Hungarian University of Agriculture and Life Sciences

TAPPING INTO THE POTENTIAL OF NEAR-INFRARED SPECTROSCOPY AND AQUAPHOTOMICS FOR THE QUALITATIVE ASSESSMENT OF POWDERED FOOD MATRICES

Balkis Aouadi
Budapest
2023

PhD School/ Program

Name: Doctoral School of Food Science

Field: Food Science

Head: **Livia Simon Sarkadi, DSc**
Department of Nutrition
Institute of Food Science and Technology
Hungarian University of Agriculture and Life Sciences

Supervisors:

Zoltan Kovacs, PhD
Department of Food Measurements and Process Control, Institute of Food Science and Technology, Hungarian University of Agriculture and Life Sciences

The applicant met the requirement of the Ph.D. regulations of the Hungarian University of Agriculture and Life Sciences and the thesis is accepted for the defense process.

....

Head of Doctoral School Supervisor

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AAE: Ascorbic Acid Equivalent

ANN: Artificial Neural Network

CA: Cluster Analysis

CGA: Chlorogenic Acid

DeTr: Detrend

DW: Dry Weight

FRAP: Ferric Reducing Antioxidant Power

GAE: Gallic Acid Equivalent

HFCS: High-Fructose Corn Syrup

HPLC: High Performance Liquid Chromatography

LDA: Linear Discriminant Analysis

LOD: Limit Of Detection

LV: Latent Variables

MLR: Multiple Linear Regression

MSC: Multiplicative Scatter Correction

M: Mol per Liter

NIRS: Near Infrared Spectroscopy

NMR: Nuclear Magnetic Resonance

OPLSR: Orthogonal Partial Least Squares Regression

PCA-LDA: Hybrid Principal Component Analysis-Linear Discriminant Analysis

PCA: Principal Component Analysis

PCR: Principal Component Regression

PC: Principal Component

PLS-DA: Partial Least Squares Discriminant Analysis

PLSR: Partial Least Squares Regression

ppb: Parts per billion

ppm: Parts per million

R^2: Coefficient of Determination

RPD: Ratio of Performance to Deviation

RMSE: Root Mean Square Error

RMSEP: Root Mean Square Error of Prediction

R: Correlation Coefficient

SEP: Square Error of Prediction

SIMCA: Soft Independent Modeling of Class Analogy

SNV: Standard Normal Variate

SOC: Soil Organic Carbon

SSC: Soluble Solids Content

SVM: Support Vector Machine

TAC: Total Antioxidant Capacity

TPC: Total Polyphenols Content

Vis-NIR: Visible Near Infrared

w/w: Weight per weight

WAB: Water Absorbance Band

WAMACS: Water Matrix Coordinates

WASP: Water Spectral Pattern

1. INTRODUCTION

Food powdering, a process that consists of converting food items into their powdered form, has been performed throughout history, spanning various cultures and corners of the globe. Delving into the historical contexts behind this now-considered essential unit operation entails an understanding of the drivers of powdering.

From ancient civilizations using mortar and pestle to grind spices for preservation and flavor enhancement, to the Industrial Revolution's mass production of powdered milk for convenience, the evolution of food powdering mirrors the development of human culinary ingenuity. Not only did the practical needs drive the adoption of powdered foods but the cultural and economic forces have also shaped their prominence in global cuisines.

The modern era of food digitization is no different, with phenomena such as the incessant globalization of the food trade and the rapid-paced lifestyle leading to notable changes in food consumption. For instance, preferences like pivoting towards convenient over traditional home-based food and the surge of ready to eat (RTE) meals and food preparations, suited for consumers' hectic schedules, show no sign of diminishing. However, this current of culinary trends raises a set of challenges pertaining to the safety and quality of marketed products. More than ever, consumers now have access to a wide array of versatile powdered food items, including but not restricted to dried herbs for seasoning, finely ground coffee for instant brewing, and tomato powders for flavoring. Warranting the integrity of these marketed goods constitutes a shared challenge for all stakeholders within the food chain. A challenge that doesn't get any easier with the upward trend in technology enabled delivery and the worldwide sourcing of ingredients.

As a means of preservation against deterioration, powdering post-drying can have, for example, despite its positive effect on incrementing the shelf life of herbs, an adverse effect on the inherent bioactive compounds, depending on how sophisticated the drying process is.

Degradation of relevant quality attributes proven to dictate the marketability of powdered products is also prevalent, mainly due to changes in storage conditions. These changes encompass among many others, discolorations, disparities of the particle size, etc.

Besides qualitative degradation of powdered products, food adulteration is another matter of concern, devaluing both branded and non-branded food items. As a practice that keeps gaining notoriety, fraud poses undue risks to consumers and undermines the economy.

While susceptibility to fraud is reliant upon factors such as the stringency of safety regulations, the effectiveness of traceability systems, and the vulnerability of food chains, the likelihood of being targeted is, on the other hand, inextricably linked to the economic value of the product, with a prime focus on the price-defining attributes of food matrices.

Benefiting from such practices, fraudsters are now tailoring the products to the liking of their consumers but are also capable of reproducing novel unauthentic products that emulate the original ones. Regrettably, in these attempts to uphold the desired quality, more refined adulteration methods are emerging, rendering the prevention of these malignant acts a not so trivial task.

Today, perpetual fraud outbreaks and constant recalls of non-genuine products not only keep reviving authenticity concerns but also prompt the question on whether another fraud scandal would occur, and when?

To keep pace with those perpetrating food while guaranteeing prime quality, the deployment of high throughput instruments and the constant betterment of existing control systems are prerequisites. Continuous efforts by researchers across disciplines acts also as an active link in mitigating any shortcomings.

Presently used techniques encompass among many others, spectroscopic, chromatographic, DNA-based methods, etc. The rather complex nature of these conventional methods impedes their adoption as real time detection tools. Picking one or the other is usually dictated either by technological aspects or by attributes inherent to the studied food matrices.

While chromatographic techniques are challenged with the inherently different polarities of the varying constituents in food matrices. DNA-based methods, on the other hand, demand a high level of technicality that hinders their usage as routine measures. Sensory based analysis is, to an extent, reliant on the subjective assessment of panelists, having disparate sensitivities and even preferences to the studied food matrix. Colorimetry, another traditionally adopted method, can be misleading since color variations can depend on the degradation rate of differing carotenoid fractions along the drying process.

Of the mentioned techniques, NIRS has been at the forefront of adopted tools, namely since it can be less energy-intensive while still offering a greater separation of the overtone and combination bands compared with the fundamental bands. Its potential has been ascertained through applications across the food chain. Over the past years, both instrumental and software advancements facilitated the pursuit of novel research endeavors. While miniaturization led to the incorporation of near infrared spectrometers for in-situ analysis, the shift in perception with

regards to relevant spectral regions as well as the mastering of multivariate data analysis tools allowed for the analysis of new matrices for uncharted applications. In essence, such upgrades granted the opportunities to analyze highly moist and liquid agrifood products, often avoided when using conventional NIRS analysis.

In this regard, aquaphotomics, a novel NIRS based technique, offers a wealth of new possibilities. By taking a holistic approach to the analysis of water rich matrices and by harnessing water's feature of mirroring the state of the studied system through induced perturbations, the method has been adopted for a panoply of applications in the food industry (Gowen *et al.*, 2012; Atanassova *et al.*, 2016; Vanoli *et al.*, 2018; Yang *et al.*, 2020; Muncan, Tei and Tsenkova, 2021) .

This being said, little is known about the effect of other perturbations such as drying on the water structures within dried herbs. Even more scarce are the studies assessing whether these structural changes can be correlated with the alterations of phytochemicals.

Restructuring of water, under stress conditions, was reported by researchers proving that a resurrection-like behavior was exclusively manifested by Haberlea rhodopensis plants (Kuroki *et al.*, 2019). The latter drastically reduced free water, accumulated dimers and tetramers and kept other water species at a constant level.

If proven existing, such knowledge can presumably, provide a major step forward towards swifter better informed decisions, namely in terms of selecting the optimal energy-efficient drying process that offers the best preservation of the herbs` bioactive compounds.

We hypothesize that aquaphotomics could be a suitable substitute to other technically demanding authentication and process control tools.

Overall, our proposed approach is a simple screening alternative requiring neither complex extraction procedures nor the use of chemicals (organic solvents), which offers reduced sampling errors compared to conventional methods. It entails shifting from the arduous determination of multiple properties to low-cost, short-time analysis where the spectral fingerprint can suffice for the characterization of samples and can be a descriptor of both the authenticity and qualitative status of the studied matrix.

2. OBJECTIVES

Seeking to extend the scope of application of aquaphotomics in food safety and quality assessment, our primary goals consisted of:

- Determining the feasibility of applying NIRS and aquaphotomics to pinpoint the admixing of Arabica coffee with inferior quality grade Robusta, both in powder and liquid form by:
    1. Building qualitative models for the varietal and geographical discrimination of the Arabica and Robusta coffee types and the classification of their blends by their level of adulteration.
    2. Building quantitative models for the estimation of the mixing ratio of the Arabica and Robusta coffee types.
    3. Determining the water spectral pattern of the tested Arabica and Robusta coffee types and their respective aqueous blends.
- Assessing the applicability of NIRS and aquaphotomics for the detection of tomato powders adulteration by coloring and bulking agents, when presented in liquid form by:
    1. Building qualitative models for the discrimination of single adulterated (coloring agent or bulking agent) and dual adulterated (coloring agent and bulking agent) mixtures by levels of adulteration.
    2. Building quantitative models for the prediction of added adulterant(s)
    3. Deciphering any subtle changes caused by the mere addition of the adulterants on the water spectral pattern of tomato extracts.
- Investigating the adequacy of NIRS and aquaphotomics in terms of revealing the effect of preservation processes on the aqueous extracts of selected herbs by:
    1. Building qualitative models for the discrimination of the herbs according to the preservation method.
    2. Building quantitative models for the estimation of the total polyphenols content (TPC) and total antioxidant capacity (TAC) in the herbal extracts

3. LITERATURE OVERVIEW

This section of the thesis highlights major literature covering the scope of our research endeavors.

3.1. Aquaphotomics: Unveiling the drivers, terminology, and applications

3.1.1. Conventional NIR vs aquaphotomics

Early near-infrared (NIR) spectroscopic studies have been impeded by the unfavorably strong water absorptions in the infrared region hindering measurements of low-concentration food constituents (Cai *et al.*, no date; Muncan *et al.*, 2021). The low intensity and spectra resolution were magnified in the presence of aqueous samples. As per the limitations with regards to the broadness of NIR bands, they called for both instrumental and data processing advancements prior to implementation into analytical applications (Otal, Iñón and Andrade, 2003).

Commonly, circumventing the hurdle in dealing with liquid samples in spectroscopy demanded performing extra sample preparation steps (Lohumi *et al.*, 2015). Núnez-Sanchez et al., (2016), for instance, resorted to oven-drying prior to analyzing milk. This eclipsing effect on weaker bands has also been reported when attempting to determine lactate concentration (Baishya *et al.*, 2020). This time, disregarding water bands has been the method of choice of the authors to ensure an accurate estimation of lactate. Selecting regions with weaker water absorption was equally performed by Villar et al., (2012).

Sample preparations intended for non-aqueous samples tend to impose certain challenges (S. Esslinger, Riedl and Fauhl-Hassek, 2014) and can incur the loss of info, when targeted at specific constituents. The non-specificity of aquaphotomics help avoiding such occurrences.

On a much larger scale, these limitations can be heavily in disfavor of the adoption of NIR measurements in the food industry.

Alternatively, and by mere alteration of the aqueous molecular system and the examination of the resulting absorption in the NIR range, aquaphotomics based research permitted the analysis of even non-infrared active components (Gowen et al., 2015; Hirschfeld, 1985; Tsenkova, 2009).

It’s this very same feature, the high sensitivity of hydrogen bonds, that is the basis of aquaphotomics based analysis where subtle changes to the studied systems can be traced back not only at the level of the absorption bands of the existing functional groups but can be equally reflected in the absorbance bands of water (Tsenkova,Muncan, Pollner, & Kovacs, 2018). Even more so, aquaphotomics approach hypothesizes that each and every molecule can have a characteristic spectral pattern, that is easily enhanced after immersion in water when subjected to electromagnetic light (Tsenkova, 2008a).

In a sense, the advent of aquaphotomics offered a different lens on the analysis of water as an often-avoided element in spectral based analysis. The rationale for this novel approach consists in closely examining the disruption or creation of existing or novel hydrogen bonds as a response to both intentional and unintentional disturbing factors. Besides being an intricate system of varying molecular conformations, water is naturally influenced by neighboring constituents and the physical field, rendering it apt of capturing the alterations of the system (van de Kraats, Munćan and Tsenkova, 2019). This feature of mirroring the state of the sample under different perturbations was proven effective in multiple application fields.

Since aquaphotomics can serve as a promising alternative, the next section lays out its basics and terminology.

3.1.2. Aquaphotomics Basics and Terminology

Argued as being merely an inert, passive medium, reductionist views of water have been dismantled with the advent of aquaphotomics.

First introduced by Professor Tsenkova, the guiding principle behind this scientific discipline lies in water's ability to serve as a characteristic fingerprint of the state of the studied biosystem (Figure 1).

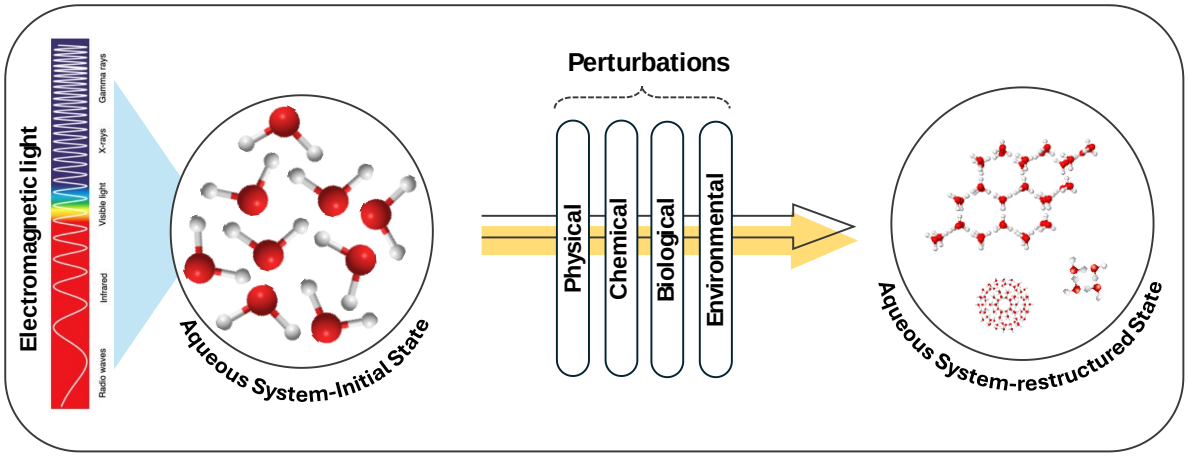

Figure 1: Water, a characteristic fingerprint of the state of biosystems

Water's active role as a medium of bioreactions and universal solvent stems from some of its peculiar properties such as its ability to form hydrogen bonds internally and with other molecules, owing to its polarity. However, disturbances exerted by water signal have long resulted in conflicting stance on the role it plays when applying conventional near infrared spectroscopy.

Additionally, the multiplicity of markers, where some studies center on specific compounds and not the other renders the idea of relying on water as a holistic descriptor all the more compelling.

The core terminology, aiding in implementing aquaphotomics to both fundamental and applied research will be laid down in this section.

This approach redefined disturbing factors, traditionally viewed as constraints (the elimination of which is prerequisite) and reframed them as essential in elucidating both basic and complex phenomena. Through aquaphotomics 'lens, these factors are no longer a nuisance but rather rich sources of information.

The methodology consists in viewing water as a multi-elemental system. Its conformations (free water, dimers, trimers etc.) are dictated both by the number and strength of hydrogen-bonds, which are typically sensitive to operational conditions and vary in the presence of different solutes. The shift from one conformation to the other usually gets reflected in the spectra. This key feature of mirroring the molecular alterations within studied systems by inspecting the interactions with surrounding water usually gets translated into what is referred to as Water Spectral Patterns WASPs (Roumiana Tsenkova, Munćan, *et al.,* 2018). These WASPs are considered integrative descriptors of all structural changes, from the subtlest to the most overt and they comprise major Water Absorbance Bands WABs.

Of quintessential importance to the remainder of analysis is the identification of these major WABs. Literature compiled an extensive database of bands, the majority of which located at 12 spectral regions referred to as water matrix coordinates (WAMACS) (Roumiana Tsenkova, Muncan, *et al.,* 2018a). Finding the correspondence between these WAMACS and specific water species derived from both empirical and theoretical studies (Headrick *et al.,* 2005; Chatani *et al.,* 2014; van de Kraats, Munćan and Tsenkova, 2019).

Being a spectroscopy- based method, the choice of the electromagnetic range that would most likely unravel the WASPs of the studied biosystems is crucial. Former research demonstrated the utility of the 1st overtone range for the purpose while more recent studies have explored and corroborated the potential of basing the analysis on the 2nd overtone range of water(Smith *et al.,* 2005).

C1, covering the 1336 to 1348nm range, was proven to correspond to the asymmetric stretching vibration (ν3 H2O). Water solvation shells [OH –(H2O)n] are the structures assigned to C2 (1360-1366nm) and can comprise 1, 2 or 4 molecules in hydroxylated water cluster. Attributed to the third coordinate C3 and covering the 1370-1376nm range are the ν1+ν3 symmetrical stretching fundamental vibration and H2O asymmetric stretching vibration. In the spectral range 1380 to 1388nm (C4), activated bands would primarily consist of water solvation shell and superoxide hydrates [O2 (H2O)n (n = 4)]. At C5 (1398-1418nm), there is mostly free water and OH^- molecules

as well as trapped water. Besides H-OH band, molecules implied in hydration dominate the C6 range, extending from 1421 to 1430nm. Spectral regions [1432-1444nm; 1458-1468nm; 1472-1482nm and 1482-1495nm], corresponding respectively to C7, C9, C10, and C11 designate water molecules comprising one, two, three, and four hydrogen bonds. As for the 8th coordinate region, 1448-1454nm, it is characterized by the abundance of ν2 + ν3, Water solvation shell, $OH\text{-}(H2O)_{4,5}$ whereas the WAMAC C12 [1506-1516nm] would typically include water species in the form of ν1, ν2, symmetrical stretching fundamental vibration, and doubly degenerate bending fundamental or strongly bound water.

These transient water states, created under certain perturbations through the disruption of existing or the creation of novel hydrogen bonds typically get reflected in the resultant water spectral pattern. Graphical representations of these patterns are termed aquagrams. Not only do aquagrams give insights about the structure-function duality but allow for simplified comparisons of the effect of forced or spontaneous changes in the system.

To ensure standardized representations of the water spectral patterns, a normalization of the absorbance is computed at relevant WABs, according to the following equation (1):

$$A'_{\lambda} = \frac{(A_{\lambda} - \mu_{\lambda})}{\sigma_{\lambda}} \quad (1)$$

While A_{λ} stands for the scatter corrected absorbance value, μ_{λ} is obtained by averaging all the spectra and σ_{λ} refers to the standard deviation of all the spectra at the perturbation-activated wavelength (λ).

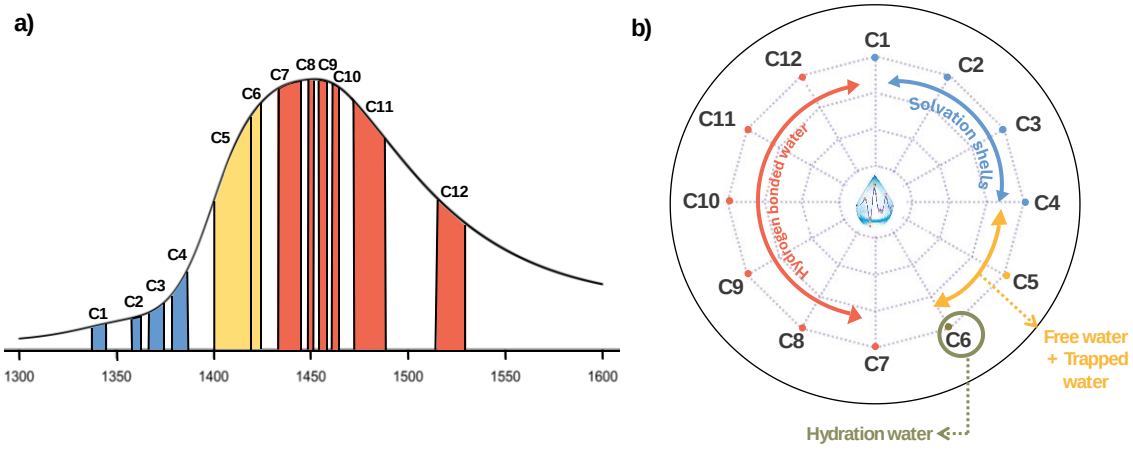

Figure 2: a) Water matrix absorbance coordinates (WAMACS) in the 1300–1600 nm range. b) Aquagrams: graphical representation of water spectral pattern (WASPs)

3.1.3. Applications of aquaphotomics

Early aquaphotomics studies shed the light on different phenomena. Theorizing that each biomolecule, when in solution is characterized by a peculiar water spectral pattern served as a tool of qualitative analysis. Tsenkova, (2008) showed that from the activated WAMACS, one can derive pivotal information regarding the different functionalities of the protein isoforms. Moreso, spectral fingerprints of low concentration solutions of serum albumin (HSA) And γ-globulin yielded high correlation coefficients R, low square errors of prediction (SEP) and good ratios of performance to deviation (RPD). Interestingly, as the studied proteins dissolved in water, the corresponding PC loadings showcased the highest variations at water bands, which in turn were consistent with prominent WAMACs from previous studies. Comparing absorbance bands of these powdered proteins to their aqueous counterparts proved that the interaction with water resulted in the appearance of other peaks that those assigned to serum albumin (1410nm) And γ-globulin (1406nm), in particular, it was confirmed that besides activating common water bands (1408 nm), the presence of each of these proteins resulted in an activation of other more specific ones (1398 nm and 1414 nm in case of γ-globulin).

Cattaneo et al., (2016), for instance, employed aquaphotomics to investigate the impact coating materials have on both cheese and melon and revealed that, for the same sample set, peculiar water spectral fingerprints were obtained by differing the coating material which is demonstrative of their distinct effect on the coated matrix (Cattaneo *et al.,* 2016). The authors also proved that the aquagrams can provide an overview of the evolution of the storage.

In another study, Li *et al.,* (2019) managed to determine minimal concentrations of methanol during fermentation, which according to their findings, have had an effect on hydrogen bonds and covalent OH (Li *et al.,* 2019) and hence precluding the need to rely on NIRS as a secondary method.

In their attempt to find a viable alternative to screen for probiotic strains in lieu of conventional phenotypic and genetic methods, Kovacs *et al.,* (2019). also emphasized the potential of aquaphotomics and NIRS (Kovacs *et al.,* 2019).

Estimating the content of soluble solids in pure apple juice, irrespective of the temperature variations is another example of aquaphotomics' effectiveness (Kaur, Künnemeyer and McGlone, 2020). The researchers pinpointed the water bands affected by changes in T° and SSC and found that they belong to the first overtone, specifically the 12 WAMACS.

A comparative study conducted by Izutsu *et al.*,(2006) and aimed at studying protein secondary structure under different physical states demonstrated the existence of bands situated at the same

regions (combination and first overtone ranges) for both aqueous and freeze-dried samples. Worth noting is the fact that freeze drying resulted in changes in α-helix and β-sheet structures. If anything these findings corroborate the adequacy of resorting to non-destructive approaches instead of arduous structure-altering freeze drying processes (Izutsu *et al.*, 2006).

Instances where the analysis of liquid samples gave better results than their non-liquid counterparts are manifold. Mamani-linares, Gallo and Alomar, (2012), for example, deployed Vis-NIR transflectance-based spectroscopy to discriminate between different meat sources (beef, llama, and horse) and have proven that the differentiation based on the corresponding juices were slightly better with 89% accurate identification of the different samples.

By studying the exerted effect of salts on water structures, Gowen, Amigo and Tsenkova, (2013) demonstrated that the technique can be even consolidated for studies where non-NIR active solutes are involved. Findings from the same study presented an interesting feature where one can relate changes in the ionic strength of considered solutions to changes in temperature. Such insights further enrich existing and traditional knowledge of structure makers or structure breakers through a visual display of their effect on corresponding aquagrams. This can be particularly beneficial in terms of constructing transferable models without the need for experimental repetitions in varying temperatures.

Other aquaphotomics based applications involved automated quality control through tracking of cheese ripening (Atanassova *et al.*, 2016), offered the opportunity for on-line real-time surveillance of yoghurt fermentation (Muncan, Tei and Tsenkova, 2021). In this later case, the observed spectral changes echoed those of chemical and physical processes occurring during fermentation.

Studies where aquaphotomics served as an authentication tool also exist, namely in case of honey adulteration with high-fructose corn syrup (HFCS). The water spectral fingerprints of analyzed Robinia honeys differed from those of HFCS (Bázár *et al.*, 2016) and the extent to which the honeys were adulterated could be estimated. Fraud detection studies also comprised tracking the tampering of tomato paste with common adulterants such as ground paprika seed, corn starch, sugar and salt (Bodor *et al.*, 2019).

Collectively, these studies iterated how shifting from reductionist approaches, which rely on the arduous measurement of specific attributes and/or parameters to the sole determination of water spectral fingerprints as holistic indicators can suffice.

Conventionally, besides having most of the bands obscured by those of water, any reductions of the wavelength range were associated with the loss of further information. This novel approach accentuated the sufficiency of relying on a narrower more informative range. For instance, the first overtone of OH water stretching vibrations, extending from 1300nm to 1600nm, has been shown to carry valuable information in terms of activated bands and the corresponding water conformation.

Throughout this thesis, opting for aquaphotomics over conventional methods was based on some technical aspects such as the minimal preparation requirements as well as the avoidance of sources of systematic variation of powdered/solid samples, namely the light scattering effects and disparities in path length, all of which can translate into spectral variations (Wold *et al.*, 1998). Another addressed factor consisted in the unavailability of some food matrices in granular/powder form. The latter is particularly relevant in the context of the fast-paced lifestyle where the majority is opting for ready to eat (RTE) food. With the prevalence of such products, new challenges came to the fore and prompted a rethinking of how samples ought to be handled.

Technically speaking, aquaphotomics enabled the measurement of components in concentration ranges of ppm ($1{:}10^6$) and ppb ($1{:}10^9$), overcoming the limitations set by using conventional NIRS (Tsenkova, 2008a; Gowen, Amigo and Tsenkova, 2013; Bázár *et al.*, 2015).

Disentangling the information contained in the overlapped bands of the NIR spectra is imperative to extract the most relevant data and make more informed decisions. This is typically ensured by resorting to chemometrics, the basis of which consists in reducing the dimensionality of the experimental data by discarding sources of redundancy, fluctuations and trivial information thus generating interpretable latent variables, that otherwise would be very hard to extract.

By Awakening the "Sleeping giant", Karl Norris paved the way for the incorporation of NIRS in applications that exceeded the initially intended agricultural use (Czarnecki *et al.*, 2015). What followed was a series of advancements at the level of the hardware where add-ons comprised novel light sources and detectors as well as the surge of handheld miniaturized instruments that helped levitating the use of NIR in applications that span many fields but also permitted remote diagnosis.

A long-standing challenge to the adoption of NIRS has been the interference caused by the pronounced background signal emanating from bulk water (Yakes *et al.*, 2017). Not only did the naissance of aquaphotomics levitate the scope of NIRS, but the method also allowed a re-assessment of the potential of NIR spectroscopy for water-rich samples.

Operating on the non-destructive principle of light interaction with analytes, near infrared spectroscopy has proven its efficacy with a track record of applications all along the food industry. Nonetheless, resorting to complementary techniques that further enhance these results by analyzing more complex matrices is recommended. Some of the presented limitations was reported by (Yakes *et al.*, 2017) who stipulated that detection estimates based solely on dry-blended samples overestimate NIR detection differences in wet-blended samples due to matrix effects arising from changes in melamine hydrogen-bonding status. NIRS spectroscopy was only used to detect and quantify water in bovine milk as long as the adulteration was below the predicted error value (Poonia *et al.*, 2017).

Another challenge in deploying conventional NIR-based analysis to powdered mixtures is the cross-sensitivity exhibited due to differences in particle size, shape and even distribution of the elemental granules. In such mixtures, better prediction accuracies were obtained for the constituents with the more abundant size range (Pasikatan *et al.*, 2001). Ensuring sample uniformity, typically through extra preparation steps or resorting to spectral scatter corrections is notably a prerequisite (S Esslinger, Riedl and Fauhl-Hassek, 2014).

One of the works also reported that rewetting of the soil prior to the NIR measurements leads to great improvements in the prediction of soil organic carbon (SOC) (Stenberg, 2010). Although the explanation for the mechanism behind increased accuracy of quantification was not specifically identified, as authors reported, the WABs were involved in improved calibrations, suggesting exactly the role of water as a sensor in this phenomenon.

Throughout these studies, aquaphotomics has opened a novel venue in terms of advancing both fundamental and applied research. Nevertheless, relatively few are the studies that examined the quality of powdered agrifood products such as ground coffee, tomato powder, and dried herbs, when presented in their aqueous format.

3.2. Relevance of the selected food matrices:

One criterion behind the choice of our samples consisted of the ubiquity of water in the end-product, be it brewed coffee, tomato extracts or herbal drinks. Another crucial denominator was the economic value of each and every studied product.

3.2.1. Tomato powders

As a culinary essential across world cuisines, tomatoes year-round demands and short shelf life have rendered their processing into other marketed products an imperative industrial step. In their powdered form, they are widely incorporated in several food and beverage preparations and constitute an indispensable ingredient of convenient food (Abul-Fadl and Ghanem, 2011; Eyiler

and Oztan, 2011; Belović *et al.*, 2018; Obadina, Ibrahim and Adekoya, 2018). Regrettably, given their economic relevance, tomato powders are not exempt from adulteration.

Yet, issues surrounding their safety have been, for the most part, disregarded. Conversely, researchers' prime focus has been essentially on varietal and genotypical determinations as well as qualitative estimations of their gustative profiles.

A criterion that has long been associated with the perceived quality of powders pertains to their vivid color. The latter not only constitutes one of their price-setting factors but also a prominent fraud target(Spence, 2015). Preventing color loss and the ensuing price depreciation is heavily dependent on regulated exposure to surrounding factors (air, temperature, moisture, light, etc.) (Obadina, Ibrahim and Adekoya, 2018). Considering the intricacy of maintaining these criteria constant, the occurrence of color alterations is eminent and often implies additions of unauthorized coloring agents.

Increasing profit margins via incorporations of bulking agents is equally problematic and well-documented. More so, in the attempt to mask the deterioration of product-intrinsic quality or neutralize any impact these bulking agents might have, coloring agents end up being added to the powders (Downham and Collins, 2000; Burrows, 2009; Sciuto *et al.*, 2017).

In assessing the authenticity of tomato-based products, researchers have resorted to a range of techniques with differing levels of efficiency(Medina *et al.*, 2019; Oliveira, Cruz-Tirado and Barbin, 2019; Zaukuu *et al.*, 2019; Li *et al.*, 2020). The complexity of some conventional authentication methods (i.e. DNA profiling) and the subjectivity of others (i.e. Sensory analysis), can however hinder their adoption as routine assessment tools(Arvanitoyannis and Vaitsi, 2007).

Conventional NIRS, on the other hand, is subject to other challenges, when analysing powdered blends. These hurdles correspond to the exhibited cross-sensitivity resulting from variations in the elemental granules in terms of their particle size, shape and even distribution. In a seemingly interesting study, (Pasikatan *et al.*, 2001) demonstrated that such differences can cause disparities in the prediction accuracy, with better performances for particles with the most abundant size range while (Yakes *et al.*, 2017) proved that estimation merely based on dry-blended samples can potentially engender over-estimations, namely due to disregarding matrix effects in wet-blended samples, occurring through alterations of the hydrogen bonds.

Typically, overcoming such limitations would require additional steps both during samples preparation (i.e. ensuring uniform particle size) and data analysis (i.e. scatter corrections)(S. Esslinger, Riedl and Fauhl-Hassek, 2014).

On-going efforts to accurately determine quality parameters prompts a rethinking of what components serve best as quality indicators. Interestingly, both early and recent studies prospected the ever-expanding scope of aquaphotomics based research for such application in the food industry, hence the choice of exploring the suitability of the method for the detection of tomato powders adulteration, when presented in their aqueous format.

3.2.2. Coffee

Long known for its cultural and organoleptic value, coffee continues to be a beverage of choice for more than 60% consumers around the globe, with exports totaling 10.12 million bags in 2022 (ICO, 2022). Considering its role as a vital source of economic exchange, coffee's vulnerability to fraudulent practices is, undeniably, eminent. These practices mainly target coffee's quality-defining features.

Coffee quality is, however, seldom explicitly defined at the point of consumption; instead, it is intricately tied to a multitude of factors including the geographical origin, the coffee variety, the cultivation methods, the bean roast, the grinding process, and more (Gloess *et al.*, 2013; Bilge, 2020; Teran, 2024). As diverse as these factors are, so are the adulteration methods, ranging from fraudulent labeling to the addition of artificial flavorings and concealing bean defects, to name a few (Toci *et al.*, 2016; Čurlej *et al.*, 2021).

While adulteration forms are innumerable, we are particularly interested in intricate cases where the two coffee varieties, Arabica, and Robusta, endowed with disparate quality grades are mixed. Such occurrences, although common, can be considered as milder cases of forgery, if not specified on the label.

In quantifying the Robusta fraction in coffee blends, (Wermelinger *et al.*, 2011) resorted to Raman spectroscopy and the considered mixtures comprised 5, 10, 25, 33, 50, and 75% w/w of Robusta. Relying on the exclusive presence of kahweol in Arabica beans, a discrimination of the two coffee types, based on their lipid fractions, was achieved. The corresponding limits of detection fluctuated between 4,9 and 7,5% w/w. Using nuclear magnetic resonance (NMR) and basing their analysis on the determination of 16-O-methylcafestol (16-OMC), (Schievano *et al.*, 2014) could also authenticate coffee blends while accurately estimating concentrations as low as 0.9%. The predictive models had respective limits of detection and quantification equal to 5 and 20mg/kg. What the authors demonstrated was the efficiency of other discriminators like tocopherols, fatty acids, or sterols, all of which would however require extra extraction steps.

