saturated fatty acids (SFAs), monounsaturated fatty acids (MUFAs) and polyunsaturated fatty acids (PUFAs) are associated with different effects on human health. MUFAs and PUFAs seem to have positive effects on human health that SFAs don't provide. These proportions are calculated by doing the sums of the FAs identified. As the unidentified FAs are not accounting in SFAs, MUFAs and PUFAs, and that wild salmon have a significantly higher proportion of them, the accuracy of these values can be discussed. Also, as the GC standard of 37 FAs was initially made for milk analysis, some long-chain polyunsaturated fatty acids are missing and possibly couldn't be identified. This could explain the higher level of unidentified components in wild salmons (26% vs. 11% and 15% for farmed types).

Results show that wild salmon seems to have a slightly higher proportion of SFAs (19%, compared to approximately 15% for conventionally and 17% for organic farmed salmon). There seems to be approximately the same proportion of PUFAs between the three groups. However, due to the higher concentration of OA in farmed salmons, the most pronounced difference is at the level of MUFAs, with proportions of 23%, 47%, and 40% for wild, conventional, and organic salmon, respectively.

Furthermore, C18:0, C16:0 and C14:0 constitute the majority of the total SFAs for the three groups, while C16:1n-9c, C18:1n-9c (OA), C20:1n-9c and C22:1n-9c are present in major quantities in MUFAs. At the level of PUFAs, EPA (7%) and DHA (19%) are the principal

components for wild salmon. However, LA (13%), ALA (5%) and DHA (4.5%) are the major ones for farmed salmons.

4.1.4 Comparison of ω-6 and ω-3 FA in wild and farmed Salmon

The ω-3 and ω-6 FAs have different biological activities on human health. Based on the results, wild and farmed salmons display similar proportions in total ω-6 and ω-3 (about 30% for both). Wild salmon is primarily composed of ω-3 fatty acids rather than ω-6 (27.5% and 3.5%, respectively) whereas farmed salmon contains approximately equal amounts of ω-3 and ω-6 fatty acids (14% and 12%, respectively). It is largely said in the literature that an ω-6 / ω-3 ratio of a diet has a direct impact on human health. As they compete for their biosynthetic enzymes and because they have distinct physiological and metabolic properties, a balanced ratio is critical for health (Simopoulos, 2016). The high imbalance has been linked to many diseases, such as obesity and coronary heart disease. The calculated ratio for farmed salmon in this study corresponds well with the findings of Aas et al. (2019) and Molversmyr et al. (2022) with a 1.15/1 ω-6/ ω-3 ratio, and the wild salmon ratio was 0.13/1, that is in the same order of magnitude than the one found by Molversmyr (0.06/1). These ratios once again show the use of vegetable oil in the farmed salmon diet. As a ω-6/ ω-3 ratio below 5/1 is beneficial for human health (Yang et al., 2016) and the actual northern Europe diet's ratio is around 15 (Simopoulos, 2016), both of the wild and farmed salmon analysed in this study could contribute to reducing the level of ω-6 in favour of ω-3.

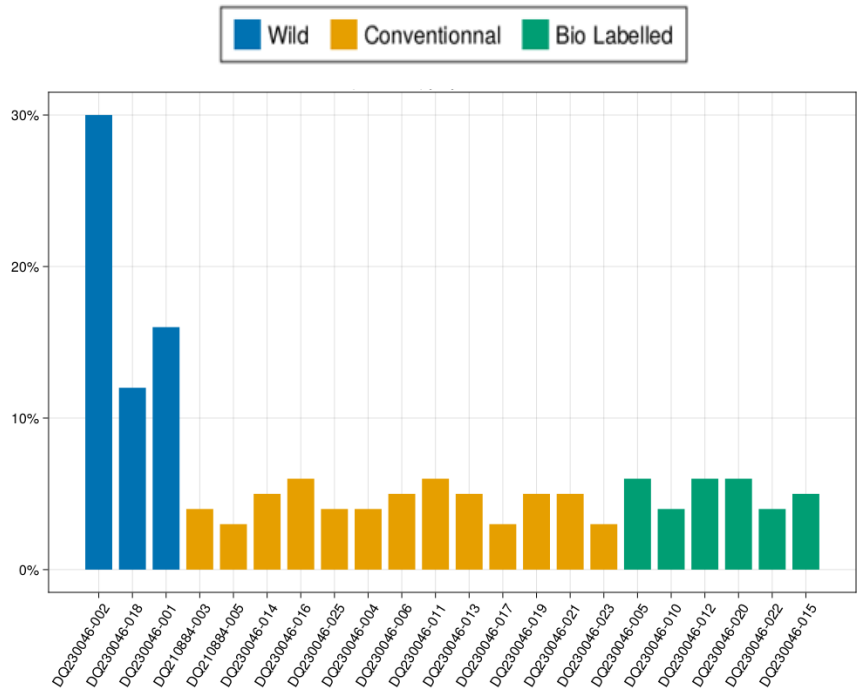


Fig. 13. DHA (C22:6-n3) proportions in wild (blue), conventionally farmed (orange) and organically farmed (green) salmon samples fat.

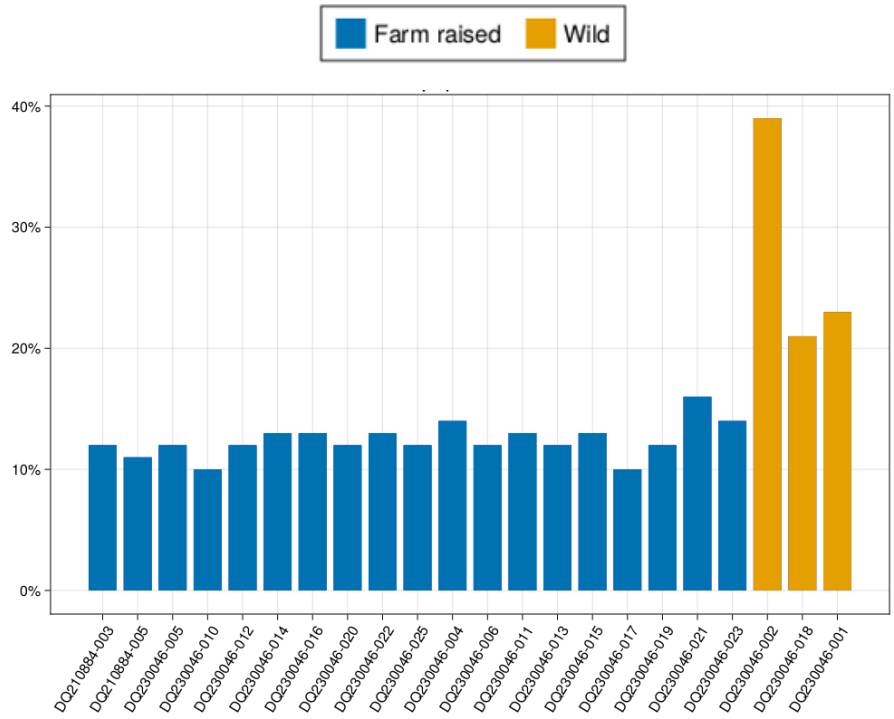


Fig. 14. ω-3 proportions in wild (orange) and farmed (blue) salmon samples fat.

4.1.5 Using PCA to detect FAs tendencies between wild and farmed Salmon.

4.1.5.1 1st PCA

A PCA was used to look at the differences between the FAs proportions in the samples. The proportion values for each fatty acid were centred and scaled to avoid giving excessive weight to the most abundant fatty acids. This approach allows us to focus on the subtle differences between the fatty acid profiles of different salmon samples. The fatty acid C12:0 was not considered because it was only detected in one of the GC columns, and the fatty acid C21:0 was not considered due to insufficient statistical evidence of its presence. In total, the dataset is composed of 26 fatty acids proportions of 24 salmon samples.

Fig.15 shows a clear separation in the scores along the PC1. The separation occurs between conventionally farmed salmon, four out of six certified organic farmed salmon, and wild salmon. PC3 also appears to separate conventional from some organic salmon. The loadings provide an overall view of the influence of fatty acids on the PCA scores.

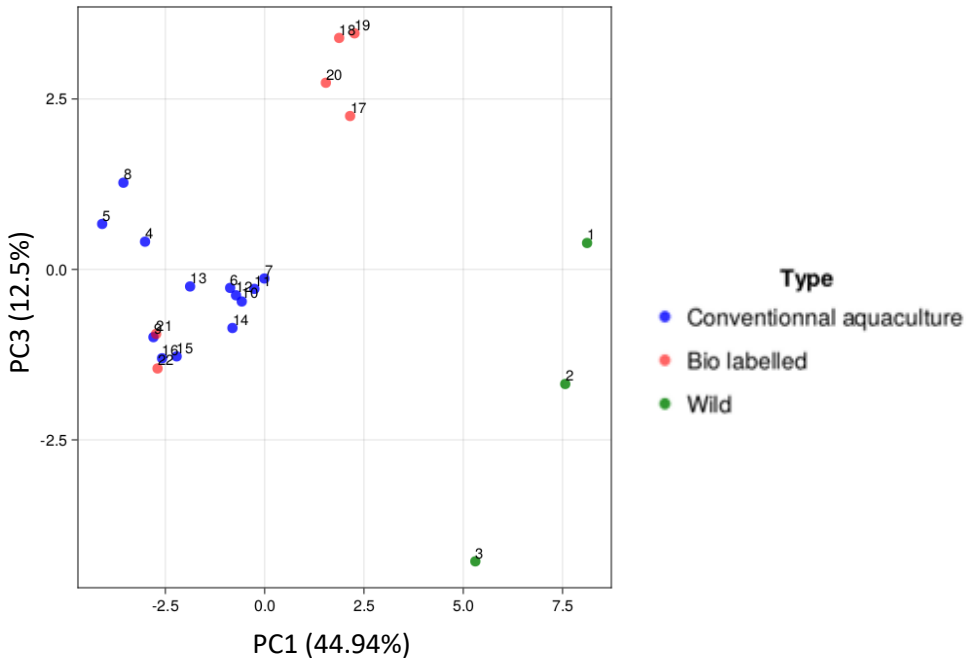


Fig. 15. 1st PCA scores on PC1 (44.94%) and PC3 (12.5%) of conventional farmed (blue), organic farmed (red) and wild (green) salmons (using 26FAs)

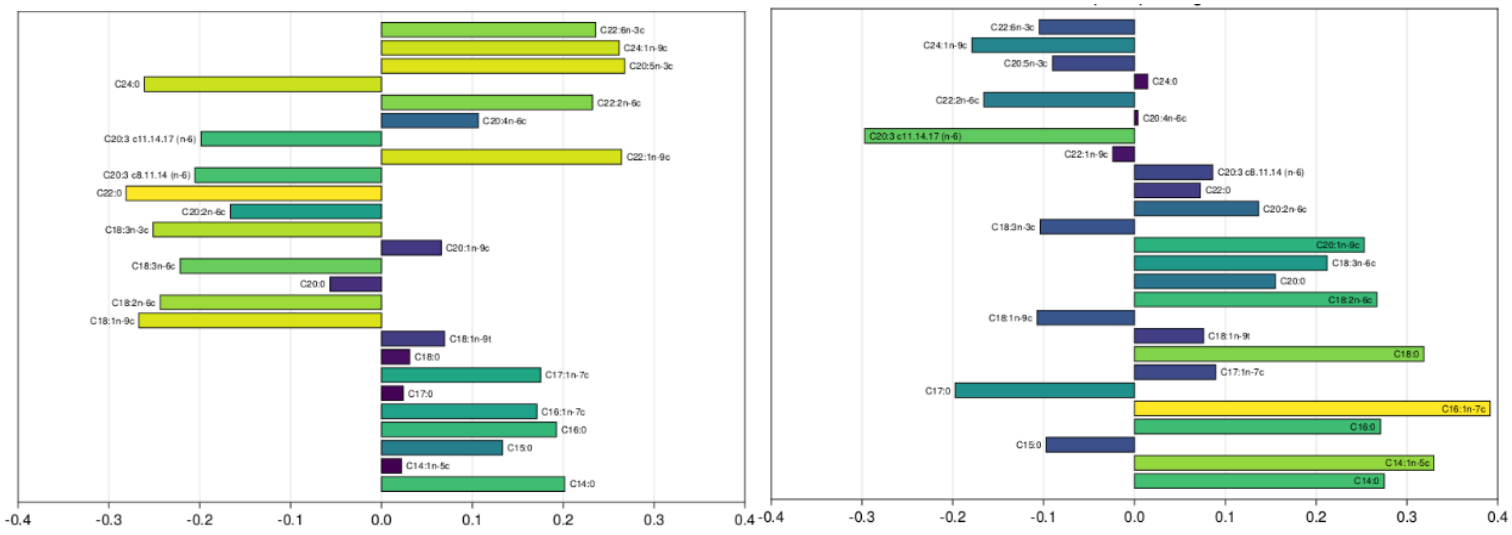


Fig. 16. PC1 and PC3 loadings of the Fig.15 PC scores graph. A FA with a longer bar has a higher influence on the PC (no matter the positive or negative direction).

4.1.5.2 2nd PCA

A second PCA was subsequently applied after restraining the data to fatty acids proportions of the seven FAs that showed the highest loadings in PC1. As a result, the PC1 of the 2nd PCA enables a separation between the three groups in the same manner as the PCA with 26 variables.

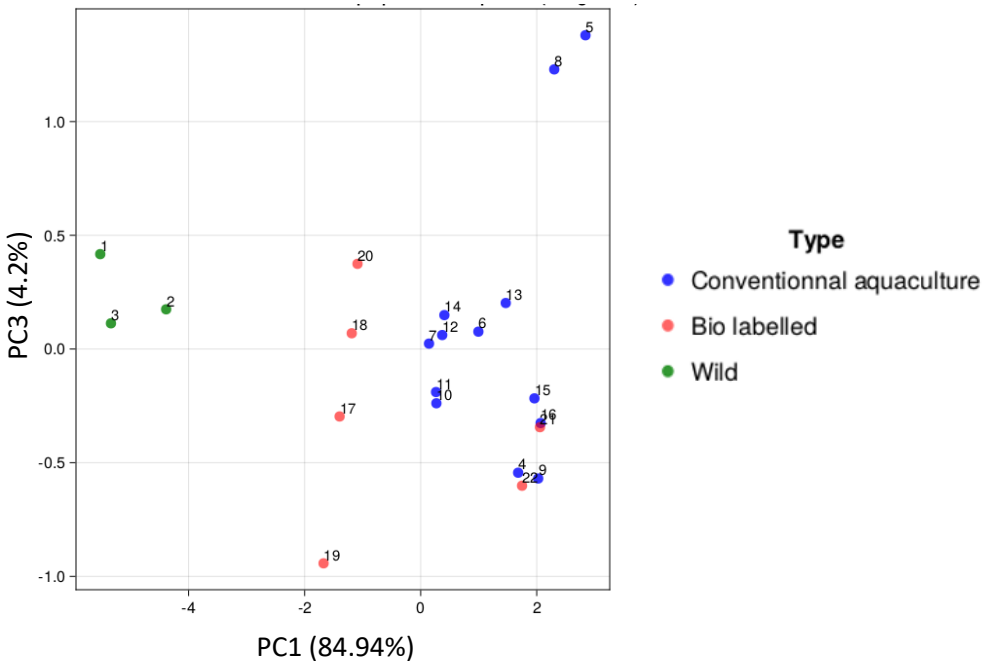


Fig. 17. 2nd PCA scores on PC1 (84.94%) and PC3 (4.2%) of conventional farmed (blue), organic farmed (red) and wild (green) salmons (using 7 FAs).

4.1.5.3 Correlation matrix and 3rd PCA

To further reduce the number of variables required for separation in our case, we can examine the existing correlation between different fatty acids. If the presence of one fatty acid is strongly correlated with another, only one of them may be used in the PCA. The signs of the loadings in Fig.16 already indicate this correlation. If two fatty acids have opposite signs, there will be a negative correlation, and if they have the same sign, the correlation will be positive. Fig.15 provides the correlation values between the seven fatty acids in the second PCA.

Fatty acids that are not correlated with others should remain as variables to ensure a good separation in the scores of the PCA. However, those that are correlated with one or more other fatty acids are not necessary. Thus, C18:1n-9c, C18:3n-3c, and C20:5n-3c are kept because they are not highly correlated with others. Additionally, C22:0 is retained as it appears to be correlated with C24:0 and C24:1n-9c, which can then be discarded.

Tab. 4. Correlation matrix of the 7 fatty acids chosen for 2nd PCA. The highest correlations values between the two FAs are in bold.

FAs	C18:1n-9c	C18:3n-3c	C22:0	C22:1n-9c	C24:0	C20:5n-3c	C24:1n-9c
C18:1n-9c	**1**	0,85	0,87	-0,83	0,81	-0,79	-0,82
C18:3n-3c	0,85	**1**	0,79	-0,78	0,75	-0,75	-0,78
C22:0	0,87	0,79	**1**	-0,88	**0,93**	-0,88	**-0,9**
C22:1n-9c	-0,83	-0,78	-0,88	**1**	-0,8	0,82	**0,91**
C24:0	0,81	0,75	**0,93**	-0,8	**1**	-0,72	-0,79
C20:5n-3c	-0,79	-0,75	-0,88	0,82	-0,72	**1**	0,84
C24:1n-9c	-0,82	-0,78	**-0,9**	**0,91**	-0,79	0,84	**1**

A third PCA, utilizing these four variables, once again demonstrates the same separation along PC1. In our case, for the separation of our samples, we were able to move from a complete characterization of the fatty acid profile to a characterization using only four fatty acids, while still achieving a robust separation.

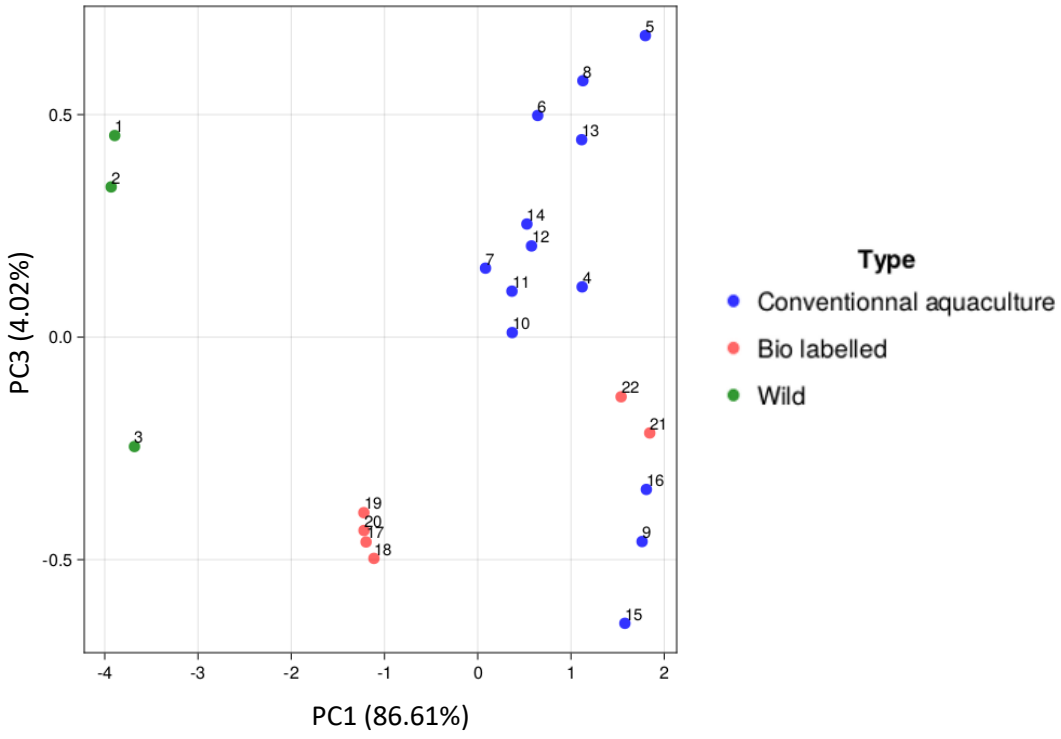


Fig. 18. 3rd PCA scores on PC1 (84.94%) and PC3 (4.2%) of conventional farmed (blue), organic farmed (red) and wild (green) salmons (using 4 FAs).

4.2 Raman Spectra of Salmon

4.2.1 General salmon spectra, major bands and interpretation

Raman spectroscopic analysis was performed using three different wavelengths for the incident light. The resulting spectra are quite different in their profile and some information seems to be visible depending on the laser used. The choice of the wavelength of incident light for Raman could therefore be chosen depending on the molecule to analyse in the sample. In the case of salmon, it seemed that carotenoids were more visible using the 532 nm laser, and lipids showed a much higher Raman scattering intensity at 1064 nm and 785 nm than at 532nm. Since the overall information is comparable between the spectra, the use of different lasers seems to allow a better focus depending on the molecule to be analysed.

Fig.19 illustrates the differences between the scatterings obtained with the three lasers. Each spectrum represents the average of all salmon spectra measured with each laser. This approach allows us to observe only the general peaks while smoothing out the peaks that are relatively absent by averaging.

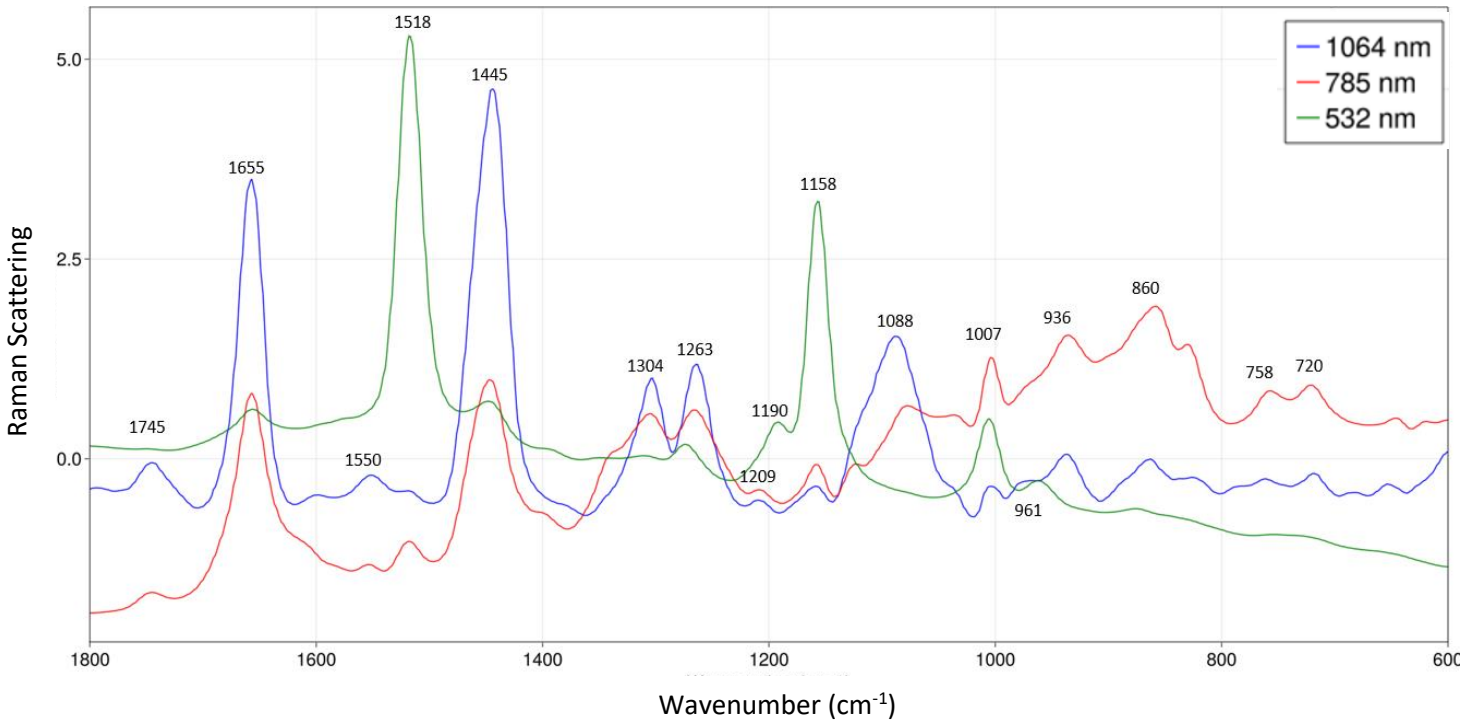


Fig. 19. Salmon fillet average Raman spectra acquired with 532nm (green), 785nm (red) and 1064nm (blue) lasers.

Based on this graph, we can identify the peaks that stand out most in the data from the three lasers. The wavenumbers at which these peaks occur provide information about the sample, and the relationships between wavenumbers, chemical bonds, and molecules are documented in the literature. By cross-referencing the average salmon spectrum for each laser with the chemical information for specific wavenumbers, it is possible, with a certain degree of certainty, to chemically interpret the information contained in the Raman spectra. Tab.5 lists the peaks and associated chemical bonds that are most likely to be visible in the spectra. These bonds are in turn associated with molecules typically found in salmon.

The 532nm Raman spectrum is dominated by bands associated with carotenoids. The main scattering bands are the C=C stretch (1518 cm^{-1}), and the polyene C-C stretch (1158 cm^{-1}). The 1064nm and 785nm spectra are dominated by bands associated with fatty acids. The main bands related to fatty acid unsaturation are the C=C stretch (1655 cm^{-1}) and the alkene C-H bend (1263 cm^{-1}). The main bands related to fatty acid saturation are the methylene scissor deformation (1445 cm^{-1}) and methylene twisting deformation (1304 cm^{-1}). Other bands of interest originate from the aromatic ring breathing of phenylalanine (1007 cm^{-1}) and the liquid aliphatic C–C stretch (1088 cm^{-1}) (Lintvedt et al., 2023). The regions before 600 cm^{-1} and after 1800 cm^{-1} are not shown in the graph because it was noisy and detrimental to model performances and therefore discarded from the further analyses. We will refer to the 600 to 1800 cm^{-1} range as the "fingerprint" area, and in most analyses, the 600-800 cm^{-1} area will also be discarded.

The main bands found in Fig.19 and cited by other works on salmon spectroscopy analysis have been listed in Tab.5 below. Each band is described with the vibration type, and the most probable component associated with the band, in the salmon muscle matrix.

Tab. 5. Major bands found in Fig. 19, their vibration types, and their most probable associated component of the salmon muscle matrix, towards literature. The relative intensities of the bands for each laser are given by vw (very weak), w (weak), s (strong) and vs (very strong).

Wavenumber (cm^{-1})	Vibration type	Associated Component	1064nm	785nm	532nm	References
720	C-C stretching	Collagen	vw	w	i	(Lafleur et al., 1972)
758	C-C Ring Breathing	Tryptophan	vw	w	i	(Stone et al., 2004)
828		Out-of-plane ring breathing tyrosine/O-P-O stretch DNA	wv	w	i	(Stone et al., 2004)
		Phosphodiester				(Ruiz-Chica et al., 2004)
856	v(C-C)	(C-C) vibration of the collagen backbone				

		Hydroxyproline (collagen type I)	v	s	vw	(Cheng et al., 2005)
936	δ(N-C=N)	Guanine	s	s	i	(Lafleur et al., 1972; Mathlouthi et al., 1986)
961	δ(=C-H) trans RHC=CHR	Lipid	i	i	w	(Afseth et al., 2006; Sánchez-Alonso et al., 2012)
1000-1007	In plane CH3 rock Aromatic ring breathing	Phenylalanine	w	s	s	(Movasaghi et al., 2007)
1088	C–C/C–N/C–O stretch	Lipids	vs	w	i	(Afseth et al., 2006)
1158	C-C polyene chain stretch	Carotenoids	w	w	vs	(Afseth et al., 2006; Hikima et al., 2017)
1190	/	Carotenoids	i	i	w	(Esmonde-White and Lewis, 2022; Mathlouthi et al., 1986)
1209	C-C6H5 stretching mode	Phenylalanine and tyrosine	vw	vw	i	(Lau et al., 2005)
1263	Alkene C-H bend	Lipids	vs	s	w	(Malini et al., 2006)
1304	Methylene twisting deformation	CH2 deformation (lipid), adenine, cytosine	vs	s	vw	(Stone et al., 2004)
1445	CH2 scissoring	Methylene-Lipid	vs	vs	w	(Afseth et al., 2006; Sánchez-Alonso et al., 2012)
1518	In phase C=C stretch	Carotenoids	vw	w	vs	(Afseth et al., 2006; Huang et al., 2003)

1550	v(C=C)	Tryptophan	w	vw	i	(Huang et al., 2003)
1655	cis C=C stretch	Lipid	vs	vs	w	(Lakshmi et al., 2002)
1745	C=O stretch	Lipids	s	w	i	(Lakshmi et al., 2002)

4.3 Salmon Spectra by Classes

Before going further, it is interesting to notice some eventual visible differences between the spectra of the analysed samples. The samples have been categorized into different classes?types. Each sample is categorized by its origin (wild or farm-raised), its geographical origin, its price, its mean of culture (organic or conventional) and its freshness.

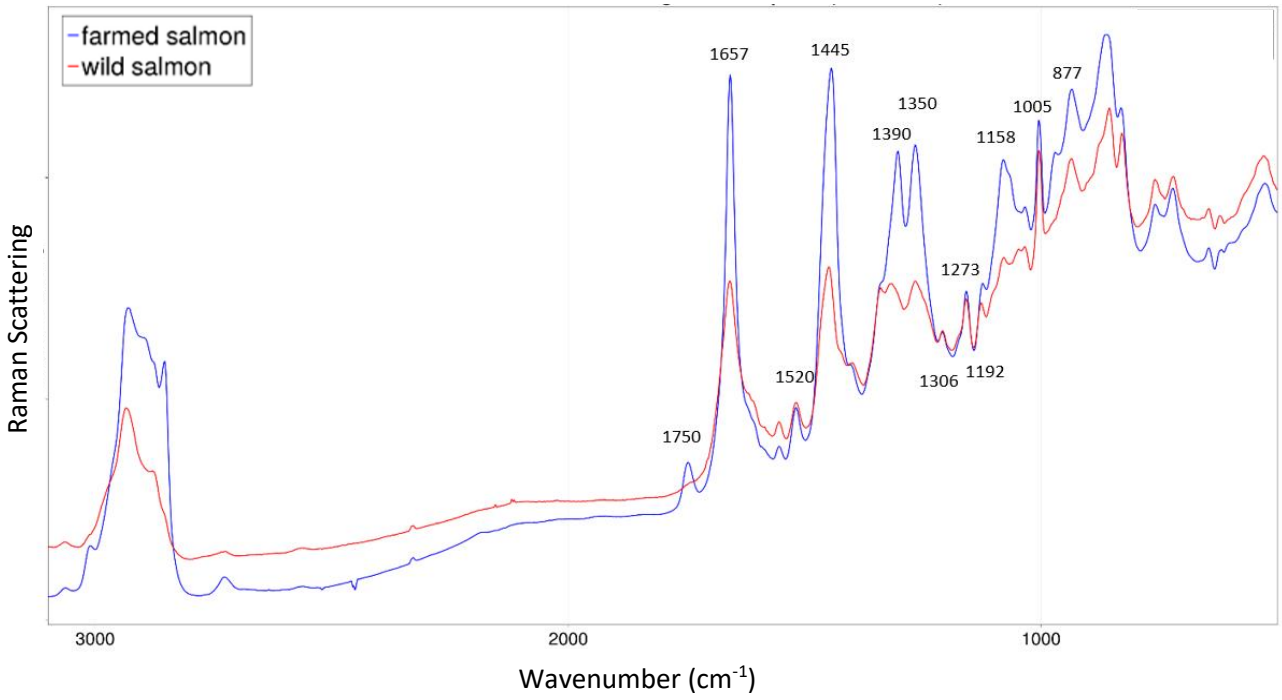


Fig. 20. Average spectra of farmed (blue) and wild (red) salmons acquired with 785nm and pre-processed with SNV.

The two spectra of Fig.20 are the mean spectra of all wild and farmed salmon samples in 785 nm. What can be noticed in the general profile of the two spectra is the absence of a peak near 1750 cm^{-1} in the wild salmon's mean spectrum, whereas it is present in the farmed salmon's mean spectrum. This peak could be related with lipids as seen in Tab.5 (with 1745

cm^{-1}). The other peaks seem to appear in both spectra but at different intensities. The 532nm spectra of wild and farmed in Fig.21, however, do not show such visual difference.

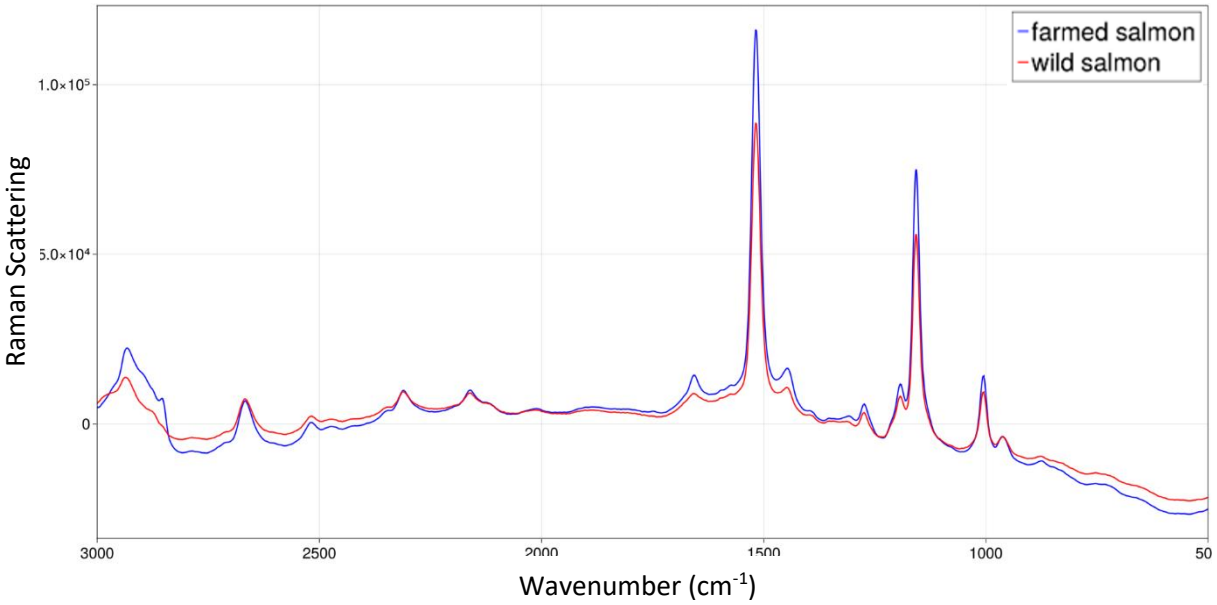


Fig. 21. Average spectra of farmed (blue) and wild (red) salmons acquired with 532nm and pre-processed with SNV.

4.4 Challenges linked to the analysis of Salmon fillets by Raman spectroscopy

4.4.1 Reproducibility and Repeatability Studies

The reproducibility and repeatability studies are used to evaluate a method's accuracy. Therefore, the two concepts don't evaluate the same thing.

A definition of repeatability could be the "narrow agreement between successive results, obtained with the same method, on identical material, under the same conditions (same operator, same measuring device, same laboratory, repetitions over a short duration)".

A definition of reproducibility could be the "narrow agreement between successive results obtained with the same method on identical material subjected to testing under different conditions (for example, with different operators or different laboratories).

In this work, this study aims to evaluate the repeatability and reproducibility of the process from the sampling to the analysis of the salmon muscle by the Raman spectrometer. To have an idea of the repeatability and reproducibility of the method, PCA was used. The graphs show the orthogonal distance and score distance between the spectra and the PCA for

each point. The higher the orthogonal and score distances (OD and SD), the more a point statistically differs from the other. This is used to demonstrate the heterogeneity level within a set of measurements.

4.4.1.1 Repeatability

To evaluate the repeatability of the measurement method used to analyse salmon with Raman spectroscopy in this work, a measurement of 4 spectra was performed six times in one day on a sample, separating each measurement by one hour. The spectra were acquired using a 785nm laser and were not pre-processed at all. A PCA with 10 principal components was then applied to the data to explore variations between measurements. The difference between the points could be due to the time parameter. However, it also could come from the intrinsic heterogeneity of the sample, as the tissues analysed are likely to be heterogeneous.

The graph of the OD on the SD shows that no group of points is very far from the other. The points of each group are disseminated amongst the others. This gives a good insight that there would be no outlying group of spectra that would be very different from the others and represent an important repeatability problem (as explained at the end of the 3.3.3 section).

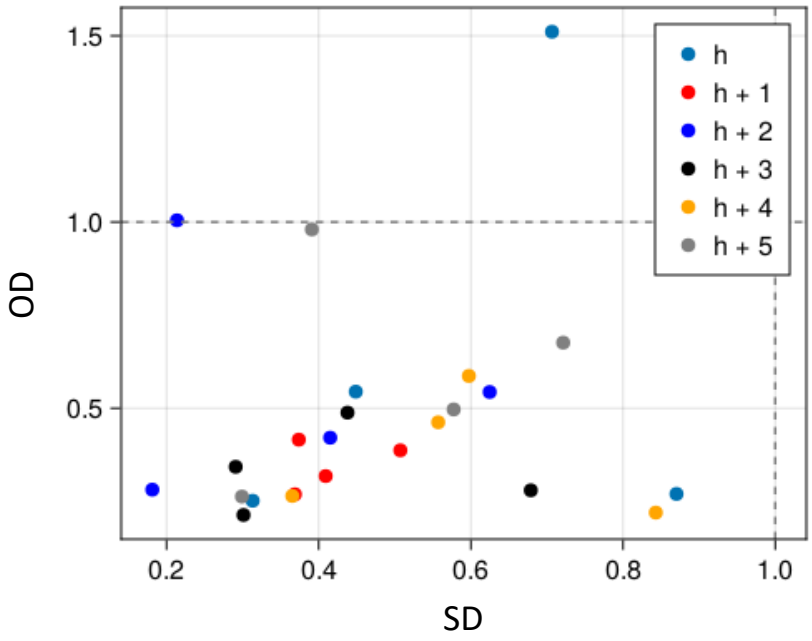


Fig. 22. Orthogonal Distance (OD) vs. Score Distance (SD) of a 10 PC PCA of multiple 785nm Raman measurements of a sample, once an hour for 5 hours (4 spectra by hour so 4 points for each hour on the graph).

4.4.1.2 Reproducibility

To assess reproducibility, a sample was analysed once a day for 4 days. Each day, 36 spectra were measured. By doing that, it allows more variation of the measurement conditions due to the longest period that the repeatability study. The operator, however, was always the same person. The spectra were acquired using a 532nm laser and were not pre-processed at all. The OD-SD graph tells us that the J group have a few points that have a high SD and a high OD compared to the other. These points are far from the other in the PC space and they are not well explained by the PCA. This means that this group contains more extreme data than the other groups. This difference could be due to a variation in the spectra due to a variation in time. However, most of the points in the graph are close to each other, so there seems to be a good reproducibility on 4 days.

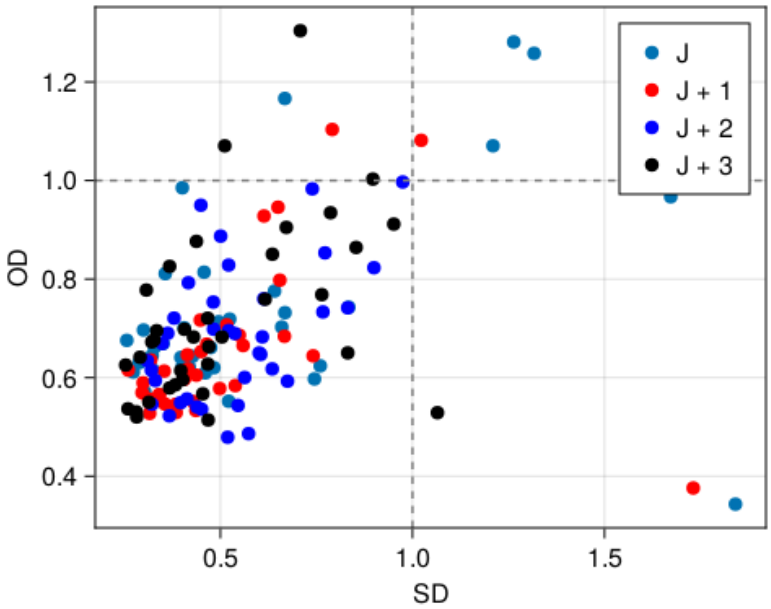


Fig. 23. Orthogonal Distance (OD) vs. Score Distance (SD) of a 10 PC PCA of multiple 532nm Raman measurement of a sample, analysed once a day for 4 days (36 spectra for a day = 36 points for a day on the graph).

4.4.2 Heterogeneity in a Salmon Fillet

Something to consider when analysing a muscle with Raman is the heterogeneity of a salmon fillet. This heterogeneity can be the seasonal variation in total fat or FAs composition, or due to the natural heterogeneity of a salmon (Chang et al., 2022; Lintvedt et al., 2023). In this study, ready-to-consume pieces of salmon have been purchased in-store. Therefore, the analysed areas are not the same for each salmon. The dates on which the salmon were caught are also not standardized, as they were caught from early August to late October, in 2021 or 2022. Samples have been caught or raised at different seasons and different years across the set of samples. This has to be taken into account whether in the spectroscopic analysis or the GC analysis, as uncertainty remains about how much the composition of the fat can vary into

a muscle in function of the localisation in the fish or the seasonality (Lintvedt et al., 2023). For example, literature already showed moderate differences between EPA and DHA composition in different areas of salmon fillets (Nanton et al., 2007). The need for further investigation of compositional variations over a fillet is therefore important and seems to be lacking in the literature (Lintvedt et al., 2023).

To measure directly on a slice is interesting because the non-destructive aspect of Raman spectroscopy is an important benefit of the method and would be more convenient for routine applications. However, grounding could resolve the potential heterogeneity problem.

This experiment was conducted to evaluate the differences between the spectra of various sections within the same fillet. To accomplish this, four sections were sampled from a fillet: two parallel sections (one closer to the centre, referred to as zone 4 in Fig. 24, and one further away, referred to as zone 1), and two perpendicular sections to the dorsal axis (one central and one further away, referred to as zones 3 and 2, respectively).