Cross-validated models based on electronic nose fingerprints proved effective in terms of determining Robusta added in 10, 20, 30, 40, 50, 60, 70, 80, and 90% w/w proportion to pure Arabica (Brudzewski, Osowski and Dwulit, 2012), with an average error of 0,21%.

Vibrational spectroscopy was the technique adopted by (Pizarro, Esteban-Díez and González-Sáiz, 2007) in their attempt to predict artificial tampering of Arabica with up to 60% w/w Robusta. The obtained root mean square error of prediction (RMSEP) corresponded to 0.79%. Similarly, Suhandy and Yulia, (2017) attempted to accurately quantitate added Robusta in the [0-60%] range, using a different spectral range, the UV-Visible (200–400 nm). What was accentuated was the role certain intervals play in optimizing the PLS models and enhancing the ratio of prediction to deviation up to 2.15.

The approach differed for (Spaniolas *et al.*, 2006), who examined the utility of a more sophisticated lab-on-a-chip system for discriminating the two coffee varieties, in which case the LOD amounted to 5%.

Both former and recent studies underscore the existence of multiple chemical markers, aiding in coffee authentication processes, but also adding to the complexity and technicality of applied methods (Daniel *et al.*, 2018; Song *et al.*, 2019; Forchetti and Poppi, 2020). Moreover, the vast majority of qualitative studies were performed either on coffee beans, powders, or chemically mediated extracts. Fewer studies used aqueous coffee solutions.

In this context, research targeting the effect of amalgamation of Arabica with Robusta on their respective water spectral pattern is, to the best of our knowledge, non-existing. Our proposed approach could prove beneficial namely in cases where access to beans, to chemical reagents or overly qualified personnel is limited.

3.2.3. Herbs

Offering a myriad of health benefits and rich gustative experiences, herb-based beverages are another category of products that are currently trending, namely for present-day consumers, which are becoming more health-conscious. The medicinal prominence and resultant marketability of these drinks are often dictated by their richness in bioactive compounds.

Typically, herb-drying has been the go-to approach to preserve herbs ‘unique aromas, decadent flavors and functionality(Thamkaew, Sjöholm and Galindo, 2021). By removing the moisture from the herbs, both microbial growth and spoilage could be prevented thus guaranteeing their long-term storage.

Traditionally, herb-drying was performed through exposure to air, a practice that is often reliant on sunlight and climate. However, as demands grew, so did the need for efficacy and consistency. Thus, spurred by demands for industrial efficiency and driven by technological advancements, the drying processes transitioned from conventional methods to more sophisticated ones (Nurhaslina, Andi Bacho and Mustapa, 2022).

From electric dehydrators and their ability to ensure controlled temperatures and airflow to the emergence of microwave drying and its significant effect on reducing drying time, extending to freeze-drying for moisture removal under vacuum conditions. Not only did these advancements elevate the quality of dried herbs but they also ensured their availability year-round, catering to a diverse array of culinary, medicinal, and industrial needs (Calín-Sánchez *et al.,* 2020).

Interestingly, extensive research investigating the effect of the varying drying processes on herbs showed that the former can have divergent effects depending on the latter (Abascal, Ganora and Yarnell, 2005; Siriamornpun, Kaisoon and Meeso, 2012; Hazarika and Gosztola, 2020; Alwafa *et al.*, 2022; Zhang *et al.*, 2022). In assessing such effects, both complex and chemically mediated methods have been adopted and often required a high level of technicality (Rahimmalek and Goli, 2013; ElNaker *et al.*, 2021; Pachura *et al.*, 2022; Chaves *et al.*, 2023).

Conversely, studies reporting on the use of aquaphotomics to correlate the alterations in total polyphenols and antioxidant capacity to structural changes of water are, to the best of our knowledge, unavailable. Nevertheless, existing literature regarding plants' innate adaptation to stress conditions such as drought, through different water management mechanisms, comprising among many others, a restructuring of intrinsic water (the reorganization of water on molecular level), have inspired the undertaking of this study. By applying different preservation methods, we're attempting to assess whether the water absorption bands can be indicative of these experimental and operational disparities.

Ensuring that herb-based drinks impart the desired functionality requires a thorough assessment of the how well bioactive compounds such as total antioxidants and total polyphenols are preserved. While studies addressing such concerns are manifold, the approach we propose tackles a novel aspect, pertaining to the study of preservation induced alterations of the water spectral pattern happening in parallel to changes in the total antioxidant capacity (TAC) and total polyphenols content (TPC).

Aquaphotomics reportedly offers holistic compositional overview of varying analytes but with little focus on low concentration mixtures of Robusta-to-Arabica coffee, on tomato extracts comprising both bulking and coloring agents and on dried herbs extracts.

Tapping into the full potential of aquaphotomics for the qualitative assessment of these food matrices necessitates the use of sophisticated chemometric tools, which will be the focus of the subsequent chapter.

3.3. Chemometrics:

Featuring mainly broad and overlapping bands that correspond to vibrational overtones and combination modes, near infrared spectral data typically poses major analytical challenges, if not processed with chemometrics (Danezis *et al.*, 2016).

Resorting to the set of tools offered by chemometrics (Figure 3) has been shown to aid in modulating the multivariate data and extracting relevant information that would otherwise be hard to discern (Bu, 2007; Trygg and Holmes, 2007; Moore, Spink and Lipp, 2012).

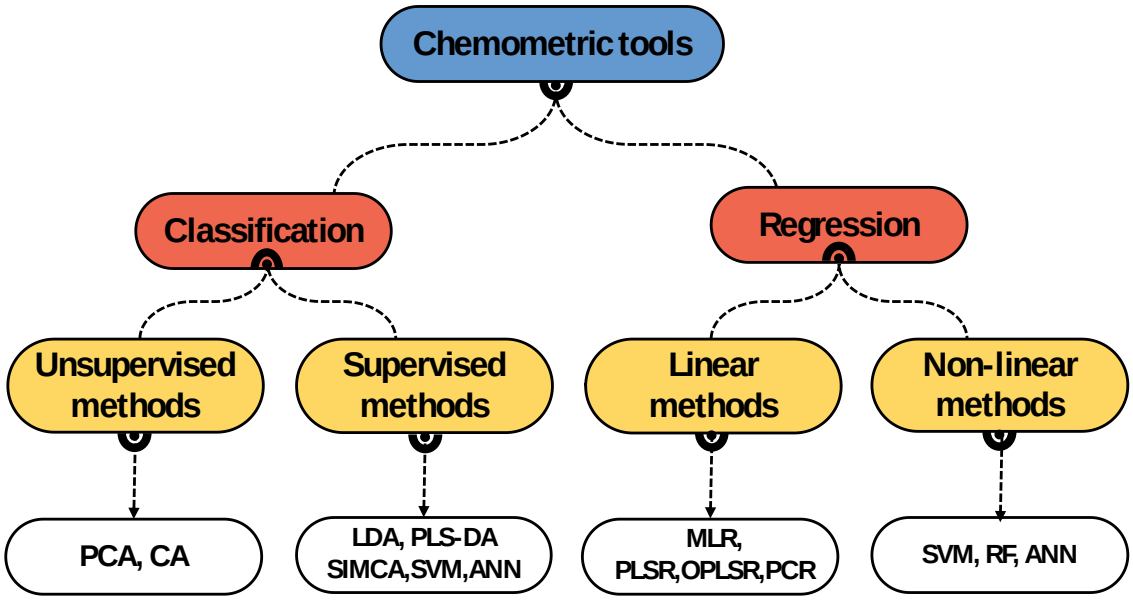

Figure 3: Basic chemometric tools utilized for multivariate spectral data analysis

By definition, this discipline "utilizes mathematical and statistical modelling to recognize patterns and relationships within highly complex data and translates them into useable analytical parameters"(McGrath *et al.*, 2018).

Typically, after spectral inspection for visibly apparent differences between the studied samples, data compression is performed. This is done by applying a pattern recognition method called principal component analysis (PCA). This non-supervised exploratory tool retrieves fewer

uncorrelated variables referred to as principal components (PCs) by linearly combining the initially multidimensional uncorrelated variables into orthogonal ones, all while preserving data variance. Essentially, PCA displays systematic variation in the data and gives an overview of any existing trends. The choice of the optimal PC number is primordial to the subsequent analysis in order to prevent any underfitting or overfitting of the data, causing either lack of information or the inclusion of systematic noise, respectively. Optimizing the accuracy of subsequent models also demands executing an outlier detection, often performed during PCA (Jollife and Cadima, 2016).

Based on the former PCs, serving as input for the supervised linear discriminant analysis (LDA) categorization, class memberships of the samples are determined. In a sense, LDA tends to maximize the ratio of the between-class to within-class variance by finding the linear combination of features that would best separate the different classes in a given dataset. Assessing the discriminatory ability of the model often requires data partition into two subsets, one for model training and the other for model testing. This subsampling of the data and its subsequent effect on the robustness of the model has been thoroughly described by (Berrar, 2018).

Partial least squares regression (PLSR) is then performed for the quantitative determination of substances of interest. This basic chemometrics tool operates by relating two data matrices through the creation of orthogonal latent vectors that maximize their covariance (Wold, Sjöström and Eriksson, 2001).

Deciding upon the model ensuring the most accurate prediction would require the evaluation of metrics such as the coefficients of determination (R^2) as well as the Root Mean Square Errors (RMSE) of calibration, cross-validation, and prediction. These metrics assess the fit of the tested data to the regression line while estimating the difference between predicted and actual values. The closer R^2 is to 1 and the lower RMSE is, the more accurate the model (Naes *et al.*, 2002).

4. MATERIALS AND METHODS

Structured into three parts, the following section outlines the sample preparation procedure, along with the experimental design and multivariate statistical methods used for the analysis of tomato, coffee, and herbs samples.

Two crucial starting steps consisted in ensuring representative sampling and adequate sample preparation. To this end, standardizing the measurement was instrumental to minimize unwanted and external sources of data variability. For an equal dispersion of the elemental particles of our mixtures, their homogenization was warranted. To account for any drift due to varying temperature and humidity changes, a datalogger was used.

4.1. Aquaphotomics for detecting adulteration in tomato extracts

4.1.1. Samples acquisition and preparation of tomato extracts

In this study, the tomato powders were provided by local Ghanaian markets and aseptically transported to the laboratory. To account for potential variability stemming from the different cultivars, three tomato varieties were considered and consisted of Navrongo (NA), Tytanium (TY), and Tuobodom (TU). The food coloring agent (FC) comprised of a binary mixture of Ponceau 4R (E124) and Tartrazine (E102) and was purchased from Dainess foods (Accra, Ghana), the other two adulterants were annatto powder (AP) and corn flour (CF), both procured from the local market and processed at the food processing unit, Department of food science and technology, at the Kwame Nkrumah University of Science & Technology, Ghana.

Subsequent handling and analysis of the samples was conducted at the department of food measurement and process control, at the Hungarian University of Agriculture and Life Sciences.

Mixtures comprised both single (AP, CF, or FC) and dual (AP+FC or CF+FC) adulterants. The latter were added in concentrations of 0,5%, 1%, 2%, 5%, 10%, 15% up to 20% w/w with the final sample weight totaling 5g per sample. Besides authentic samples (0%) and pure adulterants (100%), the prepared mixtures were solubilized. The 10 times dilution was performed by transferring 2.5 g of each sample into 25mL volumetric flasks and filling up to the mark with ambient temperature MilliQ water. Upon thorough homogenization, syringe-filtration through a 125 mm diameter filter paper (Macherey-Nagel, Germany) was ensured. Following extraction, these aqueous solutions were hermetically stored in the freezer (-20°C) prior to analysis (Figure 4).

Considering three replicates per each sample type, a total of 333 aqueous extracts were obtained. To facilitate their identification, barcode system labelling was employed.

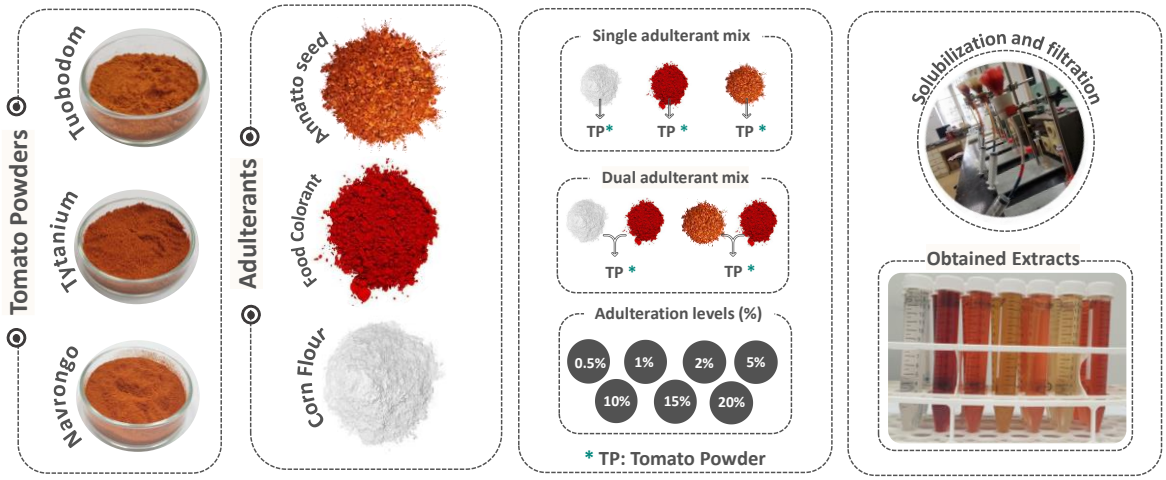

Figure 4: Authentic tomato powders, adulterants, their combinations, and aqueous extracts

4.1.2. Spectral acquisition of tomato extracts

While ensuring the randomization of scanning order, the spectra of the tomato extracts was obtained with a benchtop Rapid Liquid Analyzer (XDS-RLA, Metrohm, Herisau, Switzerland) operating in transmission mode and covering the range 400-2500 nm. A 1mm pathlength cuvette was used whereas the spectral data interval and the wavelength accuracy consisted of 0.5nm and 0.05nm respectively.

A total of 32 successive scans for each recorded spectrum were collected then averaged for each sample. This was performed three times to obtain three consecutive scans per sample. Between measurements, the cuvette was rinsed with MilliQ water to avoid cross contamination.

4.1.3. Signal preprocessing and modeling of tomato extracts

To minimize spectral variability, arising from factors other than compositional disparities between the samples, a combination of preprocessing tools was adopted. Smoothing (Savitzky Golay filter (2nd order polynomial) with adjustable window width (17, 19, 21, 23, 27, 29 points)), was performed singularly or jointly with multiplicative scatter correction (MSC), detrend (DeTr), standard normal variate (SNV), 1st or 2nd derivative (Barnes et al., 1989; Savitzky & Golay, 1964).

R software v 4.0.2, specifically aquap2 package (Kovacs & Pollner, 2016), was used for both spectral pretreatment and multivariate data analysis.

By focusing essentially on water's 1st and 2nd overtone ranges, the NIR-obtained spectra were subjected to chemometric modelling. Initially, principal component analysis (PCA) served for pattern recognition, both amongst authentic tomato extracts and pure adulterants but also considering separate groups of single and dual adulterated extracts. A hybrid (PCA-LDA) was

used for classification according to the different adulteration levels (0,5%, 1%, 2%, 5%, 10%, 15% up to 20% w/w). For each tomato variety, separate models were considered for single-adulterated samples, tampered with annatto powder (AP), corn flour (CF) or food coloring agent (FC) but also for dual adulterated samples, which comprised the (AP+FC) or (CF+FC) mixtures. The optimal number of PCs was selected based on the ability to yield the best validation accuracy while minimizing any differences between the training and validation accuracies. Partial least squares regression (PLSR) was performed to estimate the concentration of added adulterants.

PCA-LDA and PLSR models were respectively validated by three-fold and leaving one group out cross-validations. The samples left out during PLS modelling spanned various groupings, considering criteria such as replicates, the sample types, the adulteration levels as well as the type of adulterants, all conducted according to the methodology outlined in (Berrar, 2018).

To examine how reliable the predictive PLSR models were, metrics such as the coefficient of determination (R2) and the root mean squared errors (RMSE) were computed, both during the calibration and cross-validation stages.

A quintessential step consisted in visualizing the water spectral patterns portrayed by the aquagrams. Two specific ranges, corresponding to water's first overtone (1300–1600 nm) and second overtone (800-1100 nm) were considered, owing to their demonstrated relevance in terms of highlighting peculiar water conformations induced by systematic perturbations.

4.2. Aquaphotomics for coffee quality grade determination

4.2.1. Samples acquisition and preparation of coffee blends

For the purpose of our study, Arabica beans originating from Brazil (ARA1), Columbia (ARA2), and Ethiopia (ARA3) and Robusta beans sourced from Vietnam (ROB1), Uganda (ROB2), and India (ROB3) were French-roasted, ground, and procured by Bourbon café (Tahitótfalu, Hungary). Mixtures comprising 0.5, 1, 2, 3, 5, 10, 20, 35% w/w Robusta were prepared by pairing ARA3 and ROB3. The threshold above which the addition of Robusta can have a palpable effect on coffee aroma is 35% (Assis, Oliveira and Sena, 2018). Alongside these mixtures, marketed blends of different provenance and serving as test samples were considered (Figure *5*). The 1st blend B10% consisted of 90% Arabica (South American) and 10% Robusta (South-east Asian), whereas the 2nd blend B30% comprised 70% Arabica (Central and South American) and 30% Robusta (Southeast Asian). Triplicate samples were prepared for each of the mixture concentration levels.

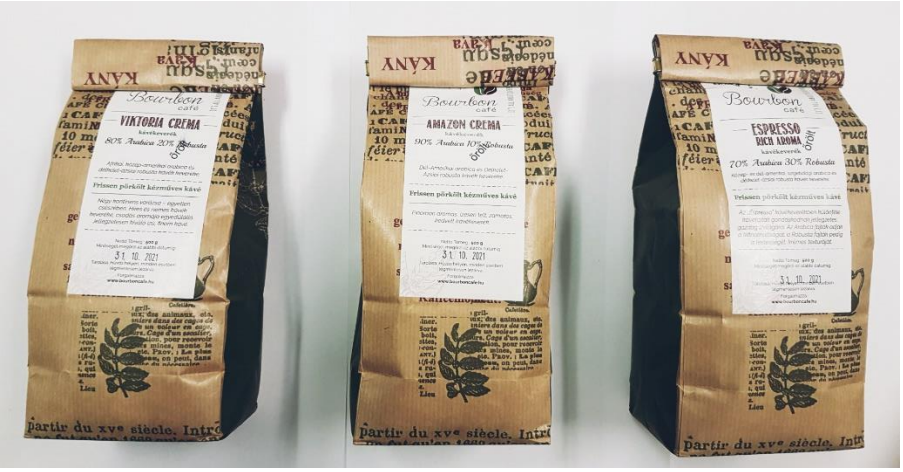

Figure 5: Marketed coffee blends serving as test samples

After formulating the mixtures, water extracts of the pure coffee varieties, the mixtures, as well as the marketed blends were prepared by pouring 100 mL Milli-Q water, heated at boiling point, onto 8 g of coffee. After five minutes, the samples were filtered using a 25 µm pore-sized quantitative filter Whatman paper. The obtained extracts were cooled to room temperature (25 °C) prior to analysis.

4.2.2. Spectral acquisition of coffee blends

A benchtop MetriNIR Spectrometer (MetriNIR Research, Development and Service Co., Budapest, Hungary) was used to collect the spectral data in the wavelength range of 740–1700 nm. For a more representative spectra of each of the ground coffee mixtures, the cuvette was rotated between the three consecutive scans of each sample type during scanning.

In the case of the aqueous samples, a thermoregulated cuvette with a sample layer thickness of 0.5 mm was used to maintain the temperature of the samples at 25 °C. The cuvette was thoroughly washed with Milli-Q water between measurements and dried for the next sample. A total of 324 spectra, made up of three consecutive scans of the three refills of the triplicates of each mixture, were acquired. For reference data, Milli-Q water spectra were taken after every 5th sample measurement.

Two modes of spectra acquisition were adopted, diffuse reflectance mode in the case of powders and transflectance mode for liquids.

4.2.3. Signal preprocessing and modeling of coffee blends

Truncating the instrument's spectral range to 800-1670nm was performed, to minimize the spectral noise. Two other spectral ranges consisting of the 1st (1300-1600nm) and 2nd (800-1100nm) overtone ranges were selected, owing to their relevance in aquaphotomics based research.

After inspecting the raw spectra for any apparent disparities amongst the samples, spectral pretreatment was conducted and its effect on the obtained results was evaluated. An initial step consisted in smoothing the spectra through fitting the points into a 2^{nd} polynomial, using Savitzky Golay filter, while also testing several window widths (11, 17, 19points). Smoothing was equally combined with other preprocessing tools such as detrend (DeTr), standard normal variate (SNV), multiplicative scatter correction (MSC) and derivatives (1^{st} or 2^{nd}).

An overview of systematic trends and data compression was ensured by applying principal component analysis (PCA). Two types of PCA models were developed, the first used the combination of coffee variety (Arabica vs. Robusta) and geographical origin as the class variable when assessing only authentic samples while the second took into consideration the mixing ratio as class variable, in the case of blends.

These same class variables were adopted for the determination of the class memberships of the samples by PCA-LDA. To examine how accurate the classification of pure Arabica and Robusta, their resulting mixtures at different blending ratios, as well as the marketed blends was, the PCA-LDA models were 3-fold cross-validated.

Subsequently, to relate the NIR spectrum to the added Robusta, partial least squares regression modelling (PLSR) was performed. Different validation methods were considered before deciding upon the model that ensures the most accurate prediction of the Robusta-to- Arabica ratio.

Next, the PLS models were assessed by computing metrics such as the measurement of fit (R2) and the root mean square errors of calibration and cross-validation (RMSEC and RMSECV). These values give an overview of the fit of the tested data to the regression line while estimating the difference between predicted and actual values.

To validate the PLS models, cross-validations of varying robustness were evaluated. The less robust being leave one sample out validation while the cross-validation based on the grouping defined by specific class variables (by repeats, by sample type, etc.) corresponded to the more robust one. Testing the accuracy of the predictive models was ensured using the two commercialized blends B10% and B30%.

To display the water spectral pattern of the aqueous pure coffees and their corresponding blends, aquagrams were examined at both the 1^{st} and 2^{nd} overtone ranges, more specifically at the selected water matrix coordinates (WAMACs) of 1342, 1364, 1374, 1384, 1412, 1426, 1440, 1452, 1462, 1476, 1488, and 1512 nm (1st overtone) and at 890, 908, 924, 946, 954, 975, 1001, 1019, 1036, 1044, and 1060 nm (2^{nd} overtone). These wavebands, according to literature, are ascribed to

peculiar water conformations and were proven to be highly practical in terms of highlighting how well determined perturbations affect systems.

Throughout the analysis, and for each type of sample, averaging of repeats, refills and consecutive scans was attempted, to reduce sources of variability.

4.3. Aquaphotomics for tracking herbs preservation processes

4.3.1. Samples acquisition and preparation of preserved herbs and their extracts

For this experiment, plant stands were established in Soroksar (Pest County, Hungary) in the research field belonging to of the Medicinal and Aromatic Plants Department.

The examination of fresh samples was conducted within three hours post-harvest, until then, the plant materials were refrigerated.

An initial preliminary study involving three herbs consisting of thyme (*Thymus vulgaris*), marjoram (*Origanum majorana*) and rosemary (*Rosmarinus officinalis*) was considered. A total of nine preservation methods were applied, ranging from traditional (sun-drying, shade-drying) to more sophisticated ones such as oven-drying (40°C and 60°C), lyophilization, microwave drying (250 W and 700 W), slow and fast freezing. The duration of each of these methods is indicated in

Table *1*.

Table 1: Duration of the preservation processes for the 1st set of herbs (h= hours, min=minutes, d=days)

	Duration					
	Sun (d)	Shade (d)	Oven-drying 40°C(h)	Oven-drying 60°C(h)	Microwave 250W(min)	Microwave 700W(min)
Rosemary	2	5	21	6	19	7
Marjoram	5	8	20	5	15	6
Thyme	5	8	21	5	20	7

The second batch of herbs comprised peppermint (*Mentha × piperita*), lovage (*Levisticum officinale*), lemon balm (*Melissa officinalis*), Greek oregano (*Origanum vulgare*), tarragon (*Artemisia dracunculus*), garden sage (*Salvia officinalis*), summer savory (*Satureja hortensis*) and sweet basil (*Ocimum basilicum*). This selection of herbs underwent the following preservation methods (lyophilization, oven-drying (40°C and 60°C), microwave drying (250 W and 700 W), and slow freezing) conducted for the duration indicated in Table *2*.

Table 2: Duration of the preservation processes for the 2nd set of herbs (h= hours, min=minutes, d=days)

	Duration			
	Oven-drying 40°C (h)	Oven-drying 60°C (h)	Microwave 250W(min)	Microwave 700W (min)
Peppermint	20	3	15	6
Greek oregano	15	3	16	6
Lemonbalm	15	3	12	4
Lovage	18	4	18	7
Basil	30	6	15	6
Garden sage	30	5	15	6
Summer savory	20	4	15	6
Tarragon	25	5	15	6

In case of sun-drying, leaves were positioned outside in a closely latticed compartment while being exposed to direct sunlight during the day before being transferred to a sheltered room at night. A datalogger (RHT 10, Extech Instruments, Nashua, USA) captured ambient air temperature, which, in average, ranged between 40 to 45°C during the day and 16 to 20°C at night.

When subjected to shade drying, the leaves were placed in a dark, sun-shielded room with adequate ventilation, where the daytime temperature averaged 20°C, with nighttime temperatures falling between 10-15°C.

To undergo microwave drying, the leaves were put in a 20L sized microwave oven (SMW 1917WH, Sencor, Opava, Czech Republic) and dried both at 250W and 700W, with constant ventilation every 3rd and 1st minute, respectively. Once ready, the fully dried leaves were kept in paper bags, at room temperature and away from sunlight, heat, and moisture fluctuations, till further analysis.

As per the lyophilization, it was performed using a lyophilizer (Scan Vac Cool Safe, LaboGene, Lillerød, Denmark) where samples were initially quickly frozen at -80° C before undergoing freeze-drying at -109°C. Lyophilized samples were later stored in polyethylene bags at 4°C.

Slow and fast freezing were conducted respectively in a 230L domestic freezer (ZRAN230FW, Zanussi-Electrolux, Stockholm, Sweden) and a Blizzard Ultra Low Temperature Freezer (NU-

99828J, NuAire, Inc., Plymouth, USA). In the former, fresh packaged leaves were kept at -18°C whereas in the latter, they were rapidly frozen at -80°C before being stored at -18°C till subsequent analysis, within 1 month of storage.

After preservation, all herbs were pulverized using a coffee grinder (Bosch TSM6A011W, München, Germany) and sieved through a stainless-steel sieve (VWR collection) having a 200mm diameter and 500µm aperture width, except for the frozen ones, which were instead finely chopped.

To obtain the aqueous herbal extracts, 100ml distilled and boiling water was poured over 1g of the previously ground plant material. After being soaked for 24 hours, the extracts were filtered through a 125 mm diameter filter paper (Macherey-Nagel, Germany) and stored (in the freezer) till further analysis.

4.3.2. Reference measurements on preserved herbs and their extracts

- Determination of total phenolic content (TPC)

Following the protocol outlined by (Singleton and Joseph A. Rossi, 1965), the total phenolic content (TPC) of the aqueous herb extracts was determined by measuring the absorbance at 760nm using a Thermo Evolution spectrophotometer (201, Labomed Inc., Los Angeles, USA). The incubation in hot water (50°C) lasted for 5min. As per the standard curve, it was obtained using gallic acid (0.3M) and was characterized by the following slope-intercept equation (2):

$$y = 0.1243x - 0.0216 \quad (2)$$

with an R^2 of 0.9998. The total polyphenol content was expressed in mg of gallic acid equivalent (GAE) per g of dry weight (d. w.) of the aqueous extract. The measurement was performed three times for each of the three replicates resulting in 9 measurements per sample type.

- Determination of total antioxidant capacity (TAC)

The total antioxidant capacity (TAC) was measured employing the methodology delineated by (Benzie and Strain, 1996), by determining ferric reducing antioxidant power (FRAP) via a measurement of the absorbance at 596nm. A Thermo Evolution 201 spectrophotometer was also used for the purpose. The calibration curve equation (3), yielding an R^2 of 0.9996 is:

$$y = 0.2527x - 0.0166 \quad (3)$$

TAC was expressed in mg ascorbic acid equivalent (AAE) per g of dry weight (d.w.) of the aqueous extract. Again, for each triplicate, the measurement was conducted three times generating 9 measurements per sample type.

- Determination of the dry matter content:

The dry matter content of the obtained herbal extracts was determined using gravimetric method. 20ml of the extract is placed in a crucible that has been previously weighed to determine its initial mass, the sample is then subjected to a heating to remove moisture and volatile components. After the heating process is complete and the sample has been allowed to cool in a desiccator to avoid moisture uptake from the atmosphere, the crucible with the dried sample is weighed again to determine the final mass.

4.3.3. Spectral acquisition of preserved herbs and their extracts

The ground herbs were scanned in reflectance mode using the benchtop XDS Rapid Content Analyzer (XDS-RCA, Metrohm, Herisau, Switzerland). On the other hand, the transmission-based module of the XDS NIR spectrometer (XDS-RLA, Metrohm, Herisau, Switzerland) was used in case of aqueous herbal extracts. Covering the spectral range 400-2500nm, both modules register 4200 data points per spectrum, with an interval of 0.5nm. Overall, 30 co-averaged spectra were collected per sample.

To operate the spectrometer, VISION 2.51 software was used, and the obtained spectra were exported as NSAS format files to R studio (v. 3.3.2, The R Foundation for Statistical Computing, Vienna, Austria). Subsequently, aquap2 package (Kovacs & Pollner, 2016) was applied for data processing.

For each sample type, three replicates were prepared resulting in a total of 81 samples (3 herbs × 9 preservation methods × 3 repeats) for the first set of herbs and 168 samples (8 herbs × 7 preservation methods × 3 repeats) for the second.

Serving as an environmental control, Milli-Q water was scanned after every 10^{th} aqueous sample. In case of ground herbs, a reference measurement was also ensured. For uniformity of the powdered herbs in the cuvette, the latter was gently tapped three times against a laboratory work bench before spectral data acquisition. Three consecutive scans were acquired for each sample, including the references.

To ensure the minimization of any potential systematic bias, the randomization of the scanning order was performed. While the scanning was done at room temperature, constant monitoring of

the temperature and relative humidity was performed using a Voltcraft DL-121TH Multi-Data logger.

4.3.4. Signal preprocessing and modeling of preserved herbs and their extracts

For each of the conducted studies, a set of chemometric tools, better aligned to the respective datasets, was applied. Herein, we cover how the data was pretreated in an attempt to optimize the qualitative and quantitative models.

Inspecting the obtained spectra was a primordial first step in order to select the wavelength range better suited for the subsequent analysis. Essentially, we focalized on the short ranges 800-1100nmand 1300-1600nm commonly adopted in case of aquaphotomics based analysis. Detecting the outliers was equally performed to discard data points not falling within the expected data distribution.

After discarding irrelevant and noise filled spectral ranges, a selection of spectral pretreatments was applied and their effect on the generated results was investigated. Savitzky-Golay smoothing ensured the fitting of the spectral points into a 2^{nd} polynomial within varied window widths (21, 23, 27, 29, 31, 33). Smoothing was conducted both individually and in combination with other preprocessing tools such as 1^{st} derivative, to resolve any overlapping peaks.

Principal component analysis (PCA) was conducted for the visualization of any pattern depending on the applied preservation method. Next, multi-class categorization was performed using hybrid principal component analysis-linear discriminant analysis (PCA-LDA) models. The three-fold cross validated models were examined in terms of their ability to correctly assign the herbal aqueous extracts to their respective groups, according to the class variable defined by the drying method.

Eventually, PLSR models were built to estimate the content of total polyphenols (TPC) and total antioxidant capacity (TAC). Leave one group out cross-validation based on the grouping defined by specific class variables (repeats, consecutive scans) were considered before deciding upon the model that ensures the most accurate prediction that relates the reference measurements of TPC and TAC to those predicted based on the NIR generated data. The contributing wavelengths were inspected, and their assignments were determined.

Aquagrams, representative tools of the water spectral pattern, were investigated with regards to their ability to highlight the effect of preservation methods on the herbal extracts.

5. RESULTS AND DISCUSSION

5.1. Aquaphotomics for detecting adulteration in tomato extracts

5.1.1. Results of the authentic tomato extracts and adulterants

The results of spectral subtraction, aimed at singling out the spectral fingerprints of control samples (pure adulterants and authentic tomato extracts) from those of pure water are presented in (Figure 6). Referring to the difference spectra, peaks situated at 1375nm and associated with the presence of carbohydrates were identifiable for the extracts of all three tomato varieties (Borba *et al.*, 2021) with different intensities of absorbance. Authentic tomato extracts also displayed a distinct peak at 1415 nm, attributed to the first overtone band of C–H stretching mode and typical peak of β-carotene (Hagos *et al.*, 2022). The spectral profile of food colorant (FC) extract mainly featured two peaks of high intensities at 1400nm and 1425nm, attributed to free water molecules (S0) and O–H second overtone respectively. A peak of lower intensity occurred at around 1460 nm, coupled with a less prominent broad band at the range 1475nm-1525nm. Studies ascribed the peak at 1475nm to the first overtone of O–H stretching vibrations but also to N–H stretch first overtone(Dumitriu *et al.*, 2015), a prominent constituent of both food colorants Tartrazine and ponceau 4R.