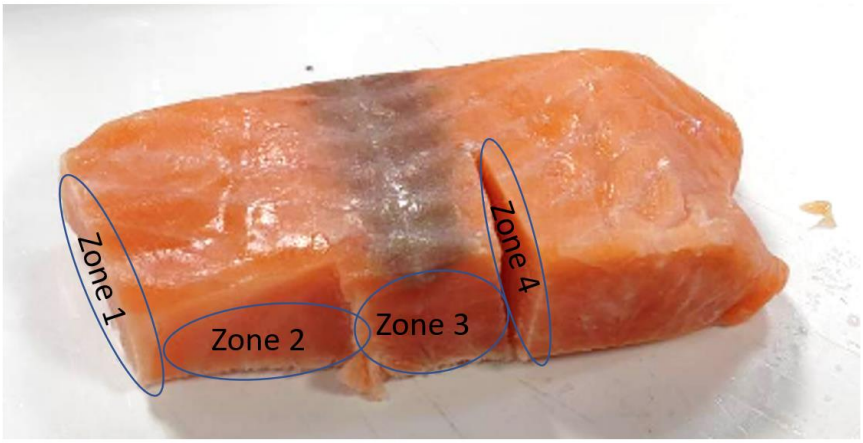


Fig. 24. Salmon piece used for the heterogeneity study. The referenced zones are the area where the four slices were cut.

To detect any significant differences between the areas, each slice was measured at 36 locations using a 532 nm laser. A principal component analysis (PCA) was then applied to the entire dataset. The pre-processing steps involved SNV followed by a 1st derivative SG. The aim was to notice any separation between the areas based on the scores.

In the Fig.25 graph of scores on PC3 and PC4, zones 2, 3, and 4 appear to overlap, while the spectra from zone 1 are relatively separated from the spectra of the other three zones based on the PC4 scores. PC4 shows a relatively high loading at wavenumbers 1650 cm^{-1}, 1525 cm^{-1}, and 1430 cm^{-1}. The wavenumbers 1650 cm^{-1} and 1430 cm^{-1} are often associated with lipids, while the wavenumber 1525 cm^{-1} is characteristic of carotenoids.

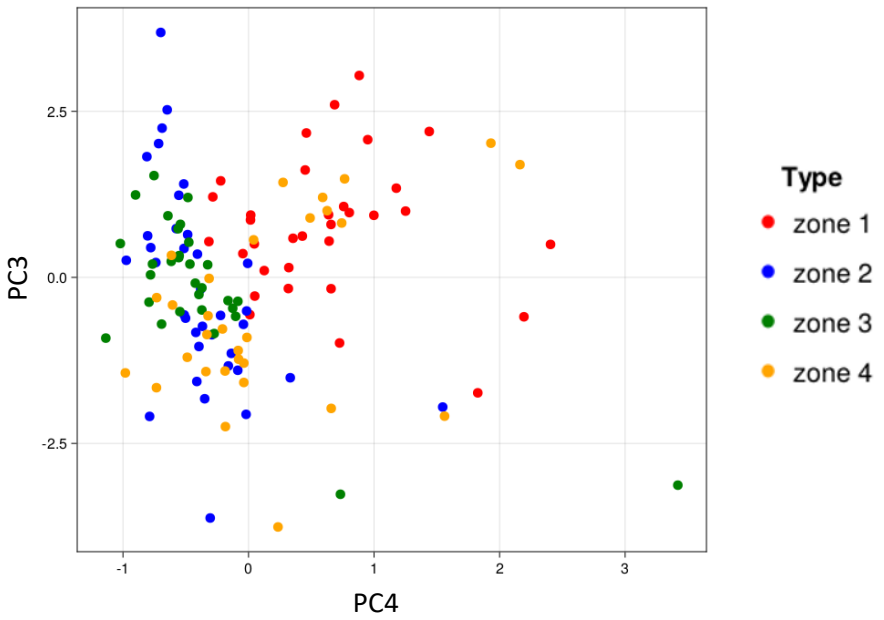


Fig. 25. Graph of PC3 and PC4 scores of a PCA on all spectra, coloured by zones of the sampling on the salmon piece of Fig.24.

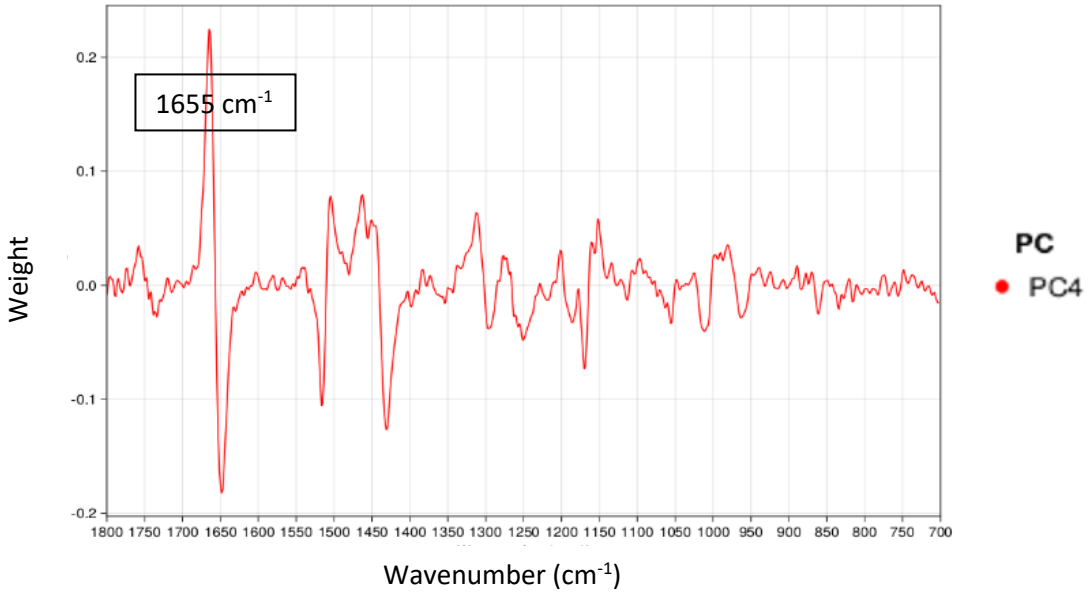


Fig. 26. Loadings of the PC4 from PCA from Fig.25 and its highest loading band annotated (around 1655 cm^{-1}).

A mean spectrum was calculated for each zone. By comparing the spectra of each zone, it becomes apparent that there is indeed a difference in the highest loading of PC4 (around 1650-1670 cm^{-1}). The wavenumber 1655 cm^{-1} is often referenced as a band excited by cis C=C stretch bonds, which is related to lipids (refer to Tab.3). A conclusion could be made that there is a lipidic difference between the different zones.

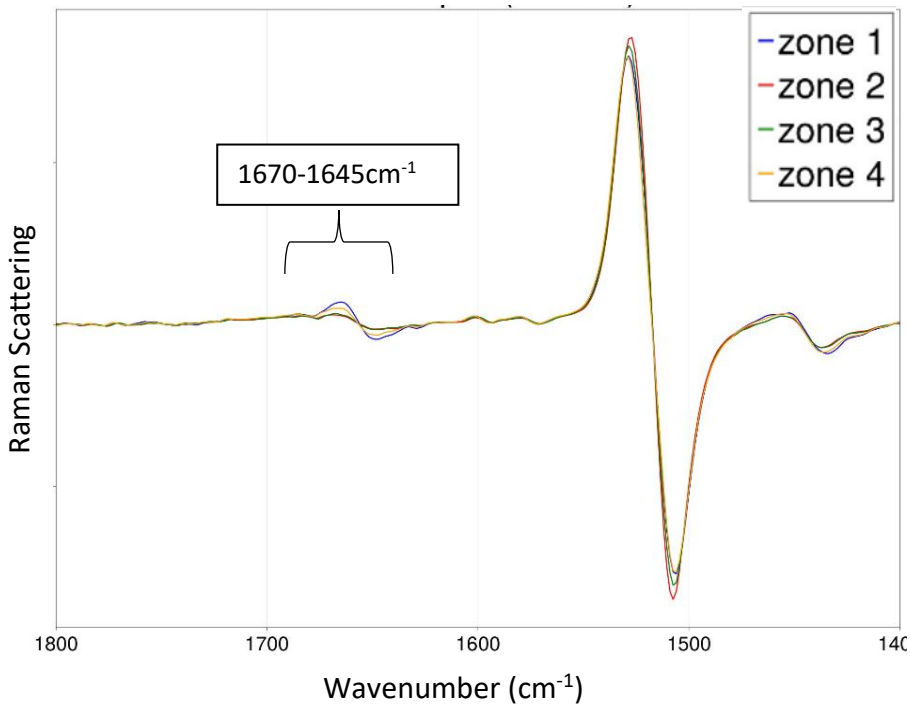


Fig. 27. Average Raman spectra of zone 1 (blue), zone 2 (red), zone 3 (green) and zone 4 (yellow) acquired with 532nm laser, from wavenumber 1800 cm^{-1} to 1400 cm^{-1}. The zone between 1670 cm^{-1} to 1645 cm^{-1} is annotated.

4.5 Fresh and thawed salmon samples Analysis

Four samples of fresh salmons were chosen randomly to assess the feasibility of discriminating between fresh and thawed salmon in an unsupervised way. The sample was then frozen overnight in a -20°C freezer. The day after, a new slice was cut and thawed for 10 minutes before being analysed again with the three lasers. Spectra were acquired with a 532 nm laser and pre-processed with SNV and SG 1st derivative (polynomial = 2, window = 7). A PCA was applied to detect possible variations due to thawing. Fig.28 shows the average spectra of the 4 samples fresh and thawed.

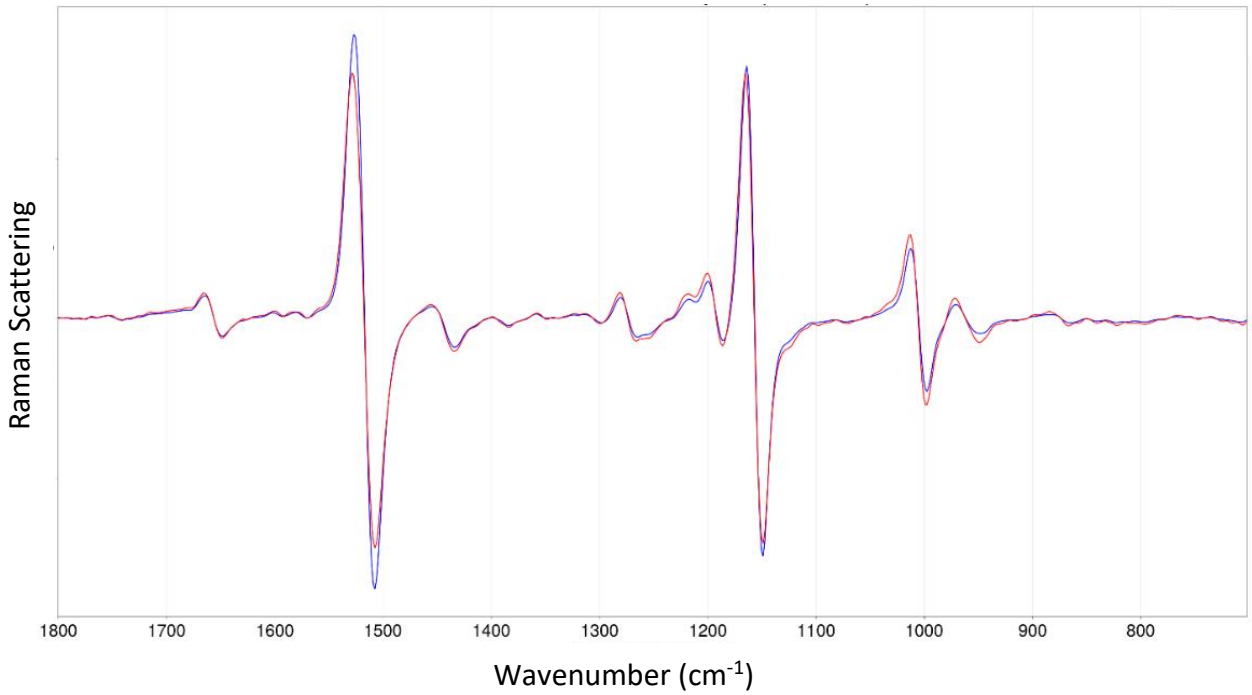


Fig. 28. Mean spectra of fresh (blue) and thawed (in red) salmon samples from 1800 cm^{-1} to 700 cm^{-1}. Both spectra were acquired with 532nm laser and pre-processed with SG 1st derivative and SNV.

When the aim is to separate two groups of spectra based on a specific factor using PCA scores, it is important to minimize the influence of other factors that may also differentiate the groups. For instance, it is possible that if two groups are taken on different days, they may be separated by time rather than solely by the freezing process itself. This is why this experiment was conducted over three days, where some of the fresh samples were analysed on the same day as some thawed samples. However, it is important to bear in mind that industrial freezing processes carried out to sell frozen products are not conducted under the same conditions as freezing performed in the laboratory.

4.5.1 Thawing influence within a single sample

A PCA was applied to the spectra on one of the 4 samples to see variations within a single sample. The acquisition parameters were: 532nm laser, 2000 ms integration time with 5 co-additions, x10 lens, 20 mW laser power and 4 cm^{-1} resolution. The spectral range used is between 1800 cm^{-1} and 800 cm^{-1} wavenumbers, and a 1st derivative SVG filtering and SNV were applied on the dataset.

Each point in the Fig.29 represents a single spectrum, 36 spectra were taken on the fresh, and 36 on the thawed sample. The fresh and thawed measures are separated by their scores along PC1 and PC3. A visualisation of the loadings of these principal components will help to understand the Raman scattering bands behind this separation.

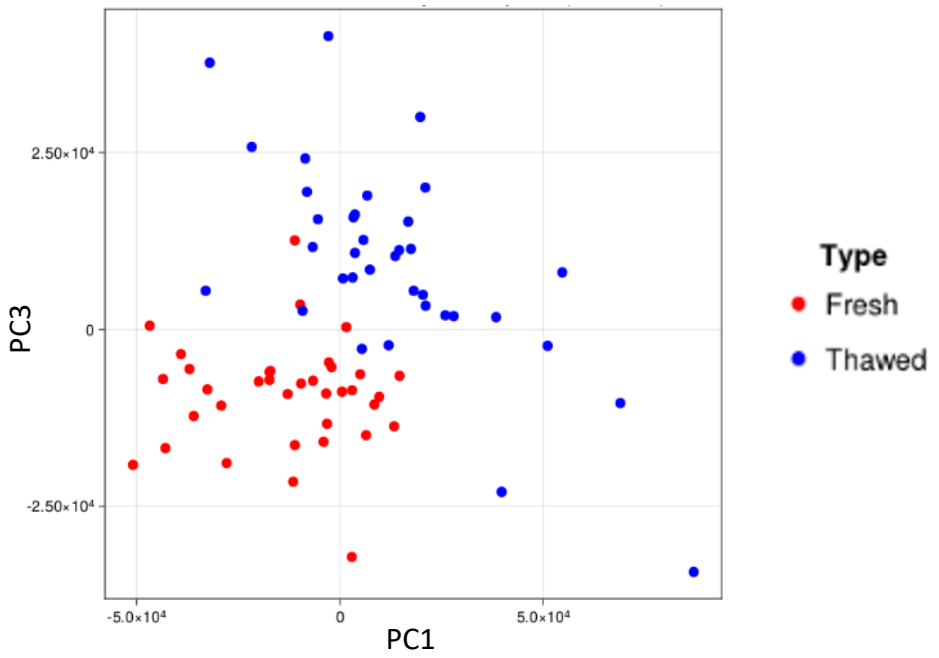


Fig. 29. Scores on PC1 and PC3 of 72 spectra of the same sample, analysed fresh (red) and thawed (blue).

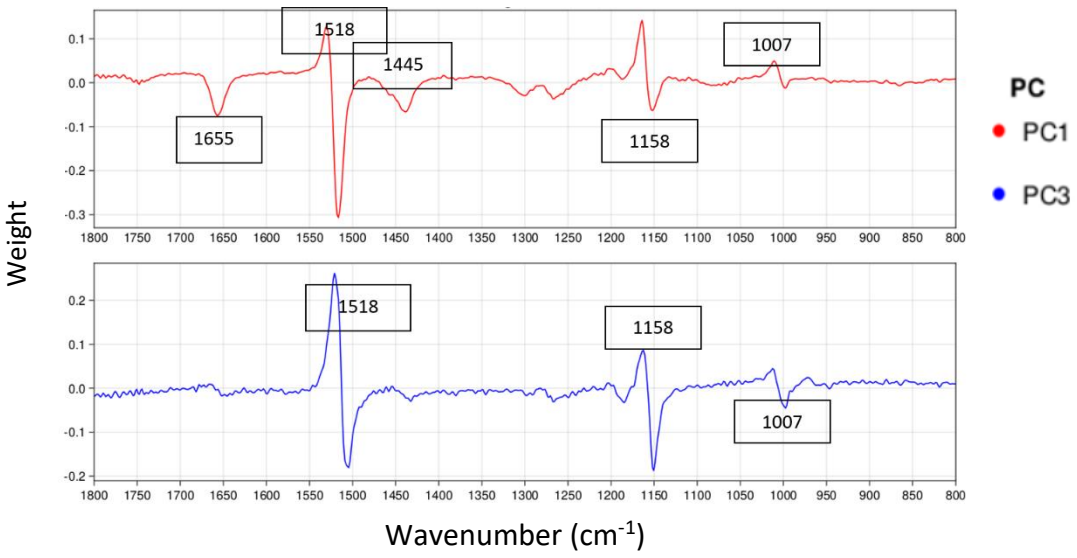


Fig. 30. Loadings of PC 1 and 3 of the PCA in Fig.29.

PC1 and PC3 look very similar. We notice that, for both, the more weighted bands are between 1500 cm^{-1} and 1550 cm^{-1} (where 1518 cm^{-1} is associated with C=C stretching vibration), and between 1150 cm^{-1} and 1170 cm^{-1} (where 1158 cm^{-1} is associated with C-C polyene chain stretch vibration). It is shown in Fig.19 that the 532nm spectrum has very

intense peaks in these areas. Looking at Tab.5, these bands would refer to carotenoids. Therefore, PC1 is a combination of these carotenoid bands, and bands around 1650 and 1430 cm^{-1}, that would rather give some spectral information about lipids. The spectral bands around 1007 linked to phenylalanine are also important in both loadings.

4.5.2 General thawing influence within 4 samples

To have a general approach, a PCA was also applied to the four samples at the same time to see if a general difference in the spectrum can be used to discriminate between fresh and thawed samples. As before, a measurement is made of 36 individual spectra, acquired with the 532nm laser. The acquisition parameters were: 2000ms integration time with 5 co-additions, x10 lens, 20 mW laser power and 4 cm^{-1} of resolution. Spectra were pre-processed in two different ways to eventually notice additional other interesting differences in the spectra. The 1st pre-processing was an SNV then a SG 1st derivative (pol = 2, derivate order = 1, window = 7 points). The 2nd was an SNV followed by a SG smoothing (pol = 1, derivate order = 0, window = 7). In the following score graphs, each measurement is represented by 4 points (average spectra). Each of these points is the average spectrum of a 3x3 mapping (9 spectra) on a square area, and each square area forms a 1 cm side square on the salmon slice. A sample is therefore represented by 8 points: 4 fresh and 4 thawed.

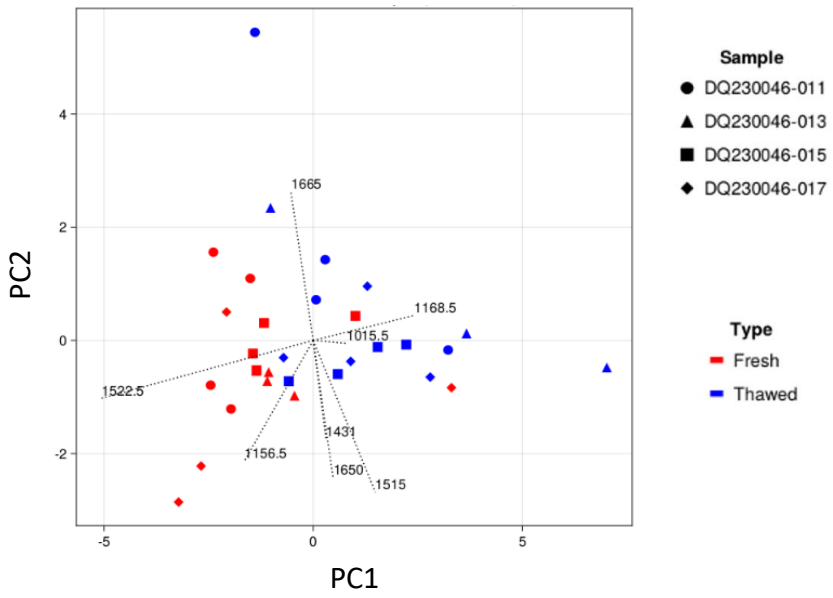


Fig. 31. Scores of 4 samples, fresh (red) and thawed (blue) on PC1 and PC2 of the PCA with the 1st pre-processing : SNV and SG 1st derivative (pol = 2, d = 1, window = 7).

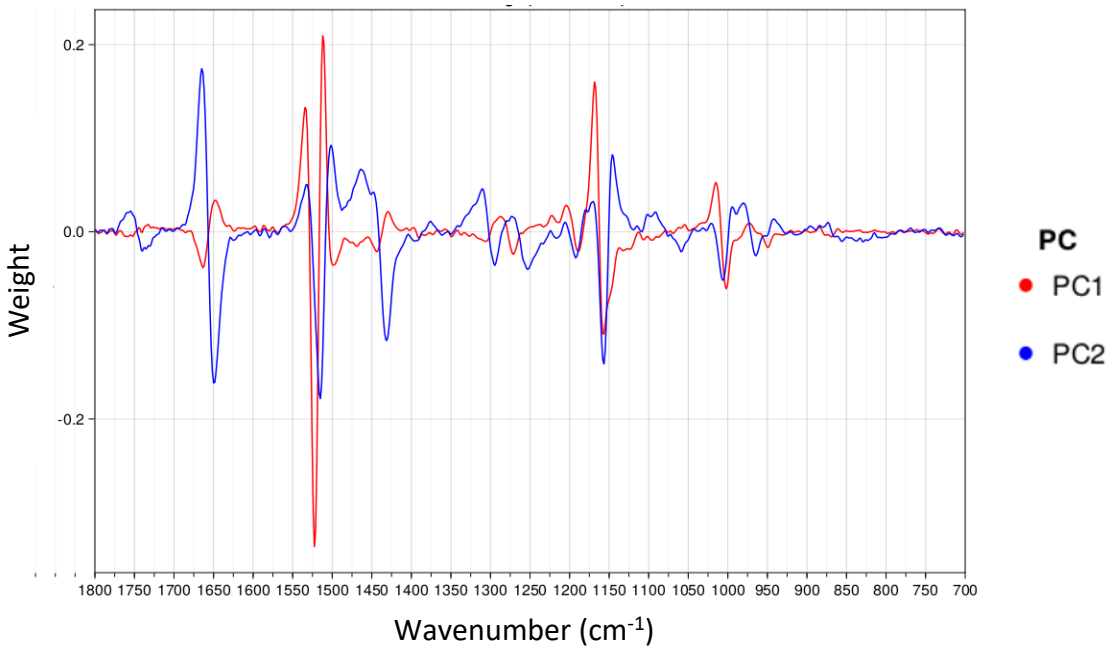


Fig. 32. Loading of PC1 and PC2 of the PCA with the 1st pre-processing : SNV and SG 1st derivative (pol = 2, d = 1, window = 7).

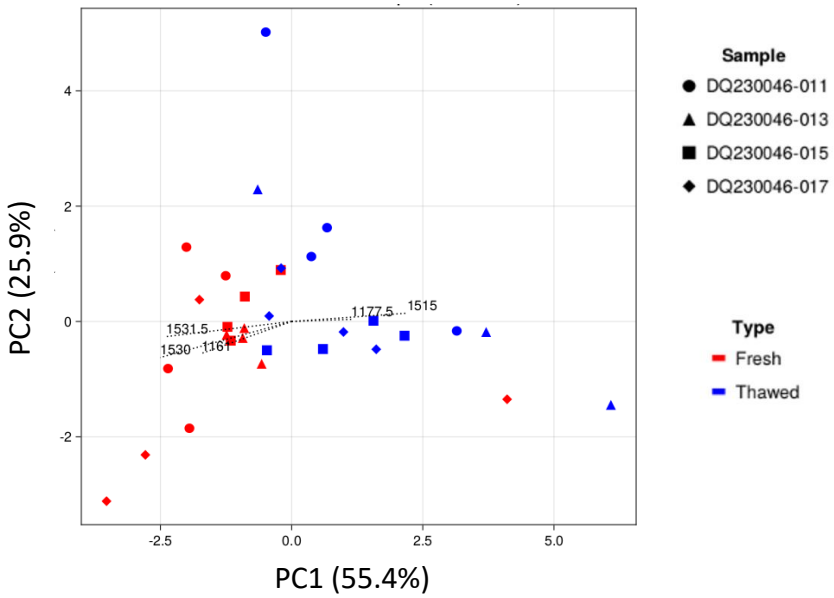


Fig. 33. Scores of 4 samples, fresh (red) and thawed (blue) on PC1 and PC2 of the PCA with the 2nd pre-processing: SNV and SG (pol = 2, d = 2, window = 21).

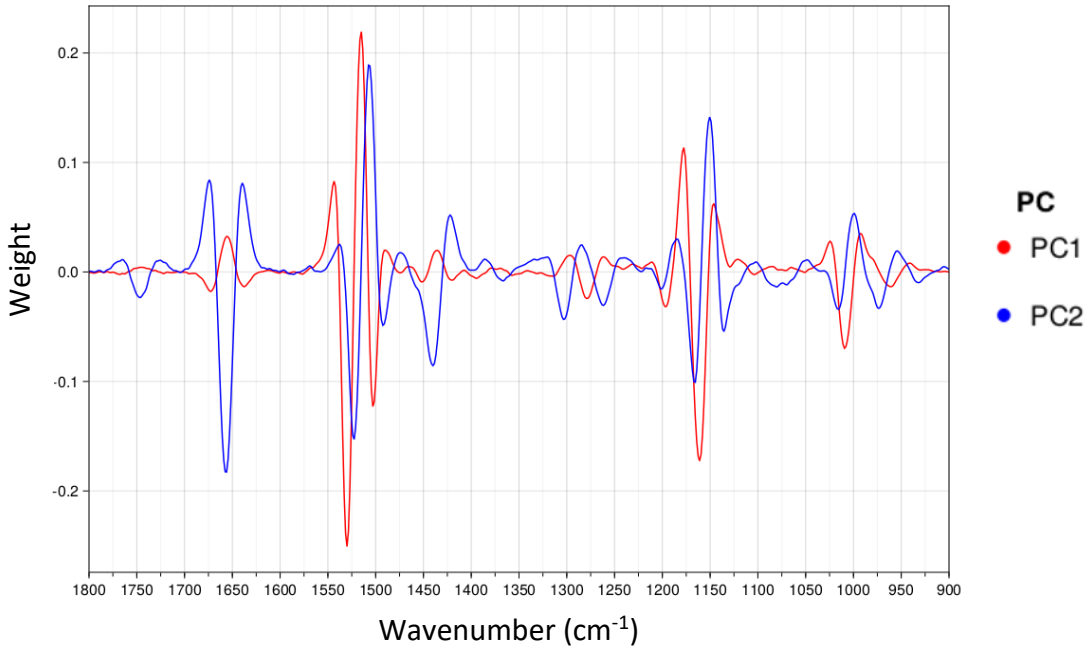


Fig. 34. Loadings of PC1 and PC2 of the PCA with the 2nd pre-processing: SNV and SG 2nd derivative (pol = 2, d = 2, window = 21).

It seems that scores along PC1 and PC2 for both PCA can slightly separate the spectra of the fresh and thawed samples. This is an interesting result as there are four different samples in the PCA. That would indicate a possible general difference in the spectrum of a fresh and a thawed salmon muscle, and so, the possibility to discriminate between fresh and thawed salmons. The most important bands that separate the groups are referenced in the loadings plots. These most weighted bands are then indicated in the PC space, on the biplot. Their direction from the origin indicates the influence these bands have on the position of a point in the biplot, providing clues for interpreting the PCA results.

The bands from 1150 to 1200 cm^{-1} seem to separate the groups (shown in the biplot with 1156, 1161, 1168 and 1171 cm^{-1} bands). These bands are typically associated with carotenoids. There also seems to be significant variation in the spectral bands from 1500 to 1550 cm^{-1} (shown in biplot with 1515, 1522, 1530 and 1531 cm^{-1} bands), which could also be related to carotenoids in the tissues. As confirmed by the two biplots representing two different pre-processing, the 1150-1200 cm^{-1} and 1500-1550 cm^{-1} areas of the 532 nm spectra seem to be interesting spots to look at for the discrimination of fresh and thawed salmon. The area around 1655 cm^{-1} has high importance on PC2 for both pre-processing. Therefore, it does not seem to effectively separate the fresh and thawed spectra as seen on the Fig.31 biplot (represented by 1665 and 1650 cm^{-1} bands).

4.5.3 Fresh and thawed samples PLS-DA

4.5.1.1 Fresh and thawed PLS-DA: train and test sets

The separation between the fresh and thawed salmon spectra on the PCA gives a good opportunity for a discriminant model building. To build the model, SNV and SG 1st derivative filtering were applied to the spectral data of 288 spectra, from four different samples.

Three of these samples were used to build the model and one independent sample was used to test its performance on an unknown sample, so 75% of the total spectra went in the train set and 25% in the test set. No matter the sample, all the spectra of the train set was split randomly into calibration and validation sets, sharing 80% and 20% of the train set respectively as shown in Tab.6.

Tab. 6. Distribution of the data to build the fresh/thawed PLS discrimination model.

Sample	Fresh	Thawed
Sample 1	36	36
Sample 2	36	36
Sample 3	36	36
Sample 4	36	36
Total	144	144

Train set (75%) → Calibration set (80%), Validation set (20%)

Test set (25%)

First, it is important to see how similar the test sample from the train samples is. Fig.35 shows the repartition of both sets of a PCA. The PC1 and PC2 have been chosen by default because they represent the most variations across the spectra. In our example, there is a zone on the top left that is not well represented by the train set. This is interesting because if a model predicts well in a space that is not well represented by the train set, it proves its robustness.

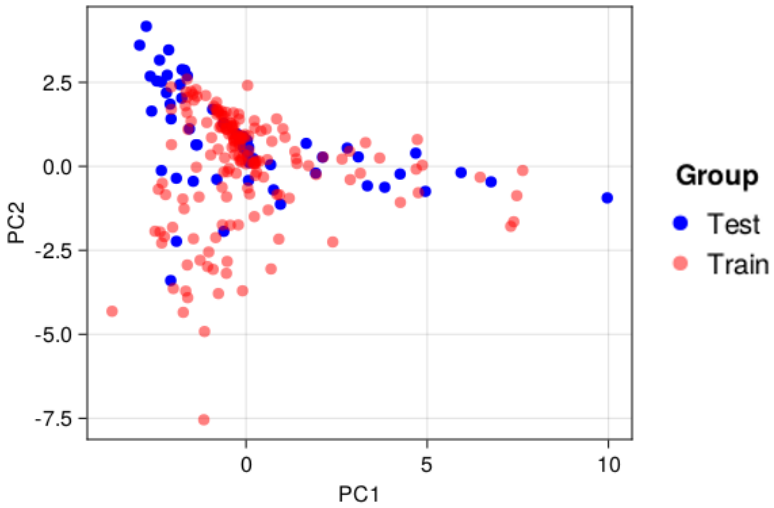


Fig. 35. Scores of train (red) and test (blue) sets on PC1 and PC2 of a PCA applied to the whole dataset. Each point is an individual spectrum.

4.5.1.2 Fresh and thawed PLS-DA : the model

The model used is a PLS-DA model. It is built using the train set. To optimize the model, the model is built with the calibration set. The validation set is then given to the model to be predicted. Fig.36 shows the Error $(1 - {Accuracy}/{100})$ of the model built with the calibration set and tuned with the validation set. In this case, the validation set indicates that the model with three latent variables gives the smallest error and, thus, the best performance of the model.

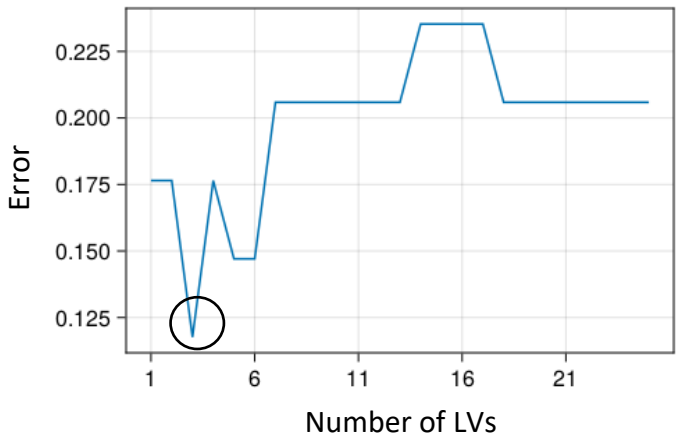


Fig. 36. Performance of the fresh/thawed discrimination model based on the Error (1-Accuracy/100) of models composed of 1 to 24 latent variables. The lowest error is given with 3 LV.

Fig.37 is the confusion matrix and tells us how the model predicted the test set. As a reminder, the test set is made of spectra of a single sample (Sample 4), that were not used to build the model (see Tab.6). The confusion matrix indicates that the Accuracy of the model of the test set is 88,2%. In total, 2 of the 29 spectra from the thawed sample were predicted as fresh, and 5 of the 32 spectra from the fresh sample were predicted as thawed. This model is an example of good accuracy with few latent variables (not too complex) to predict the test set.

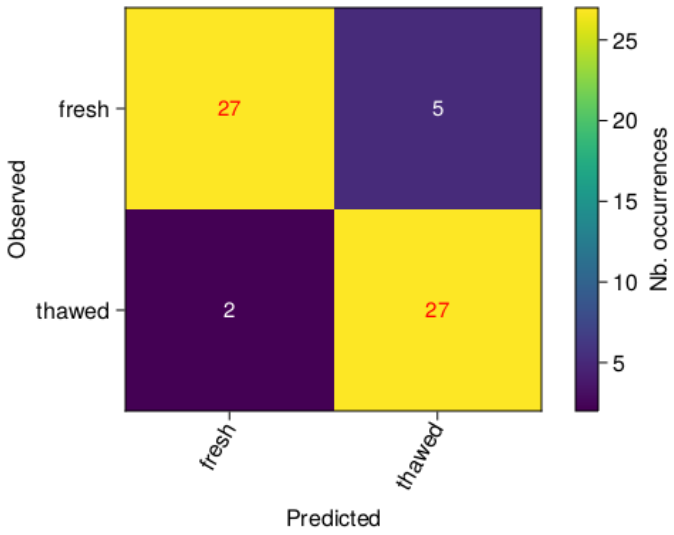


Fig. 37. Confusion matrix of the prediction of the test set by the fresh/thawed PLS-DA model. An occurrence is an individual sample.

To have a closer look at the latent variables that allow the two groups to be discriminated, it is possible to look at their loadings. The more weight a band has, the more important it is for discrimination. Fig.38 shows the loadings of the three latent variables in the model. The most important bands for LV1 and LV2 are around 1525cm^{-1}, 1150cm^{-1} and 1000cm^{-1}. These bands are generally associated with carotenoids. They are the same bands that were found in the PCA. LV3 gives better accuracy to the model and has important weights in areas around 1650 cm^{-1} and 1425 cm^{-1}, which would be more likely to be related to lipids.

As the model predicts accurately based on these bands, a conclusion can be made that a general difference exists in the spectra of fresh and thawed samples at the level of carotenoids related bands. A impact of thawing on lipidic related band also seems to exist as it allow a better prediction of the fresh or thawed state of a sample.

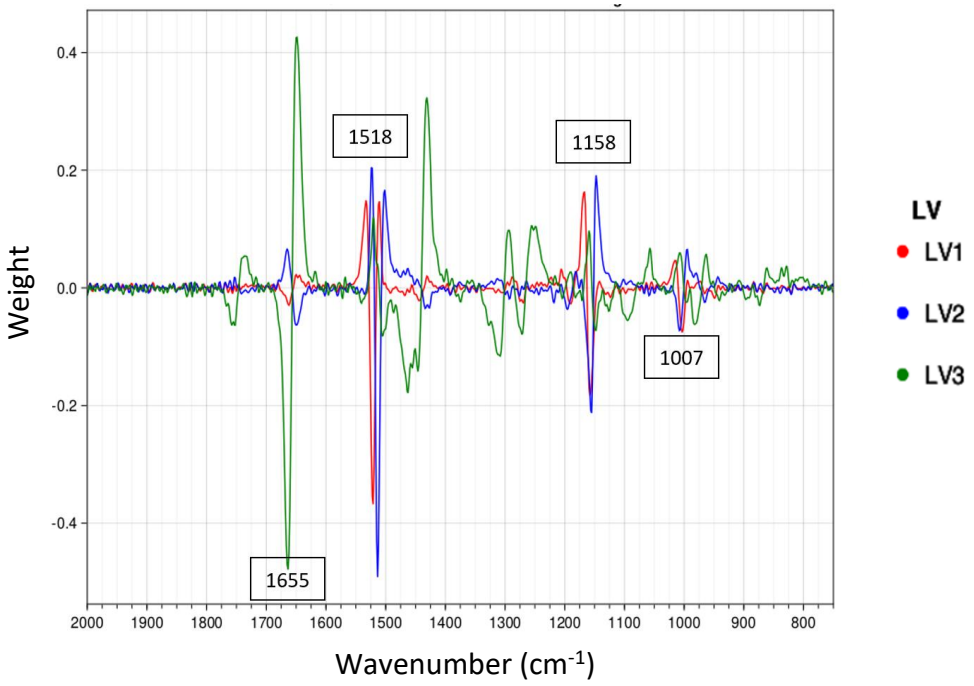


Fig. 38. Loadings of the LV1 (red), LV2 (blue) and LV3 (green) of the PLS-DA model to discriminate between fresh and thawed spectra, built from 3 samples.

4.6 Wild and Farmed salmon analysis

4.6.1 PCA of wild, conventionally and organic farmed salmon spectra

The part about wild and farmed salmon analysis with Raman spectroscopy mainly aims to look at its potential to discriminate between wild and farm-raised salmons. To add some potentially interesting observations to the study, the farm-raised group has been split into the organic salmon, and the conventionally farmed salmon. Their differences have been first studied in the GC part,

Before trying any predicting models, it is interesting to have an idea of the spectral data of wild, organic, and farmed salmon. PCA was applied on spectra measurement directly on sample slices, as well as on spectra measured on ground aliquots of the sample.

4.6.1.1 PCA on spectra from Ground samples

This experiment was conducted with a dataset of three spectra per sample with 24 ground salmon samples. The ground samples aliquots used here were taken from the ground samples acquired during the GC analysis preparation. Each point on the PCA is a single-measure spectrum, and three spectra by sample were measured. The spectral range used was 800-1800 cm^{-1} and spectra were acquired with a 785 nm incident laser, the laser giving the most interesting information on the data for wild and farm-raised salmons. As pre-processing, SNV and SG (window = 9, 1^{st} derivative) were applied to each spectrum.

Due to the difficulties in finding wild salmon samples, the study encountered a lack of data for this group. The PCA in Fig.39 shows that wild samples seem to form a separate cluster from the farm-raised samples along PC4.

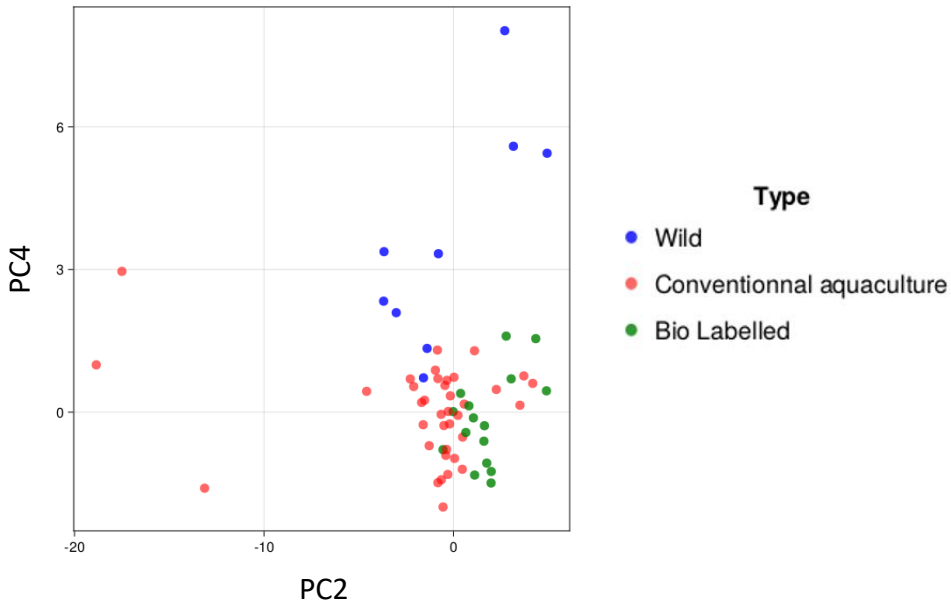


Fig. 39. Scores on PC2 and PC4 of conventionally farmed (red), organically farmed (green) and wild (blue) salmons from PCA on 785 nm Raman spectra from ground samples aliquots.

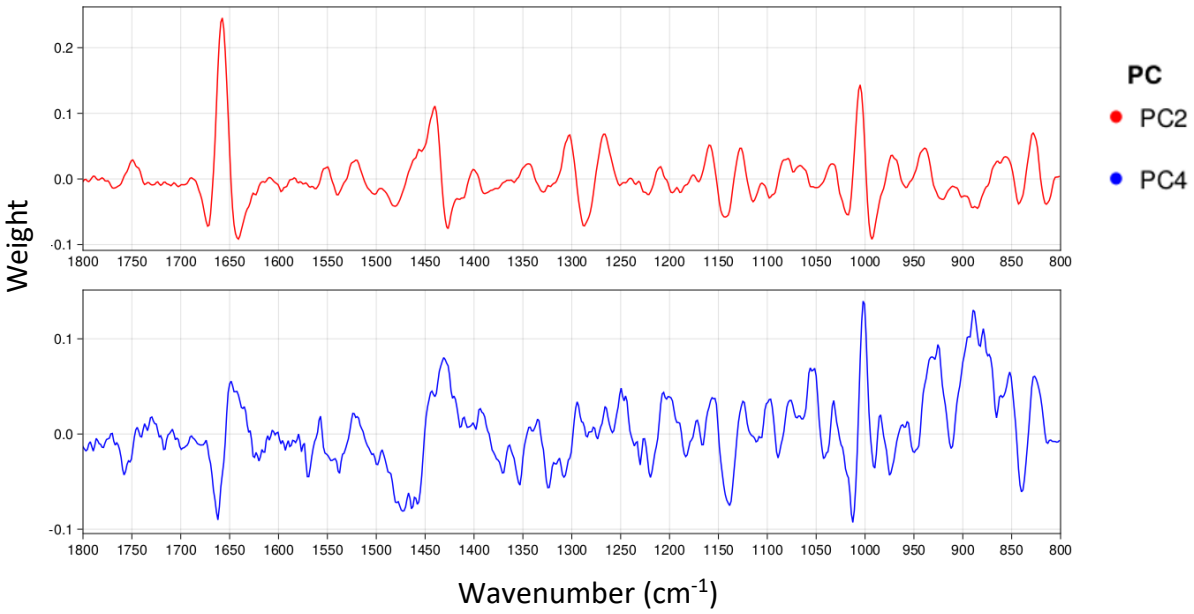


Fig. 40. Loadings of the PC2 (red) and PC4 (blue) from Fig.39.