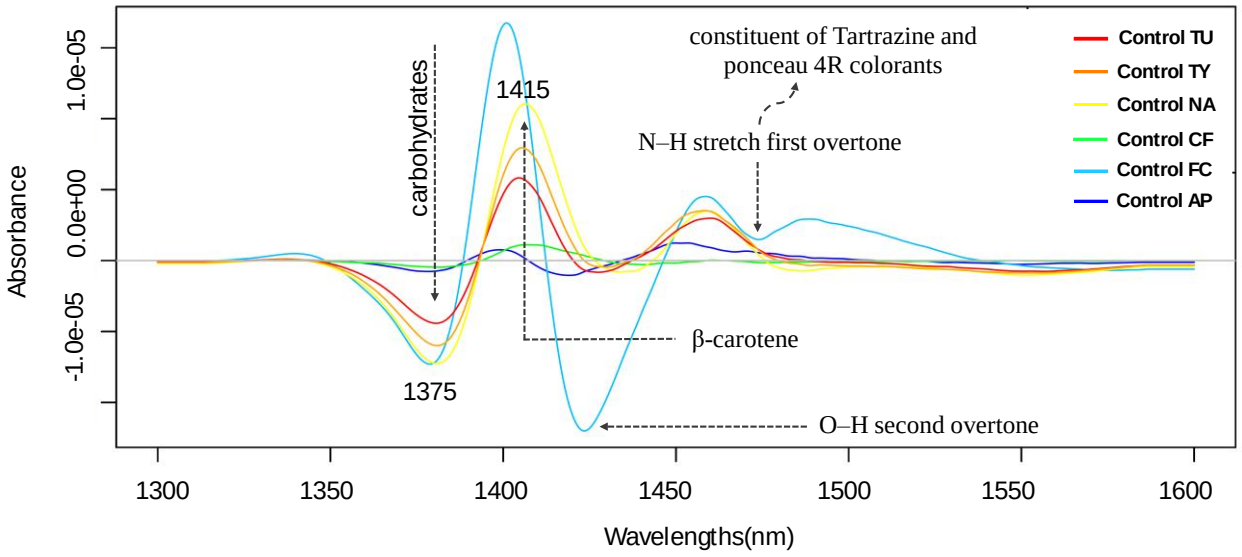

Figure 6: Difference spectra of Navrongo (Control NA), Tuobodom (Control TU) and Tytanium (Control TY) extracts and pure adulterants (Corn flour (Control CF), Food Colorant (Control FC), Annatto Powder (Control AP)), obtained by subtraction from the average pure water spectrum in the spectral range 1300-1600 nm, spectral preprocessing: 2nd derivative

Results of principal component analysis, carried out to assess whether any spectral pattern is distinguishable among control samples (pure adulterants and authentic tomato extracts), are

presented in (Figure 7). As depicted in the scores plot, PC1 and PC2 account for 96.55% of the overall data variance and show that despite the overlapping between pure tomato extracts, they were nonetheless distinctively separated from the pure adulterants. Subsequently, the extracts of the three authentic tomato varieties (Navrongo (NA), Tuobodom (TU) and Tytanium (TY)) and those of the pure adulterants (Annatto powder (AP), Corn flower (CF) and Food colorant (FC)) were classified by PCA-LDA modelling. Figure 7b displays the corresponding classification plot, obtained after 3-fold cross-validation.

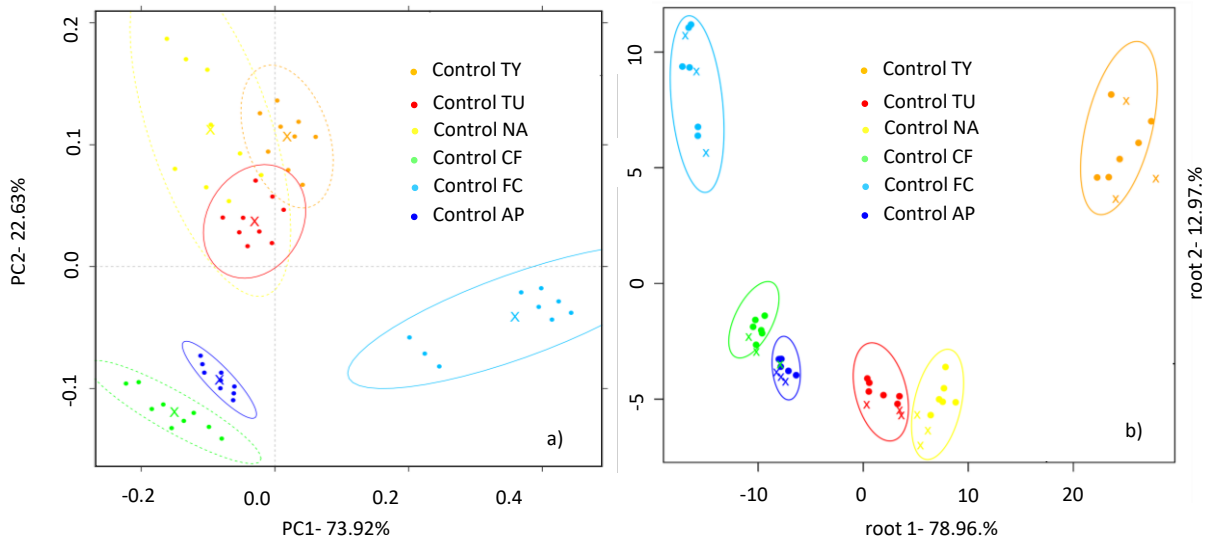

Figure 7: a) PCA on pure tomato powder extracts (Control NA, Control TU, Control TY) and pure adulterants (Control CF, Control FC, Control AP) in the 1300–1600 nm range. N=54. b) PCA-LDA classification plot of pure tomato powder extracts and pure adulterants in the 1300–1600 nm spectral range. N=54

Accuracies of the model, as indicated in the confusion matrix (Table 3) amounted to 98.16% and 98.13% for the recognition and prediction, respectively. Corn flour extracts were misclassified as Annatto extracts in 11.17 % of the cases. As per the between group variance expressed in the root1 of the LDA plot, it consisted of 78.96%, thus showcasing the main separation of the tested sample groups.

Table 3: PCA-LDA classification model on the NIRS data of control samples (authentic tomato extracts and pure adulterants) after three-fold cross-validation (N = 54).

Recognition Accuracy (98.16%)							
		Control Samples					
		Annatto	Corn	Food colorant	NA	TU	TY
Control Samples	Annatto	**100**	11	0	0	0	0
	Corn	0	**89**	0	0	0	0
	Food colorant	0	0	**100**	0	0	0
	NA	0	0	0	**100**	0	0
	TU	0	0	0	0	**100**	0
	TY	0	0	0	0	0	**100**
Validation Accuracy (98.13%)							
		Control Samples					
		Annatto	Corn	Food colorant	NA	TU	TY
Control Samples	Annatto	**100**	11.17	0	0	0	0
	Corn	0	**88.83**	0	0	0	0
	Food colorant	0	0	**100**	0	0	0
	NA	0	0	0	**100**	11	0
	TU	0	0	0	0	**89**	0
	TY	0	0	0	0	0	**100**

NA: Navrongo Variety; TU: Tuobodom Variety; TY: Tytanium Variety

When examining the water spectral pattern of the same extracts at the first water overtone range (1300-1600nm), the displayed pattern validated the previous results where authentic TU, TY and TU tomato extracts had similar yet distinguishable water spectral patterns. What the aquagrams also showed is that the food colorant adulterant was markedly distanced from the other adulterants and tomato extracts, with higher absorbance value at spectral regions associated with less hydrogen bonded water structures (Figure 8).

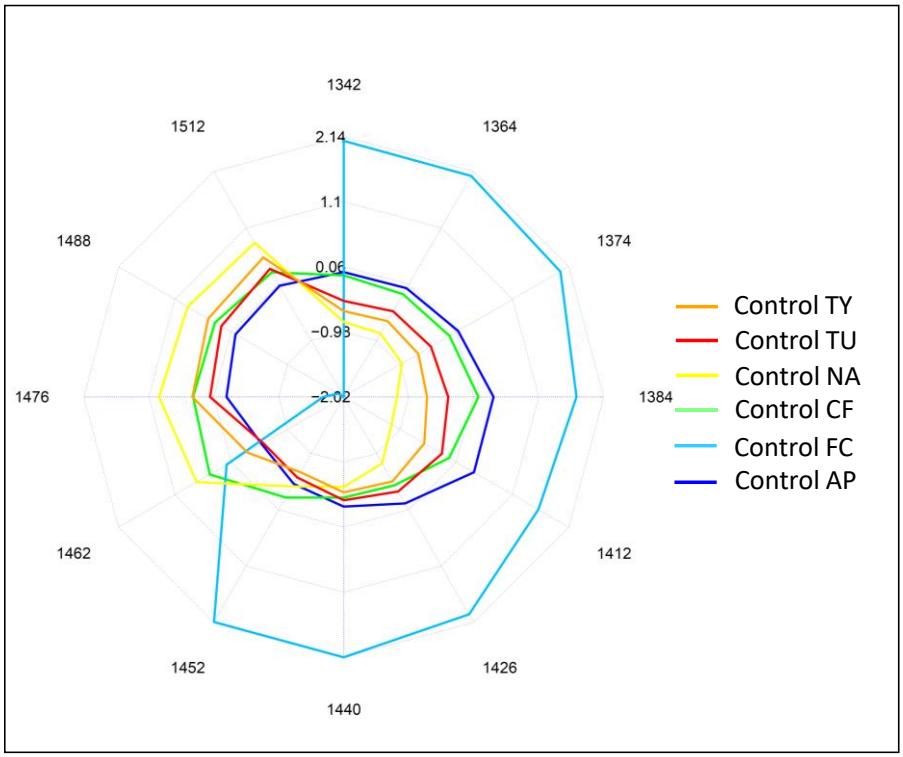

Figure 8: Aquagrams of pure TP extracts (NA, TU, TY) and pure adulterants (CF, FC, AP), N=54, Spectral range: 1300-1600nm

5.1.2. Results of the adulterated tomato powders extracts

Figure 9 depicts the results of exploratory principal component analysis (PCA) on pure and adulterated tomato powder extracts within each of the studied tomato varieties (Tuobodom (Figure 9a), Navrongo (Figure 9b), and Tytanium (Figure 9c)) performed according to the mixture type (single adulterated (AP, CF and FC)or dual adulterated(FC+AP)or (CF+FC) or pure (Control AP, Control FC, Control CF and Control TP)).

Across all three PCA scores plot, the first two principal components (PC1 and PC2) account for most of the data variance. For Tuobodom extracts, PC1 and PC2 represent 74.54% of the data variance whereas for Navrongo, the first two components represent 92.46%. of the data variance. For Tytanium extracts, on the other hand, components PC1 and PC2 are representative of 87.179% of the variability.

No apparent trend was observed with regards to the mixture type (single-adulterated and dual-adulterated) across the three varieties with an overlapping of all mixture types. Nevertheless, what is noticeable is the separation of pure adulterants from the rest of the considered samples.

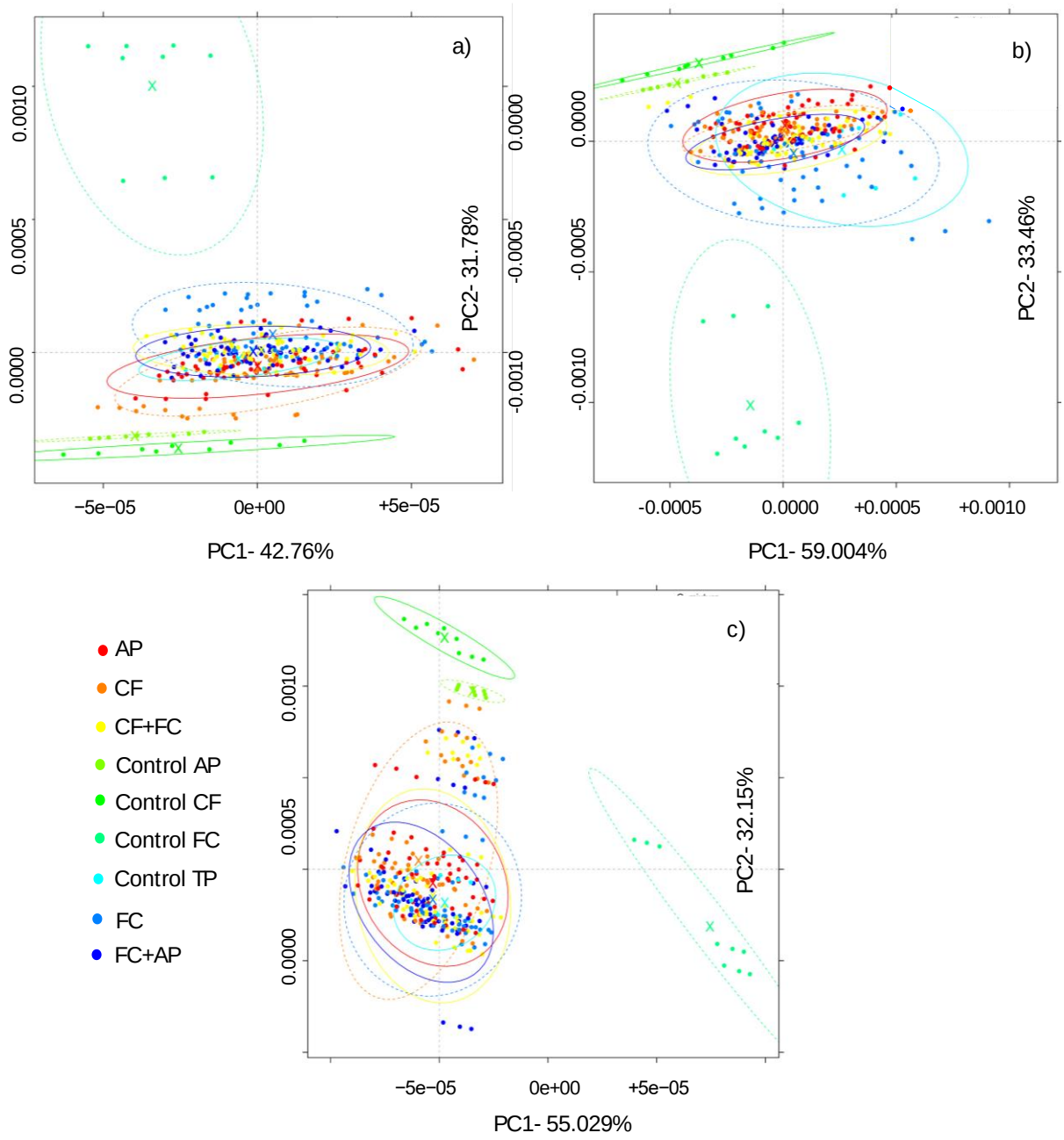

Figure 9: PCA on adulterated tomato powder extracts (CF, AP, FC, CF+FC, FC+AP), pure adulterants (Control CF, Control FC, Control AP) and authentic tomato powder extracts (Control TP: TU variety (a), NA variety (b), TY variety (c)) in the 1300–1600 nm range. N=351.

The preliminary inspection of the aquagrams of the whole dataset (Figure 10) depicted a distinctive pattern specifically at the region [1342nm-1374nm] with increasing absorbances occurring concurrently with the augmentation of added adulterants. Such variation is subtly noticeable at the lower adulteration levels but markedly apparent at the highest levels of adulteration (10%,15% and 20%). Obtaining this water spectral pattern, despite the varietal differences of the studied tomato powders, ascertains the efficiency of the method in detecting the adulteration level and indicates its potential suitability for authenticating other tomato varieties, in their liquid format.

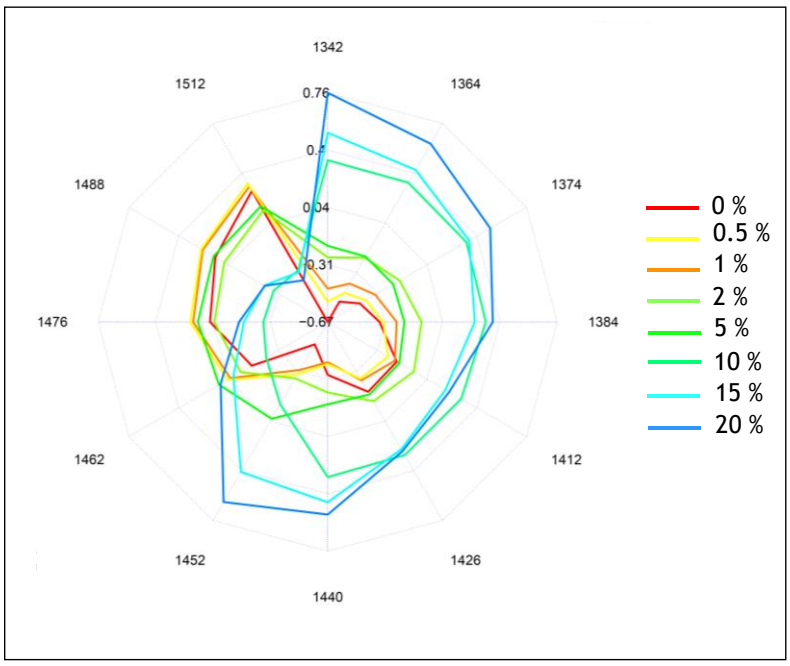

Figure 10: Aquagrams of all adulterated mixtures (0.5% to 20%) and Pure tomato powder extracts (0%), N=972, Spectral range: 1300-1600nm

Once assured of the method's adequacy in terms of distinguishing authentic tomato extracts from the studied bulking and coloring agents, the following step consisted in evaluating its performance with regards to the classification of the differently tampered samples (single and dual adulteration), and the estimation of the extent of adulteration (% of added adulterant).

5.1.2.1. Tomato extracts with single adulterants:

Over the 800-1100nm range, corresponding to the second overtone range of water, the PCA-LDA model built to classify corn flour adulterated TU samples based on the concentration level presented a 100% accurate recognition and predict rates (Figure 11).

Misclassifications occurred when basing the analysis on the first overtone range (1300-1600nm) resulting in lower performance of the model (98% for recognition and 81% for prediction).

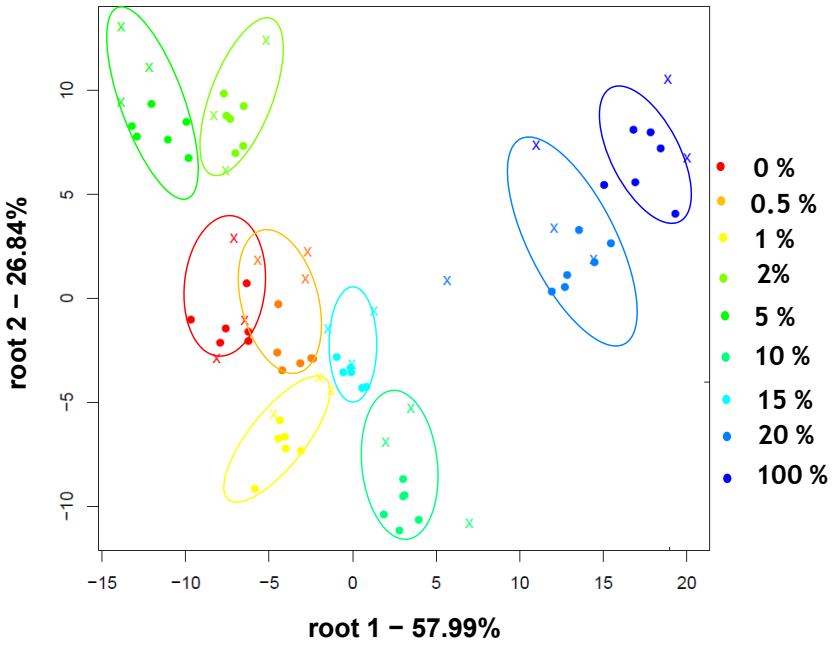

Figure 11: Classification plot of Corn Flour-adulterated Tuobodom (TU) extracts by adulteration levels, N=81 ,3-fold Cross-Validation, Spectral pre-processing: Savitzky Golay (SG,17pts), Spectral range: 800-1100nm

Table 4 gives an overview of the PCA-LDA classification accuracies of the corn flour adulterated tomato powder extracts for all the investigated tomato varieties, in the two considered water spectral ranges. While the performance of the models slightly differed across the varying varieties, their overall efficiency corroborates the potential of the method in distinguishing the different levels of adulteration.

Table 4: PCA-LDA classification accuracies of the corn flour adulterated TP extracts in the ranges 800-1100nm and 1300-1600nm:

Tomato Variety	Spectral Pretreatment	1st Overtone (1300-1600nm)		Spectral Pretreatment	2nd Overtone (800-1100nm)	
Tytanium	None (Raw Spectra)	R	85.18%	2nd Derivative	R	100%
		P	82.69%		P	97.55%
Tuobodom	sgol@2−17−0	R	98.00%	sgol@2−17−0	R	100%
		P	81.00%		P	100%
Navrongo	sgol@2−17−0 + Detrend	R	91.98%	2nd Derivative	R	100%
		P	60.50%		P	100%

R= recognition accuracy; P=prediction accuracy; sgol@2−19−0=Savitzky-Golay filter (2nd polynomial, 19points); SNV= Standard normal variate.

While average accuracies of prediction were achieved when assessing mixtures containing solely corn flour (R2 equal to 0.6, at best), the water spectral pattern as illustrated by the aquagram highlighted the extent of adulteration (Figure 12). What we can clearly observe is the abundance of highly hydrogen bonded water structures in the case of pure tomato extracts (0%) comprising water molecules with four hydrogen bonds (S4) at 1488nm, with three hydrogen bonds located at around 1476nm and those with two hydrogen bonds (S2) (1462nm). As the level of added corn flour increases, less absorbance values are noticed at the formerly stated wavelengths while higher ones are detected at regions with less hydrogen bonded water. Interestingly, the aquagram highlights prominent peaks of pure corn flour extracts (100%) at 1342nm, 1364 nm attributed to the first overlap regions of OH and CH with CH3, CH2 chemical groups (Aires *et al.*, 2018) but also features a peak at 1452nm, which corresponds to symmetric and asymmetric stretching of the 1st overtone of water.

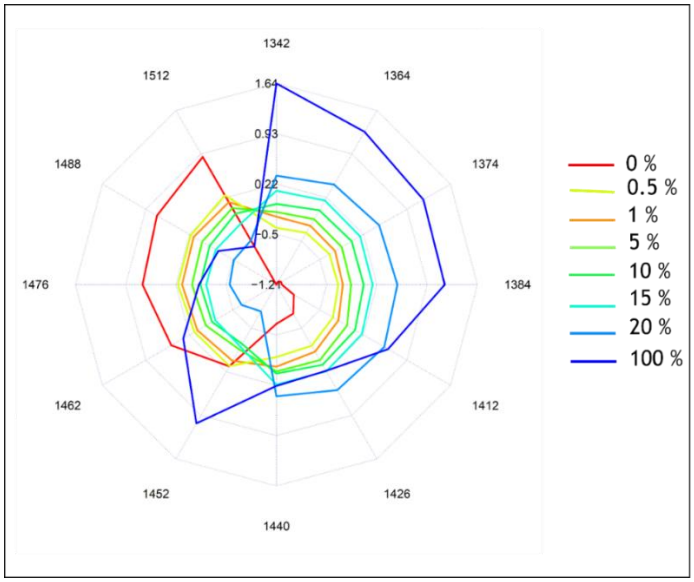

Figure 12: Aquagrams of pure tomato powder extract (0%), pure CF (100%) and Corn Flour (CF) adulterated tomato powder extracts in the concentration range of 0.5–20% and, N=9 each, Spectral range: 1300-1600nm

In categorizing tomato powder extracts adulterated with food colorant, all the PCA-LDA models demonstrated a high classification accuracy, achieving average recognition rates in the range [82.27% - 100%]. Such outcomes depended on the tested variety, the applied range, and the spectral pretreatment. The prediction rates, on the other hand, ranged from 77.00 to 97.55%. Once again, the models obtained after analyzing the spectra acquired in the second overtone were of better performance. The subsequent table provides a more comprehensive overview of the results obtained (Table 5).

Table 5: PCA-LDA classification accuracies of the food colorant adulterated tomato powder extracts scanned using the XDS liquid analyzer in the ranges 800-1100nm and 1300-1600nm.

Tomato Variety	Spectral Pretreatment	1st Overtone (1300-1600nm)		Spectral Pretreatment	2nd Overtone (800-1100nm)	
Tytanium	sgol@2−17−0	R	82.70%	None (Raw Spectra)	R	100%
		P	81.56%		P	97.51%
Tuobodom	sgol@2−19−0 + SNV	R	94.00%	2nd Derivative	R	100%
		P	77.00%		P	93.85%
Navrongo	sgol@2−19−0	R	82.27%	sgol@2−17−0 + SNV	R	100%
		P	79.08%		P	97.55%

R=Average recognition accuracy; P=Average prediction accuracy; sgol@2−19−0 =Savitzky-Golay filter (2nd polynomial, 19points); SNV= Standard normal variate.

When predicting the concentration of food colorant added within the ranges of 0.5% to 20% w/w, the PLSR models delivered a reliable estimation of the degree of adulteration for all three tomato varieties. While the R^2CV values ranged from 0.94 to 0.98, the errors of cross-validation RMSECV fell within the range of 0.79 to 1.59 % w/w. Table 6 provides a comprehensive overview of the various PLSR models, the number of latent variables (LV), the applied pretreatments, and the type of validation.

Table 6. PLSR models developed in the range 1300-1600nm to predict added food colorant to the tomato powder extracts scanned using the XDS liquid analyzer.

Tomato Variety	Spectral Pretreatment	R2C	RMSEC (%w/w)	R2CV	RMSECV (%w/w)	LV	Validation type
Tuobodom	1st Derivative	0.99	0.48	0.98	0.79	5	leave one replicate out
Navrongo	sgol@2−17−0 +MSC	0.95	1.43	0.94	1.59	3	
Tytanium	sgol@2−17−0 +Detrend	0.99	0.70	0.96	1.33	7	

sgol@2−17−0 =Savitzky-Golay filter (2nd polynomial, 17points); MSC= Multiplicative scatter correction

As indicated in both Table 6 and Figure 13, the most accurate predictive model was obtained in case of *Tuobodom* (TU) variety, the corresponding regression vector displayed in (Figure 13) reveals that the wavelengths predominantly influencing the attained prediction rate are 1418nm, 1450nm and 1487nm. Amongst these peaks is the OH stretching vibration (1450nm) which has

been demonstrated by Sheelarani et al., (2022) to be a symbolic conformation indicative of pure Ponceau 4R dye.

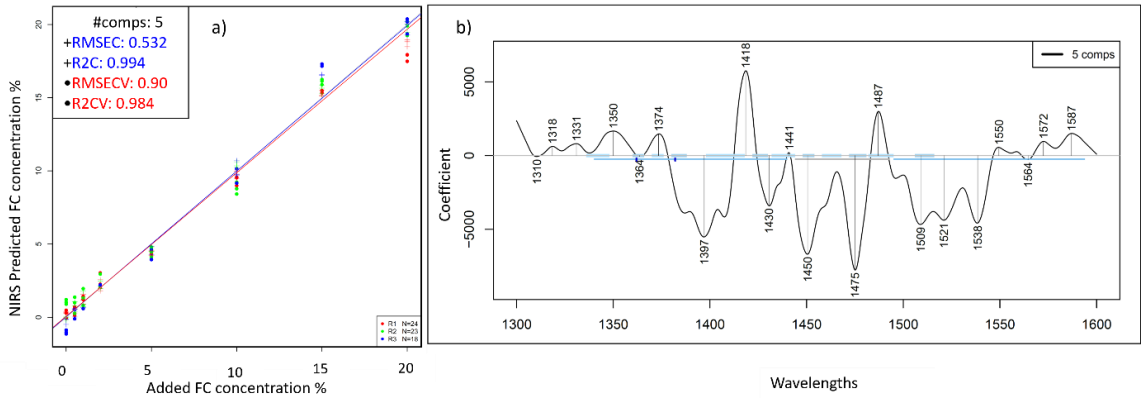

Figure 13.(a) PLSR analysis of the *Tuobodom* (TU) tomato extracts derived from the smoothed (SG smoothing (29 points)) 1st derivative spectra for the prediction of added food colorant (% w/w); (b) Corresponding regression vector, Spectral range: 1300-1600nm.

In our attempt to assign each tomato powder extract to its respective group, depending on the degree of annatto adulteration, the most accurate PCA-LDA model was the one obtained for *Tytanium* (TY) variety after smoothing the spectra generated in the second overtone range, coupled with SNV. The yielded recognition and prediction rates were estimated at 100% and 98.77%, respectively. Misclassifications only occurred in 11% of the cases, namely between extracts containing 1% and 5% of annatto.

Table 7 illustrates the optimal PCA-LDA classification models with their corresponding accuracies, for various spectral pretreatments in both the 1st and 2nd overtone regions and for each of the investigated varieties.

Table 7. PCA-LDA classification accuracies of the annatto powder adulterated tomato powder extracts scanned using the XDS liquid analyzer in the ranges 800-1100nm and 1300-1600nm.

Tomato Variety	Spectral Pretreatment	1st Overtone (1300-1600nm)		Spectral Pretreatment	2nd Overtone (800-1100nm)	
Tytanium	sgol@2−19−0	R	98.22%	sgol@2−19−0 + SNV	R	100%
		P	69.50%		P	98.77%
Tuobodom	sgol@2−17−0 + SNV	R	71.67%	None (Raw Spectra)	R	97.76%
		P	68.59%		P	88.59%
Navrongo	sgol@2−17−0 +Detrend	R	88.32%	2nd Derivative	R	100%
		P	76.57%		P	96.33%

5.1.2.2. Tomato extracts with dual adulterants (bulking agent + coloring agent)

Models developed to predict the added food colorant in extracts of tomato powder mixtures comprising the dual adulterants (FC+CF) exhibited highly accurate predictive performances with coefficients of determination in the range [0.85- 0.95] and low errors of cross-validation fluctuating between 0.81 and 1.37 %w/w (Table 8). Figure 14 depicts the predictive model obtained for *Tuobodom* (TU) tomato variety extract along with its respective regression vector. Prominent peaks included 1364nm, 1402nm and 1436nm, falling within the water matrix coordinates (WAMACS) C2 to C7.

Table 8**.** PLSR models developed in the range 1300-1600nm to predict added food colorant to the (FC+CF) tampered tomato powder extracts scanned using the XDS liquid analyzer.

Tomato Variety	Spectral Pretreatment	R^2C	RMSEC (%w/w)	R^2CV	RMSECV (%w/w)	LV	Validation type
Tuobodom	sgol@2−19−0	0.97	0.51	0.95	0.83	5	leave one replicate out
Navrongo	sgol@2−23−0*	0.98	0.49	0.94	0.81	5	
Tytanium	sgol@2−19−0 + Detrend	0.94	0.81	0.85	1.37	5	

*sgol@2−23−0 =Savitzky-Golay filter (2nd polynomial, 23points)

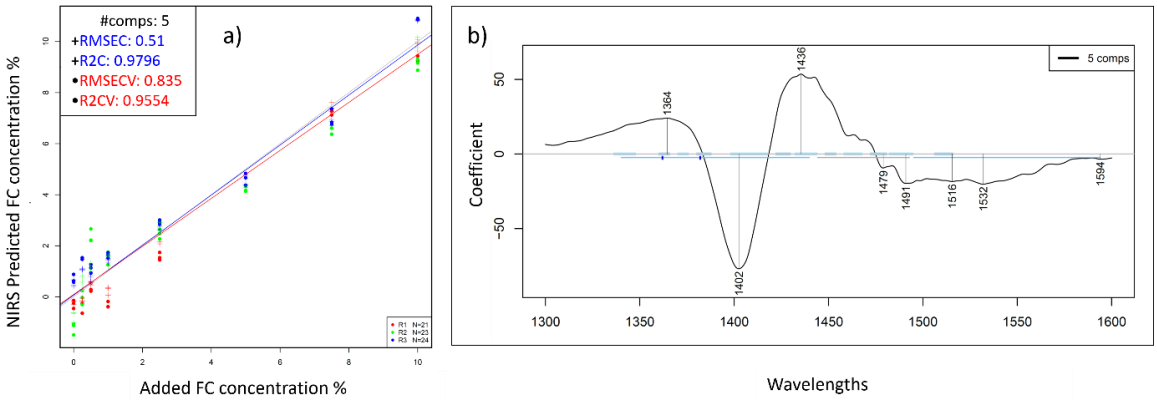

Figure 14**.**(a) PLSR analysis of the (FC+CF) tampered TU tomato extracts for the prediction of added food colorant (% w/w); Regression vector (b). Spectral preprocessing: Savitzky Golay smoothing (27 points), LV=5, Spectral range: 1300-1600nm.

Estimating the amount of added food colorant in dual adulteration mixtures (AP+FC), containing both annatto powder and food colorant, was as efficient as in the first dual mixture (CF+FC). This time, the best model was the one generated for *Navrongo* variety after smoothing and detrending of the spectral data (Table 9). The coefficient of determination R^2CV equaled 0.95 and the error of cross-validation RMSECV was estimated at 0.79% w/w (Figure 15). The regression vector of both models (Figure 14 and Figure 15) showcased some peaks at adjacent wavelengths [(1402, 1436, 1445, 1479, for (CF+FC) mixture and (1409, 1434, 1447, 1476) for (FC+AP) mixture],

emphasizing the consistency of the corresponding conformations as reliable markers of adulteration.

Table 9. PLSR models developed in the range 1300-1600nm to predict added food colorant to the tomato powder in AP+FC adulterated mixtures scanned using the XDS liquid analyzer.