The loadings of PC4 and PC2 in Fig.40 show the importance of some bands in the PCA scores. PC2 shows important weights for bands around 1650 cm^{-1}, 1440 cm^{-1}, between 1300

cm^{-1} and 1250 cm^{-1} and around 1000 cm^{-1}. These bands are often related to lipids, except for the range around 1000 cm^{-1}, which is well known to give information on carotenoids. Differences in lipids and carotenoids thus exist in the dataset and would be detectable by Raman spectroscopy.

4.6.1.2 PCA on spectra from sample slices.

For this experiment, a database of spectra from 27 salmon samples was used. The spectra were acquired with the 785 nm laser. The measurement parameters were: 10 000 ms integration time with 5 co-additions, x10 lens, 100 mW laser power and 4cm^{-1} resolution. The spectra were taken randomly over a period of more than one month, as the samples of the study were acquired at different moments. Each point on the PCA is average spectrum of 4 square areas of 2x2 spectra each (16 spectra for 1 average spectrum), taken at different random places on a slice. The dataset was then pre-processed using a SNV transformation and a SG 1st derivative (pol = 1, derivative order = 1, window = 7). By analysing spectra that were measured randomly over a long period of time, an unsupervised separation of groups based on scores is less likely to be influenced by factors such as time and more by intrinsic chemical and physical differences within the salmon samples. In addition, the spectra were obtained from slices, which allows the PCA to capture the intrinsic variability within an unprocessed salmon slice.

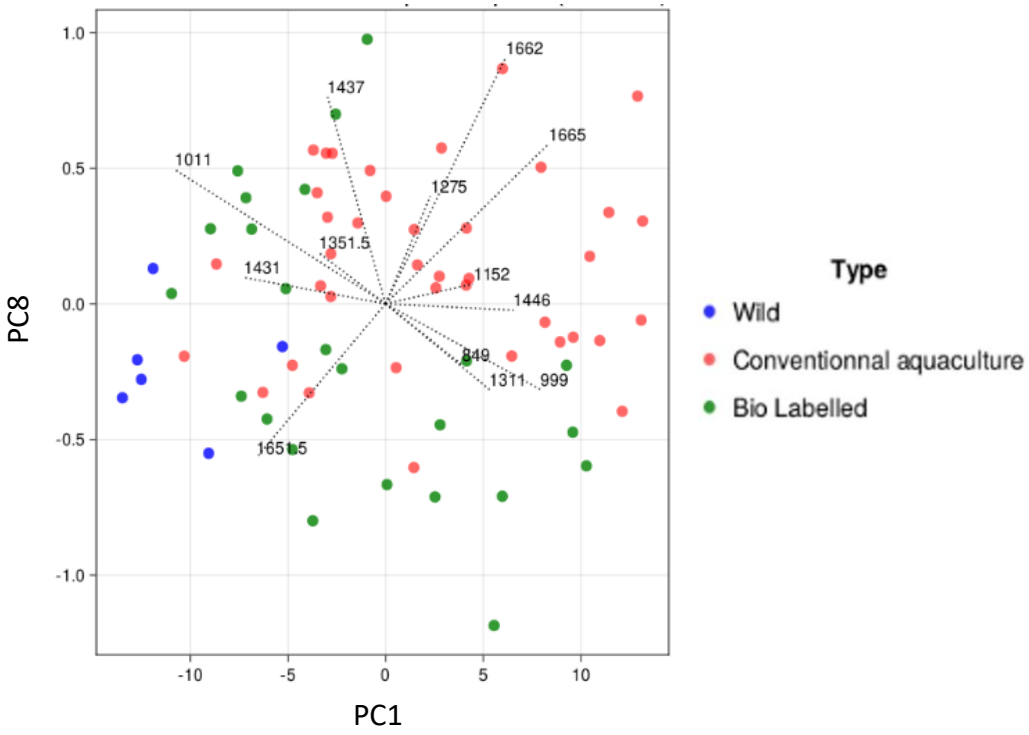


Fig. 41. PCA biplot of spectra acquired on salmon slices with 785 nm laser. Points are coloured by classes: wild (blue), conventionally farmed (red) and organically farmed (green) salmon.

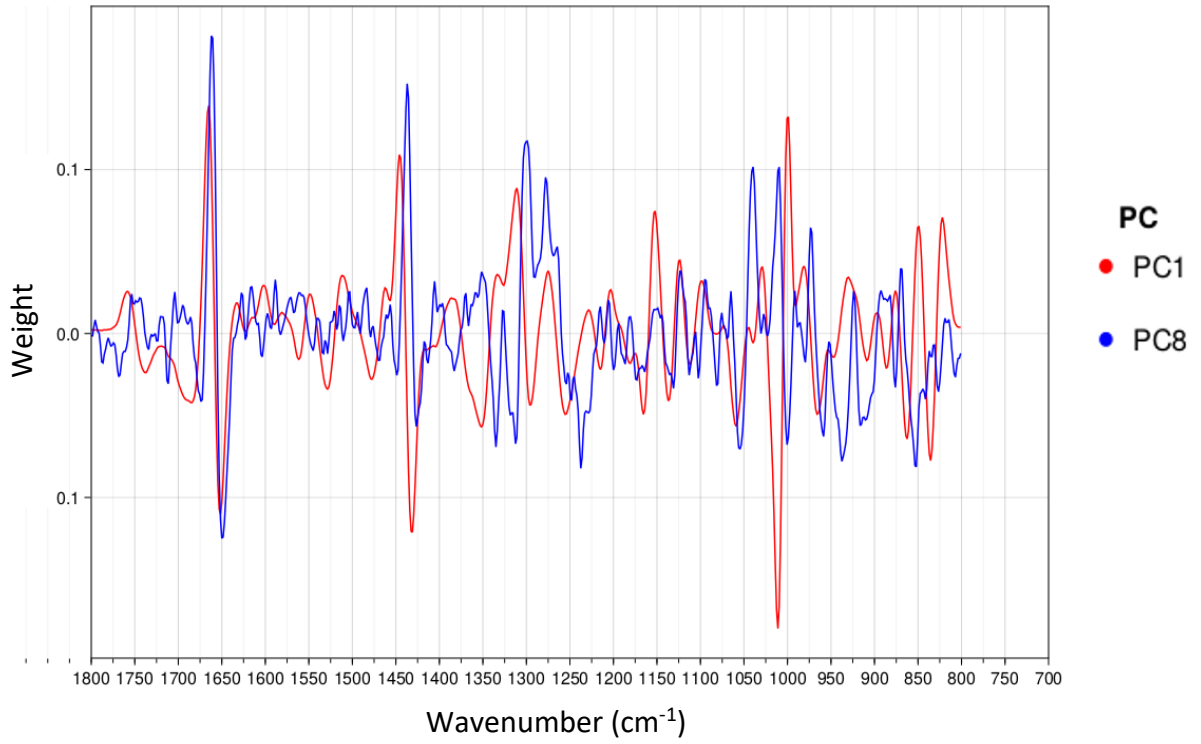


Fig. 42. Loadings of PC1 (red) and PC8 (blue) from Fig. 41.

The PCA does not show a good separation of the three groups along the PC1 and PC8 score axis. However, some tendencies seem to exist: wild samples are located on the left of the graph, and conventionally farmed samples are rather on the right. Organic samples are disseminated between the two groups, in the middle. While PC1 represents a significant proportion of the sample variability, PC8 is more likely to represent noise rather than true chemical information of the sample.

A total of 13 bands that have the greatest influence on PC1 and PC8 scores (bands with the highest loadings) are represented on the PCA biplot. The groups appear to be separated along an axis between PC1 and PC8. Bands at 1651 cm^{-1}, 1662 cm^{-1}, and 1665 cm^{-1} seem to have a considerable influence along this axis. The band 1655 cm^{-1} is often associated with lipidic unsaturation, showing the presence of lipids. The biplot indicates that these bands' weights are in the spatial direction that separates the wild and the farmed samples, so they play an important role. Also, bands around 1445 cm^{-1} (shown in the biplot by 1446 cm^{-1} and 1431 cm^{-1} bands) are related to lipidic saturation, separate between wild and farmed samples and have high loadings in PC1. The wild and conventional samples, therefore, seem to be separated by their lipidic quantity or nature. This result would be confirmed by the GC data.

Another band that goes in the PC1 direction is the 1152 cm^{-1} band in the biplot. This band is very near the 1158 cm^{-1} band that is related to carotenoids C-C polyene chain stretch (referring to Tab.5). Fig.42 shows that this zone has an important influence on the PC1 scores. The carotenoids could therefore represent a significant difference between wild and farmed salmons.

As for the loadings of the ground samples PCA, the bands around 1000 cm^{-1} (showed by 1011 cm^{-1} and 999 cm^{-1}) also seem to be an important area of variance in the spectral data, with a relatively important influence on PC1. These bands are associated with phenylalanine (reference band at 1007 cm^{-1}).

The band at 1275 cm^{-1}, which comes from PC8, also exerts an influence along this axis. According to Tab.5, the 1275 cm^{-1} band could be attributed to lipid variation. However, it is dangerous to interpret the PC8 of a PCA as it explains a very small part of the variation.

4.6.2 PLS-DA model for wild, conventionally, and organic farmed salmon spectra.

A discrimination model was built to determine if organic, wild, and conventionally farm-raised salmons could be classified by a PLS-DA model. A total of 27 samples were used. In this experiment, to maximize variability, each spectrum is not an average but a single measurement of a sample. The data used here is the large database of all spectra measured over a period of one month, there is no standard number of spectra for each sample. The spectra were acquired by measuring directly on a salmon muscle slice and then pre-processed using a SNV transformation and a SG 1^{st} derivative (window = 21). The test set consists of 5 samples, including 2 conventionally raised (87 spectra), 2 organic (76 spectra), and 1 wild (27 spectra). The training set includes the remaining 22 samples, totalling 3 wild (50 spectra), 8 organic (236 spectra), and 11 conventionally raised (532 spectra).

The best performance was achieved with a PLS-DA model using 21 latent variables, making it a highly complex model that may be less realistic. The model's performance on the test data resulted in 72.1% accuracy. As observed in the confusion matrix (Fig.43), the organic spectra are the best predicted with 77.6% (59 good classifications on 73 spectra) accuracy, followed by conventionally raised with 69% (60 on 87) and wild spectra with 66% (18 on 27). With the rather poor prediction accuracies and the high complexity of the model even if it uses all the spectrum from 800 to 1800 cm^{-1} to predict, this model did not give significant results. Even if some clear differences exist between the fatty acid profiles of the wild, organic and conventionally farmed samples as shown by the GC results, the model did not seem to be able to classify them easily. The classification between these classes, therefore, seems to be rather complex.

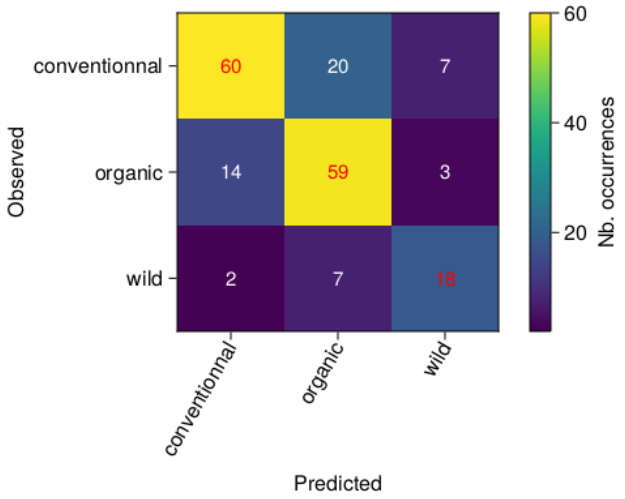


Fig. 43. Confusion matrix of classification of the test set by the PLS-DA model for wild, conventionally farmed, and organically farmed salmon. Each occurrence is an individual spectrum.

4.6.3 A potential way to discriminate between wild and farmed salmon: Prediction of FAs concentration using Raman spectroscopy.

Gas chromatography (GC) is currently the method of choice for analysing the fatty acid composition of foods. However, this method is destructive and involves laborious wet chemistry for sample preparation. The results obtained are accurate, but the procedure is often too costly and time-consuming for industrial quality evaluation. In the case of the study, the potential of Raman spectroscopy for the characterisation of the FA profile in salmon was evaluated.

4.6.4 PLS model for FAs prediction

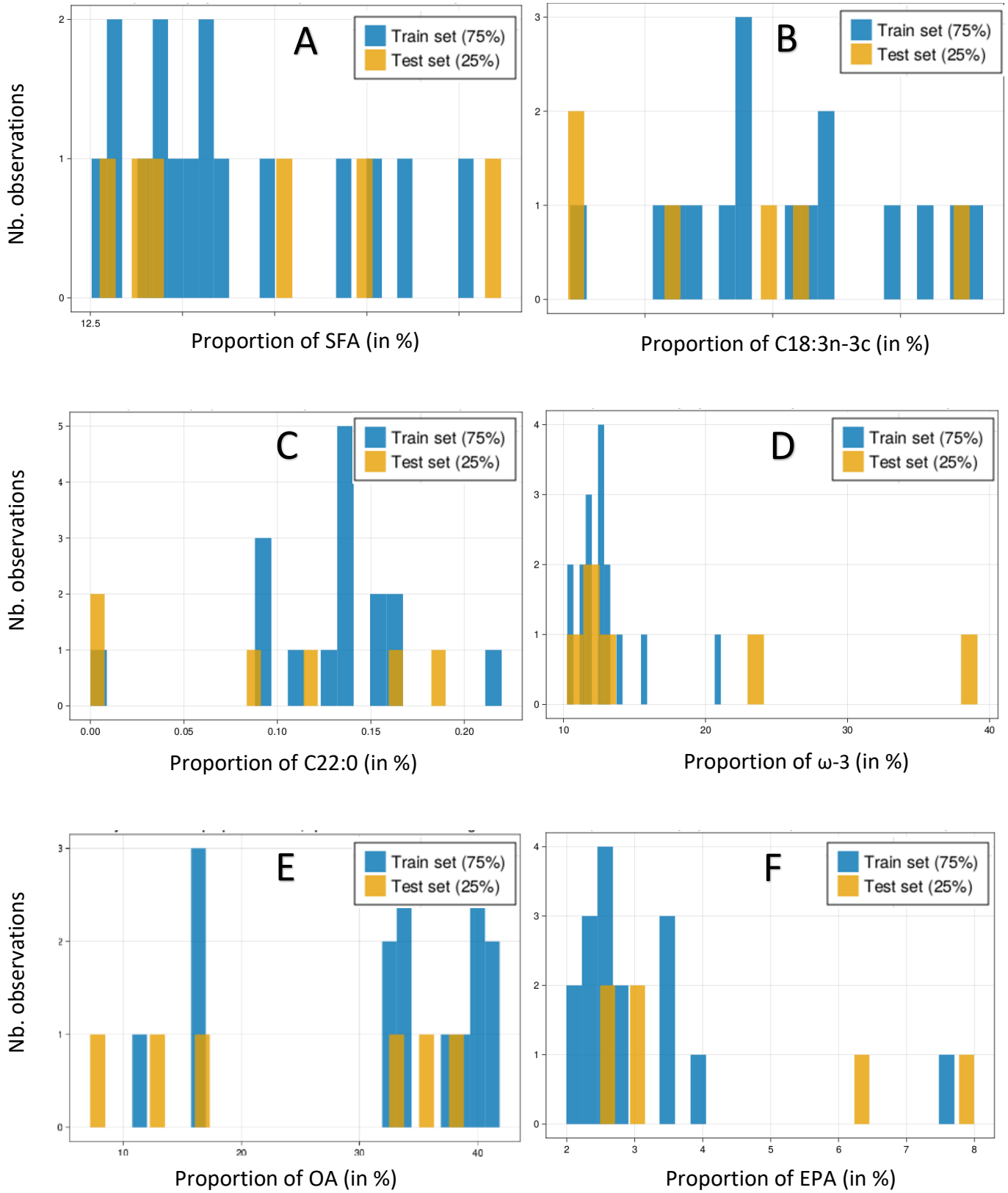
In this section of the study, a model was built to predict the proportions of the four key fatty acids identified in section 4.1.5 using Raman spectra, namely oleic acid (C18:1n-9c), α-linolenic acid (C18:3n-3c), C22:0, and eicosapentaenoic acid (C20:5n-3c). Additionally, the prediction of proportions for MUFA, SFA and ω-3 were also evaluated.

The spectra obtained from the 785nm and 532nm incident lasers of the Senterra instrument, as well as the spectra from the portable Rigaku instrument at 1064nm, were used. For the four fatty acids, the most interesting predictive results in terms of accuracy were obtained using the 785nm laser. For this experiment, ground salmon samples were used. They were obtained by the same process as for the GC part, by firstly frozen in liquid nitrogen, then ground. Each sample is represented by one spectrum: an average spectrum of three spectra taken directly on the ground muscle. The acquisition parameters were: 10^3 ms integration time with 5 co-additions, x10 lens, 100 mW laser power and 4cm^{-1} resolution. The spectra were subjected to a SG derivative (f=31, pol=3, d=2) and a SNV correction.

The regression models employed were all partial least square regression models (PLS-R) model. To build them, a set of 24 salmon samples with known proportions of fatty acids were used (from the GC analysis). For all subsequent models, wavenumbers ranging from 800 to 1800 cm^{-1} were employed. Indeed, the region between 1800 and 3000 cm−1 contains almost no spectral information and can therefore be safely discarded, and the region below 800 cm^{-1} didn't offer any improvement in the prediction performances for the models. The set of samples was split into a train set and a test set, with 18 of the samples assigned to the training set for calibration (14) and validation (4), while the 6 remaining samples were designated as the test set.

The selection criteria used for the train-test split was the Kennard-Stone algorithm. This algorithm ensures that the test set contains a diverse range of fatty acid proportion values, as it selects the most distant objects within the dataset. It allows the model to cover a wide range of data variability and evaluate its robustness. The Kennard-Stone algorithm proved especially useful since the distribution of values for some fatty acids was not uniform. Constructing a model based on non-uniformly distributed values carries the risk of inadequate representativeness and generalization errors, potentially compromising the prediction quality. Notably, the fatty acid values obtained from GC exhibited some lack of uniformity, for example in the case of omega-3 fatty acids.

The following histograms display the distribution of proportion values in the train and test sets. Generally, the K-S algorithm ensures that the distribution of the test data is consistent with the distribution of the train set while maximizing the variability of the test dataset.


A
Train set (75%)
Test set (25%)
Nb. observations
Proportion of SFA (in %)
B
Train set (75%)
Test set (25%)
Proportion of C18:3n-3c (in %)
C
Train set (75%)
Test set (25%)
Nb. observations
Proportion of C22:0 (in %)
D
Train set (75%)
Test set (25%)
Proportion of ω-3 (in %)
E
Train set (75%)
Test set (25%)
Nb. observations
Proportion of OA (in %)
F
Train set (75%)
Test set (25%)
Proportion of EPA (in %)

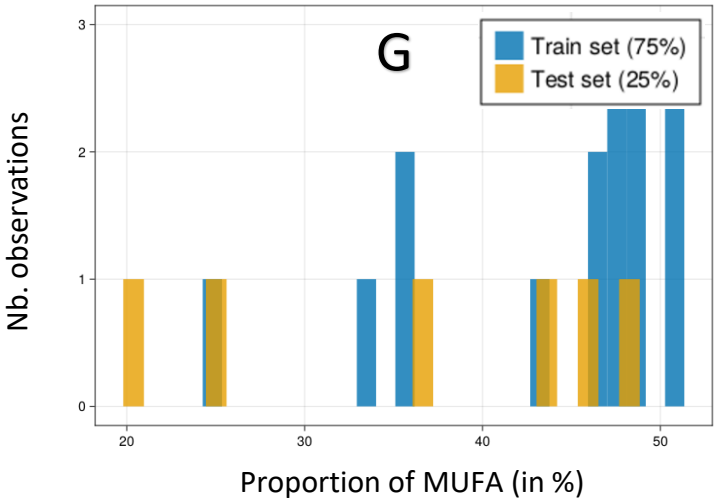


Fig. 44. Histograms of the density of the FAs data for train (blue) and test (yellow) sets after the Kennard-Stone algorithm splitting for SFA (A), ALA (B), C22:0 (C), ω-3 FAs (D), OA (E), EPA (F) and MUFA (G).

Due to the limited amount of data in the train set, the K-fold method was employed during the model-tuning process to select the optimal number of latent variables. The K-fold method involves dividing the train set into K equal parts (in this case, K = 3) and repeating the process n times (here, n = 10). Subsequently, two out of the three folds are used for model calibration, while the remaining fold is used for validation. The average performance for each fold is then computed, and the optimal model is the one that yields the lowest RMSEP for the number of latent variables. The graph in Fig.44 gives the RMSEP on the validation dataset for different models (changing in the number of latent variables). Each curve is the RMSEP for a single validation set. As 10 different validation sets were created by the K-Fold method, there are 10 RMSEP curves on the graph. The RMSEP of the 10 validation sets are averaged, and the best number of latent variables is the one that has the lowest averaged RMSEP value. In Fig.45, it is the 3 LVs model.

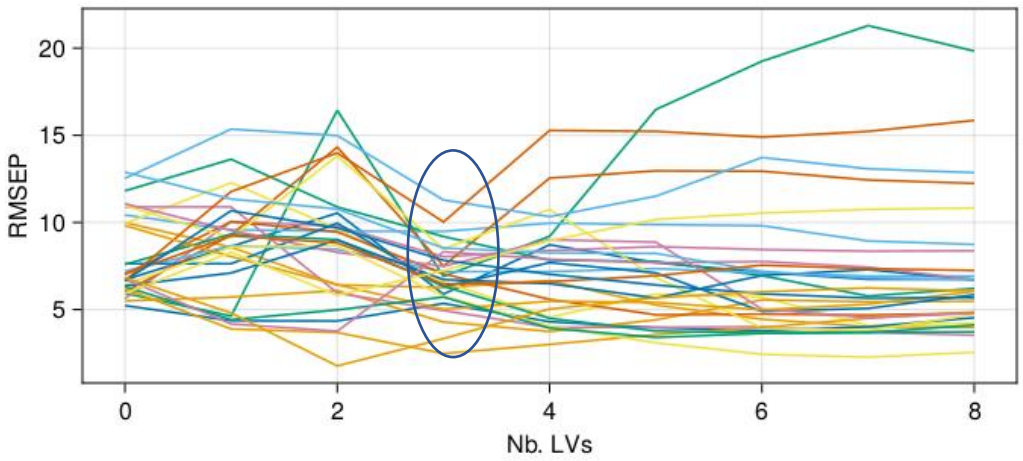


Fig. 45. RMSEP on 10 validation datasets vs. the number of latent variables (LVs). The model with 3 LVs has the lowest mean RMSEP.

With the K-fold method, all the data in the train set can be part of both the calibration set and the validation set. The importance of choosing the number of latent variables (LV) is illustrated here with the regression model for the proportion of saturated fatty acids (SFA). In Tab.7 below, two models, one with 1 latent variable and the other with 2 latent variables, are compared. Having 2 latent variables offers a dual advantage as it improves the R2 value and reduces the RMSEP. Thus, the model would perform better. This means that the second latent variable slightly enhances the prediction performance. Fig.45 illustrates the two latent variables loadings. The LV2 give interesting results with higher loadings of some zones, as LV1 is rather at a constant level of loading all over the spectrum.

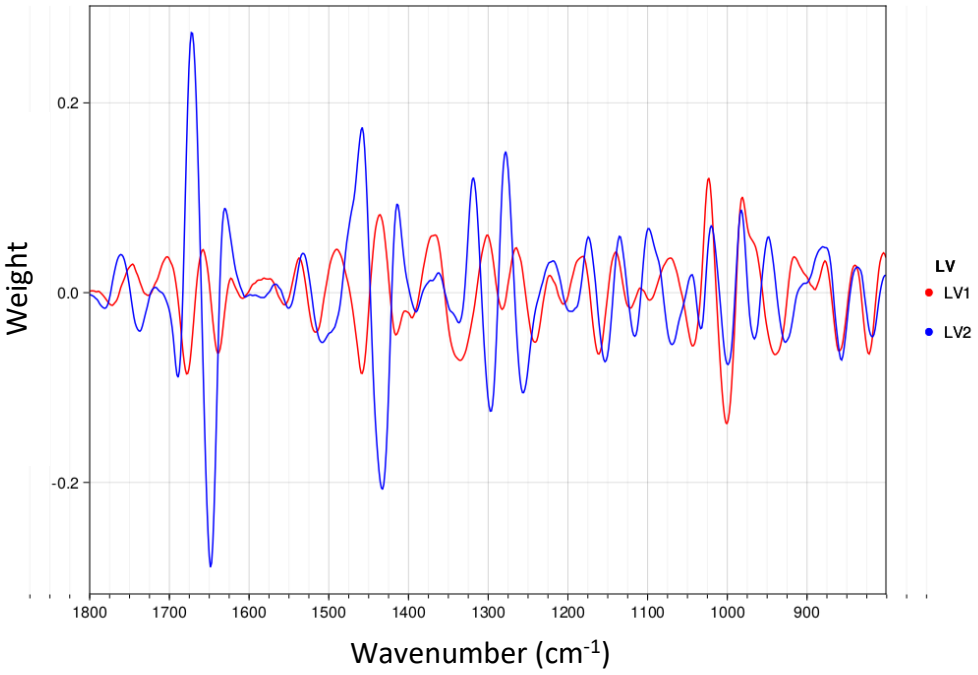


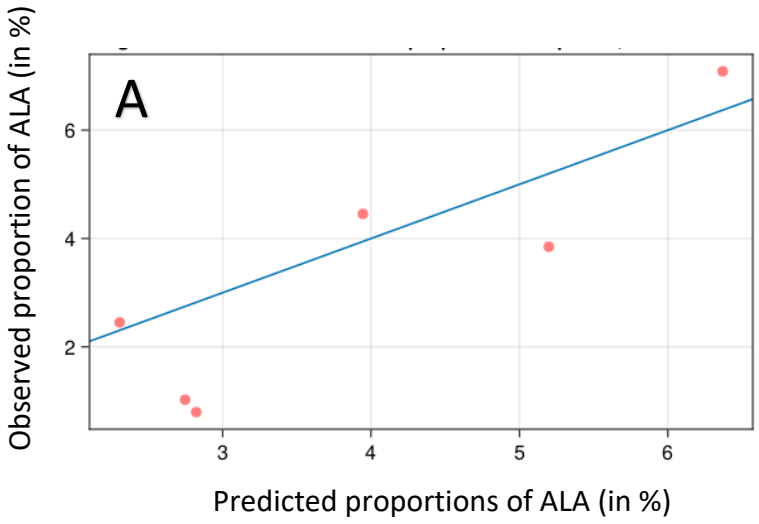
Fig. 46. Loadings of the LV1 and LV2 of the SFA prediction model with 2LV.

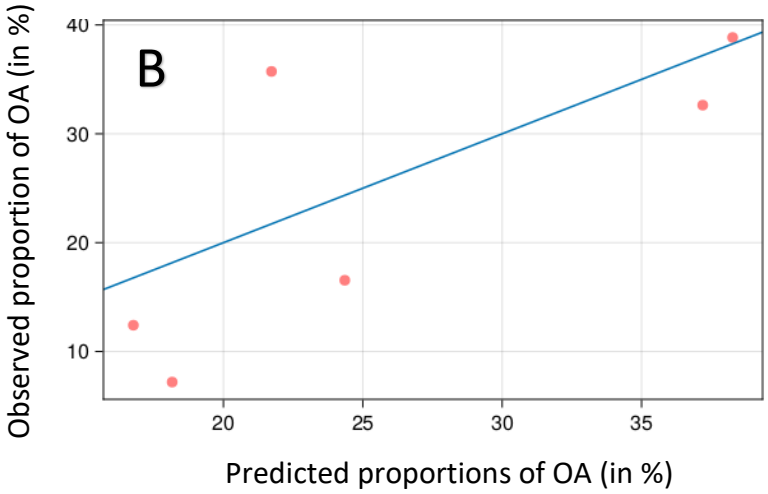
Tab.7 provides the performance indices of the optimized models for each regression of fatty acid proportions from the 785nm Raman spectra of the test set.

Tab. 7. PLS-R models: Their performance indicators (RMSEP and R^2) and the mean of the values to predict on the test set.

PLSR model performances			
Target FAs	**RMSEP**	**R^2**	**Mean value**
C18:1n-9c (4LV) PLSR	8.34	0.54	23.89

C18:3n-3c (2LV) PLSR	1.27	0.65	3.27
C20:5n-3c (1LV) PLSR	0.97	0.78	4.26
SFA (1LV) PLSR	1.43	0.85	17.88
SFA (2LV) PLSR	1.16	0.90	17.88
MUFA (3LV) PLSR	5.44	0.75	36.57





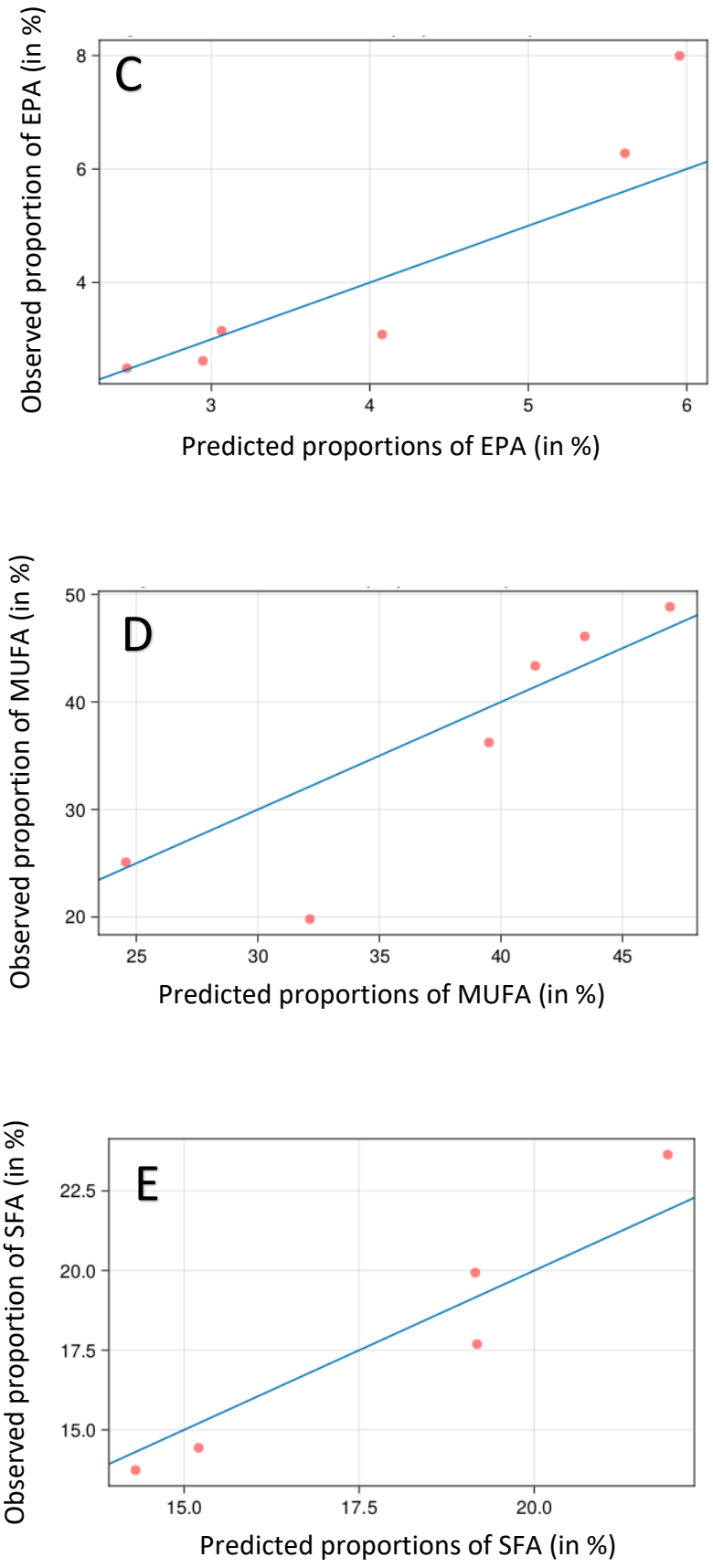


Fig. 47. Illustrations of the PLS-R model observed values (Y axe) vs. predicted values (X axe) of FA concentration for ALA (A), OA (B), EPA (C), MUFA (D) and SFA (E). The blue line is the "ideal regression line", passing by the origin with a slope of 1.

Amongst the four fatty acids to be predicted, three of them yielded interesting results: C18:1n-9c, C18:3n-3c, and C20:5n-3c. However, C22:0 did not show satisfactory results. To further explore the analysis, a regression was performed to predict the proportions of mono-unsaturated fatty acids (MUFA) and saturated fatty acids (SFA) in the samples. The prediction for SFA demonstrated a good coefficient of determination (R2 = 0.9) and a small RSMSEP (1,16% on a mean value of 17,88%). The proportions of SFA and MUFA, which exhibit significant differences between wild and farmed salmon, as shown in Tab.3, gave satisfactory prediction results for our samples (with respective R2 values of 0.9 and 0.75). The loadings of the latent variables (LV) for these two models can be compared to identify the Raman bands used to predict the proportions of SFA and MUFA.

Let's have a look at the loadings for SFA (2LV) and MUFA (3LV) PLSR models in Fig.47 and Fig.48. The principal areas used for the predictions of SFA and MUFA (if we base on the LV1 and LV2 for both models) are the 1650-1675cm^{-1} area (LV2), the 1425-1475 cm^{-1} area (LV1 and LV2), the area between 1325 and 1250 cm^{-1} (LV2) and the area between 1025 and 975 cm^{-1} (LV1). From the literature, the first 3 zones could be related to lipids, giving an insight that the prediction models are robust and not based on spurious correlations. However, the 1025-975 cm^{-1} zone is not known to be directly related to lipids, but to phenylalanine.

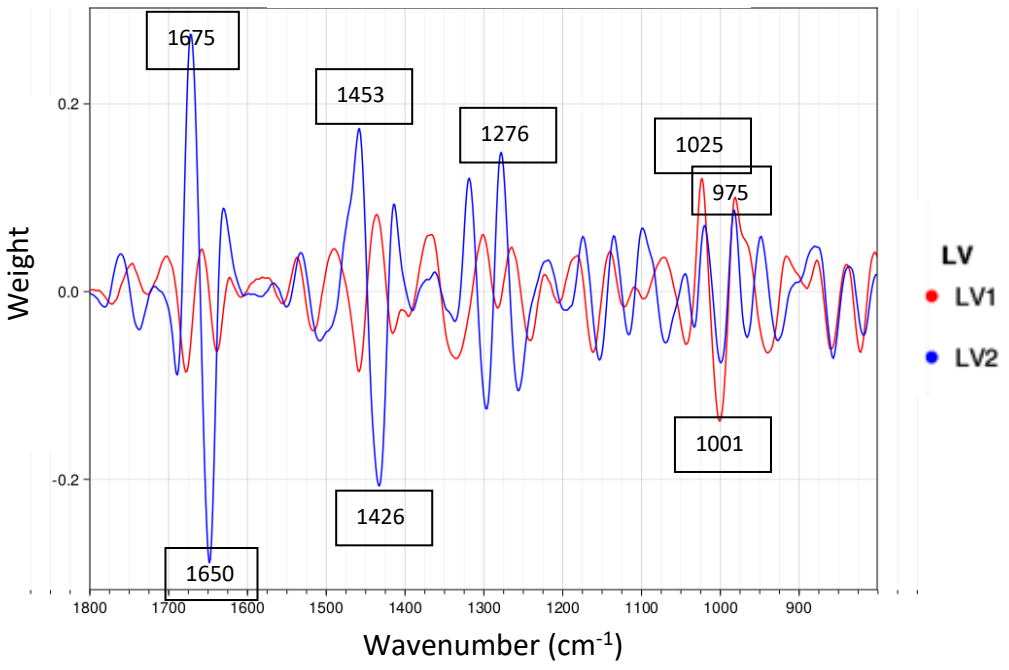


Fig. 48. Loading plot of LV1 (red) and LV2 (blue) of SFA PLS-regression.

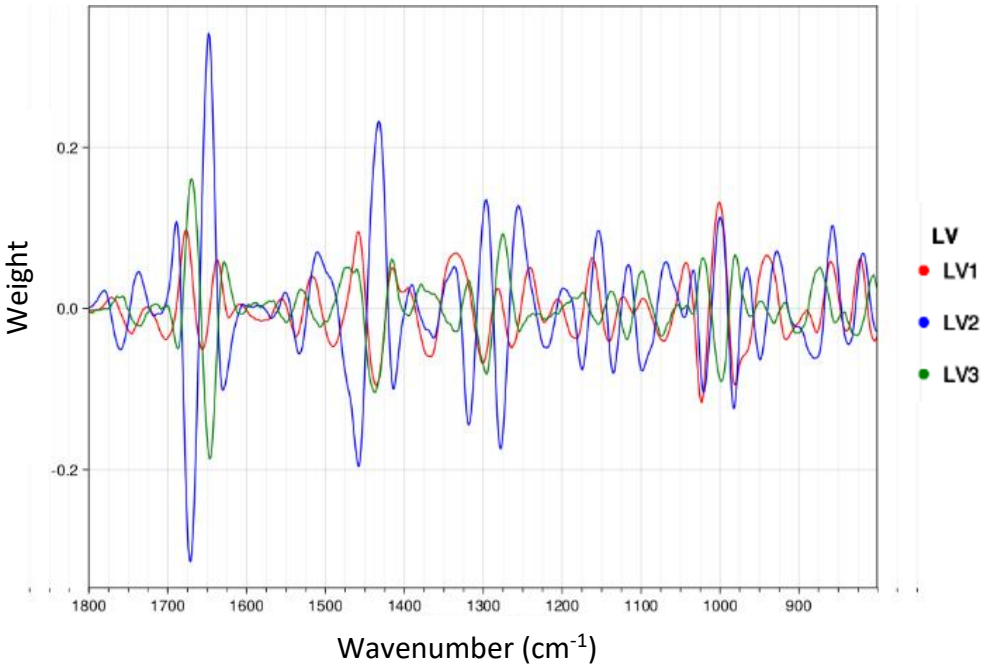


Fig. 49. Loading plot of LV1 (red), LV2 (blue) and LV3 (green) of MUFA PLS-Regression.

A few potential sources of error are important to mention, however. Firstly, there is uncertainty in the reference values obtained from GC (gas chromatography) as referenced in Tab.3. These reference values were used to build the models so they would be a direct error source.

In the present experiment, the aim was to evaluate a FAs prediction model using Raman spectra obtained from the analysis of ground samples of salmon muscle. In this case, 24 samples are a bit too low to take solid conclusions. Indeed, the more samples available to build a model, the more robust it would be. Nonetheless, some results show promising avenues that could be explored with a larger number of samples to yield interesting outcomes.

For instance, the detection of the quantity of monounsaturated fatty acids (MUFA) and saturated fatty acids (SFA) with a simple spectral measurement directly on the fillet would be of great interest for routine quality control. Furthermore, as found in section 4.5.1 of the study, differences in fatty acid profiles could potentially serve as a means of discriminating between wild and farm-raised salmon. This adds a highly significant application to the rapid and non-destructive characterization of salmon using Raman spectroscopy.

5 General Discussion

5.1 Feasibility of Raman measurement of salmon fillets.

The literature already explores salmon analysis, focusing on predicting the freshness of salmon fillets and determining their origin, whether wild-caught or farm-raised. The potential

applications are likely to expand, and their precision will improve with the availability of new data.

In this study, exploratory experiments were conducted using relatively simple statistical tools. The unique feature of this study is the examination of slices from unstandardized market-purchased salmon pieces.

The initial step aimed to determine the feasibility of meaningfully comparing salmon spectra. Taking measurements from random positions on the salmon muscle carries a risk due to fillet heterogeneity. To assess this heterogeneity, the spectra were analysed using PCA to identify noticeable differences. General disparities emerged between different measurement locations on a single salmon piece, demonstrating the existence of heterogeneity, which constitutes a primary challenge in spectroscopic analysis. To address this, samples can be homogenized through grinding. Alternatively, averages can be computed for a salmon slice to produce a more generalized spectrum. However, this method is debatable since it smoothens certain spectral information.

A logical approach for studying reproducibility and repeatability in Raman spectroscopy involves assessing the performance of prediction models derived from the data. Variability in repeatability or reproducibility to some extent can be ignored if it does not lead to a decline in performance. PCA serves as a robust tool for detecting key variations in spectral data. To effectively gauge repeatability or reproducibility, the tool used should be tailored based on a threshold, determined by the effectiveness that is needed. Therefore, other tools, more specific and easier to interpret than PCA, should be considered.

5.2 Wild, organically, and conventionally farmed samples

This part of the study showed limited results due to a difficulty to find fresh wild salmon samples. This is explained by the fact that wild salmon is rather found fresh on the market during the summer, and the samples in this study were acquired during October. Also, it is difficult to find fresh wild salmon in Belgium, even in fisheries, as it seems that they prefer to sell cheaper farmed salmons.

So, on the first hand, due to the difficulty to find other species, all the wild samples in the study are therefore from the *Oncorhynchus Gorbuscha* species. Also, all wild salmons found on the market that were not *Oncorhynchus Gorbuscha* were smoked salmons. On the second hand, the farmed salmons of this study are all from *Salmo Salar* species, that became very rare in the wild, but widely used for aquaculture. This means that the discrimination between wild and farmed in this study is also a discrimination between the two species.