Tomato Variety	Spectral Pretreatment	R^2C	RMSEC (%w/w)	R^2CV	RMSECV (%w/w)	LV	Validation type
Tuobodom	sgol@2−19−0*+ MSC	0.97	0.57	0.95	0.86	5	leave one replicate out
Navrongo	sgol@2−21−0+ Detrend	0.97	0,51	0.95	0.79	4	
Tytanium	sgol@2−17−0+ SNV	0.91	0.99	0.85	1.3	3	

*sgol@2−19−0=Savitzky-Golay filter (2nd polynomial, 19points), SNV=Standard normal variate

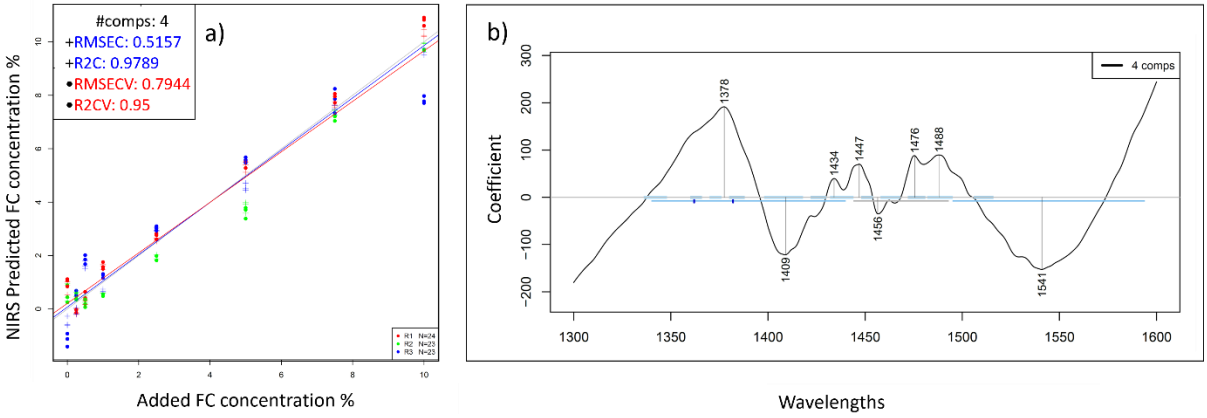

Figure 15. PLSR analysis of the (AP+FC) tampered NA tomato extracts for the prediction of added food colorant (% w/w) (a). Corresponding regression vector(b). Spectral Preprocessing: Savitzky Golay smoothing (21 points) + Detrend, LV=4, Spectral range: 1300-1600nm.

In quantifying added corn flour in the dual adulterated tomato powder extracts (CF+FC), the predictive model for *Tytanium* variety was the best, obtained after performing spectral smoothing (19 points), detrend and standard normal variate. Referring to the regression vector (Figure 16), the contributing peaks corresponded to 1374, 1400, 1434, 1452, 1462 and1534 nm. The majority of these peaks fall within the range 7000-6000 cm-1(≈1428-1667 nm) affirmed by Chen et al., (2017) to correspond to absorbances of corn flour. The latter also depicts a broad band around 1450 nm typically ascribed to water bound to other molecules. Noah et al., (1997) was also among those who highlighted the significance of bands located at 1400nm and 1600 nm, which are associated with the 1st overtone of hydroxyl groups, when analyzing starch.

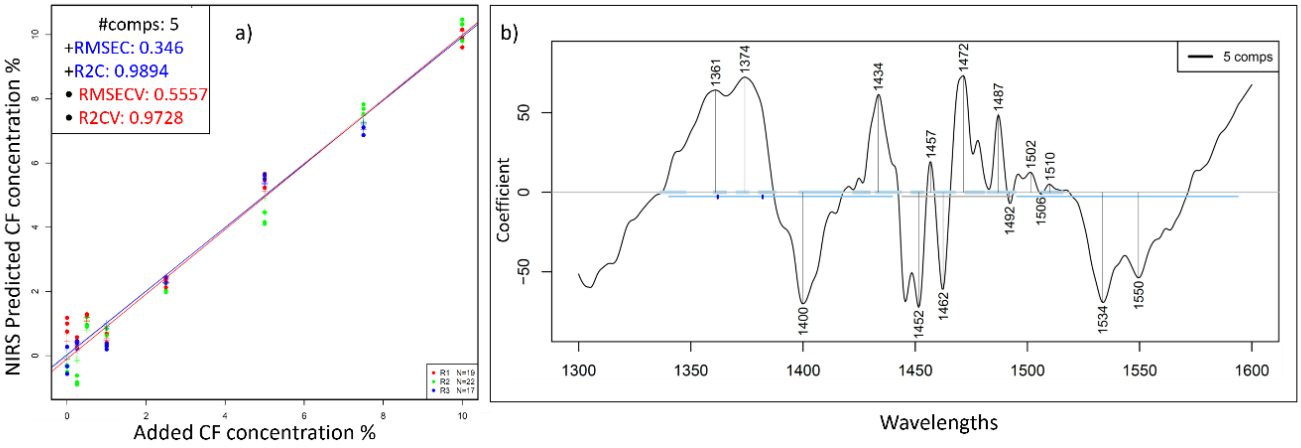

Figure 16.PLSR analysis of the CF+FC tampered TY tomato extracts for the prediction of corn flour (% w/w) (a); Corresponding regression vector (b), Spectral Preprocessing: Savitzky Golay smoothing (19 points) + Detrend+SNV, LV= 5, Spectral range: 1300-1600nm.

As per the rest of the predictive models, their accuracies are specified in

Table 10. In function of the variety, the spectral pretreatments and the number of latent variables, the models R^2 values fluctuated between 0.95 and 0.97 and RMSECV values ranged from 0.55%w/w to 0.79%w/w. Given these slight disparities, we can infer that the tomato variety's influence on prediction performances is not substantial, hence widening the scope of the method to encompass other marketed tomato powder varieties.

Table 10. PLSR models developed in the range 1300-1600nm to predict added corn flour to the tomato powder in CF+FC adulterated mixtures scanned using the XDS liquid analyzer.

Tomato Variety	Spectral Pretreatment	R^2C	RMSEC (%w/w)	R^2 CV	RMSECV (%w/w)	LV	Validation type
Tuobodom	sgol@2−19−0*+ Detrend	0.99	0.31	0.96	0.69	7	leave one replicate out
Navrongo	sgol@2−23−0+ MSC	0.99	0.32	0.95	0.79	7	
Tytanium	sgol@2−19−0+ Detrend	0.98	0.34	0.97	0.55	5	

*sgol@2−19−0=Savitzky-Golay filter (2nd polynomial, 19points), MSC=multiplicative scatter correction

Efficient determination of the added annatto powder in mixtures tampered with both annatto powder and food colorant was demonstrated, notably for *Tuobodom* and *Tytanium* varieties. Their corresponding models exhibited good coefficients of determination and minimal errors (Table 11). In a similar study addressing paprika adulteration with annatto, Oliveira et al., (2020) identified prominent peaks of pure annatto samples around 1475 nm. When attempting to estimate the added amounts, the regression vector revealed peaks at both 1475 nm and 1600nm.

Table 11. PLSR models developed in the range 1300-1600nm to predict added annatto powder to the tomato powder in AP+FC adulterated mixtures scanned using XDS liquid analyzer.

Tomato Variety	Spectral Pretreatment	R^2C	RMSEC (%w/w)	R^2CV	RMSECV (%w/w)	LV	Validation type
Tuobodom	sgol@2−19−0*	0.98	0.43	0.95	0.75	7	leave one replicate out
Navrongo	sgol@2−17−0+ Detrend	0.99	0.34	0.95	0.74	7	
Tytanium	sgol@2−17−0+ MSC	0.96	0.63	0.86	1.33	5	

*sgol@2−19−0 =Savitzky-Golay filter (2nd polynomial, 19points), MSC=multiplicative scatter correction

In the analysis of tomato powder extracts adulterated with annatto powder and food colorant, we observed that, in alignment with the earlier mentioned results, basing the analysis on the second overtone range delivered better results in comparison with the first water overtone. Average recognition rates of 100% and prediction rates in the range [93.85% - 98.77%] were obtained (Table 12).

Table 12. PCA-LDA classification accuracies of the AP+FC adulterated tomato powder extracts and scanned using XDS liquid analyzer in the ranges 800-1100nm and 1300-1600nm.

Tomato Variety	Spectral Pretreatment	1st Overtone (1300-1600nm)		Spectral Pretreatment	2nd Overtone (800-1100nm)	
Tytanium	sgol@2−17−0	R	76.54%	sgol@2−17−0 + SNV	R	100%
		P	70.46%		P	98.77%
Tuobodom	sgol@2−17−0 + SNV	R	75.94%	None (Raw Spectra)	R	100%
		P	65.49%		P	93.85%
Navrongo	sgol@2−17−0	R	67.30%	None (Raw Spectra)	R	100%
		P	62.94%		P	98.77%

R=Average recognition accuracy; P=Average prediction accuracy; sgol@2−17−0 (Savitzky-Golay filter (2nd polynomial, 17points); SNV: Standard normal variate

For the *Navrongo* variety, both recognition and prediction of 100% accuracy rates were obtained when PCA-LDA models were developed to classify the extracts tampered simultaneously with corn flour and food colorant in concentrations ranging from 0.5% to 20% and in the presence of pure extracts (authentic tomato (0%) and pure adulterants (100%)). The corresponding classification plot, displayed in Figure 17, showcases the between group variance expressed in the root1 and root2, totaling 48.64% and 21.86%, respectively.

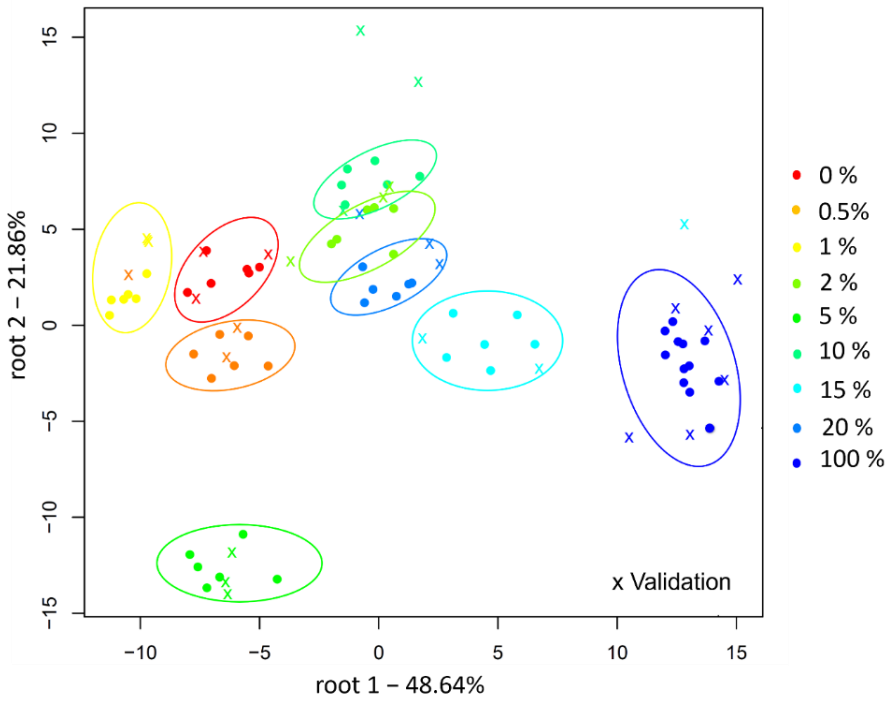

Figure 17: Classification plot of CF+FC-adulterated tomato powder extracts by adulteration levels, N=90, 3-fold CV, Spectral pre-processing: SG smoothing (19 points) +1st derivative

Regarding the remaining results (Table 13), they depended on the spectral range and applied preprocessing but also varied from one variety to the other. The second-best model was the one corresponding to *Tytanium* variety (TY), obtained after applying some smoothing, jointly with SNV. With an average prediction rate of 98.77%, the misclassifications occurred only between levels 1% and 2%, with the latter mistakenly classified as the former in 11% of the cases.

Table 13. PCA-LDA classification accuracies of the tomato powder extracts adulterated with CF+FC scanned using the XDS liquid analyzer in the ranges 800-1100nm and 1300-1600nm.

Tomato Variety	Spectral Pretreatment	1st Overtone (1300-1600nm)		Spectral Pretreatment	2nd Overtone (800-1100nm)	
Tytanium	None (Raw Spectra)	R	85%	sgol@2−17−0 + SNV	R	100%
		P	64.25%		P	98.77%
Tuobodom	sgol@2−17−0 + MSC	R	88%	1st Derivative	R	100%
		P	54%		P	95.10%
Navrongo	None (Raw Spectra)	R	85%	1st Derivative	R	100%
		P	60%		P	100%

R: Average recognition accuracy; P: Average prediction accuracy; sgol@2−17−0 (Savitzky-Golay filter (2nd polynomial, 17points); MSC: Multiplicative scatter correction

To conclude, discrepancies amongst the presented results were influenced by varying factors, including inherent sample characteristics, such as the nature of the adulterant (bulking or coloring

agent), mixture type (single or dual adulteration) but also to operational aspects such as the applied pretreatments and the spectral range. The qualitative and quantitative models (PCA-LDA and PLSR) gave commendable outcomes across the studied tomato cultivars, thus corroborating the potential of aquaphotomics in authenticating other varieties prone to adulteration. Further investigations are essential to enhancing the reliability of the method and explore its potential implementation on an industrial scale. This entails the inclusion of other prevalent adulterants and tomato varieties as well as addressing the complexity of accounting for additional ingredients in the tested food matrix.

5.2. Aquaphotomics for coffee quality grade determination

5.2.1. Near Infrared Analysis of Pure Ground Coffee

A primary step consisted in assessing whether or not the applied method could discriminate between the pure varieties of ARA1, ARA2, ARA3, ROB1, ROB2, and ROB3 in the form of ground coffee. Analysis of pure ground coffee samples of differing varieties by means of principal component analysis, as showcased in Figure 18, demonstrates a pattern of separation along the axis of PC1, which together with PC2 accounts for 99% of the data variability. The efficacy of NIRS in terms of separating the samples was not only based on their respective varieties, but also on their provenance as the different samples came from different sources: Brazil (ARA1), Columbia (ARA2), Ethiopia (ARA3), Vietnam (ROB1), Uganda (ROB2), and India (ROB3). This trend suggests the compositional variability within each of the evaluated varieties. Indeed, studies have shown the role of geographical origin in conferring a specific chemical composition to coffee. This is in accordance with the findings reported by (Giraudo *et al.*, 2019) , who proved that intra-varietal differences of coffee beans originating from different countries and continents can be traced by their respective NIR spectral patterns. According to the corresponding loadings vector (Figure 18c), the wavebands 1390, 1408, 1438, 1452, and 1512 nm contributed the most to this separation. The PCA-LDA model scores yielded 100% recognition and prediction of the pure Arabica and Robusta varieties (Figure 18b).

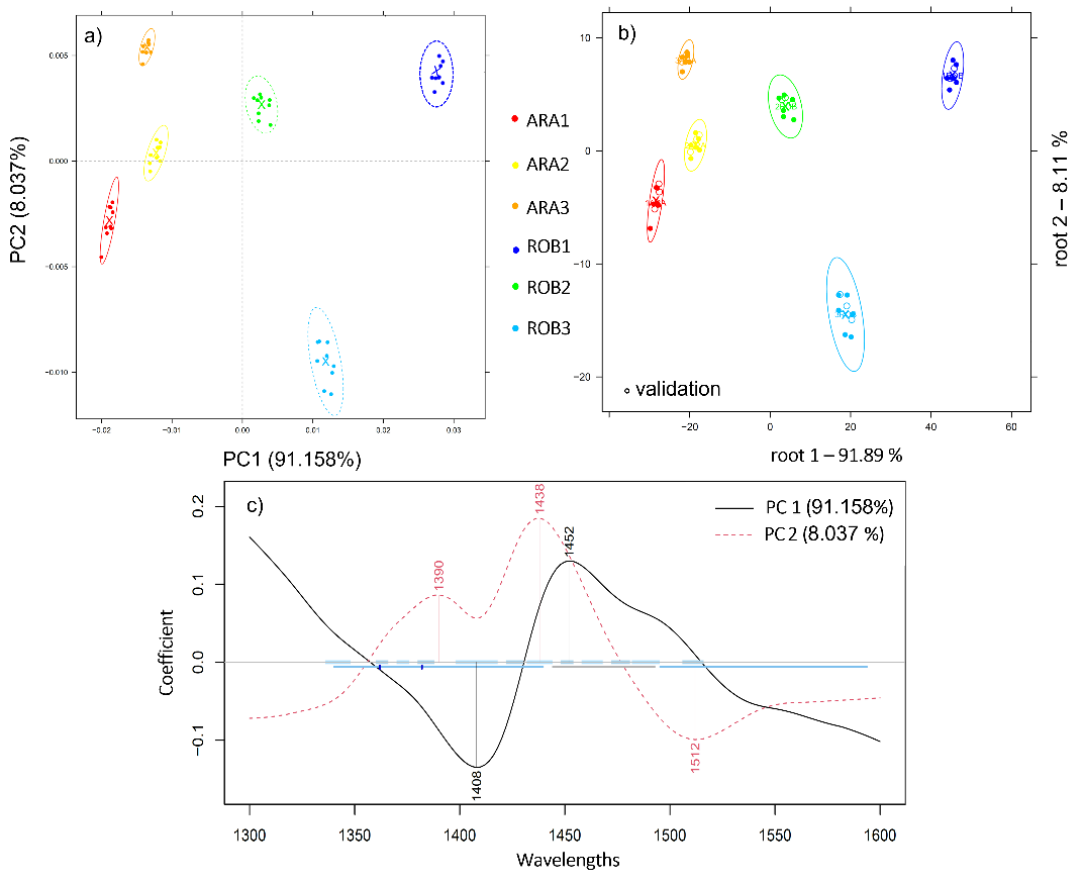

Figure 18: (a) PCA on pure arabica (ARA1, ARA2, ARA3) and Robusta (ROB1, ROB2, ROB3) ground coffee in the range 1300–1600 nm, N = 54, SG smoothing (19 points) and MSC; (b) PCA-LDA classification plot of pure Arabica (ARA1, ARA2, ARA3) and Robusta (ROB1, ROB2, ROB3) ground coffee samples, N = 54;(c) The respective loading plot.

5.2.2. Near Infrared Analysis of Ground Coffee Mixtures

For the remainder of our study, we focused on the mixtures prepared by mixing the pair (ARA3, ROB3). To determine if a recognizable pattern is ascribable to the adulterated Arabica depending on the added Robusta, principal component analysis (PCA) was performed in the first overtone (1300–1600 nm), 2nd overtone (800–1100 nm), as well as in the truncated spectral range of the instrument, 800–1670 nm.

The model illustrating the most distinctive pattern was obtained in the range 800–1670 nm using the smoothed and MSC pretreated spectra (Figure 19a). According to the loadings plot (Figure 19c), the wavelengths responsible for the variance in the data are mostly those located at 970, 1106, 1126, 1266, 1298, 1318, and 1464 nm. Previous studies have attributed bond vibrations at 1126 nm of the 2 × C-H stretching and 2 × C-H deformation and (CH2)n C-H stretching second overtone to coffee fatty acids and chlorogenic acid (CGA) (Esteban-Díez, González-Sáiz and Pizarro, 2004). Indeed, these constituents have already proven to be good discriminators of the varietal origin of coffee (Kamiloglu, 2019).

Relying solely on the visual inspection of the separated samples, the truncated range 800–1670 nm served better for the pattern recognition of the mixtures with principal component analysis (PCA). The analysis of the samples by means of linear discriminant analysis (LDA), however, proved better when performed at the first overtone 1300–1600 nm.

An accurate recognition and prediction of 95.87% and 94.45% were obtained, respectively, using the raw spectra. The misclassifications occurred mainly between sample pairs (1% and 3%; 3% and 5%) whereas those comprising at least 10% Robusta were 100% accurately classified.

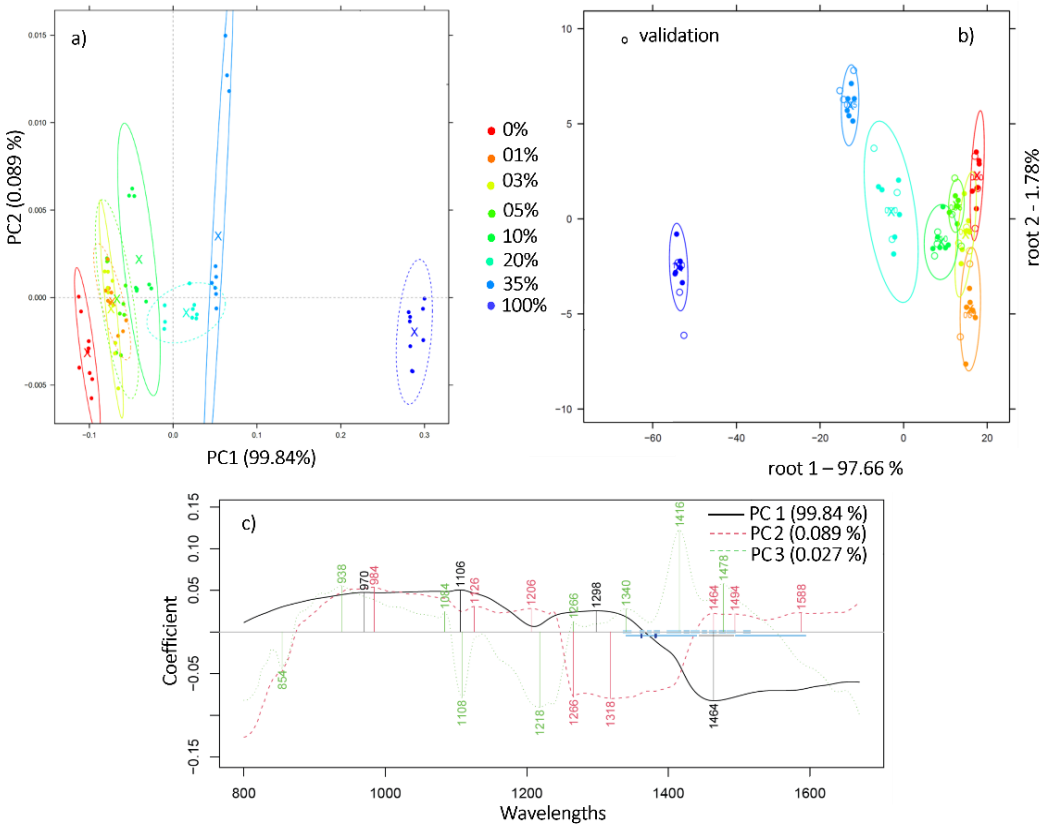

Figure 19: (a) PCA scores plot of robusta adulterated ground coffee in the concentration range 1%–35%, N = 72, wavelength range 800–1670 nm, Savitzky-Golay smoothing (window size 19) and MSC; b) PCA-LDA classification plot of the robusta adulterated ground coffee in the concentration range 1%–35%, wavelength range 1300-1600 nm, N = 72; (c) PCA loadings plot.

Once the mixtures were correctly classified, PLSR models were built in order to assess the feasibility of near infrared spectroscopy in predicting the Robusta to Arabica ratio. By leaving one group out (three consecutive scans of the same replicate) cross-validation, the model built on the smoothed first derivative of the spectra enabled a coefficient of determination (R2CV) of 0.99 and an error (RMSECV) of 2.4% (Figure 20a). Similar results (R2 > 0.99 and RMSE below 1.2% w/w)

were found when evaluating Arabica-Robusta mixtures in the range of 0–60% (Pizarro, Esteban-Díez and González-Sáiz, 2007).

The corresponding regression vector showcases the most significant wavelengths in terms of accurately determining the added Robusta. These peaks are located at 1324, 1374, 1402, 1422, 1444, 1470, 1498, 1518, 1540, and 1556 nm (Figure 20b). Prior studies have assigned wavelengths in the 1400–1600 nm range to some typical components of coffee, such as caffeine, sugar, and chlorogenic acids (Ribeiro, Ferreira and Salva, 2011). The addition of Robusta, naturally richer in chlorogenic acid and caffeine content (Adnan *et al.*, 2020), could explain the prominence of these wavebands when predicting the added Robusta.

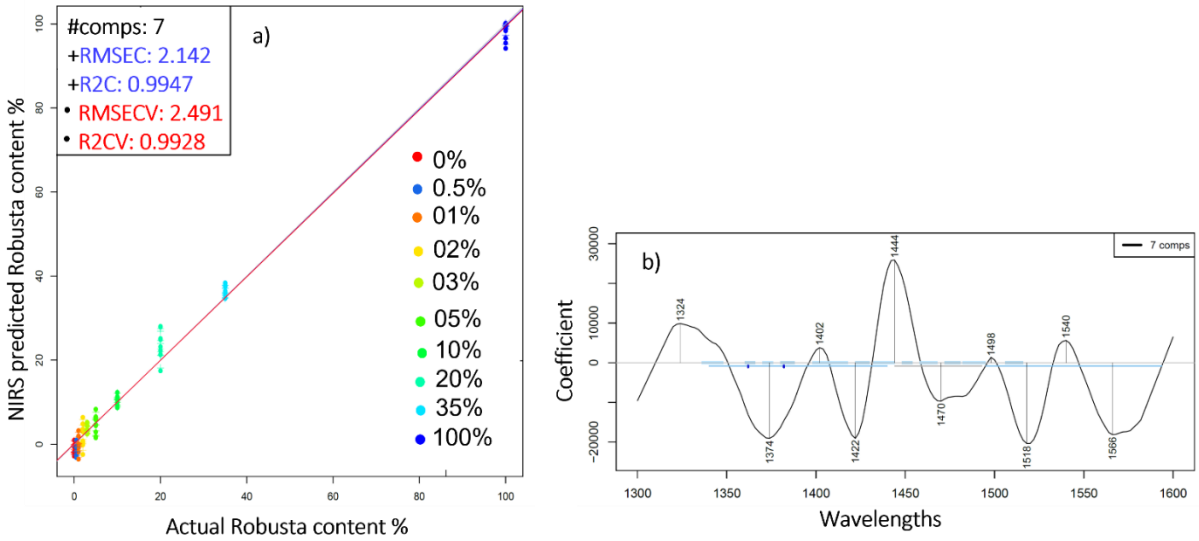

Figure 20: a) PLSR analysis of the ground coffee mixtures derived from the smoothed (Savitzky-Golay smoothing (window size 19)) first derivative spectra for the prediction of added Robusta (% w/w); (b) The respective regression vector in the range 1300-1600nm.

The PCA-LDA classification and PLSR prediction of the Robusta-to-Arabica ratio were also performed on the marketed blends B10% and B30%. Figure 21a illustrates the obtained results where B10% and B30% were discriminated from the pure Arabica and Robusta ground coffee samples with 100% accuracies of recognition and prediction.

The regression model built to predict the Robusta content and cross-validated by leaving three consecutive scans of each replicate at a time enabled an estimation of added Robusta with R2CV and RMSECV values of 0.97 and 3.93% w/w, respectively (Figure 21b).

This slight difference compared to the model constructed only on the mixtures could be due to the different composition of the marketed blends B10% and B30%, obtained by combining other Arabica and Robusta varieties, from a geographical origin other than that of ARA3 and ROB3.

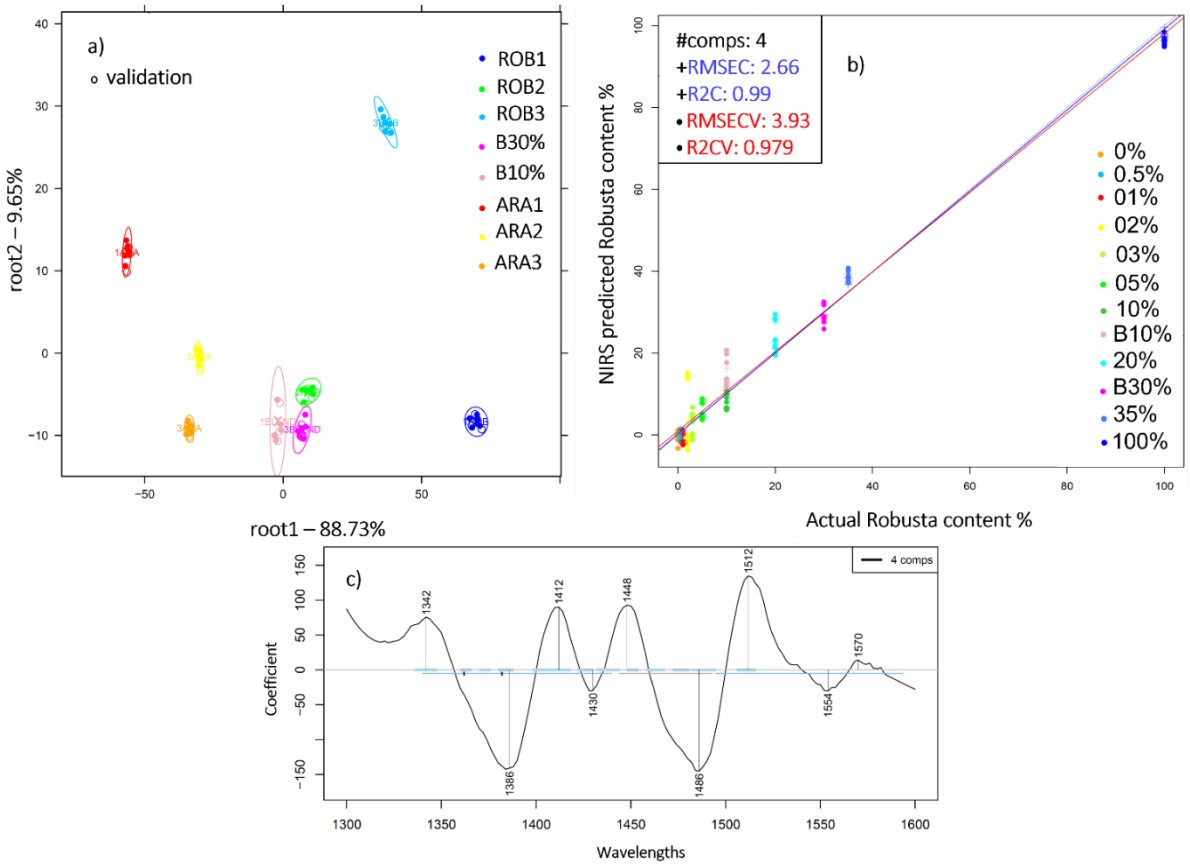

Figure 21: a) PCA-LDA Classification of pure ground Arabica (ARA1, ARA2, ARA3), Robusta, Robusta (ROB1, ROB2, ROB3) and marketed blends B10%, B30% in the range 1300–1600 nm, N = 72, spectral pretreatment: Savitzky-Golay smoothing (window size 19) and 1st derivative; (b) PLSR analysis of the ground coffee mixtures and B10% and B30%, spectral pretreatment: Savitzky-Golay smoothing (window size 17) and SNV; (c) The respective regression vector in the 1300–1600 nm.

5.2.3. Near Infrared Analysis of Pure Liquid Coffee Extracts

Performing principal component analysis (PCA) on the pure Arabica (ARA3) and Robusta (ROB3) liquid extracts in the short wavelength range of 800–1100 nm revealed a pattern of separation into two respective clusters depending on the coffee variety (Figure 22). Combined, PC1 and PC2 accounted for more than 99% of the data variance and the bands contributing the most to this separation were positioned at 950, 982, and 1034 nm.

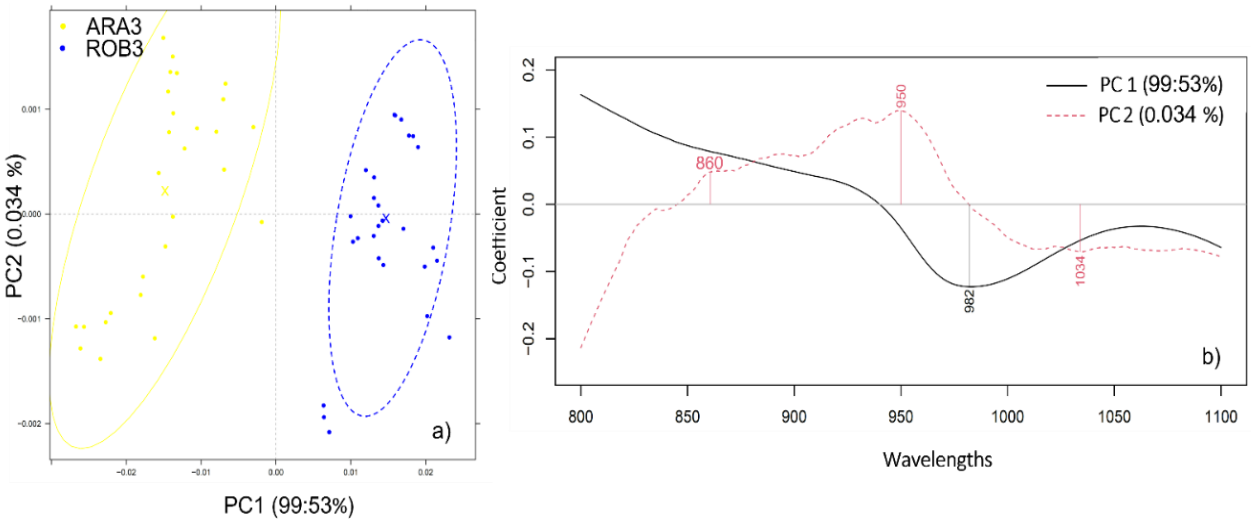

Figure 22: (a) PCA analysis applied to the spectra of pure Arabica (ARA3) and Robusta (ROB3) liquid extracts in the 800–1100 nm range, N = 54, spectral pretreatment: Savitzky-Golay smoothing (window size 19) and MSC; (b) The respective PCA loadings plot.

Next, absorbances of ARA3 and ROB3 aqueous samples as projected on the aquagram were investigated in 12 characteristic wavelength ranges in the second water overtone in NIR region. What the aquagram accentuated is that Robusta coffee extracts, contrarily to Arabica, were majorly characterized by water molecules that are structured into water shells (908 nm), V1 and V2 bonded water while Arabica has high hydrogen bonded water structures (1060 nm) and is rich in water clusters with two, three, and four hydrogen bonds (1018, 1036, and 1044 nm) (Figure 23a).(Wu, He and Feng, 2008) are among those who investigated the compositional analysis of milk in the short NIR wavelength range (800–1050 nm) and reported the potential assignment of the 1018 and 1042 nm to the interaction of fat–water. The fact that Arabica is naturally richer in fat content can explain the high absorbance observed at these bands (Speer and Kölling-Speer, 2006).

The incorporation of marketed blends B10% and B30% into the aquagram calculation is presented in (Figure 23b). Notably, the resulting water spectral pattern followed a logical sequence. Out of the two studied blends, the one with the highest Arabica content (B10%) had a similar pattern to pure Arabica, with slightly lower absorbance values. When the percentage of Robusta increased, as is the case of B30% blend, higher absorbances in the wavelength range of 890–954 nm were emphasized.

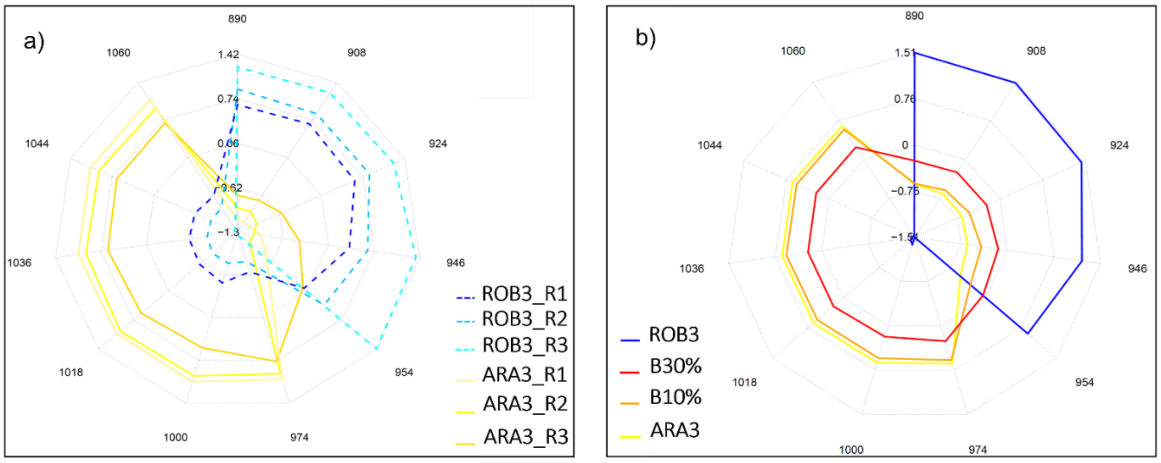

Figure 23: a) Aquagram representation of the spectral pattern of pure Arabica and Robusta extracts at the 2nd overtone (800–1100 nm), N= 9 each; b) Water spectral pattern of marketed blends B10% and B30%, pure Arabica (ARA3) and pure Robusta (ROB3), N= 27 each

Likewise, PCA-LDA was proven performant when assigning the samples ARA3, B10%, B30%, and ROB3 to their specific classes, with an accurate recognition of 91.26% of the samples while at the prediction level 83.39% were correctly categorized. The separation is most apparent along the axis of the first discriminant factor (Figure 24). The misclassifications occurred mostly between samples with proximate composition. Thus, B10% was primarily misclassified in 14.78% of the cases to the group 0% (ARA3) while B30% was identified as B10% in 11.11% of cases.

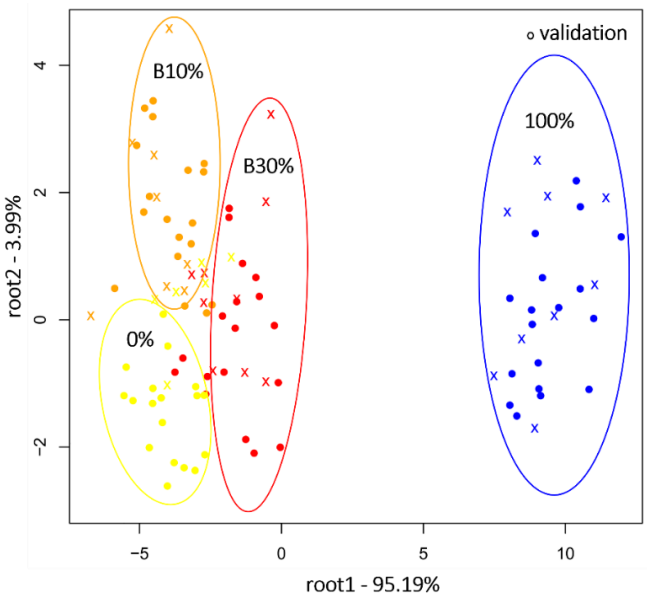

Figure 24: PCA-LDA classification of marketed blends B10% and B30%, pure Arabica (0%) and pure Robusta (100%) at the second overtone region (800–1100 nm), N = 108, NrPCs = 18.

5.2.4. Near Infrared Analysis of Liquid Coffee Mixtures

When all adulteration levels were considered, LDA models destined at regrouping the samples into their corresponding groups enabled a 100% recognition of the different mixtures, and the prediction rate amounted to 71.32%. The following table recaps where misclassification occurred presumably due to the low Robusta concentrations or to the proximity of certain levels Table 14. The classification was even less efficient when the model blends (B10%; B30%) were included into the construction of the predictive model (55.58% prediction rate). The blend B10% was misidentified as belonging to the group containing 5% Robusta in 11.14% of the studied cases and misclassified with 3.67% to the following groups: B30%, 35%, 20%, 10%. B30%, on the other hand, was wrongly categorized as belonging to the groups B10%, 35%, and 3% in 7.44% of the cases. While the blending ratio of these model blends fits into the range covered by our study, their heterogenous composition could have an effect on the classification accuracy. Indeed, the effect of the blend composition on the accuracy of the classification model has already been proven by (Tavares *et al.*, 2016), who, basing their study on the analysis of the lipid extracts by HPLC, proved that proportions as high as 10% of maize and 20% of coffee by-products are required to identify the adulteration of coffee by means of PCA and LDA.

Table 14: PCA-LDA classification model on the NIRS data of pure and adulterated mixtures after three-fold cross-validation (N = 268).

		Validation Accuracy %									
		Robusta-to-Arabica Ratio									
		0%	0.5%	1%	2%	3%	5%	10%	20%	35%	100%
Robusta-to-Arabica ratio	0%	**66.74**	8.03	0	7.44	0	11.11	11.11	0	0	0
	0.5%	14.79	**51.92**	7.45	7.44	14.78	7.44	7.44	0	3.67	0
	1%	0	3.96	**70.41**	0	7.44	0	0	11.11	0	0
	2%	0	8.03	0	**81.44**	7.44	0	0	0	0	0
	3%	3.67	20.02	3.67	0	**55.56**	7.44	3.67	0	0	0
	5%	3.67	0	3.67	3.67	11.11	**66.67**	3.67	7.44	3.67	0
	10 %	7.45	8.03	0	0	3.67	3.67	**66.67**	7.44	3.67	0
	20%	3.67	0	14.79	0	0	3.67	7.44	**70.33**	3.67	0
	35%	0	0	0	0	0	0	0	3.67	**85.32**	0
	100%	0	0	0	0	0	0	0	0	0	**100**

Averaging the consecutive scans and the parallel spectra of each of the studied mixtures and those of the controls (pure Arabica and pure Robusta) was proven effective when it comes to improving the accuracy of the predictive PLSR model. The optimal cross-validated model was the one built in the second overtone region, 800–1100 nm, and was characterized by R2CV and RMSECV

values of 0.95 and 6,35 % w/w, respectively (Figure 25a). Again, the blends lowered the accuracy of the regression model (R2CV=0.9). The most prominent wavelengths corresponded to 840, 870, 954, and 990 nm (Figure 25c). Interestingly, the band situated at 954nm was already proven relevant when differentiating pure Arabica and Robusta based on their water spectral patterns (Figure 23). Similar results (R2=0.95) were obtained by (Núñez *et al.*, 2020) when examining the HPLC-UV fingerprints of brewed Arabica coffee containing Robusta in proportions ranging from 15% to 85%.

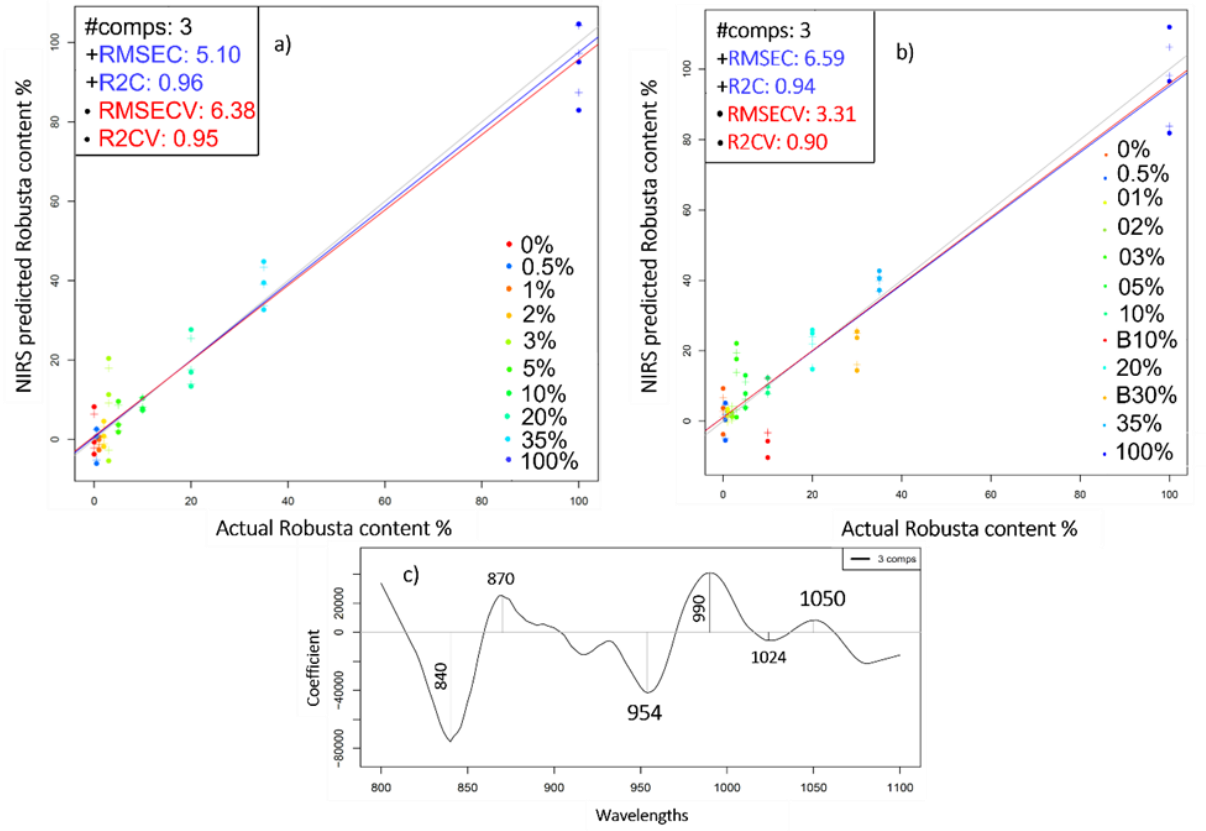

Figure 25: a) Y-fit graph of the prediction of Robusta concentration in liquid mixtures N = 30; b) PLSR analysis on the dataset containing marketed blends B10% and B30% and the mixtures, N=36; c) Regression vector of the predictive model (b) in the 800–1100 nm range. Spectral preprocessing: Averaging consecutive scans, Savitzky-Golay smoothing (window size 19), 1st derivative.

The complexity of differentiating between the mixtures with the lowest adulteration levels was evidenced primarily by their respective water spectral pattern where an overlapping of blends containing Robusta fractions as low as 0.5%, 1%, and 2% occurred. Notably, above these concentration levels, the higher the ratio Robusta to Arabica was, the higher the absorbance in the wavelengths that are characteristic of pure Robusta. Inversely, the lower the added Robusta, the higher the absorbance in the wavelengths characteristic of pure Arabica (Figure 26a).

Assessing whether or not the inclusion of marketed blends B10% and B30% can still be translated into distinctive water spectral patterns, in the presence of lab generated mixtures covering both low (1%) and high blending ratios (35%), was also at-tempted and confirmed the adequacy of the analysis from an aquaphotomics stand-point in terms of highlighting the respective composition of the studied samples. Once again, the intricacy of detecting the lowest blending ratio was reflected by a slight overlapping with pure Arabica extract (Figure 26b).

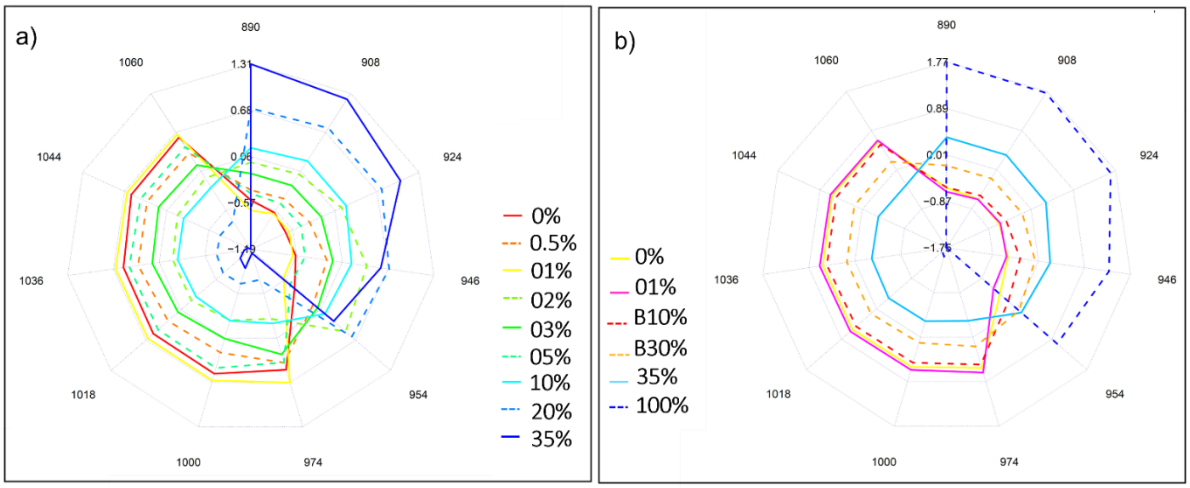

Figure 26: (a) Aquagrams of aqueous solutions of Robusta-Arabica blends in the concentration range of 0.5–35% and pure Arabica (0%) in the second overtone region; (b) Aquagrams of the dataset comprising pure Arabica (0%), pure Robusta (100%), Robusta-Arabica blends (01% and 35%) and marketed blends (B10% and B30%).

5.3. Aquaphotomics for tracking herbs preservation processes

5.3.1. PCA-LDA analysis of ground herbs

Performing linear discriminant analysis on ground samples, both in case of the 1st and 2nd set of herbs has proven effective in terms of categorizing the different varieties, regardless of the applied preservation method, with 100% accurate recognition and prediction rates. Slightly lower but nonetheless good accuracies were obtained when assigning the samples into their corresponding groups, when the class variable used for classification consisted of the performed drying. In this regard, for the 1st set of herbs comprising of rosemary, thyme and marjoram, the PCA-LDA models allowed for respective recognition and prediction rates estimated at 88.88% and 85.72%, in the 2nd overtone and equal to 94.18% and 85.7% in the 1st overtone (Figure 27).

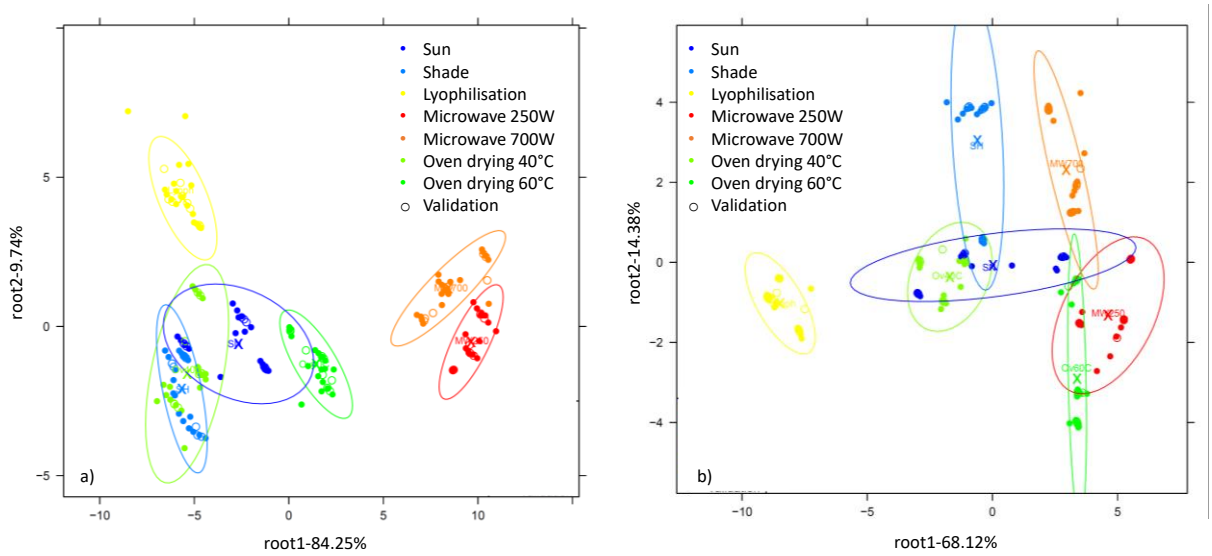

Figure 27: (a) PCA-LDA classification plot of the 1st set of preserved and ground herbs (rosemary, thyme and marjoram) at the 1st overtone region (1300–1600 nm) N = 189 , NrPCs =22, (b) PCA-LDA classification plot at the 2nd overtone region (800–1100 nm), N = 189, NrPCs =16 .

In case of the 2nd set of herbs, the respective recognition and cross-validation accuracies amounted to 98.88% and 98.6% when basing the analysis on the spectral data obtained in the 1st water overtone region and 98.74% and 98.61% when based on the data generated in the 2nd overtone (Figure 28). PCA-LDA models built on each herb individually gave 100% accurate recognition and cross-validation rates in both ranges.

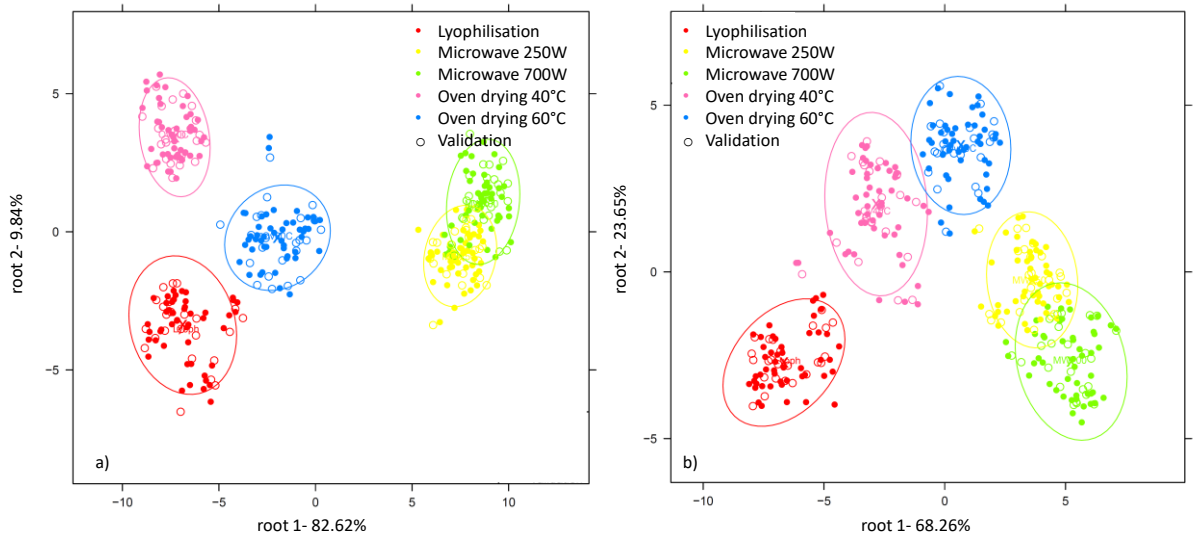

Figure 28: (a) PCA-LDA classification of the 2nd set of preserved and ground herbs at the 1st overtone region (1300–1600 nm) N = 360, NrPCs =22, (b) PCA-LDA classification plot at the 2nd overtone region (800–1100 nm), N = 360, NrPCs = 23.

When inspecting where misclassification might have occurred in the 2nd overtone range, we observed that in 1.38% of the cases, herbs that were microwave dried at 250W were misclassified as belonging either to the group of herbs oven dried at 60°C or to those microwave-dried at 700W. Other misclassifications also occurred in case of microwave dried herbs (700W) which were assigned to the group of microwave-dried herbs (250W) in 4.17 % of the cases. On the other hand, data analysed in the 1st overtone range showcased that 6.96% of the herbs dried using microwave drying at 250W were attributed to the group of herbs dried at the higher microwave power of 700W.

Table 15: PCA-LDA classification model on the NIRS data of 2nd set of preserved herbs after three-fold cross-validation (N = 360).

		Validation Accuracy				
		Preservation method				
		LYO	MW 250W	MW 700W	OV 40°C	OV60°C
Preservation method	LYO	**100%**	0%	0%	0%	0%
	MW 250W	0%	**97.25%**	4.17%	0%	0%
	MW 700W	0%	1.38%	**95.83%**	0%	0%
	OV 40°C	0%	0%	0%	**100%**	0%
	OV 60°C	0%	1.38%	0%	0%	**100%**

LYO: Lyophilisation; MW 250W: Microwave drying(250W); MW 700W: Microwave drying(700W); OV 40°C: Oven drying(40°C); OV 60°C: Oven drying(60°C)

5.3.2. Spectral inspection of the herbal extracts:

The first step in assessing the spectral fingerprints of the herbal extracts in the NIR range and more specifically in the 1st overtone range of water consisted of subtracting the average water spectrum from the average spectra of the fresh herbs. The obtained subtracted spectra are portrayed in Figure 29. Essentially, the different herbs were characterized by a peak at around 1413 nm and a broad band over the 1450-1550nm, all of which having differing absorbance intensities. While most of the herbs shared a similar spectral pattern, Greek oregano had a peculiar one.

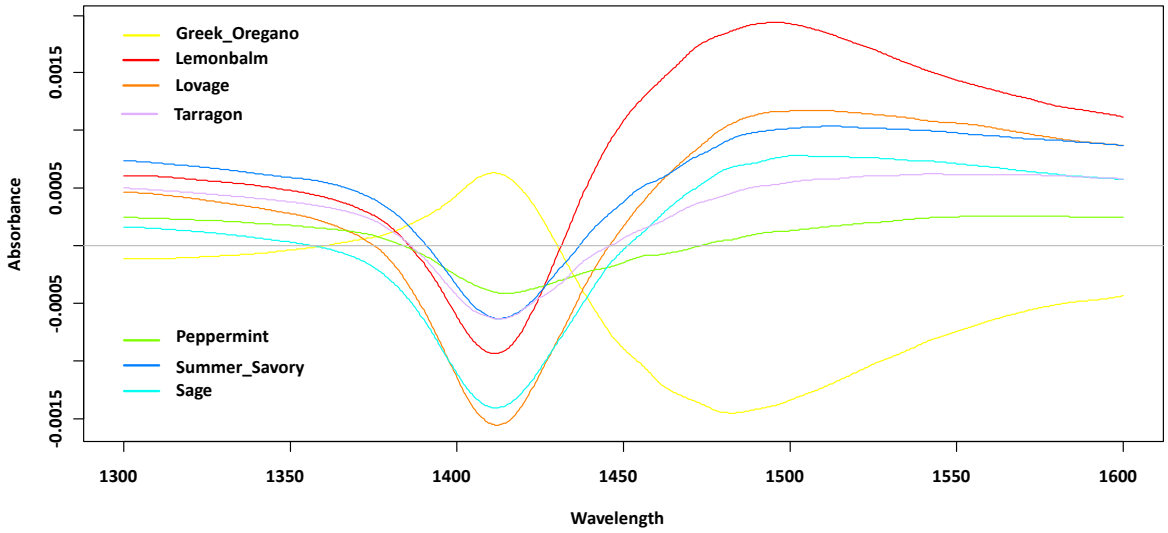

Figure 29: Difference spectra of the fresh herbal extracts obtained by subtraction from the average pure water spectrum in the spectral range 1300–1600 nm.

5.3.3. Determination of dry matter content of the herbal extracts

PLSR predictive model constructed to determine if the herbal extracts actual dry matter content can be correlated to that predicted based on their spectral fingerprint is presented in Figure 30 .

Considering that the displayed model takes into account all studied herbs from the 2nd set we can presume that the disparities in the dry matter content between the herbs preserved using the same preservation method which can be highlighted in the Y-fit graph of the prediction are due to the high variability in terms of inspected herbs.

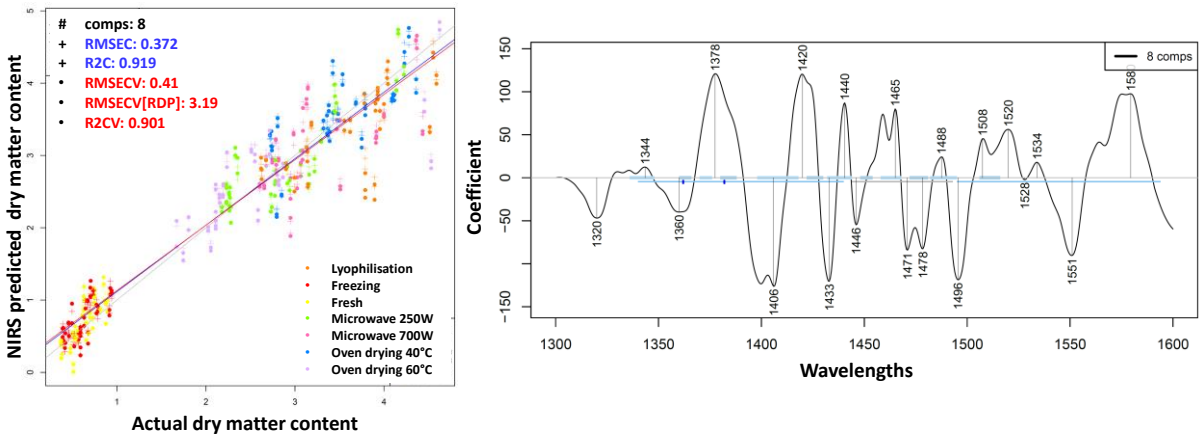

Figure 30: PLSR analysis on the 2nd set of herbal extracts for the prediction of the dry matter content (mg/ml); Regression vector of the predictive model in the 1300–1600 nm range. N=345, LV=8, Spectral Preprocessing: Savitzky-Golay smoothing (window size 37).

The model is characterized by an R^2CV amounting to 0.901, an RDP of 3.19 and an error of cross validation RMSECV of 0.41mg/ml thus attesting to the good correlation between the dry matter content estimated using the reference method to that predicted using near infrared spectroscopy. According to the corresponding regression vector, the contributing wavelengths consisted of 1320nm,1360nm, 1406nm, 1420nm, 1433nm, 1440nm, 1446nm, 1465nm, 1471nm, 1478nm, 1488nm, 1496nm, 1508nm, 1520nm, 1534nm, 1551nm, and 1580nm.

While the dry water content of freeze dried and fresh samples falls within the same range, what the water spectral pattern indicates is that the water contained in these samples has distinctive structure. The aquagrams of the whole sample set arranged according to their preservation method proves the prevalence of less hydrogen-bonded water structures in freeze-dried samples as opposed to more hydrogen-bonded water in fresh herbal extracts (Figure *31*).

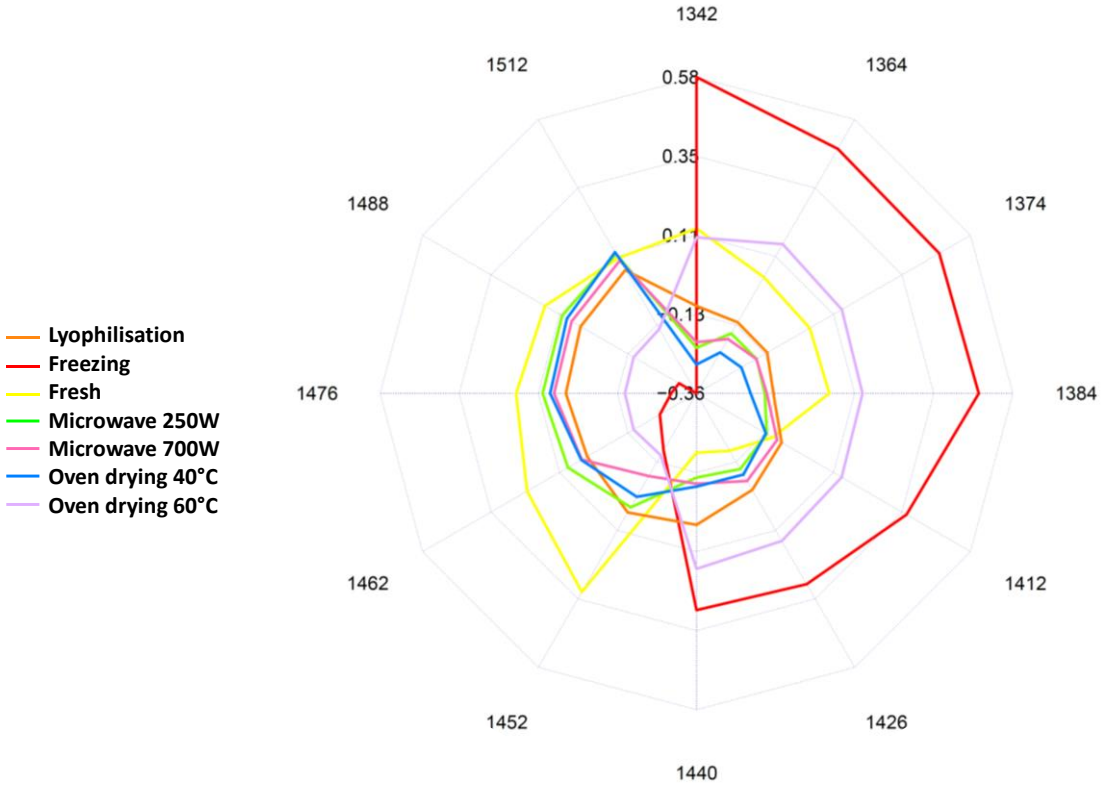

Figure 31: Aquagrams of the second set of herbal extracts according to their preservation method at the first overtone of water (1300-1600nm).

When assessing the aquagrams of the first set of herbs (marjoram, thyme and rosemary), peculiar water spectral patterns were obtained for each, confirming the intricacy of each of the studied herbs when subjected to the same preservation method.

For instance, marjoram's aquagram, portrayed in Figure 32, presented the most obvious variation in the water species affected by each preservation process. In this regard, lyophilized samples were

characterized by the abundance of less hydrogen bonded water whereas frozen samples presented higher absorbance values at wavelengths associated with water structures having strong hydrogen bonds.

Remarkably, within each preservation pair [(fast and slow freezing), (oven drying at 40°C and 60°C), (sun and shade drying), (microwave drying at 250W and 700W)] the water spectral pattern presented similarities, with slight differences in the intensity of absorbance.

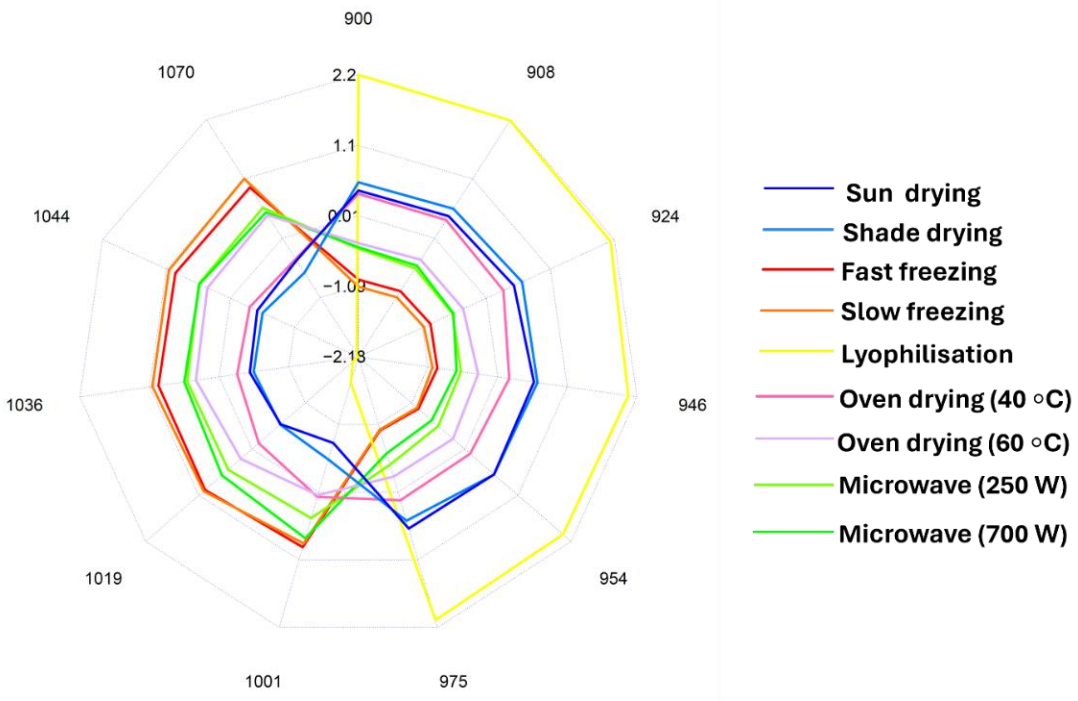

Figure 32: Aquagrams of marjoram extracts according to the applied preservation method in the 2nd overtone 800-1100 nm, N=81

For thyme and rosemary, the aquagrams presented in Figure 33, reflected similarities between the water spectral patterns of extracts prepared from oven dried samples and those that underwent microwave drying. Again, congruent with the results obtained for marjoram, frozen herbs were predominantly marked by higher absorbances, specifically at the range from 1001nm to 1070nm, proven to be linked to strongly hydrogen bonded water (McGlone and Kawano, 1998). Other studies involving the application of super chilling have equally ascertained shifts in the range 900-1070 nm in the spectra of studied matrices (Grassi *et al.*, 2024).