The PCA on wild organic and conventionally farmed samples showed some tendencies in the data, but there wasn't a significant separation of the groups. The potential differences between the bands between these groups were at the level of the lipids (and the carotenoids bands, but also at the level of phenylalanine-related bands (1007 cm-1).

Indeed, wild, organic, and conventionally farmed salmon showed different average fatty acid profiles and some generalities for the sample set have been noticed with the GC data. Wild salmon seemed to have a lower recuperation percentage after the extraction of fatty acid and therefore would be less fatty than the farmed salmons. All the GC data are

expressed in proportion of fatty acid in the total fatty acid profile. Thus, it does not give the concentration of FAs in a piece of salmon, but it gives an idea of the fat nutritional quality. A big issue that was encountered in GC was the proportion of unidentified FAs for salmon that were way higher for the wild than both farmed salmons (25%, to 10% and 15%). These FA were probably highly unsaturated FAs that were not present in the GC standard used for this study, as it was initially used for milk analysis.

Using PCA and correlation table, a few FAs were identified to be the most significant differences between samples. Only these FAs were further investigated as they could be the key FAs for easier and less complex application, if for example, we would like to classify salmons accurately by predicting only these FAs. Of course, these were the FAs found in the sample set of this study, and they could not be sufficient for discriminating with other samples. Therefore, the general method seemed to work here.

The prediction of the selected FAs with the PLSR model based on the Raman spectroscopy has been done by carefully choosing the training and the test set to have an ideal variance in the two sets (with the Kennard-Stone algorithm for example). This was necessary because of the lack of data to build the models. Other configurations that are not shown in the study gave less accurate performances, only the best results are shown here. Furthermore, the evaluation of the models and the prediction based on Raman spectra are linked directly to the accuracy of the GC part. As the GC data were expressed in FAs proportion, and not in absolute FAs concentration in the muscle, the predictive performance could be affected. A more logical way to do this would be to predict absolute quantities of each fa based on the spectra, and applications would be interesting too.

Data used to predict FAs proportions were the 785nm data and the 800 to 1800 cm^{-1} spectral zone. A model predicts one FA proportion. In general, prediction showed good R^2 and low RMSEP for EPA, SFA and MUFA with R^2 of 0.78, 0.90 and 0.75 respectively. Even if oleic acid was in high quantity in the samples its prediction was not accurate (RMSEP of 8.34 on 23.89 average value to predict). The models for other FAs did not show good result as the distribution of their proportions were very low or very high, making it difficult for the model to predict accurately. Therefore, it would be interesting to, for example, classify the samples by their ω-3 concentration and make PLS-DA models.

A good model is not too complex and is robust with new samples. In the study K-Fold method shows how a model could be tuned efficiently by splitting the train set in multiple validation and calibration datasets and giving performances to choose the best number of latent variables for any model. If we look at the loadings of the latent variables for MUFA and SFA PLS-R models, the spectral areas that seem to explain the most variation of SFA and MUFA proportions are the 975-1025, 1325-1250, 1425-1475 and 1650-1675 cm^{-1} areas. Except for the 975-1025 area that is linked to phenylalanine, the three others are typically linked to lipids. The loadings for MUFA and SFA seem to be the same and no big difference can be spotted visually.

5.3 Fresh and thawed samples

To evaluate if thawing was detectable on the Raman spectral data of salmon muscle, 532nm laser spectra from 800 to 1800 cm^{-1} were used. Good separation between spectra of

one sample, fresh and thawed, indicates there is a difference between before and after the freezing and thawing. The loadings indicate that a variation due to thawing exists in the carotenoids C-C polyene chain stretch at 1158 cm^{-1} and C=C stretch around 1518 cm^{-1} but also at the level of lipids and phenylalanine. A general difference between before and after a thawing could therefore be somewhere in the combination of these spectral features.

The same result is found with the PCA on multiple samples, aiming to detect a general difference due to thawing. Loadings graphs (Fig.32 and Fig.34) tell that carotenoid bands around 1158 cm^{-1} and 1518 cm^{-1} are the most efficient to separate between fresh and thawed samples on a PCA score graph. The second principal component has high loadings on the bands around 1655 cm^{-1} that are linked to C=C stretch and more precisely to fatty acid unsaturation, but as seen on biplots, its efficacity to discriminate fresh and thawed samples is not truly demonstrated.

To verify the hints given by the PCA analysis on fresh and thawed samples a PLS-DA model was built. The loadings of the latent variables of a PLS model show the bands that were the most useful to discriminate between the classes. The model showed an 88,2% accuracy to predict the fresh or thawed nature of spectra from a new sample, with only 3 LVs. The loadings of LVs told that carotenoid bands, around 1518 cm^{-1}, were the more important in discrimination, followed by bands around 1158 cm^{-1}, and then the phenylalanine band (around 1007 cm^{-1}) for the first LV. Bands around 1655 cm^{-1} and 1445 cm^{-1} related to lipids became important in the LV3, showing that they are also useful to refine the performance of prediction (the Err went from 0.175 to 0.125 when using the LV3 on the validation set).

6 Conclusion

This study allowed to demonstrate that:

- Salmon muscle is a heterogeneous product, and it is to be considered in its analysis with Raman spectroscopy. However, averaging multiple spectra to have a zone spectrum showed good results in several data analysis in this study.

- The three lasers do not have the same abilities for prediction and analysis. The 532 nm laser is better suited for carotenoids-related bands, while the 785 and 1064 nm lasers are better for lipids-related bands. These lasers could therefore be complementary in some cases. Also, the acquisition parameters, such as the integration time, laser power and co-addition have very important impacts on the spectra and must be optimised.

- Some differences in the lipidic profile as well as for the carotenoids exist between the wild, organic, and conventionally farmed salmons, and these differences could be spotted by Raman spectroscopy. Just a PCA already allows us to see interesting tendencies at the level of carotenoids and lipidic related bands, when analysing wild compared to farmed sample Raman spectra (acquired with the 785nm laser).

- Raman spectroscopy (especially with a 785 nm laser) was very effective to classify fresh and thawed samples. The bands that seem to be useful to do so seem to be bands related to lipids and carotenoids. It has also shown a good potential to predict

individual FAs proportions in the salmon flesh, but the models could certainly be refined.

- Raman spectroscopy is a very sensitive method. Variability due to external factors must be avoided strongly. Standardisation of measurement methods to compare samples is not an easy thing, especially in the case of measurement directly on a salmon muscle.

- Raman spectroscopy seems to be well suited for the analysis of salmon flesh, as it is not influenced by water, and spectra don't have important noise problems.

Considering all these factors, Raman spectroscopy is believed to hold significant potential for the rapid, cost-effective, easily implementable, and non-destructive evaluation of salmon freshness, authenticity, and nutritional quality.

7 Perspectives

Salmon analysis with Raman spectroscopy is wide and certainly be explored during the next years, especially with the pressure due to mislabelling in the sector. This exploratory study enlightened important things about salmon analysis with Raman, that must be optimised if it comes to be extendedly used in the salmon industry.

Firstly, differences in the Raman spectra appear to exist within the same salmon depending on the measured location. To ensure minimal variability between measurements, the method must be standardized. To achieve this, the same piece of salmon should be sampled and measured consistently across all samples. The Norwegian Quality Cut, already used in whole salmon quality analysis, could be a suitable candidate for this purpose. Additionally, certain sample locations could represent strategic points for spectral analysis, such as naturally fatty areas for predictive models using spectral data related to lipids or more colourful regions if the model's target bands associated with carotenoids. Therefore, spatial studies of a salmon using Raman spectroscopy to spot these differences would be of great interest.

The acquisition parameters also must be standardised, as they change the spectrum. An interesting thing would be to use different wavelengths of incident lasers complementary, as they give different spectral profiles. By using both 532nm and 785nm lasers data, for example, the compounds of interest could be more focused on.

Moreover, the database could be significantly expanded to increase variability, particularly concerning the discrimination of wild and farmed salmon. Including an equal number of wild and farmed salmon would allow for a more nuanced understanding and the creation of discrimination models. This study only encompassed two species of salmon, while others also exist. Different species of wild and farmed salmon may exhibit distinct characteristics. It is crucial to prepare models for all potential variations encountered in the market.

The prediction of salmon fatty acid profiles is also of significant interest. With more refined prediction models, it might be possible to determine the nutritional quality of salmon through routine testing instead of relying on GC analysis. Similarly, as demonstrated in this study with a limited number of samples, it is likely feasible to predict the freshness of a salmon. Distinguishing between fresh and previously frozen salmon is essential due to their varying market values, making rapid testing crucial for preventing fraud. However, the freezing process used in this study differs from industrial food practices, necessitating an exact reproduction of the industry-standard procedure. Additionally, a very good understanding of the chemical and physical effects of freezing on salmon flesh would be needed.

An aspect that should have been further investigated in this work is the analysis of carotenoids using Raman spectroscopy. As demonstrated in the context of fresh/thawed classification, carotenoids play a significant role. Theoretically, it might be possible to discriminate between wild and farmed salmon based on the nature of their carotenoids, which are subtly different. These subtle differences could potentially be detected at lower resolutions than the 4 cm^{-1} used in this study, but it must be verified. A thorough understanding of carotenoid and lipid variations within salmon flesh would improve its analysis by Raman spectroscopy.

Another potential study could focus on contaminants detection with Raman spectroscopy, as both wild and farmed salmon could contain some of them.

8 References

Aas, T.S., Ytrestøyl, T., Åsgård, T., 2019. Utilization of feed resources in the production of Atlantic salmon (Salmo salar) in Norway: An update for 2016. Aquac. Rep. 15, 100216. https://doi.org/10.1016/j.aqrep.2019.100216

Afseth, N.K., Bloomfield, M., Wold, J.P., Matousek, P., 2014. A novel approach for subsurface through-skin analysis of salmon using spatially offset Raman spectroscopy (SORS). Appl. Spectrosc. 68, 255–262. https://doi.org/10.1366/13-07215

Afseth, N.K., Dankel, K., Andersen, P.V., Difford, G.F., Horn, S.S., Sonesson, A., Hillestad, B., Wold, J.P., Tengstrand, E., 2022. Raman and near Infrared Spectroscopy for Quantification of Fatty Acids in Muscle Tissue—A Salmon Case Study. Foods 11, 962. https://doi.org/10.3390/foods11070962

Alert, I., 2007. Government Fails Consumers, Falls Short on Seafood Inspections. Food Water Watch May 30th.

Ambati, R.R., Phang, S.-M., Ravi, S., Aswathanarayana, R.G., 2014. Astaxanthin: Sources, Extraction, Stability, Biological Activities and Its Commercial Applications—A Review. Mar. Drugs 12, 128–152. https://doi.org/10.3390/md12010128

Amigo, J., Martí-Aluja, I., Gowen, A., 2013. Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality. pp. 343–370.

Angiulli, F., Pizzuti, C., 2002. Fast Outlier Detection in High Dimensional Spaces. Presented at the Proceedings of the Sixth European Conference on the Principles of Data Mining and Knowledge Discovery, pp. 15–26. https://doi.org/10.1007/3-540-45681-3_2

Atlantic salmon, 2022. . Wikipedia.

Balaji, S., Tchoukanova, N., Djaoued, Y., 2010. Investigation on Natural and Synthetic Astaxanthin and its Isomers by a Combined Raman/IR Microprobe with Samespot Technology. AIP Conf. Proc. 1267, 553–554. https://doi.org/10.1063/1.3482672

Barrionuevo, A., 2007. Food imports often escape scrutiny. N. Y. Times 1.
Beddington, J.R., Kirkwood, G.P., Muir, J., 2005. Managing to harvest? Perspectives on the potential of aquaculture. Philos. Trans. R. Soc. B Biol. Sci. 360, 191–218. https://doi.org/10.1098/rstb.2004.1572
Bjerkeng, B., 1997. Chromatographic Analysis of Synthesized Astaxanthin—A Handy Tool for the Ecologist and the Forensic Chemist? Progress. Fish-Cult. 59, 129–140. https://doi.org/10.1577/1548-8640(1997)059<0129:CAOSAA>2.3.CO;2
Blanchet, C., Lucas, M., Julien, P., Morin, R., Gingras, S., Dewailly, É., 2005. Fatty acid composition of wild and farmed Atlantic salmon (Salmo salar) and rainbow trout (Oncorhynchus mykiss). Lipids 40, 529–531. https://doi.org/10.1007/s11745-005-1414-0
Brodersen, K., Bremner, H.A., 2001. Exploration of the Use of NIR Reflectance Spectroscopy to Distinguish and Measure Attributes of Conditioned and Cooked Shrimp (Pandalus borealis). LWT - Food Sci. Technol. 34, 533–541. https://doi.org/10.1006/fstl.2001.0802
Capuano, E., Lommen, A., Heenan, S., de la Dura, A., Rozijn, M., van Ruth, S., 2012. Wild salmon authenticity can be predicted by 1H-NMR spectroscopy. Lipid Technol. 24, 251–253. https://doi.org/10.1002/lite.201200235
Carcea, M., Brereton, P., Hsu, R., Kelly, S., Marmiroli, N., Melini, F., Soukoulis, C., Wenping, D., 2009. Food authenticity assessment: ensuring compliance with food legislation and traceability requirements. Qual. Assur. Saf. Crops Foods 1, 93–100. https://doi.org/10.1111/j.1757-837X.2009.00011.x
Cawthorn, D.-M., Steinman, H.A., Witthuhn, R.C., 2012. DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. Food Res. Int. 46, 30–40. https://doi.org/10.1016/j.foodres.2011.11.011
Chang, W.-H., Ling, Y.S., Wang, K.-C., Nan, F.-H., Chen, W.-L., 2022. Discrimination of Atlantic salmon origins using untargeted chemical fingerprinting. Food Chem. 394, 133538. https://doi.org/10.1016/j.foodchem.2022.133538
Cheng, J.-H., Sun, D.-W., Zeng, X.-A., Liu, D., 2015. Recent Advances in Methods and Techniques for Freshness Quality Determination and Evaluation of Fish and Fish Fillets: A Review. Crit. Rev. Food Sci. Nutr. 55, 1012–1225. https://doi.org/10.1080/10408398.2013.769934
Christiansen, R., Struksnæs, G., Estermann, R., Torrissen, O.J., 1995. Assessment of flesh colour in Atlantic salmon, Salmo salar L. Aquac. Res. 26, 311–321. https://doi.org/10.1111/j.1365-2109.1995.tb00919.x
Cline, E., 2012. Marketplace substitution of Atlantic salmon for Pacific salmon in Washington State detected by DNA barcoding. Food Res. Int. 45, 388–393. https://doi.org/10.1016/j.foodres.2011.10.043
Corbo, T., Lovera, P., Clemente, F., Lincoln, T., 2005. Tabled Lables: Consumers Eat Blind While Congress Feasts on Campaign Cash.
Cozzolino, D., Murray, I., 2012. A Review on the Application of Infrared Technologies to Determine and Monitor Composition and Other Quality Characteristics in Raw Fish, Fish Products, and Seafood. Appl. Spectrosc. Rev. 47, 207–218. https://doi.org/10.1080/05704928.2011.639106
Dai, Q., Cheng, J.-H., Sun, D.-W., Pu, H., Zeng, X.-A., Xiong, Z., 2015. Potential of visible/near-infrared hyperspectral imaging for rapid detection of freshness in unfrozen and frozen prawns. J. Food Eng. 149, 97–104. https://doi.org/10.1016/j.jfoodeng.2014.10.001
D'Amico, P., Armani, A., Gianfaldoni, D., Guidi, A., 2016. New provisions for the labelling of fishery and aquaculture products: Difficulties in the implementation of Regulation (EU) n. 1379/2013. Mar. Policy 71, 147–156. https://doi.org/10.1016/j.marpol.2016.05.026
Das, R.S., Agrawal, Y.K., 2011. Raman spectroscopy: Recent advancements, techniques and applications. Vib. Spectrosc. 57, 163–176. https://doi.org/10.1016/j.vibspec.2011.08.003
Davinelli, S., Nielsen, M.E., Scapagnini, G., 2018. Astaxanthin in Skin Health, Repair, and Disease: A Comprehensive Review. Nutrients 10, 522. https://doi.org/10.3390/nu10040522

Dempson, J.B., Power, M., 2004. Use of stable isotopes to distinguish farmed from wild Atlantic salmon, Salmo salar. Ecol. Freshw. Fish 13, 176–184. https://doi.org/10.1111/j.1600-0633.2004.00057.x

Ding, R., Huang, X., Han, F., Dai, H., Teye, E., Xu, F., 2014. Rapid and nondestructive evaluation of fish freshness by near infrared reflectance spectroscopy combined with chemometrics analysis. Anal. Methods 6, 9675–9683. https://doi.org/10.1039/C4AY01839G

DIRECTION GENERALE DES POLITIQUES ECONOMIQUES, n.d. Cahier des charges concernant le mode de production et de préparation biologiques des espèces aquacoles et leurs dérivés. [WWW Document]. URL https://docplayer.fr/20792283-Cahier-des-charges-concernant-le-mode-de-production-et-de-preparation-biologiques-des-especes-aquacoles-et-leurs-derives.html (accessed 6.6.23).

Efremov, E.V., Ariese, F., Gooijer, C., 2008. Achievements in resonance Raman spectroscopy: Review of a technique with a distinct analytical chemistry potential. Anal. Chim. Acta 606, 119–134. https://doi.org/10.1016/j.aca.2007.11.006

Ellis, D.I., Muhamadali, H., Haughey, S.A., Elliott, C.T., Goodacre, R., 2015. Point-and-shoot: rapid quantitative detection methods for on-site food fraud analysis – moving out of the laboratory and into the food supply chain. Anal. Methods 7, 9401–9414. https://doi.org/10.1039/C5AY02048D

Elmadfa, I., Majchrzak, D., 1998. Carotinoide und Vitamin A in Fischproben. Eur. J. Nutr. 37, 207–210. https://doi.org/10.1007/s003940050018

Ermakov, I.V., Ermakova, M.R., Gellermann, W., 2006. Quantitative detection of astaxanthin and cantaxanthin in Atlantic salmon by resonance Raman spectroscopy.

Esmonde-White, K., Lewis, M., 2022. The role of Raman spectroscopy in food and beverage analysis 18.

FAO 2023, 2004. Fisheries and Aquaculture - Cultured Aquatic Species - Salmo salar [WWW Document]. fao.org. URL https://www.fao.org/fishery/en/culturedspecies/salmo_salar/en (accessed 12.5.22).

Fiorino, G.M., Losito, I., De Angelis, E., Arlorio, M., Logrieco, A.F., Monaci, L., 2019. Assessing fish authenticity by direct analysis in real time-high resolution mass spectrometry and multivariate analysis: discrimination between wild-type and farmed salmon. Food Res. Int. 116, 1258–1265. https://doi.org/10.1016/j.foodres.2018.10.013

Foran, J.A., Good, D.H., Carpenter, D.O., Hamilton, M.C., Knuth, B.A., Schwager, S.J., 2005. Quantitative Analysis of the Benefits and Risks of Consuming Farmed and Wild Salmon. J. Nutr. 135, 2639–2643. https://doi.org/10.1093/jn/135.11.2639

Friesen, E.N., Higgs, D.A., Devlin, R.H., 2015. Flesh nutritional content of growth hormone transgenic and non-transgenic coho salmon compared to various species of farmed and wild salmon. Aquaculture 437, 318–326. https://doi.org/10.1016/j.aquaculture.2014.11.035

Galvez, J.F., Mejuto, J.C., Simal-Gandara, J., 2018. Future challenges on the use of blockchain for food traceability analysis. TrAC Trends Anal. Chem. 107, 222–232. https://doi.org/10.1016/j.trac.2018.08.011

Griffiths, A.M., Sotelo, C.G., Mendes, R., Pérez-Martín, R.I., Schröder, U., Shorten, M., Silva, H.A., Verrez-Bagnis, V., Mariani, S., 2014. Current methods for seafood authenticity testing in Europe: Is there a need for harmonisation? Food Control 45, 95–100. https://doi.org/10.1016/j.foodcont.2014.04.020

Guerin, M., Huntley, M.E., Olaizola, M., 2003. Haematococcus astaxanthin: applications for human health and nutrition. Trends Biotechnol. 21, 210–216. https://doi.org/10.1016/S0167-7799(03)00078-7

Hansen, H., Onozaka, Y., 2011. When Diseases Hit Aquaculture: An Experimental Study of Spillover Effects from Negative Publicity. Mar. Resour. Econ. 26, 281–291. https://doi.org/10.5950/0738-1360-26.4.281

Hatlen, B., Jobling, M., Bjerkeng, B., 1998. Relationships between carotenoid concentration and colour of fillets of Arctic chair, Salvelinus alpinus (L.), fed astaxanthin. Aquac. Res. 29, 191–202. https://doi.org/10.1046/j.1365-2109.1998.00956.x

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321. https://doi.org/10.1098/rspb.2002.2218

Hellberg, R.S.R., Morrissey, M.T., 2011. Advances in DNA-Based Techniques for the Detection of Seafood Species Substitution on the Commercial Market. JALA J. Assoc. Lab. Autom. 16, 308–321. https://doi.org/10.1016/j.jala.2010.07.004

Hikima, J., Ando, M., Hamaguchi, H., Sakai, M., Maita, M., Yazawa, K., Takeyama, H., Aoki, T., 2017. On-site Direct Detection of Astaxanthin from Salmon Fillet Using Raman Spectroscopy. Mar. Biotechnol. 19, 157–163. https://doi.org/10.1007/s10126-017-9739-7

Hites, R.A., Foran, J.A., Carpenter, D.O., Hamilton, M.C., Knuth, B.A., Schwager, S.J., 2004. Global Assessment of Organic Contaminants in Farmed Salmon. Science 303, 226–229.

Hofherr, J., Martinsohn, J., Cawthorn, D., Rasco, B., Naaum, A.M., 2016. Chapter 3 - Regulatory Frameworks for Seafood Authenticity and Traceability, in: Naaum, A.M., Hanner, R.H. (Eds.), Seafood Authenticity and Traceability. Academic Press, San Diego, pp. 47–82. https://doi.org/10.1016/B978-0-12-801592-6.00003-6

Hu, Y., Huang, S.Y., Hanner, R., Levin, J., Lu, X., 2018. Study of fish products in Metro Vancouver using DNA barcoding methods reveals fraudulent labeling. Food Control 94, 38–47. https://doi.org/10.1016/j.foodcont.2018.06.023

Jacquet, J.L., Pauly, D., 2008. Trade secrets: Renaming and mislabeling of seafood. Mar. Policy 32, 309–318. https://doi.org/10.1016/j.marpol.2007.06.007

Kamal, M., Karoui, R., 2015. Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review. Trends Food Sci. Technol. 46, 27–48. https://doi.org/10.1016/j.tifs.2015.07.007

Kappel, K., Schröder, U., 2016. Substitution of high-priced fish with low-priced species: Adulteration of common sole in German restaurants. Food Control 59, 478–486. https://doi.org/10.1016/j.foodcont.2015.06.024

Kudelski, A., 2008. Analytical applications of Raman spectroscopy. Talanta 76, 1–8. https://doi.org/10.1016/j.talanta.2008.02.042

Landry, J.D., Torley, P.J., Blanch, E.W., 2020. Detection of Biomarkers Relating to Quality and Differentiation of Some Commercially Significant Whole Fish Using Spatially Off-Set Raman Spectroscopy. Molecules 25, 3776. https://doi.org/10.3390/molecules25173776

Leal, M.C., Pimentel, T., Ricardo, F., Rosa, R., Calado, R., 2015. Seafood traceability: current needs, available tools, and biotechnological challenges for origin certification. Trends Biotechnol. 33, 331–336. https://doi.org/10.1016/j.tibtech.2015.03.003

Lesnoff, M., 2023. Jchemo.jl.

Li, J., Zhu, D., Niu, J., Shen, S., Wang, G., 2011. An economic assessment of astaxanthin production by large scale cultivation of Haematococcus pluvialis. Biotechnol. Adv. 29, 568–574. https://doi.org/10.1016/j.biotechadv.2011.04.001

Li-Chan, E.C.Y., 1996. The applications of Raman spectroscopy in food science. Trends Food Sci. Technol. 7, 361–370. https://doi.org/10.1016/S0924-2244(96)10037-6

Lintvedt, T.A., Andersen, P.V., Afseth, N.K., Heia, K., Lindberg, S.-K., Wold, J.P., 2023. Raman spectroscopy and NIR hyperspectral imaging for in-line estimation of fatty acid features in salmon fillets. Talanta 254, 124113. https://doi.org/10.1016/j.talanta.2022.124113

Liu, D., Zeng, X.-A., Sun, D.-W., 2013. NIR Spectroscopy and Imaging Techniques for Evaluation of Fish Quality—A Review. Appl. Spectrosc. Rev. 48, 609–628. https://doi.org/10.1080/05704928.2013.775579

Lundebye, A.-K., Lock, E.-J., Rasinger, J.D., Nøstbakken, O.J., Hannisdal, R., Karlsbakk, E., Wennevik, V., Madhun, A.S., Madsen, L., Graff, I.E., Ørnsrud, R., 2017. Lower levels of Persistent Organic Pollutants, metals and the marine omega 3-fatty acid DHA in farmed compared to wild Atlantic salmon (Salmo salar). Environ. Res. 155, 49–59. https://doi.org/10.1016/j.envres.2017.01.026

Lura, H., Saegrov, H., 1991. A Method of Separating Offspring from Farmed and Wild Atlantic Salmon (Salmo salar) based on Different Ratios of Optical Isomers of Astaxanthin. Can. J. Fish. Aquat. Sci. 48, 429–433. https://doi.org/10.1139/f91-056

Lv, H., Xu, W., You, J., Xiong, S., 2017. Classification of freshwater fish species by linear discriminant analysis based on near infrared reflectance spectroscopy. J. Infrared Spectrosc. 25, 54–62.

Martinez, I., Aursand, M., Erikson, U., Singstad, T.E., Veliyulin, E., van der Zwaag, C., 2003. Destructive and non-destructive analytical techniques for authentication and composition analyses of foodstuffs. Trends Food Sci. Technol. 14, 489–498. https://doi.org/10.1016/j.tifs.2003.07.005

Mathlouthi, M., Seuvre, A.M., Koenig, J.L., 1986. F.T.-I.R. and laser-raman spectra of guanine and guanosine. Carbohydr. Res. 146, 15–27. https://doi.org/10.1016/0008-6215(86)85020-0

Megdal, P.A., Craft, N.A., Handelman, G.J., 2009. A Simplified Method to Distinguish Farmed (*Salmo salar*) from Wild Salmon: Fatty Acid Ratios Versus Astaxanthin Chiral Isomers. Lipids 44, 569–576. https://doi.org/10.1007/s11745-009-3294-6

Miller, D., Jessel, A., Mariani, S., 2012. Seafood mislabelling: comparisons of two western European case studies assist in defining influencing factors, mechanisms and motives. Fish Fish. 13, 345–358.

Moghadasian, M.H., Shahidi, F., 2017. Fatty Acids, in: International Encyclopedia of Public Health. Elsevier, pp. 114–122. https://doi.org/10.1016/B978-0-12-803678-5.00157-0

Molkentin, J., Lehmann, I., Ostermeyer, U., Rehbein, H., 2015. Traceability of organic fish – Authenticating the production origin of salmonids by chemical and isotopic analyses. Food Control 53, 55–66. https://doi.org/10.1016/j.foodcont.2015.01.003

Molversmyr, E., Devle, H.M., Naess-Andresen, C.F., Ekeberg, D., 2022. Identification and quantification of lipids in wild and farmed Atlantic salmon (Salmo salar), and salmon feed by GC-MS. Food Sci. Nutr. 10, 3117–3127. https://doi.org/10.1002/fsn3.2911

Moretti, V., Turchini, G., Bellagamba, F., Caprino, F., 2003. Traceability Issues in Fishery and Aquaculture Products. Vet. Res. Commun. 27 Suppl 1, 497–505. https://doi.org/10.1023/B:VERC.0000014207.01900.5c

Movasaghi, Z., Rehman, S., Rehman, I.U., 2007. A LIRE !!!!!!! Raman Spectroscopy of Biological Tissues. Appl. Spectrosc. Rev. 42, 493–541. https://doi.org/10.1080/05704920701551530

Muñoz-Colmenero, M., Juanes, F., Dopico, E., Martinez, J.L., Garcia-Vazquez, E., 2017. Economy matters: A study of mislabeling in salmon products from two regions, Alaska and Canada (Northwest of America) and Asturias (Northwest of Spain). Fish. Res. 195, 180–185. https://doi.org/10.1016/j.fishres.2017.07.012

Nagwa, n.d. Lesson Explainer: Lipids | Nagwa [WWW Document]. URL https://www.nagwa.com/en/explainers/313147519531/ (accessed 8.5.23).

Nanton, D.A., Vegusdal, A., Rørå, A.M.B., Ruyter, B., Baeverfjord, G., Torstensen, B.E., 2007. Muscle lipid storage pattern, composition, and adipocyte distribution in different parts of Atlantic salmon (Salmo salar) fed fish oil and vegetable oil. Aquaculture 265, 230–243. https://doi.org/10.1016/j.aquaculture.2006.03.053

Ostermeyer, U., 2004. 4.9 DIFFERENTIATION OF WILD SALMON, CONVENTIONALLY AND ORGANICALLY FARMED SALMON. Nutr. FOOD Dep. FISH Qual.

Ottavian, M., Facco, P., Fasolato, L., Novelli, E., Mirisola, M., Perini, M., Barolo, M., 2012. Use of Near-Infrared Spectroscopy for Fast Fraud Detection in Seafood: Application to the Authentication of Wild European Sea Bass (Dicentrarchus labrax). J. Agric. Food Chem. 60, 639–648. https://doi.org/10.1021/jf203385e

Pardo, M.Á., Jiménez, E., Pérez-Villarreal, B., 2016. Misdescription incidents in seafood sector. Food Control 62, 277–283. https://doi.org/10.1016/j.foodcont.2015.10.048

Power, A., Cozzolino, D., 2020. How Fishy Is Your Fish? Authentication, Provenance and Traceability in Fish and Seafood by Means of Vibrational Spectroscopy. Appl. Sci. 10, 4150. https://doi.org/10.3390/app10124150

Prendergast, A.F., Higgs, D.A., Beames, D.M., Dosanjh, B.S., Deacon, G., 1994. Searching for substitutes: canola. North. Aquac. 10, 15–20.

Ramaswamy, S., Rastogi, R., Shim, K., 2000. Efficient algorithms for mining outliers from large data sets, in: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, SIGMOD '00. Association for Computing Machinery, New York, NY, USA, pp. 427–438. https://doi.org/10.1145/342009.335437

Rašković, B., Heinke, R., Rösch, P., Popp, J., 2016. The Potential of Raman Spectroscopy for the Classification of Fish Fillets. Food Anal. Methods 9, 1301–1306. https://doi.org/10.1007/s12161-015-0312-6

Regulation, E.U., 2013. No 1380/2013 of the European Parliament and of the Council of 11 December 2013 on the Common Fisheries Policy, amending Council Regulations (EC) No 1954/2003 and (EC) No 1224/2009 and repealing Council Regulations (EC) No 2371/2002 and (EC) No 639/2004 and Council Decision 2004/585/EC. 2013. J Eur Union 354, 22–61.

Ren, L., Tian, Y., Yang, X., Wang, Q., Wang, L., Geng, X., Wang, K., Du, Z., Li, Y., Lin, H., 2023. Rapid identification of fish species by laser-induced breakdown spectroscopy and Raman spectroscopy coupled with machine learning methods. Food Chem. 400, 134043. https://doi.org/10.1016/j.foodchem.2022.134043

Rinnan, Å., Berg, F.V.D., Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal. Chem. 28, 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Rostron, P., Gerber, D., 2016. Raman Spectroscopy, a review. Int. J. Eng. Tech. Res. 6, 50–64.

Saraiva, C., Vasconcelos, H., de Almeida, J.M.M.M., 2017. A chemometrics approach applied to Fourier transform infrared spectroscopy (FTIR) for monitoring the spoilage of fresh salmon (Salmo salar) stored under modified atmospheres. Int. J. Food Microbiol. 241, 331–339. https://doi.org/10.1016/j.ijfoodmicro.2016.10.038

Savitzky, A., Golay, M.J.E., 2002. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. [WWW Document]. ACS Publ. https://doi.org/10.1021/ac60214a047

Sharafi, Y., Majidi, M.M., Goli, S.A.H., Rashidi, F., 2015. Oil Content and Fatty Acids Composition in Brassica Species. Int. J. Food Prop. 18, 2145–2154. https://doi.org/10.1080/10942912.2014.968284

Sikder, M.N.K., Batarseh, F.A., 2023. Outlier detection using AI: a survey, in: AI Assurance. Elsevier, pp. 231–291. https://doi.org/10.1016/B978-0-32-391919-7.00020-2

Simopoulos, A.P., 2016. An Increase in the Omega-6/Omega-3 Fatty Acid Ratio Increases the Risk for Obesity. Nutrients 8, 128. https://doi.org/10.3390/nu8030128

Smith, E., Dent, G., 2019. Modern Raman spectroscopy: a practical approach. John Wiley & Sons.

Sprague, M., Dick, J.R., Tocher, D.R., 2016. Impact of sustainable feeds on omega-3 long-chain fatty acid levels in farmed Atlantic salmon, 2006–2015. Sci. Rep. 6, 21892. https://doi.org/10.1038/srep21892

Storebakken, T., Foss, P., Austreng, E., Liaaen-Jensen, S., 1985. Carotenoids in diets for salmonids: II. Epimerization studies with astaxanthin in Atlantic salmon. Aquaculture 44, 259–269.

Storebakken, T., No, H.K., 1992. Pigmentation of rainbow trout. Aquaculture, The Rainbow Trout 100, 209–229. https://doi.org/10.1016/0044-8486(92)90372-R

Thawornchinsombut, S., Park, J.W., Meng, G., Li-Chan, E.C.Y., 2006. Raman Spectroscopy Determines Structural Changes Associated with Gelation Properties of Fish Proteins Recovered at Alkaline pH. J. Agric. Food Chem. 54, 2178–2187. https://doi.org/10.1021/jf0518958

Thomas, F., Jamin, E., Wietzerbin, K., Guérin, R., Lees, M., Morvan, E., Billault, I., Derrien, S., Moreno Rojas, J.M., Serra, F., Guillou, C., Aursand, M., McEvoy, L., Prael, A., Robins, R.J., 2008. Determination of Origin of Atlantic Salmon (Salmo salar): The Use of Multiprobe and Multielement Isotopic Analyses in Combination with Fatty Acid Composition To Assess Wild or Farmed Origin. J. Agric. Food Chem. 56, 989–997. https://doi.org/10.1021/jf072370d

Tolasa, S., Cakli, S., Ostermeyer, U., 2005. Determination of astaxanthin and canthaxanthin in salmonid. Eur. Food Res. Technol. 221, 787–791. https://doi.org/10.1007/s00217-005-0071-5

Torrissen, O.J., 1986. Pigmentation of salmonids — A comparison of astaxanthin and canthaxanthin as pigment sources for rainbow trout. Aquaculture 53, 271–278. https://doi.org/10.1016/0044-8486(86)90357-1

Torrissen, O.J., Hardy, R.W., Shearer, K.D., 1989. Pigmentation of salmonids-carotenoid deposition and metabolism. CRC Crit Rev Aquat Sci 1, 209–225.

Turujman, S.A., Wamer, W.G., Wei, R.R., Albert, R.H., 1997. Rapid Liquid Chromatographic Method to Distinguish Wild Salmon from Aquacultured Salmon Fed Synthetic Astaxanthin. J. AOAC Int. 80, 622–632. https://doi.org/10.1093/jaoac/80.3.622

Velioğlu, H.M., Temiz, H.T., Boyaci, I.H., 2015. Differentiation of fresh and frozen-thawed fish samples using Raman spectroscopy coupled with chemometric analysis. Food Chem. 172, 283–290. https://doi.org/10.1016/j.foodchem.2014.09.073

Villaró, S., Ciardi, M., Morillas-España, A., Sanchez Zurano, A., Acién-Fernández, G., Lafarga, T., 2021. Microalgae Derived Astaxanthin: Research and Consumer Trends and Industrial Use as Food. Foods 10, 2303. https://doi.org/10.3390/foods10102303

Ward, R.D., Hanner, R., Hebert, P.D.N., 2009. The campaign to DNA barcode all fishes, FISH-BOL. J. Fish Biol. 74, 329–356. https://doi.org/10.1111/j.1095-8649.2008.02080.x

Warner, K., Timme, W., Lowell, B., Hirschfield, M., 2013. Oceana study reveals seafood fraud nationwide. Oceana Washington, DC.

Weeranantanaphan, J., Downey, G., Allen, P., Sun, D.-W., 2011. A Review of near Infrared Spectroscopy in Muscle Food Analysis: 2005–2010. J. Infrared Spectrosc. 19, 61–104. https://doi.org/10.1255/jnirs.924

West European Fish Technologist's Association, 2004. Proceedings of the 34th WEFTA Conference 12 - 15 September 2004, Lübeck, Germany. Federal Research Centre for Nutrition and Food, Dep. for Fish Quality, Hamburg.

Whyte, J.N.C., Travers, D., Sherry, K.L., 1998. Deposition of astaxanthin isomers in chinook salmon (Oncorhynchus tshawytscha) fed different sources of pigment, Canadian technical report of fisheries and aquatic sciences. Fisheries and Oceans Canada, Nanaimo, B.C.

Wold, J.P., Marquardt, B.J., Dable, B.K., Robb, D., Hatlen, B., 2004. Rapid Quantification of Carotenoids and Fat in Atlantic Salmon (*Salmo salar L.*) by Raman Spectroscopy and Chemometrics. Appl. Spectrosc. 58, 395–403.

Wong, E.H.-K., Hanner, R.H., 2008. DNA barcoding detects market substitution in North American seafood. Food Res. Int. 41, 828–837. https://doi.org/10.1016/j.foodres.2008.07.005

Workman, J.J., Mobley, P.R., Kowalski, B.R., Bro, R., 1996. Review of Chemometrics Applied to Spectroscopy: 1985-95, Part I. Appl. Spectrosc. Rev. 31, 73–124. https://doi.org/10.1080/05704929608000565

World Aquaculture 2020 – A brief overview, 2022. . FAO. https://doi.org/10.4060/cb7669en

Wu, T., Zhong, N., Yang, L., 2018. Identification of Adulterated and Non-adulterated Norwegian Salmon Using FTIR and an Improved PLS-DA Method. Food Anal. Methods 11, 1501–1509. https://doi.org/10.1007/s12161-017-1135-4

Yang, L.G., Song, Z.X., Yin, H., Wang, Y.Y., Shu, G.F., Lu, H.X., Wang, S.K., Sun, G.J., 2016. Low n-6/n-3 PUFA Ratio Improves Lipid Metabolism, Inflammation, Oxidative Stress and Endothelial Function in Rats Using Plant Oils as n-3 Fatty Acid Source. Lipids 51, 49–59. https://doi.org/10.1007/s11745-015-4091-z

Yao, G., Guo, S., Yu, W., Muhammad, M., Liu, J., Huang, Q., 2021. DFT and Raman study of all-trans astaxanthin optical isomers. Spectrochim. Acta. A. Mol. Biomol. Spectrosc. 262, 120143. https://doi.org/10.1016/j.saa.2021.120143

Yuan, J.-P., Peng, J., Yin, K., Wang, J.-H., 2011. Potential health-promoting effects of astaxanthin: A high-value carotenoid mostly from microalgae. Mol. Nutr. Food Res. 55, 150–165. https://doi.org/10.1002/mnfr.201000414

Zheng, J., He, L., 2014. Surface-Enhanced Raman Spectroscopy for the Chemical Analysis of Food. Compr. Rev. Food Sci. Food Saf. 13, 317–328. https://doi.org/10.1111/1541-4337.12062

Zhong, N., Li, Y.P., Li, X.Z., Guo, C.X., Wu, T., 2021. Accurate prediction of salmon storage time using improved Raman spectroscopy. J. Food Eng. 293, 110378. https://doi.org/10.1016/j.jfoodeng.2020.110378

Zhu, F., Zhang, D., He, Y., Liu, F., Sun, D.-W., 2013. Application of Visible and Near Infrared Hyperspectral Imaging to Differentiate Between Fresh and Frozen–Thawed Fish Fillets. Food Bioprocess Technol. 6, 2931–2937. https://doi.org/10.1007/s11947-012-0825-6

Ziegler, H., Osmond, C.B., Stichler, W., Trimborn, P., 1976. Hydrogen isotope discrimination in higher plants: Correlations with photosynthetic pathway and environment. Planta 128, 85–92. https://doi.org/10.1007/BF00397183

Zumelzu, E., Sanz, A., Medina, J., Rull, F., 2018. Determination of Processed Salmon Components Sticking to Polyethylene Terephthalate Coatings of Containers by FT-IR and Raman Vibrational Spectroscopy. J. Spectrosc. 2018, e8798239. https://doi.org/10.1155/2018/8798239

Integrated approach to assess the freshness and authenticity of salmon fillets: a combination of Gas Chromatography and Raman Spectroscopy.

Authom Alexis

We are living in a world where food security is a global challenge, and where consumers' choices are continuously changing in terms of specific food quality. Routine methods for characterising and authenticating globally traded products like salmon are necessary. Still, today, controlling the fraud on the salmon market is difficult. In this context, Raman spectroscopy is rapidly advancing; it is among the fast, non-destructive, and cost-effective techniques highlighted in quality control. Coincidently, reliable tools for processing large databases become more accessible in open source.

In this exploratory study, several tests were conducted with Raman spectroscopy to assess the freshness, authenticity, and lipidic profiles associated with 28 salmon samples from different stores in Belgium. The acquisition parameters for spectral data and the feasibility of reliably analysing salmon flesh using Raman spectroscopy, have been discussed. The study also explored some data processing tools that can extract a wealth of interesting information from the salmon spectrum. A part of this work consisted of extracting lipids from the salmon fillets and analysing them using Gas Chromatography (GC). Both spectral and GC data were used to analyse the fatty acid profiles of the samples. The different types of salmon have been discussed, and predictive models targeting authenticity (wild, organically farmed, or conventionally farmed) and freshness (fresh or thawed) were developed and evaluated.

This end of study work shows the potential of Raman spectroscopy for the assessment of the freshness of a salmon fillet. The results also show the possibility to discriminate between wild, conventionally farmed, and organically farmed salmons according to their lipidic profiles, and the potential to evaluate some lipidic features of a salmon fillet with Raman spectroscopy.

This work was realised with the collaboration of the Centre wallon de Recherches Agronomiques (CRA-W), giving access to the spectroscopic and chromatographic instruments.

This work is open to perspectives for improvement and can serve as a foundation for further studies.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des bioingénieurs
Croix du Sud, 2 bte L7.05.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/agro**