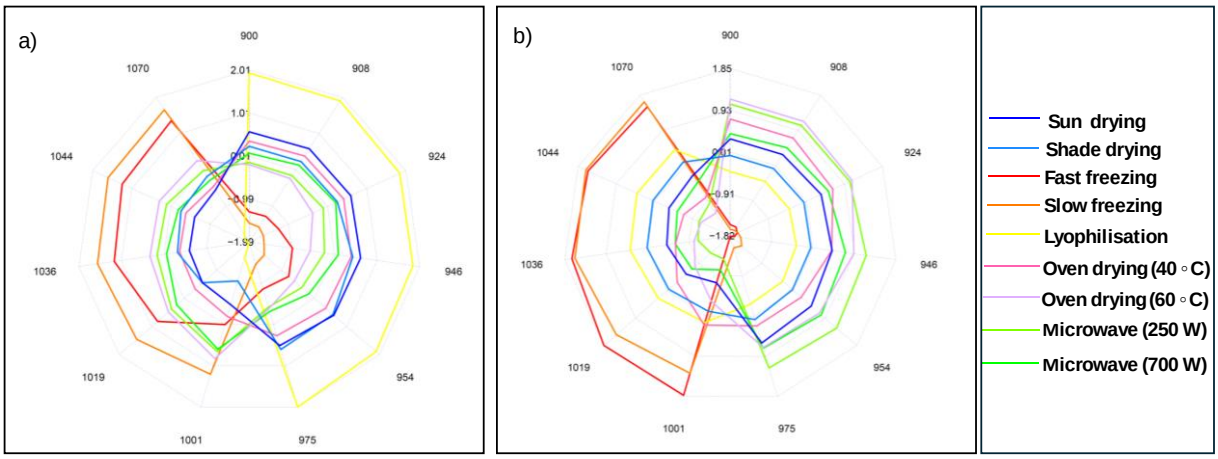

Figure 33:Aquagrams of thyme (a) and rosemary (b) in the 2nd overtone range 800-1100nm, according to the applied preservation method, N=81

Lyophilisation, on the other hand had a more pronounced effect on thyme (Figure 33a) than rosemary (Figure 33b), which translated in a similar manner to its effect on marjoram in the prevalence of less-hydrogen bonded water conformations.

If anything, these results attest to the how dynamic is the water contained in herbs, in response to perturbations such as the different preservation methods and their intensities and how each herb has a peculiar behaviour to such systematic alterations.

5.3.4. PCA-LDA analysis of the herbal extracts

Hybrid PCA-LDA classification of basil extracts based on their spectral fingerprint in the 1st water overtone range and according to the applied preservation method showed an overlapping of fresh and frozen samples but also between the samples that underwent microwave drying at 250W and 700W. The overall recognition and prediction accuracies corresponded to 91.26% and 90.47% respectively.

The corresponding classification plot is portrayed in Figure 34 and showcases the between group variance expressed in the root1 and root2, amounting to 95.71% and 3.62%, respectively.

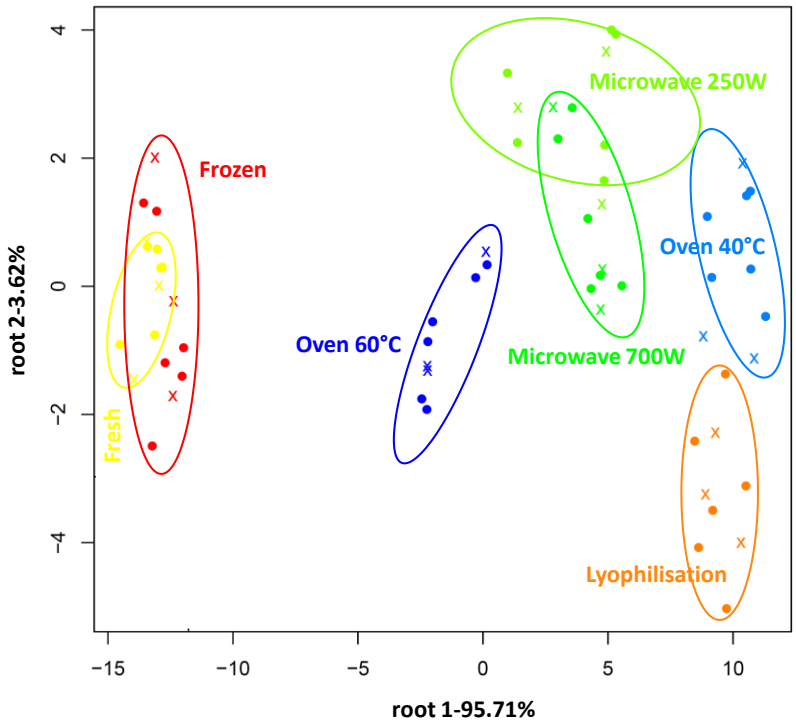

Figure 34: PCA-LDA classification plot of the basil extracts at the first water overtone region (1300-1600 nm), N = 63, NrPCs = 5

The following table (Table *16*) traces down the occurrence of misclassification, both at the validation and cross-validation levels. During the model validation, 27.83% of the fresh basil extracts were erroneously classified as frozen while 16.67% of frozen basil extracts were thought to be prepared from fresh basil.

The very same number of samples (equivalent to 13.67%) was attributed to microwave dried samples (700W) when in fact these samples were microwave dried at lower power of 250W. During cross-validation, fresh basil extracts were misclassified as frozen in 33.33% of the cases whereas extracts prepared using frozen basil were considered as fresh based extracts in 22.33% of the studied cases.

Consistently with the overlapping showcased in the classification plot, 11% of herbal extracts that underwent microwave drying at 700W were miscategorized as those dried at 250W.

Table 16: PCA-LDA classification model on the NIRS data of basil extracts after three-fold cross-validation (N =63, NrPCs =5).

Recognition Accuracy (91.26%)								
Preservation method								
Preservation method		FSH	FRZ	LYO	MW250W	MW700W	OV 40°C	OV 60°C
	FSH	**72.17**	16.67	0	0	0	0	0
	FZN	27.83	**83.33**	0	0	0	0	0
	LYO	0	0	**100**	0	0	0	0
	MW 250W	0	0	0	**100**	16.67	0	0
	MW 700W	0	0	0	0	**83.33**	0	0
	OV 40°C	0	0	0	0	0	**100**	0
	OV 60°C	0	0	0	0	0	0	**100**
Validation Accuracy (90.47%)								
Preservation method		FSH	FRZ	LYO	MW250W	MW700W	OV 40°C	OV 60°C
	FSH	**66.67**	22.33	0	0	0	0	0
	FZN	33.33	**77.67**	0	0	0	0	0
	LYO	0	0	**100**	0	0	0	0
	MW 250W	0	0	0	**100**	11	0	0
	MW 700W	0	0	0	0	**89**	0	0
	OV 40°C	0	0	0	0	0	**100**	0
	OV 60°C	0	0	0	0	0	0	**100**

FSH: Fresh; FZN: Frozen; LYO: Lyophilised; MW 250W: Microwave-dried(250W); MW 700W: Microwave-dried(700W); OV 40°C: Oven dried(40°C); OV 60°C: Oven dried(60°C)

Better classification performance was obtained in case of Greek-oregano extracts with 96.02% and 95.23% average recognition and prediction accuracies of the PCA-LDA models. Only groups corresponding to frozen and fresh herbal extracts were overlapped in about 22.33% and 11% of the cases, respectively. The corresponding classification plot is shown in Figure *35*. The between group variance expressed in the root1 and root2, amounts to 91.5% and 7.18%, respectively. Better separation can be observed along the axis of the first root.

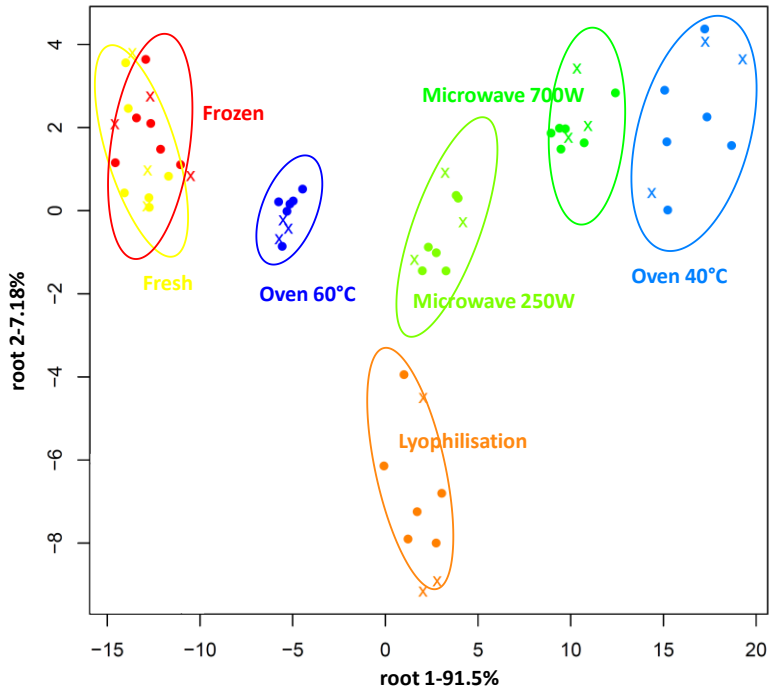

Figure 35: PCA-LDA classification plot of Greek oregano extracts at the 1st overtone region (1300–1600 nm), N = 63, NrPCs =9

For lemon balm extracts, average performance of the PCA-LDA model was obtained and more overlapping of the drying pairs was observed. The corresponding recognition and prediction accuracies were estimated at 80.16% and 74.62%, respectively.

As per the between group variance expressed in the root1 and root2, it amounted to 93.14% and 4.46%, respectively.

The classification plot (Figure *36*) showcases the peculiarity of lyophilised samples, quite distanced from the rest of the samples but also reveals the separation of the samples into two separate groups. The first group comprises extracts prepared from frozen and fresh lemon balm while the second is composed of oven-dried and microwave dried extracts. Presumably, freezing preserves the most lemon balm herbs, hence the similarity between these two sample sets.

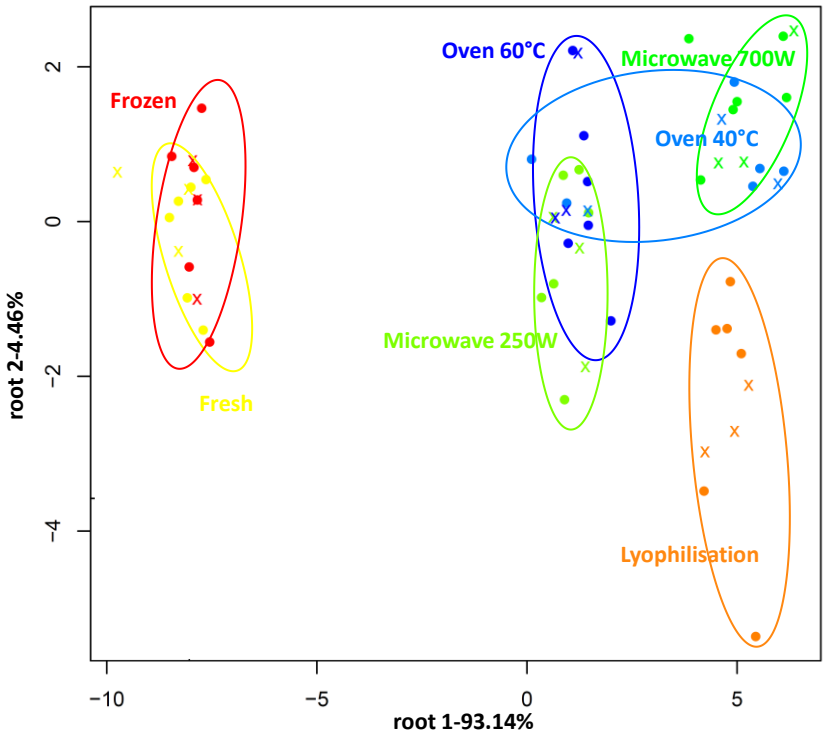

Figure 36: PCA-LDA classification plot of Lemon balm extracts at the 1st overtone region (1300–1600 nm), N = 63, NrPCs =6

Based on the spectral data of lovage extracts in the 1st overtone range, the PCA-LDA models were characterized by an even more pronounced overlapping of the different preservation methods translated in recognition and prediction rates of 84.13% and 81.01%.

In Table 17, a detailed overview of the degree of misclassification between the seven different sample groups is provided.

During the initial phase of PCA-LDA model setting, the highest percentage of overlapping occurred between fresh and frozen lovage extracts (33.33 %) whereas lovage samples undergoing oven drying at 40° C were 100% accurately classified. Extracts obtained from microwave dried samples (250W and 700W) were reciprocally misclassified, in 16.67% and 11.17% of the cases, respectively.

Table 17: PCA-LDA classification model on the NIRS data of lovage extracts after three-fold cross-validation (N =63, NrPCs =6).

Recognition Accuracy (84.13%)							
	Preservation method						
Preservation method		FSH	FRZ	LYO	MW250W	MW700W	OV40°C	OV 60°C
	FSH	**66.7**	16.7	0	0	0	0	0
	FZN	33.3	**83.3**	0	0	0	0	0
	LYO	0	0	**88.98**	0	0	0	5.5
	MW250W	0	0	0	**83.33**	11.17	0	16.67
	MW700W	0	0	0	16.67	**88.83**	0	0
	OV40°C	0	0	5.51	0	0	**100**	0
	OV60°C	0	0	5.51	0	0	0	**77.83**
Validation Accuracy (81.13%)							
	Preservation method						
Preservation method		FSH	FRZ	LYO	MW250W	MW700W	OV40°C	OV60°C
	FSH	**77.67**	22.33	0	0	0	0	0
	FZN	22.33	**77.67**	0	0	0	0	0
	LYO	0	0	**77.93**	0	0	0	11.04
	MW250W	0	0	11.04	**89**	11.04	0	11.04
	MW700W	0	0	0	11	**77.93**	11	0
	OV40°C	0	0	11.04	0	0	**89**	0
	OV60°C	0	0	0	0	11.04	0	**77.93**

FSH: Fresh; FZN: Frozen; LYO: Lyophilised; MW 250W: Microwave-dried(250W); MW 700W: Microwave-dried(700W); OV 40°C: Oven dried(40°C); OV 60°C: Oven dried(60°C)

For peppermint, PCA-LDA classification proved good in both the 1st and 2nd overtone ranges with respective recognition and prediction accuracies of 98.4% and 95.27% for the former, 97.61% and 96.81% for the latter. 22.33% of fresh peppermint extracts were mistakenly categorized as frozen. As indicated in the classification plot, intergroup distances hint at similarities between extracts prepared from microwave dried peppermint at 700W and from oven dried peppermint at 40°C but also between those oven-dried at 60°C and microwave dried at 250W.

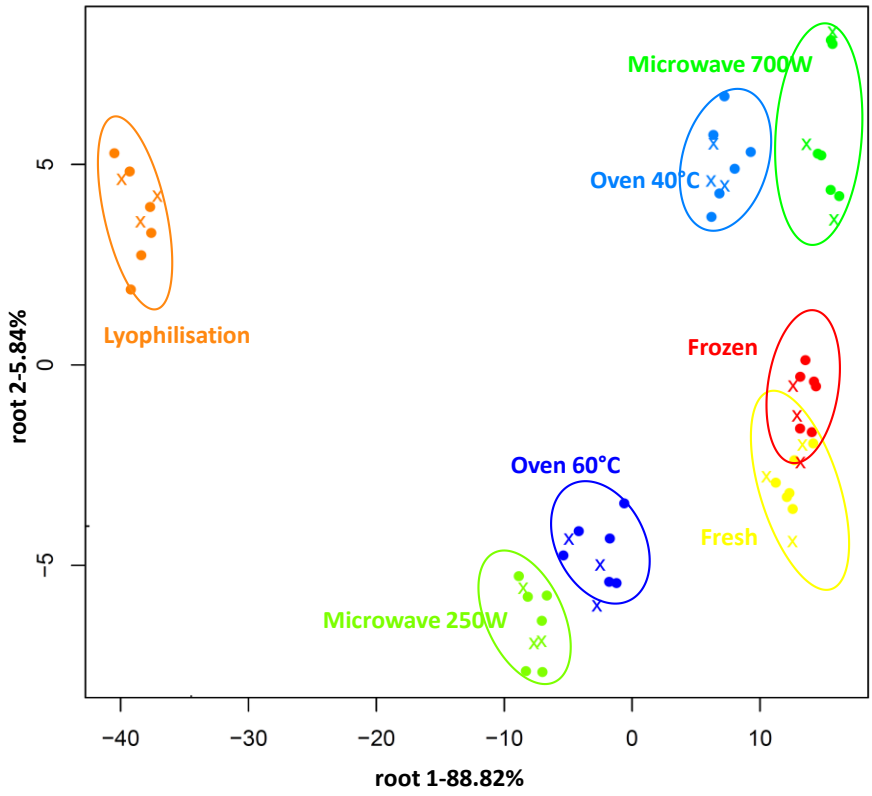

Figure 37: PCA-LDA classification plot of peppermint extracts at the 1st overtone region (1300–1600 nm), N = 63, NrPCs = 11

Differentiating the different sage-based extracts according to the applied preservation method was more effective in the 2nd water overtone range where the average recognition and prediction rates amounted to 94.45% and 93.66% compared to 85.71% and 84.19% when basing the analysis on the spectral features obtained in the 1st water overtone range.

The corresponding classification plot signals an overlapping of the pairs (fresh and frozen), (microwave drying 700W and microwave drying 250W) and (Oven drying 60°C and Oven drying 40°C). Nonetheless, what the confusion matrix indicated was solely a misclassification between fresh and frozen and microwave dried at 700W and at 250W herbal extracts in 33.33% 11% of the cases respectively, during cross-validation.

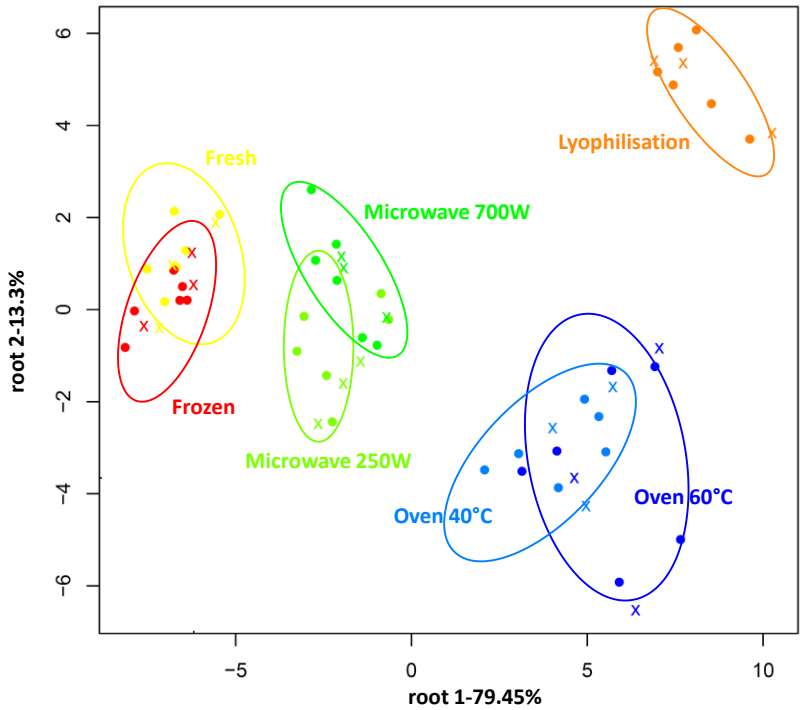

Figure 38:PCA-LDA classification plot of sage extracts at the 1st overtone region (1300–1600 nm), N = 63, NrPCs =8

Table 18: PCA-LDA classification model on the NIRS data of sage extracts after three-fold cross-validation (N =63, NrPCs =8).

		Recognition Accuracy (94.45%)						
		Preservation method						
		FSH	FRZ	LYO	MW250W	MW70W	OV40C	OV 60°C
Preservation method	FSH	**66.67**	5.5	0	0	0	0	0
	FZN	33.33	**94.5**	0	0	0	0	0
	LYO	0	0	**100**	0	0	0	0
	MW250W	0	0	0	**100**	0	0	0
	MW700W	0	0	0	0	**100**	0	0
	OV40°C	0	0	0	0	0	**100**	0
	OV60°C	0	0	0	0	0	0	**100**
		Validation Accuracy (93.66%)						
		FSH	FRZ	LYO	MW250W	MW700W	OV40C	OV 60°C
Preservation method	FSH	**66.67**	0	0	0	0	0	0
	FZN	33.33	**100**	0	0	0	0	0
	LYO	0	0	**100**	0	0	0	0
	MW250W	0	0	0	**89**	0	0	0
	MW700W	0	0	0	11	**100**	0	0
	OV40°C	0	0	0	0	0	**100**	0
	OV60°C	0	0	0	0	0	0	**100**

FSH: Fresh; FZN: Frozen; LYO: Lyophilised; MW 250W: Microwave-dried(250W); MW 700W: Microwave-dried(700W); OV 40°C: Oven dried(40°C); OV 60°C: Oven dried(60°C)

When evaluating tarragon aqueous extracts, 100% accurate assignment of the samples to their corresponding groups was obtained in both water overtone ranges. This was ascertained during the inspection of the classification plot with no apparent overlapping of the different groups of samples (Figure *39*).

The groups of fresh, frozen, lyophilized, and microwave dried tarragon extracts, although not intersecting, are positioned in close proximity, with minimal interclass distances. This wasn't the case for oven dried tarragon extracts, which were clearly distanced from the remaining extracts.

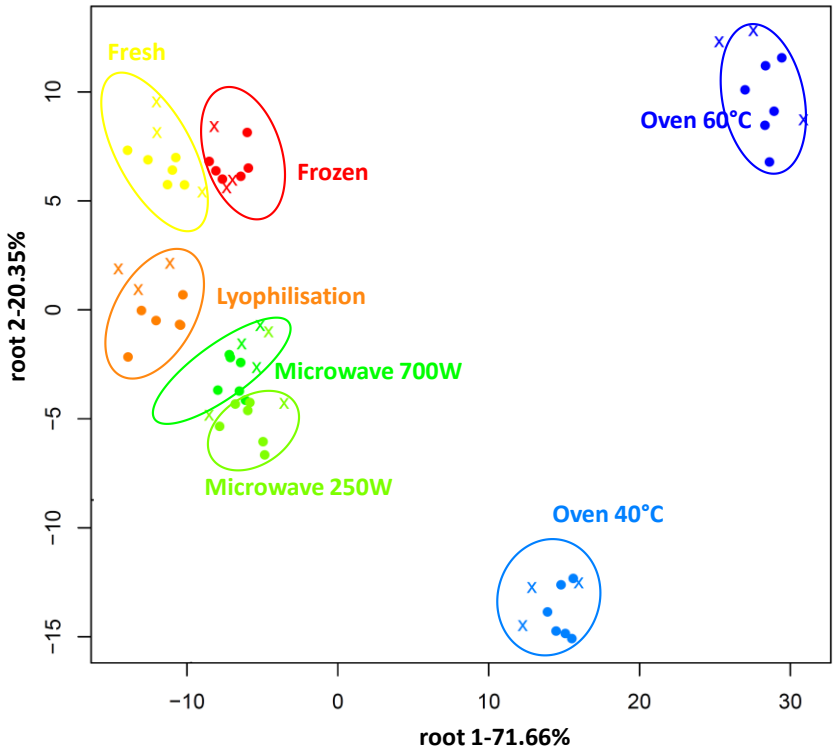

Figure 39: PCA-LDA classification plot of tarragon extracts at the 1st overtone region (1300–1600 nm), N = 63, NrPCs =15

In case of extracts obtained after preserving summer savory, what we have noticed is that, apart from extracts of frozen and fresh herbs, all other extracts were accurately assigned to their respective groups. 11% of the fresh summer savory extracts sere misidentified as prepared from the frozen herb. The overall prediction rate consisted of 98.42%.

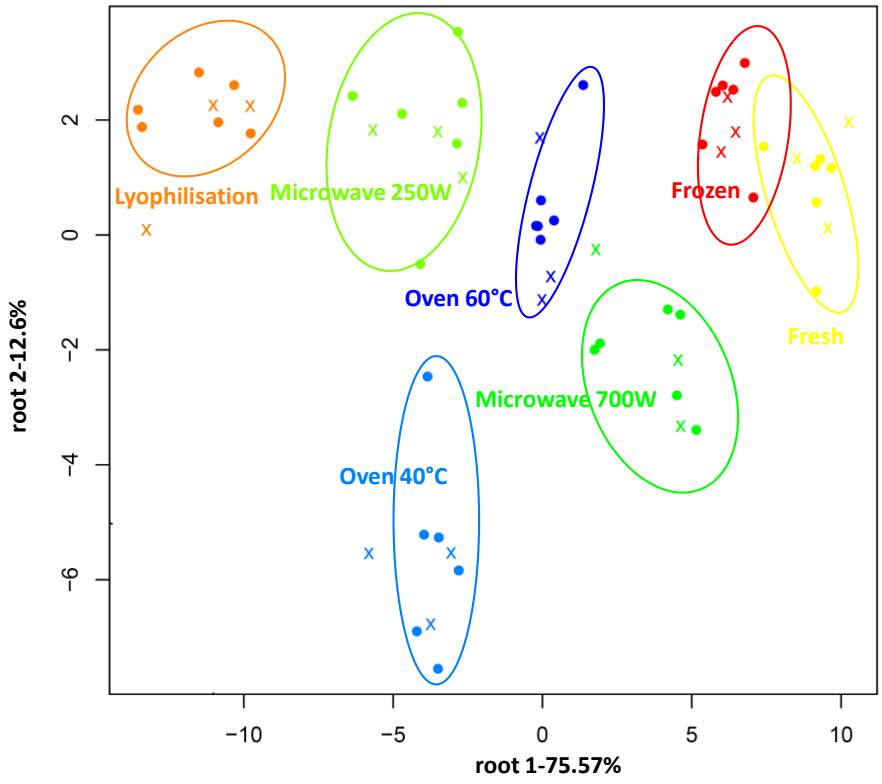

Figure 40: PCA-LDA classification plot of summer savory extracts at the 1st overtone region (1300–1600 nm), N = 63, NrPCs =13

5.3.5. PLSR prediction of TPC and TAC

Regardless of the applied preservation method, an increase in the total antioxidant capacity (TAC) was observed, in the case of Greek oregano extracts. Oven drying this perennial herb at 40°C was found to induce the highest increase in TAC, followed by both microwave drying processes, lyophilization, oven drying at 60°C and freezing.

The PLSR model confirmed that the NIR predicted antioxidant capacity could be correlated with the one measured with the reference method. The R^2 of the cross-validated model corresponded to 0.97 while the error of cross-validation was estimated at 11.06 mg AAE/g d.w. The ratio of performance to deviation amounted to 5.88, corroborating the usefulness of the method for quantitative determination of the TAC of Greek oregano extracts.

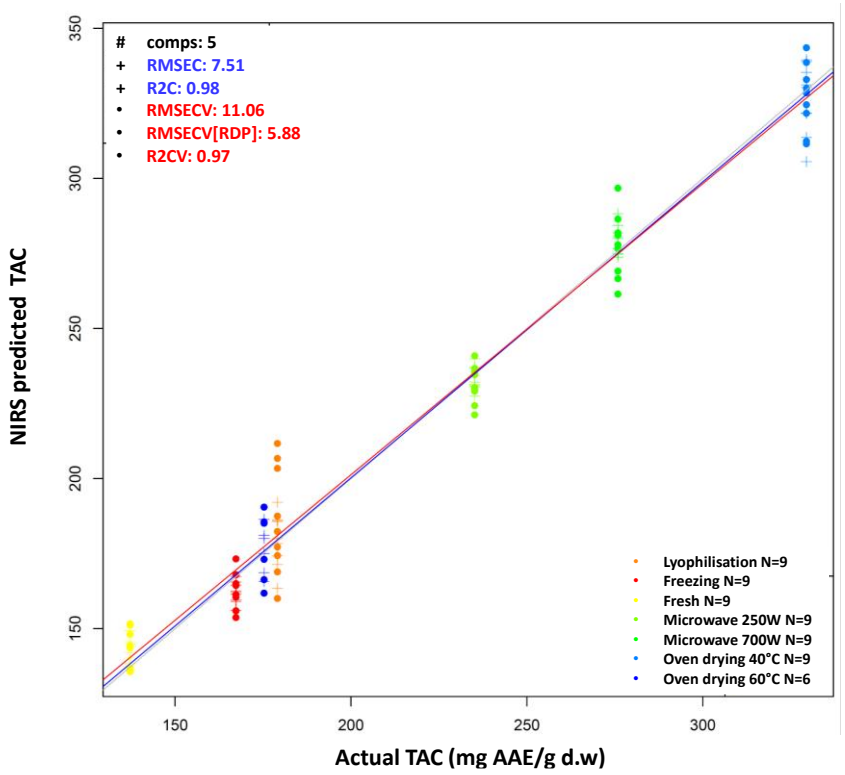

Figure 41: PLSR analysis on Greek oregano extracts for the prediction of TAC (mg AAE/g d.w); N=60, LV=5, Spectral Preprocessing: Savitzky-Golay smoothing (window size 31).

As per the wavelengths contributing the most to the obtainment of this model, they consisted of 1412nm, proven to be particularly sensitive to temperature changes (Matija *et al.*, 2012), 1430nm, which is particularly related with the water hydration band, 1448nm, 1476nm, conferred to semi-crystalline state of cellulose and 1500nm and 1566nm (Muncan *et al.*, 2022).

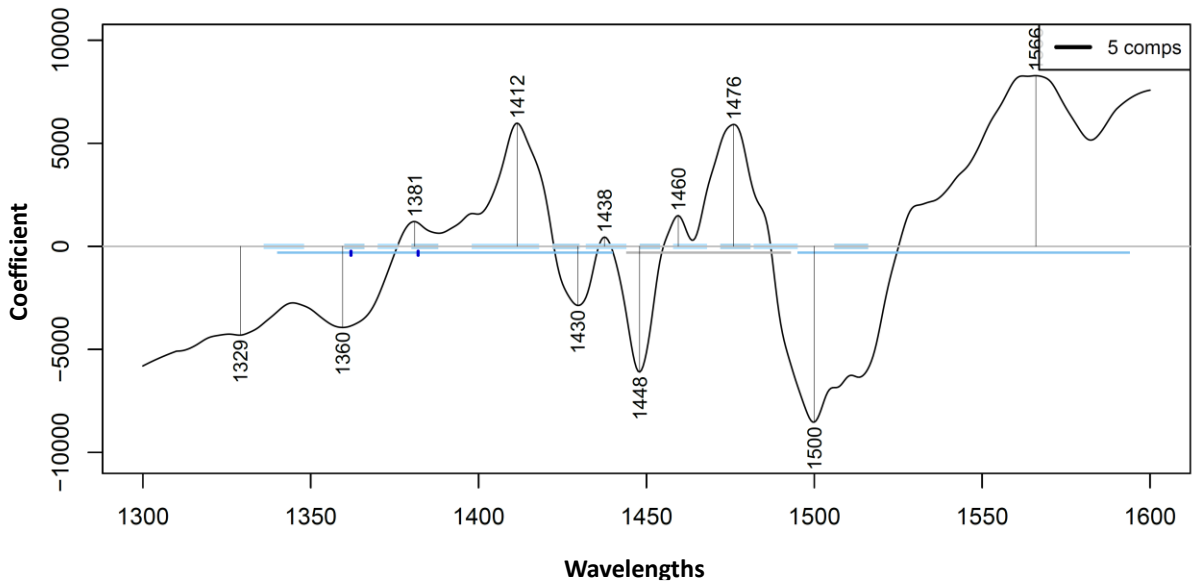

Figure 42: Regression vector of the predictive model of TAC in Greek oregano in the first overtone 1300–1600 nm range.

While the estimation of total polyphenols by means of PLSR was less accurate, it did indicate the potential correlation between NIRS predicted TPC values and those measured conventionally with an R^2 CV of 0.93 and RMSECV of 8.2 mg GAE/g d.w).Once again, of all the applied preservation processes, the effect of oven drying (40°C) on Greek oregano was the most pronounced whereas lyophilization had a minimal effect on the content of total polyphenols in the extracts.

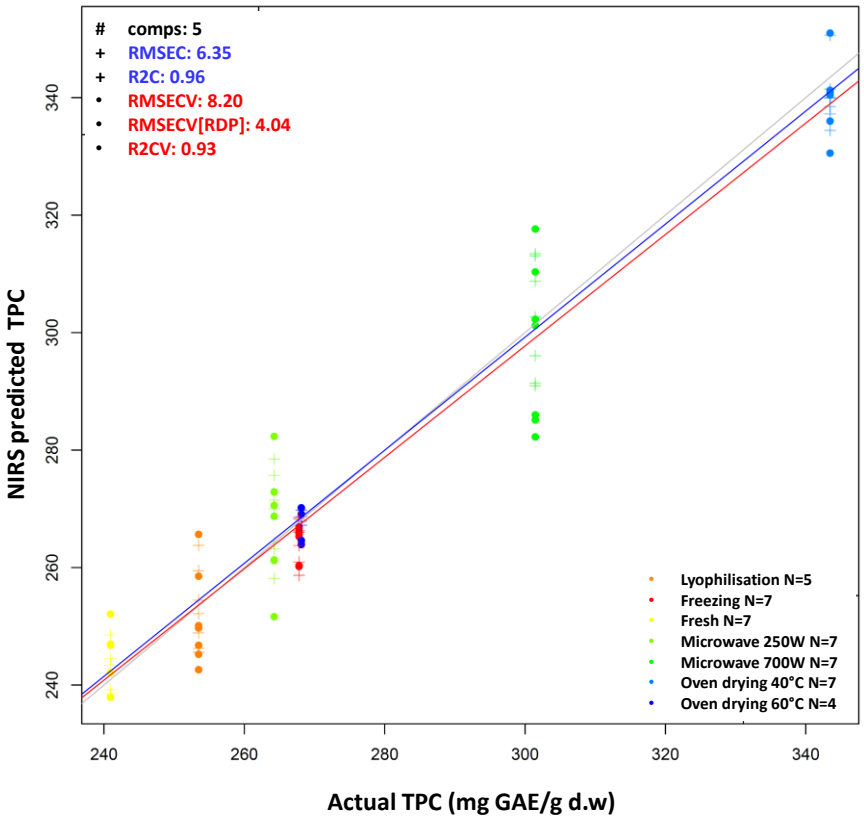

Figure 43: PLSR analysis on Greek oregano extracts for the prediction of the TPC (mg GAE/g d.w); N=44, LV=5, Spectral Preprocessing: Savitzky-Golay smoothing (window size 31).

The regression vector portrayed in Figure 44 showcases the most influential wavelengths in predicting the total polyphenols in Greek oregano extracts. Amongst these, we find bands, which, according to literature were found to be related to free water, free OH- or trapped water (1398nm, 1410nm), others corresponding to water hydration (1422nm), others characteristic of $\nu 2 + \nu 3$ and water solvation shell structures (1448nm) but also more hydrogen bonded water at 1468nm, 1478nm, 1494nm and 1528nm (Xantheas, 1995; Roumiana Tsenkova, Muncan, *et al.*, 2018b).

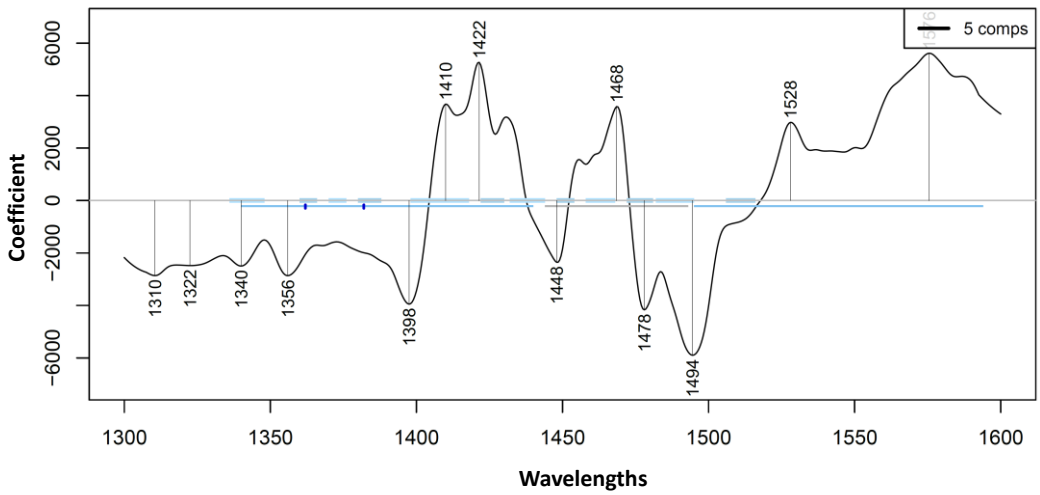

Figure 44: Regression vector of the predictive model of TPC in Greek oregano in the first overtone 1300–1600 nm range.

For lemon balm, the NIR-predicted polyphenols content was also found to be strongly correlated with the values obtained by the reference method, as validated by the PLSR model. The cross-validated model exhibited an R^2CV of 0.96, and the cross-validation error was estimated at 15.98 mg GAE/g d.w. The performance-to-deviation (RDP) ratio stood at 5.42, underscoring the method's efficacy in quantitatively determining the total polyphenols content (TPC) of lemon balm extracts. A discernable increase of the total polyphenols content was evident in all the lemon balm extracts except for those that underwent freezing (Figure 45).

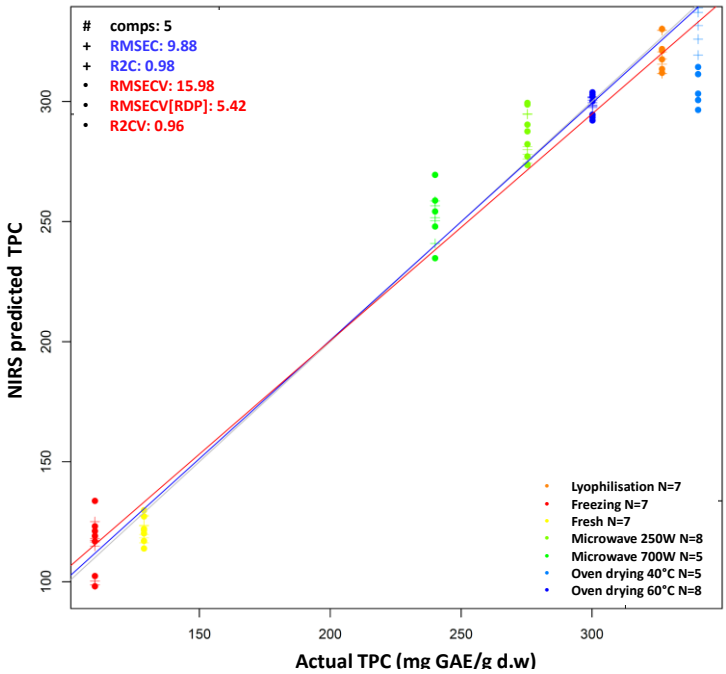

Figure 45: PLSR analysis on lemon balm extracts for the prediction of the TPC (mg GAE/g d.w); N=, LV=5, Spectral Preprocessing: Savitzky-Golay smoothing (window size 31).

By examining the corresponding regression vector (Figure 46), we can identify the contributing wavelengths to be associated to different water conformations such as 1340nm, 1364nm, 1400nm, 1438nm, 1458nm,1476nm, 1490nm, 1504nm, 1521nm, 1552nm, 1570nm.

Research has shown that the band located at 1476nm is related to the semi-crystalline regions in cellulose, a major constituent of plant material that plays a crucial role in maintaining their structure and contributing to their nutritional fiber content(Muncan *et al.*, 2022). The band at 1438nm is representative of third state of water, whose concentration slightly changes with temperature(Segtnan, Isaksson and Ozaki, 2001).

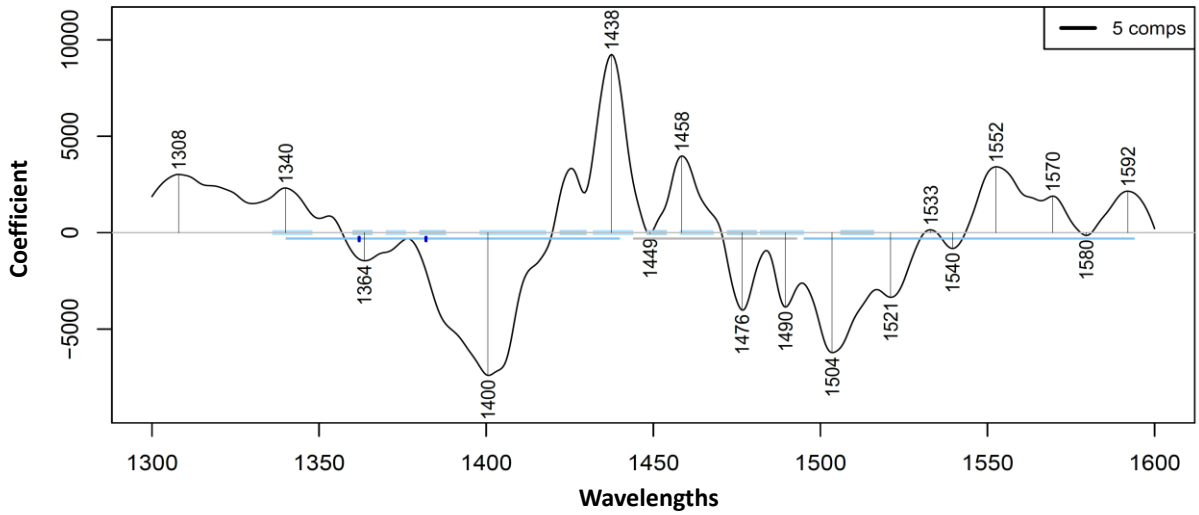

Figure 46: Regression vector of the predictive model of TPC in lemon balm in the first overtone 1300–1600 nm range.

In terms of the effect of the differing preservation processes on the total antioxidant capacity of lemon balm extracts, what the y-fit graph (Figure 47) showcases is a decrease in the TAC, following the preservation by oven-drying at 60°C or freezing, compared to the initial content in fresh herb extracts. On the other hand, carrying out microwave drying at low (250W) or high (700W) wattage, oven drying at 40°C or lyophilization had the opposite effect, by increasing the total antioxidant capacity to reach at best 400 mg AAE/g d.w.

The PLSR model corroborated the correlation between the total antioxidant capacity predicted based on the NIR spectral fingerprint of the tested extracts and the values calculated using the reference method.

The cross-validated model was characterized by an R^2 CV of 0.966 whereas the RMSECV corresponded to 24.05 mg AAE/g d.w.

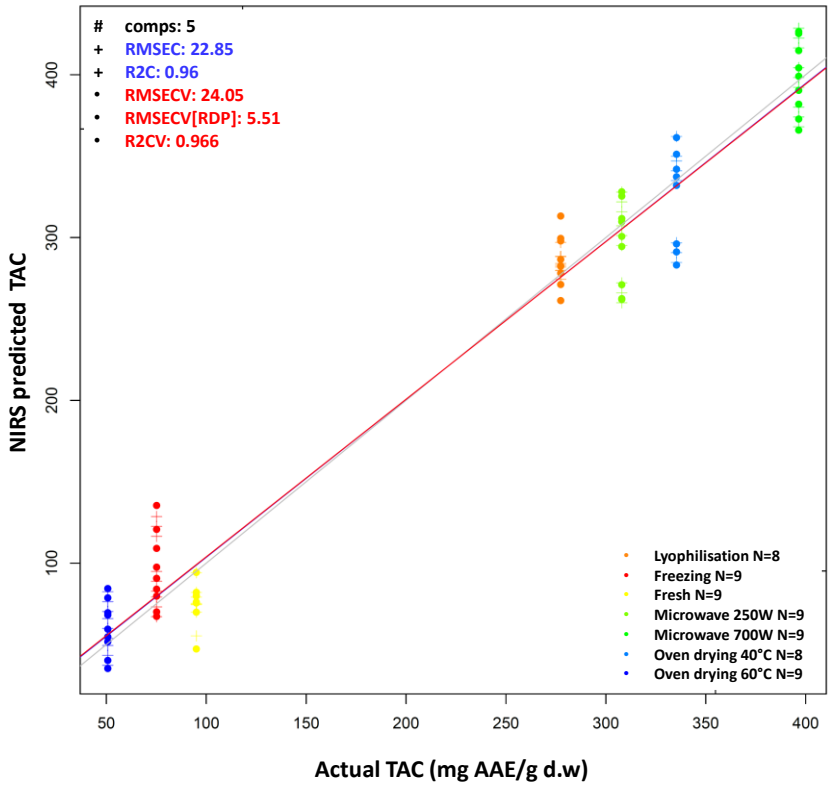

Figure 47: PLSR analysis on lemon balm extracts for the prediction of the TAC (mg AAE/g d.w); N=61, LV=5, Spectral Preprocessing: SG smoothing (window size 27), 1st derivative.

Most contributing wavelengths can be identified in the regression vector (Figure 48) as 1410nm, 1436nm, 1446nm, 1456nm, 1468nm, 1476nm, 1493nm, 1510nm, 1510nm, 1520nm, 1530nm, 1540nm. Aquaphotomics based research has already assigned these peaks to specific water matrix coordinates WAMACS associated with free, weakly hydrogen bonded and strongly hydrogen bonded water structures.

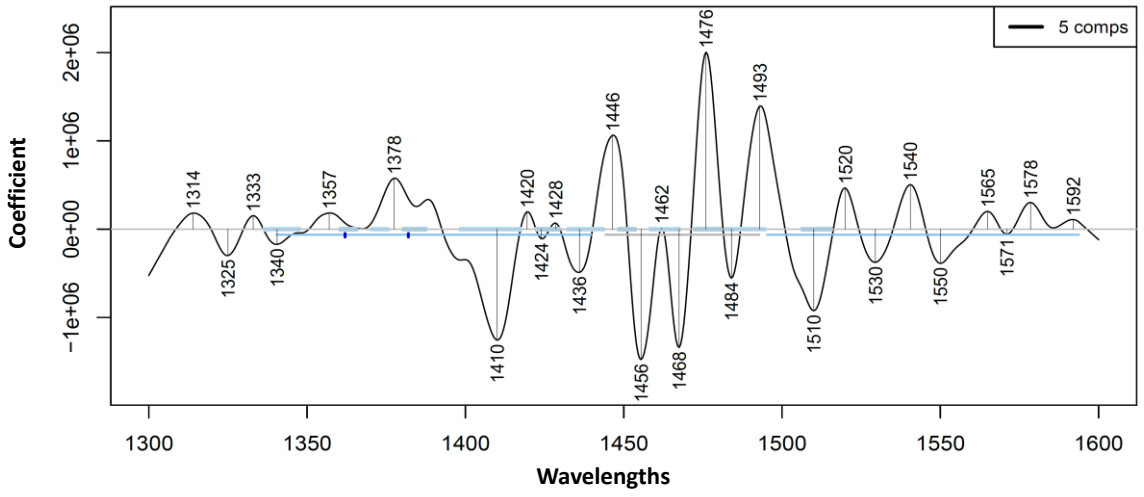

Figure 48: Regression vector of the predictive model of TAC in lemon balm in the first overtone1300–1600 nm range.

Fresh peppermint, another one of the tested herbs, presented the lowest total polyphenols content, estimated at 11.4 mg GAE/g d.w.

As evidenced in Figure 49, subjecting peppermint to oven drying (40°C and 60°C), microwave drying (250W and 700W) as well as lyophilization did increase the TPC in each of the corresponding extracts. On the other hand, freezing reduced the total polyphenols in the resultant extract.

In comparison to fresh peppermint extracts, oven drying at 40°C produced the highest rise in total polyphenols while lyophilization and microwave drying at 250W both resulted in a proportional elevation of TPC content.

PLSR model metrics confirmed the correlation between actual TPC values with those predicted from NIR generated data, with an R^2 of cross validation of 0.98 and an RMSECV amounting to 0.87 mg GAE/g d.w.

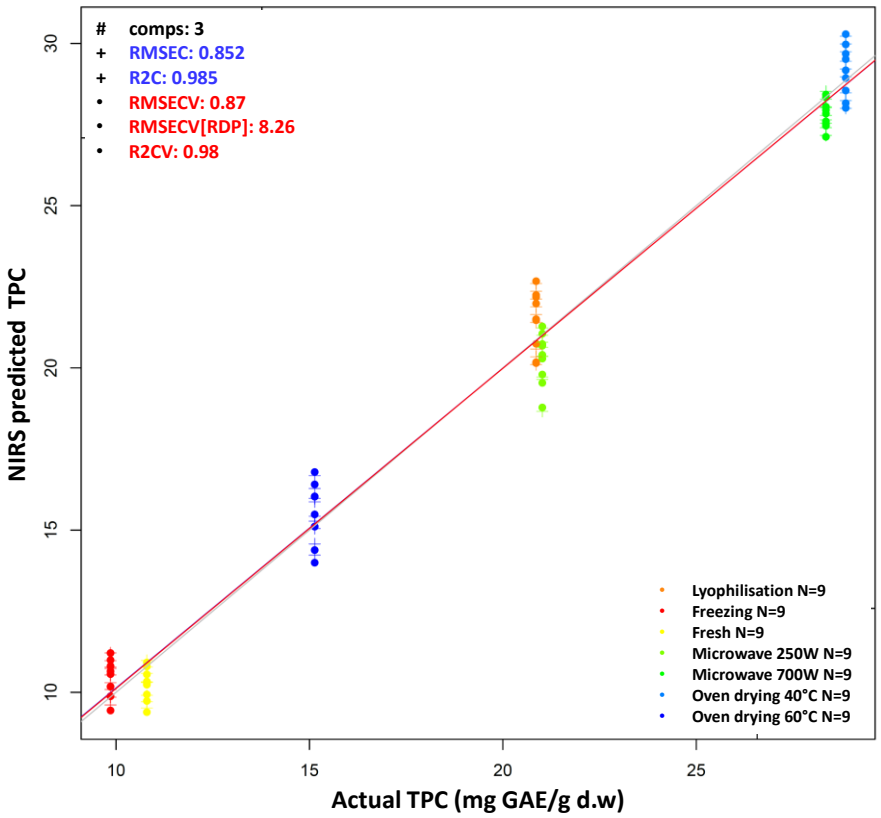

Figure 49: PLSR analysis on peppermint extracts for the prediction of the TPC (mg GAE/g d.w); N=63, LV=5, Spectral Preprocessing: Savitzky-Golay smoothing (window size 23).

Freezing had a similar effect on the total antioxidant capacity in peppermint aqueous extracts, reducing it to an average value of 4.03 mg AAE/g d.w. The rest of the preservation processes induced an increase in TAC, that was highest in case of samples microwave dried at 700W. The

R^2CV and RMSECV of the corresponding PLSR model were 0.93 and 1.72 mg AAE/g d.w respectively.

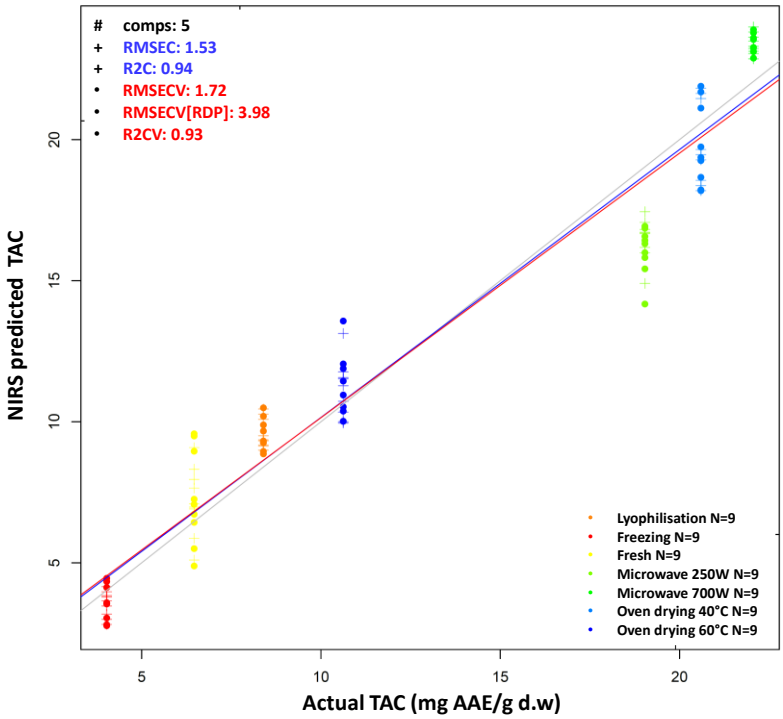

Figure 50: PLSR analysis on peppermint extracts for the prediction of the TAC (mg AAE/g d.w); N=63, LV=5, Spectral Preprocessing: Savitzky-Golay smoothing (window size 31).

The most influential wavelengths with regards to the predictive ability of the model are indicated in (Figure 51).

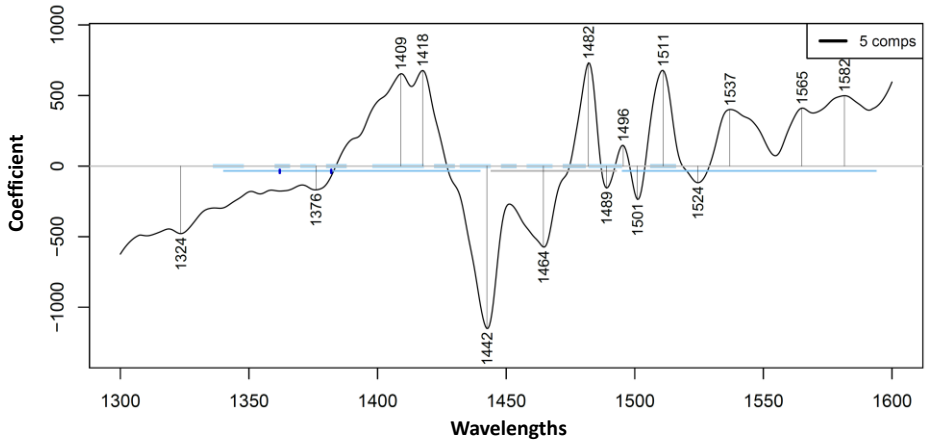

Figure 51: Regression vector of the predictive model of TAC in peppermint in the first overtone 1300–1600 nm range.

These bands consisted of 1418nm, associated with free water structures, and shown to be sensitive to changes in temperature and strongly correlated to changes in water activity (Muncan and Koruga, 2013; Hayashi, 2021), 1442nm, 1464nm, 1482nm,1511nm, which correspond to regions

of strongly bound water (Matija et al., 2012), bands situated at 1537nm, 1565nm and 1582nm were also contributing wavelengths.

For sage, the change induced by the preservation processes consisted in an increase in total antioxidant capacity that was minimal when carrying out freezing and maximal when performing both lyophilization and microwave drying at 250W. The obtained PLSR model pinpointed a good correlation between actual TAC values of the different sage extracts and the values estimated based on their spectral fingerprint in the first water overtone. The corresponding R^2 of cross-validation amounted to 0.96 whereas the root mean squared error of cross-validation was 15.6 mg AAE/g d.w.

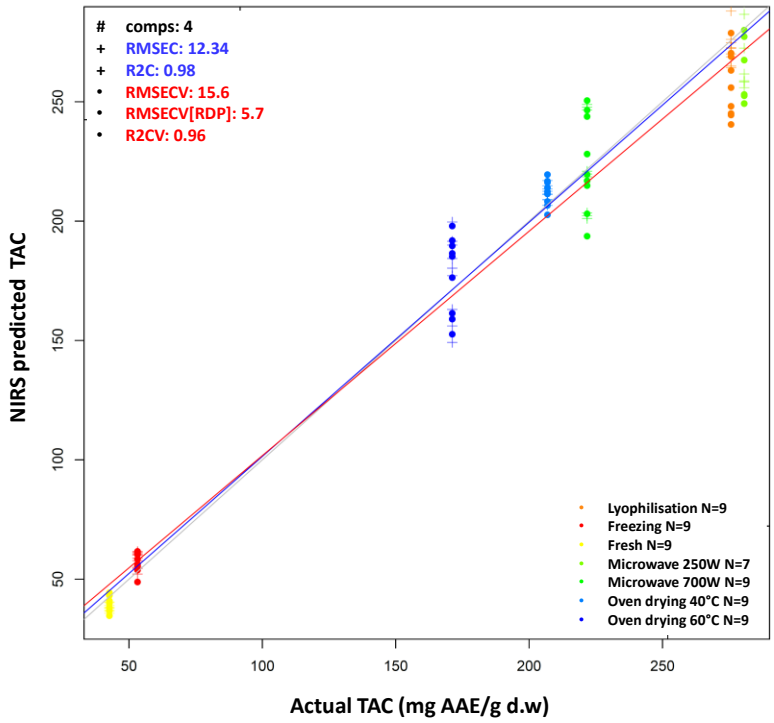

Figure 52: PLSR analysis on sage extracts for the prediction of the TAC (mg AAE/g d.w); N=61, LV=4, Spectral Preprocessing: Savitzky-Golay smoothing (window size 35).

The most relevant wavelengths in the predictive ability of the model consisted mainly of the bands at 1369nm, 1430nm, 1446nm, 1461nm, 1471nm, 1502nm. According to literature, water hydration, H-OH bend and O-O structures are assigned to 1430nm, whereas 1461nm corresponds to the first overtone of symmetric and asymmetric stretch of water molecule (Hayashi, 2021). The band at 1502nm was shown to be correlated with alterations in the hydrogen bond strength in water that would arise from temperature changes (Gowen *et al.*, 2015). At 1369nm, hydration of protons has been reported (Malegori *et al.*, 2022). This change in total antioxidants and the abundance of different water structures reflects the dynamism of water within herb tissue during preservation and how it affects the bioactive compounds in herbs.

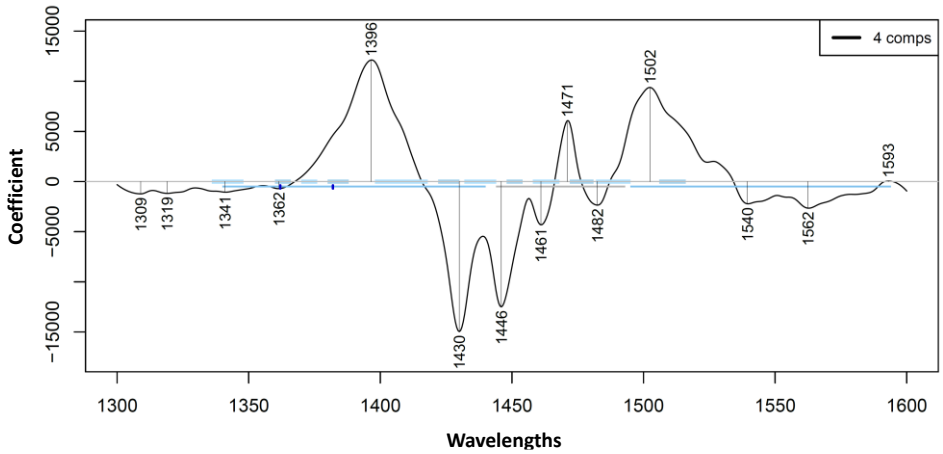

Figure 53: Regression vector of the predictive model of TAC in sage in the first overtone 1300–1600 nm range.

The preservation processes had a similar effect on the total polyphenols content (TPC) as they did on the total antioxidant capacity (TAC) of sage extracts. The increase in TPC was most notable in lyophilized samples as well as in samples subjected to microwave drying at 250W. When it comes to linking the reference values to those derived from the spectral data, the PLSR model (Figure 54) validated their correlation. The corresponding R^2 CV reached 0.97 while the error of cross-validation RMSECV was at 3.63 mg GAE/g d.w.While assessing the regression vector depicted in (Figure 55), we see the prevalence of bands at wavelengths related to different water conformations, such as 1396nm, 1408nm, 1430nm, 1452nm, 1460nm, 1478nm, 1509nm, 1532nm, 1582nm and 1592nm.

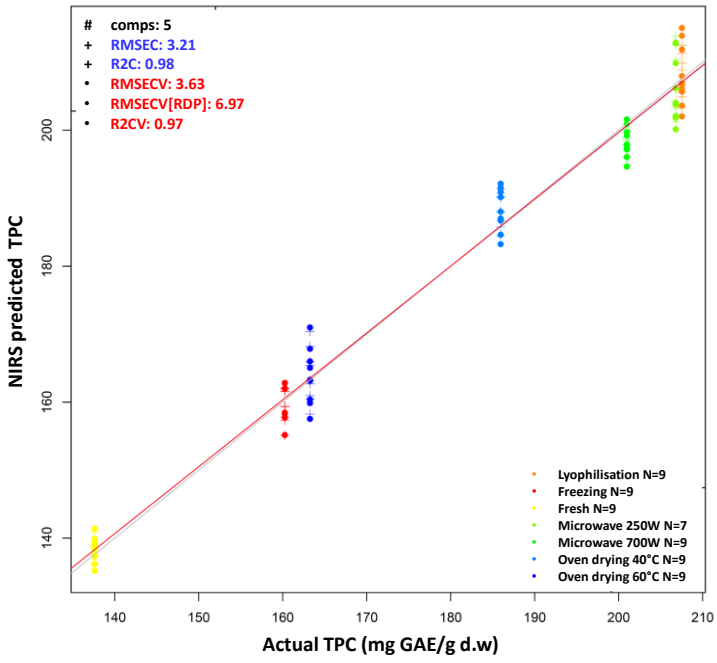

Figure 54: PLSR analysis on sage extracts for the prediction of the TPC (mg GAE/g d.w); N=61, LV=5, Spectral Preprocessing: Savitzky-Golay smoothing (window size 31).

While bands at 1396nm and 1408nm correspond to free hydrogen bonded structures (S0), 1430nm is related to hydration water, other bands related to more hydrogen bonded water corresponded to 1452nm, 1460nm, 1478nm, 1509nm, 1532nm, 1582nm and 1592nm.

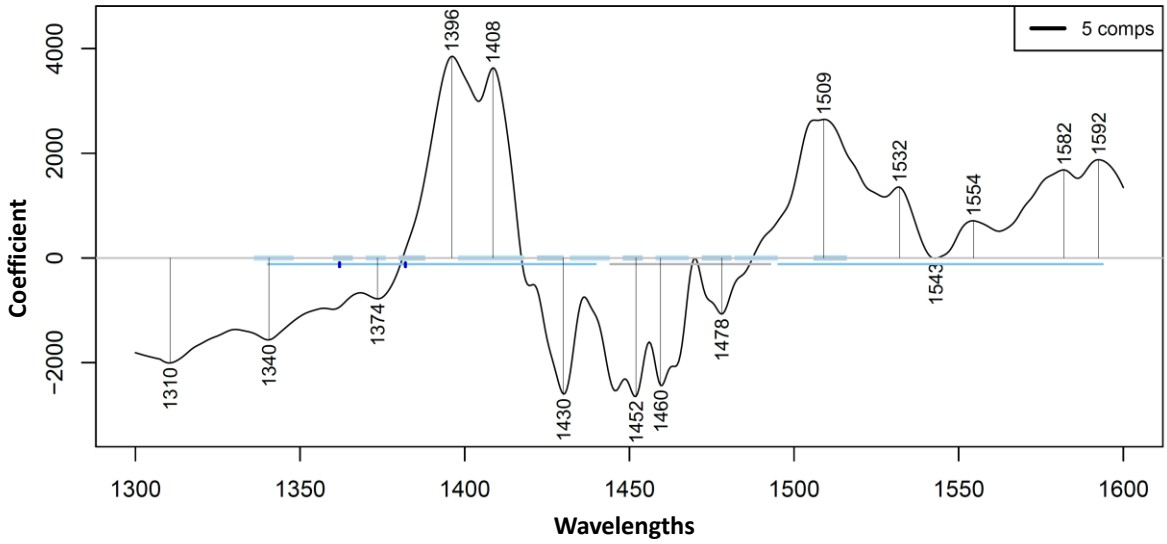

Figure 55: Regression vector of the predictive model of TPC in sage in the 1300–1600 nm range.

Figure 56 showcases the Y-fit graph relating NIRS predicted TAC values to the actual TAC values in preserved summer savory aqueous extracts. Except for freezing, which resulted in a decrease in the total antioxidant capacity of the tested extract, the rest of the applied treatments enhanced the total antioxidant capacity of their respective extracts.

Lyophilized summer savory extracts exhibited the highest total antioxidants capacity estimated at 182.27 mg AAE/g d.w in contrast to extracts derived from frozen summer savory which averaged 43.72 mg AAE/g d.w. On the other hand, microwave drying at 250W and oven drying at 40°C exerted a similar effect on the herb with an augmentation of the TAC of their corresponding extracts to similar extents, valued at 123.02 mg AAE/g d.w for the former and 125.4 mg AAE/g d.w for the latter. The PLSR model metrics consisted of an R^2 and RMSE values of 0.98 and 6.79 mg AAE/g d.w during model calibration and 0.97 and 6.92 mg AAE/g d.w during cross-validation. The performance-to-deviation (RDP) ratio was at 7.09, attesting to the reliability of the method in terms of predicting preservation induced alterations of the total antioxidants in summer savory aqueous extracts.

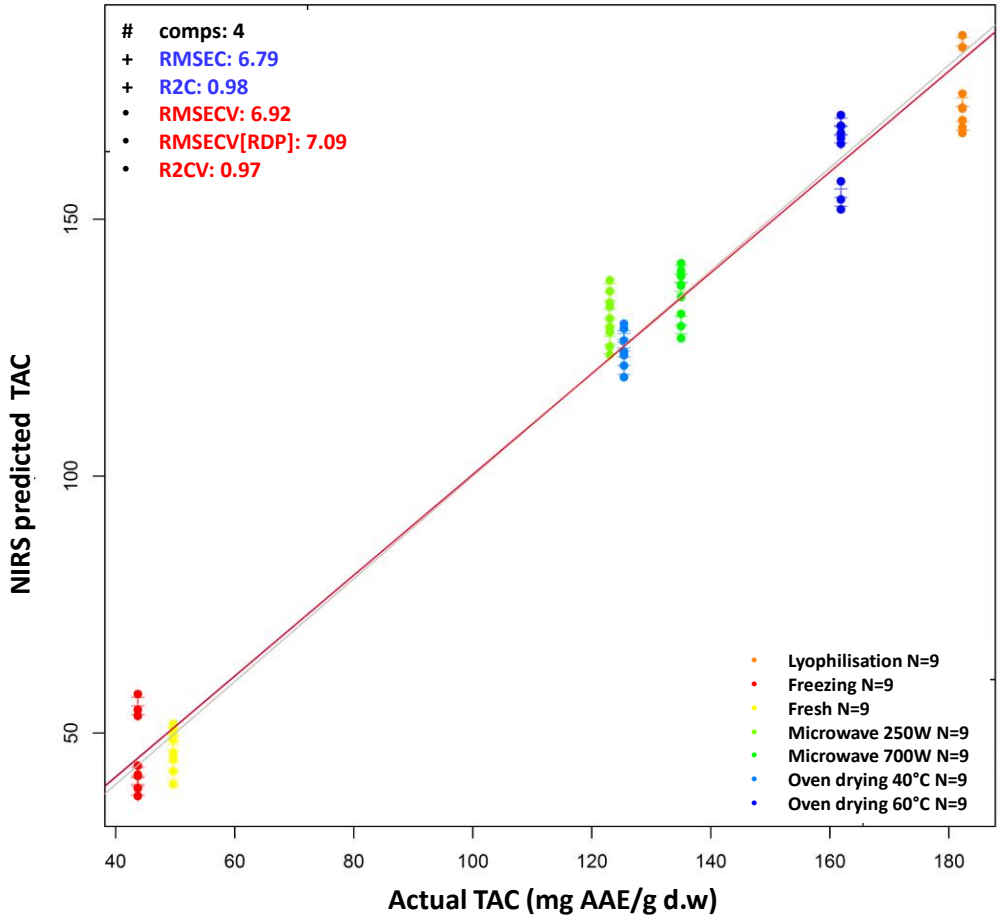

Figure 56: PLSR analysis on summer savory extracts for the prediction of the TAC (mg AAE/g d.w); N=63, LV=4, Spectral Preprocessing: Savitzky-Golay smoothing (window size 31).

The key wavelengths contributing to the accuracy of prediction were the following: 1400nm, 1416nm, 1429nm, 1442nm, 1455nm, 1498nm.

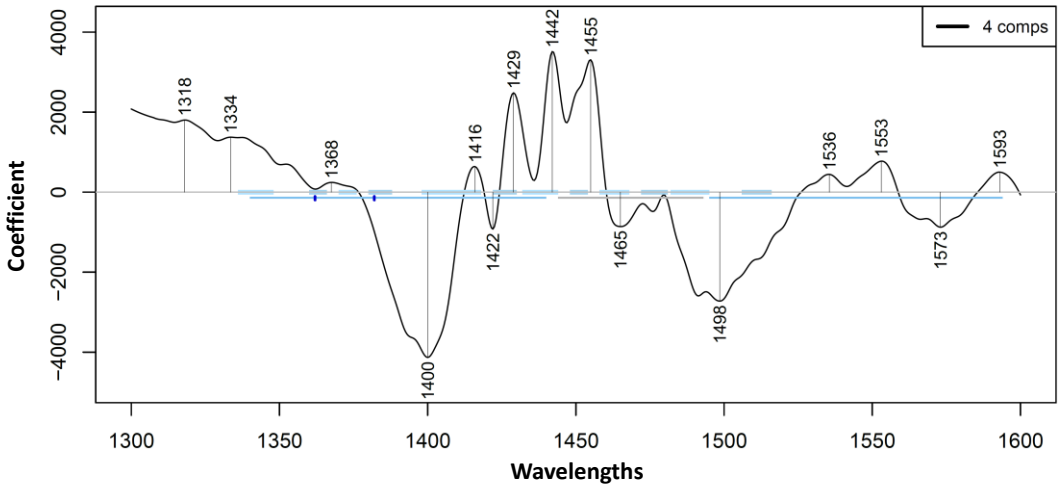

Figure 57: Regression vector of the predictive model of TAC in summer savory in the first overtone 1300–1600 nm range.

With regards to the preservation induced changes in total polyphenols of summer savory extracts, we noticed an increase that was most prominent in case of oven dried samples, followed by lyophilized and microwave dried ones. Again, accurate prediction of the TPC was ensured with an R^2CV of 0.98 and RMSECV of 5.43 mg GAE/g d.w.

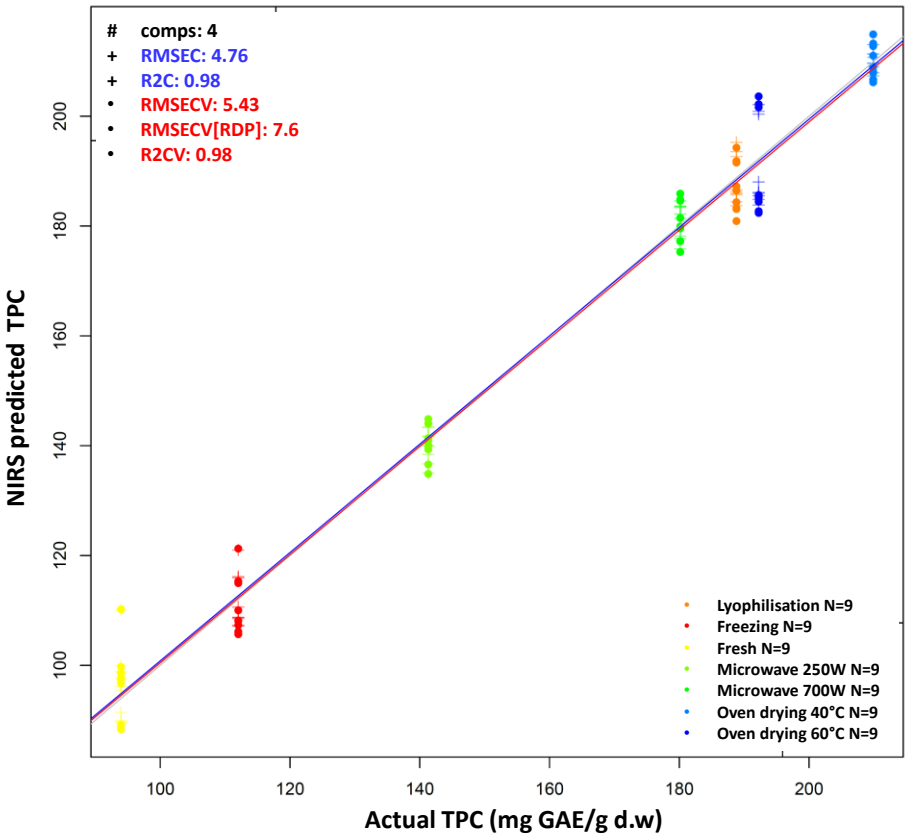

Figure 58: PLSR analysis on summer savory extracts for the prediction of the TPC (mg GAE/g d.w); N=63, LV=5, Spectral Preprocessing: Savitzky-Golay smoothing (window size 23).

According to the corresponding regression vector, portrayed in Figure 59, water assigned bands were the major ones activated by these preservation methods. For instance, the bands at 1363nm and 1384nm have been ascribed to structures such as water solvation shell, OH-$(H2O)_{1,2,4}$, OH-$(H2O)_{1,4}$, to superoxide or O2-(H2O)4 (Xantheas, 1995), 1396nm, 1418nm to trapped or free water, 1466nm to water molecules with 2 hydrogen bonds while 1486nm corresponds to water molecules with 4 hydrogen bonds (Tsenkova, 2007, 2009; R. Tsenkova *et al.*, 2018)

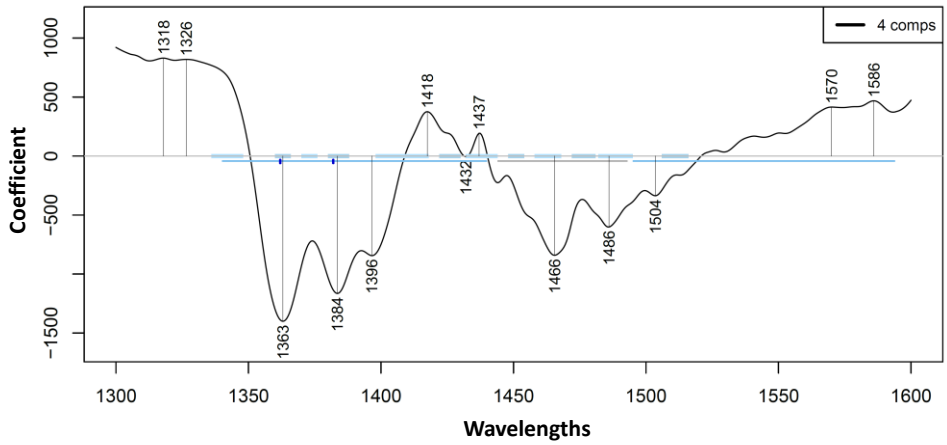

Figure 59: Regression vector of the predictive model of TPC in summer savory in the first overtone 1300–1600 nm range.

Following the different preservation processes, the increase in total antioxidant capacity of tarragon extracts was equally substantial. Initially valued at 50.91 mg AAE/g d.w, TAC nearly tripled after lyophilization, microwave drying at 700W and microwave drying at 250W reaching 144.89 mg AAE/g d.w, 143.56 mg AAE/g d.w and 140.89 mg AAE/g d.w, respectively. The built PLSR models showed a clear correlation between the previously mentioned values, determined by the reference method and those predicted using the NIR spectral fingerprints of the studied tarragon aqueous extracts. The cross-validated model presented an R^2 CV of 0.94 and an RMSECV of 7.64 mg AAE/g d.w while RDP equalled 4.46.

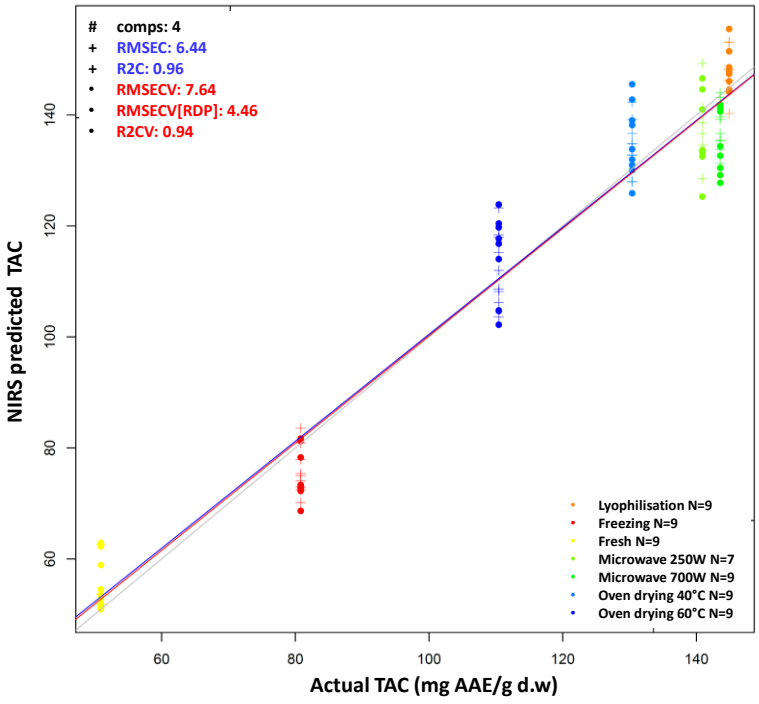

Figure 60: PLSR analysis on tarragon extracts for the prediction of TAC (mg AAE/g d.w); N=61, LV=5, Spectral processing: Savitzky-Golay smoothing (window size 29), 1st derivative.

The most prominent wavelengths (Figure 61) in terms of accurately predicting TAC were shown to be 1384nm, 1396nm, 1406nm, 1416nm, 1438nm, 1450nm, 1458nm, 1477nm, 1500nm, 1512nm, 1529nm, 1548nm, 1590nm. In accordance with other aquaphotomics based studies (Kraats, Munćan and Tsenkova, 2019), these wavebands correspond to different water conformations, from less hydrogen bonded water to strongly hydrogen bounded water.

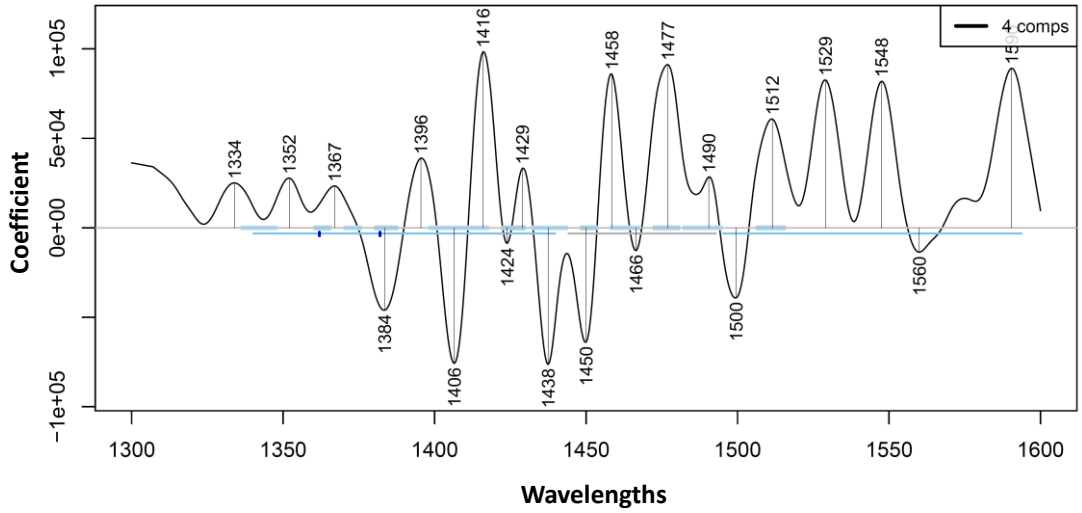

Figure 61: Regression vector of the predictive model of TAC in tarragon in the first overtone 1300–1600 nm range.

What these results confirm is that preservation induced alterations are not only coupled with changes in water contained in the herb but also infers the role played by water in estimating TAC variations.

The alterations in total polyphenols content (TPC) of tarragon aqueous extracts echoed those occurring with antioxidants (TAC).

While lyophilization was the process inducing the highest increase in the TAC, it is microwave drying that resulted in the most considerable increase of TPC, from an initial value of 107.64 mg GAE/g d.w to 171.63 mg GAE/g d.w.

The predictive model was able to accurately determine the TPC of the tested extracts with a coefficient of determination of 0.96 and an error of cross-validation of 3.79 mg GAE/g d.w (Figure *62*).

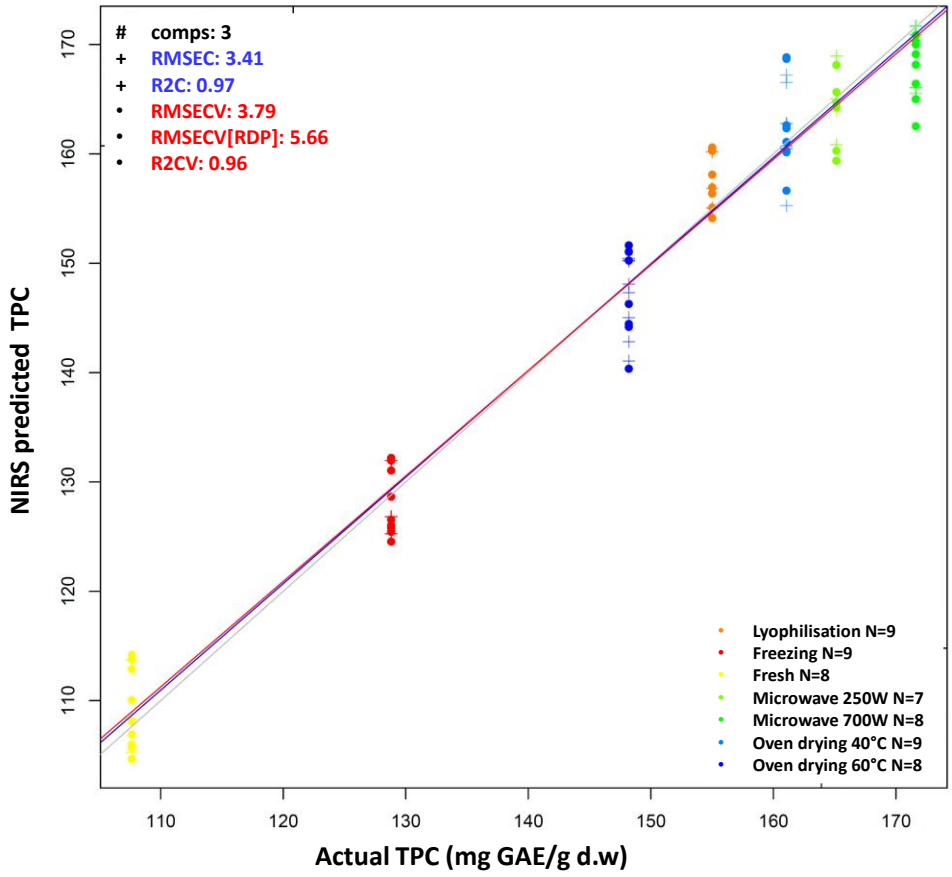

Figure 62: PLSR analysis on tarragon extracts for the prediction of the TPC (mg GAE/g d.w); N=56, LV=3, Spectral processing: Savitzky-Golay smoothing (window size 21),1st derivative.

In terms of the wavelengths with the highest contribution to the predictive ability of the model, they consisted of 1381nm, 1416nm, 1424nm, 1430nm, 1436nm, 1442nm, 1449nm,1458nm, 1491nm, 1505nm, 1524nm, 1548nm and 1591nm (Figure 63). Water structures of varying hydrogen bonding strengths were identified at these bands, ranging from Water hydration, H-OH bend and O…O (at 1424nm and 1430nm), $\nu 2 + \nu 3$, Water solvation shell and OH-(H2O)4,5 (1449nm), water molecules with 4 hydrogen bonds (S4) (1491nm), strongly bound water (1505nm) and water molecules with 2 hydrogen (1458nm) (Segtnan, Isaksson and Ozaki, 2001; Roumiana Tsenkova, Muncan, *et al.*, 2018a). If anything, these findings accentuate the dynamic behaviour of water occurring during alterations to the polyphenols in the vicinity of the studied herb.

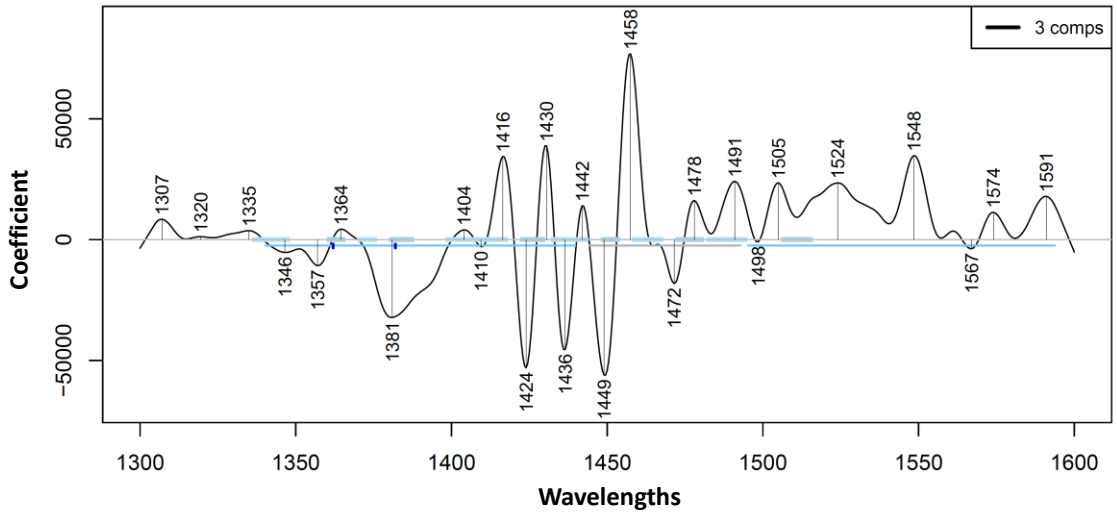

Figure 63: Regression vector of the predictive model of TPC in tarragon in the 1300–1600 nm range.

In the fresh format, the TPC of lovage aqueous extracts stood at 68.61 mg GAE/g d.w. Freezing the herb diminished the polyphenols content of the respective extract to 52.66 mg GAE/g d.w while the rest of the tested drying processes led to an increase in TPC, to range from 72.02 mg GAE/g d.w to 93.9 mg GAE/g d.w. The latter was induced by oven drying at 40°C while the former was obtained after oven drying at 60°C. A correlation between reference and NIR predicted TPC values was also demonstrated through PLSR modelling. Both calibration and cross-validation metrics validated the goodness of fit, with respective R^2 and RMSE of [0.96 and 2.38 mg GAE/g d.w] and [0.95 and 2.76 mg GAE/g d.w] (Figure 64).

As per the most prominent wavelengths (Figure 65), they were identified as the following: 1372nm, 1406nm, 1424nm, 1449nm, 1476nm,1516nm, 1532nm, 1548nm and 1592nm. Existing research has attributed some of these wavelengths to different water structures, for instance, bands at 1406nm, 1424nm, 1449nm and 1476nm were proven to be linked to free water, water hydration, $\nu 2 + \nu 3$, Water solvation shell, OH-(H2O)4,5 and water molecules with 3 hydrogen bonds, respectively whereas the band at 1548nm was shown to be assigned to the first overtone of O–H stretching. On the other hand, 1516nm corresponds to ν1, ν2, symmetrical stretching fundamental vibration, doubly degenerate bending fundamental or strongly bound water (Segtnan, Isaksson and Ozaki, 2001; Roumiana Tsenkova, Muncan, *et al.*, 2018a).

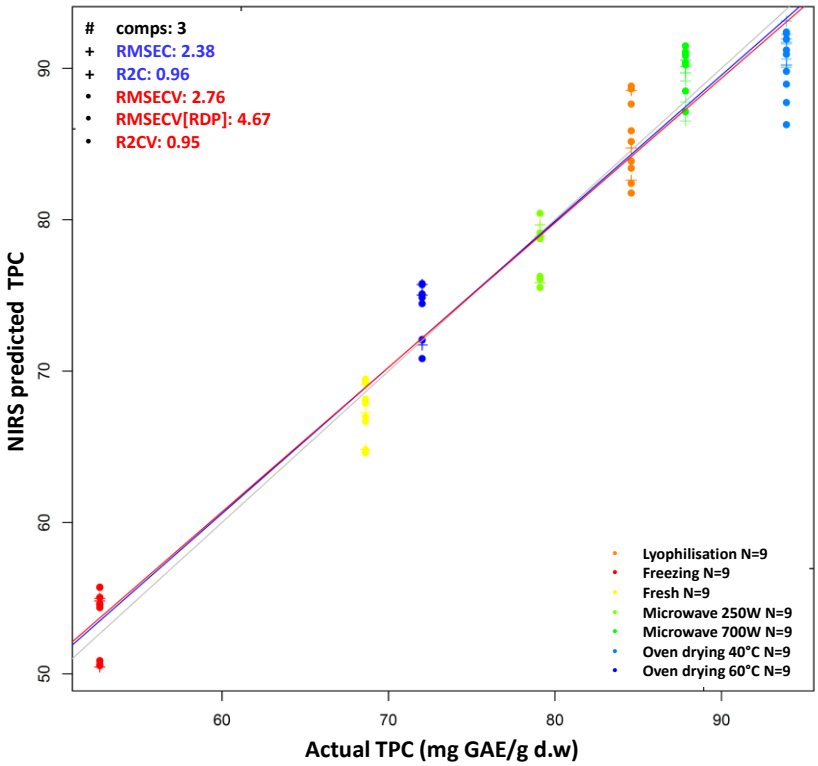

Figure 64: PLSR analysis on lovage extracts for the prediction of the TPC (mg GAE/g d.w); N=63, LV=5, Spectral processing: Savitzky-Golay smoothing (window size 31),1st derivative.

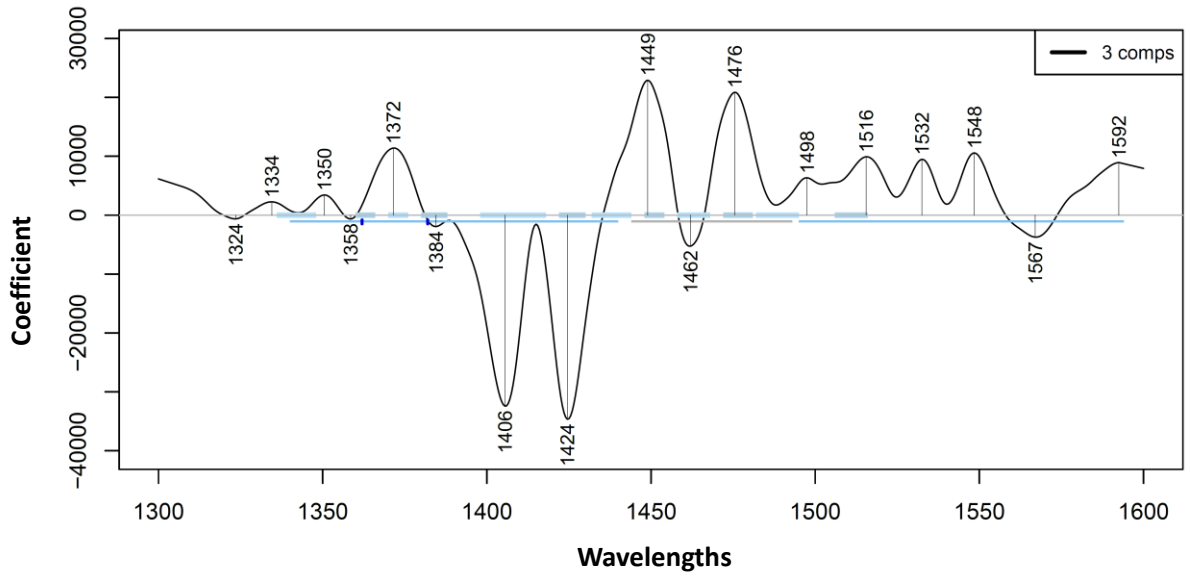

Figure 65: Regression vector of the predictive model of TPC in lovage in the first overtone 1300–1600 nm range.

Contrarily to their effect on the total polyphenols content of lovage aqueous extracts, freezing led to a rise in the total antioxidant capacity while lyophilization caused its decline.

For both microwave dried and oven dried samples, TAC values depended on the intensity. The higher the intensity, the higher the total antioxidant capacity (TAC) values.

Good performances were obtained in terms of correlating the TAC values predicted from the spectral data in the 1st overtone range and those calculated using the reference method. R^2 of cross-validation was estimated at 0.96 while the RMSECV was at 1.69 mg AAE/g d.w.

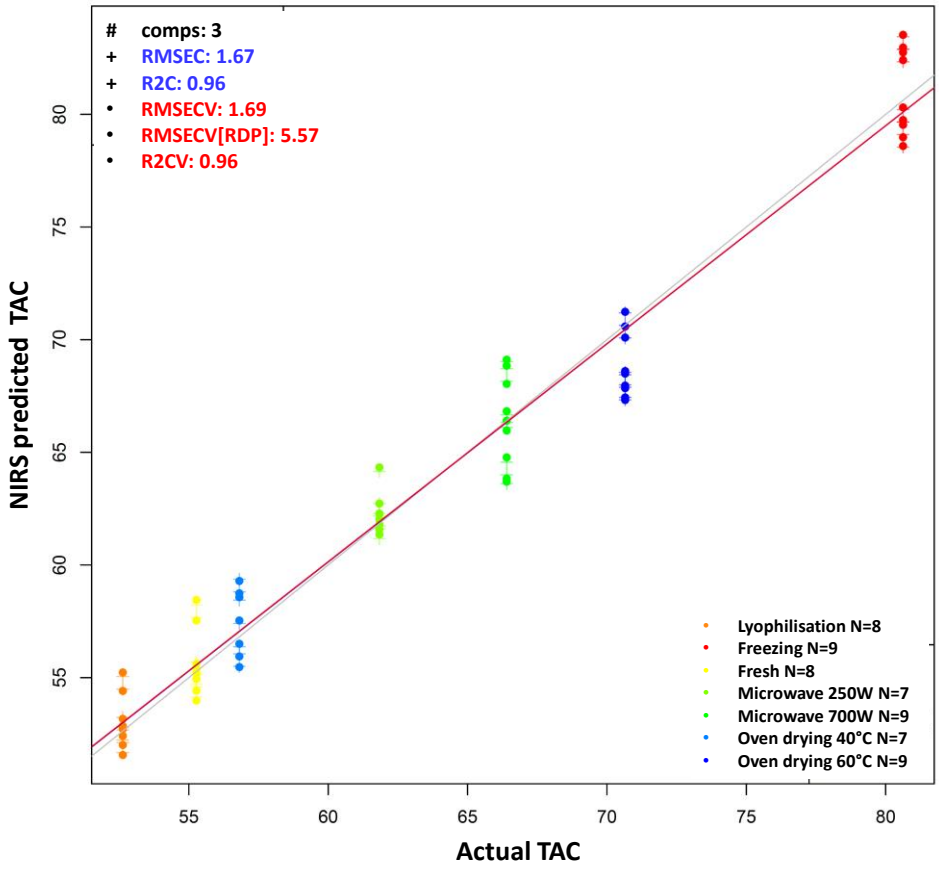

Figure 66: PLSR analysis on lovage extracts for the prediction of the TAC (mg AAE/g d.w); N=57, LV=5, Spectral Preprocessing: Savitzky-Golay smoothing (window size 33).

Two broad bands were characteristic of the corresponding regression vector (Figure 67). The first of higher intensity was situated at 1412nm while the second was located at1594nm. In a study conducted to assess changes generated due to induction heat in aqueous media, (Kakuta *et al.*, 2018) proved that the band at 1412nm was the most temperature sensitive, the very same band was presumably indicative of changes in lipids within the samples (Ni *et al.*, 2024) but also linked to water(Martínez-Valdivieso *et al.*, 2014).

Considering the role lipids play in conferring antioxidant capacity to plant based foods (Pellegrini *et al.*, 2003), we can assume that the occurrence of the 1412nm peak in our herbal extract can be used as biomarker for TAC determination.

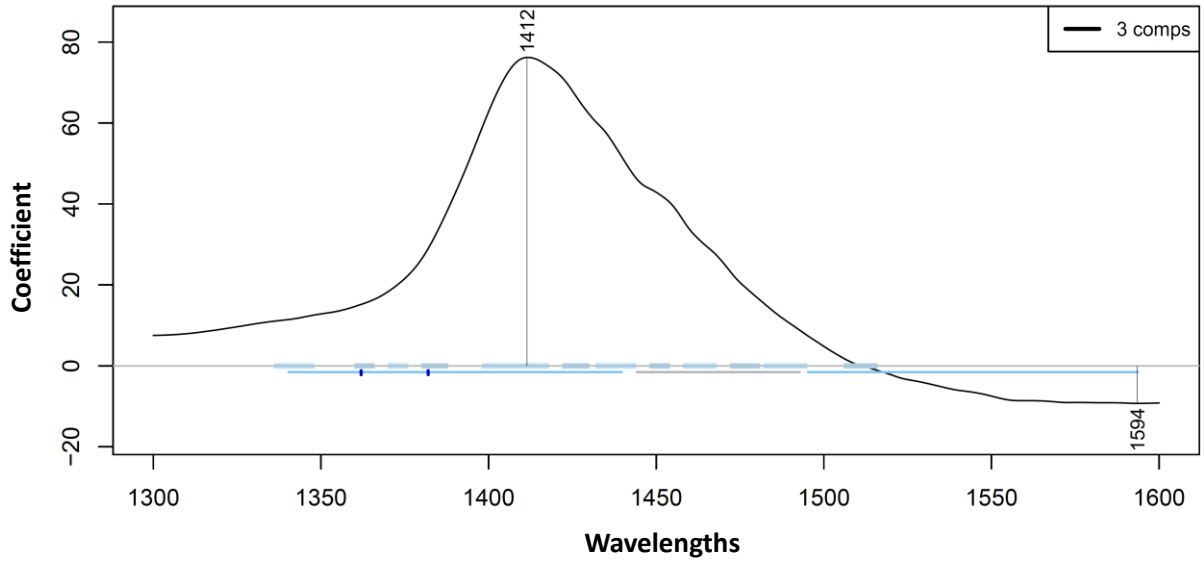

Figure 67: Regression vector of the predictive model of TAC in lovage in the first overtone 1300–1600 nm range.

For basil aqueous extracts, the total antioxidant capacity depended not only on the type of preservation process but equally on its intensity.

Basically, increasing the wattage of microwave drying, from 250W to 700W, resulted in an increase of TAC values from 183.37 mg AAE/g d.w. to 285.4 mg AAE/g d.w. Conversely, an elevation in the temperature of oven drying, from 40°C to 60°C, led to a decrease in the TAC values from 154.73 mg AAE/g d.w. to 52.97 mg AAE/g d.w.

Both lyophilization and freezing induced a reduction in the total antioxidant capacity. Initially valued at 116.11 mg AAE/g d.w, the TAC of lyophilized basil extracts decreased to 63.43 mg AAE/g d.w whereas frozen basil extracts presented a TAC value of 88.85 mg AAE/g d.w.

PLSR predictive model did prove the correlation between reference values and those derived from the spectral data, ascertaining once again the adequacy of determining preservation induced alterations of the bioactive compounds of basil extracts by relying on their spectral fingerprints. The model metrics consisted of R^2 and RMSE values of [0.96 and 14.06 mg AAE/g d.w] during calibration and [0.95 and 15.31 mg AAE/g d.w.] for cross-validation (Figure 68). The wavelengths ensuring the accuracy of the model consisted of 1398nm, 1430nm,1449nm, 1482nm, 1500nm, 1532nm and 1553nm (Figure 69).

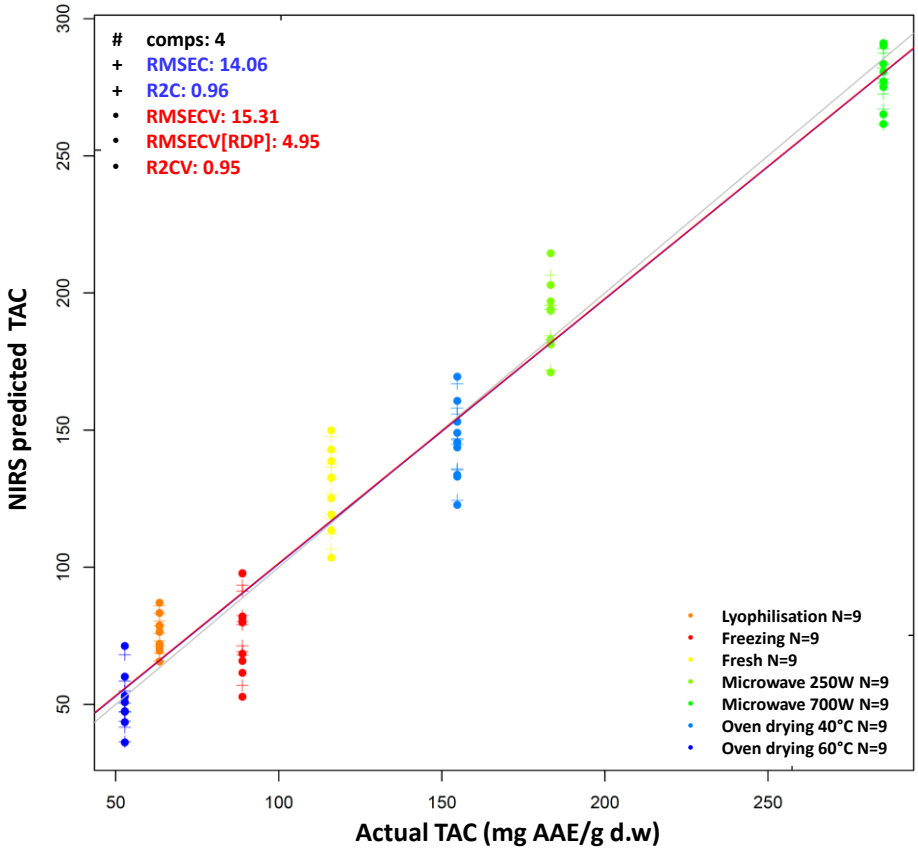

Figure 68: PLSR analysis on basil extracts for the prediction of the TAC (mg AAE/g d.w); N=63, LV=5, Spectral Preprocessing: Savitzky-Golay smoothing (window size 35).

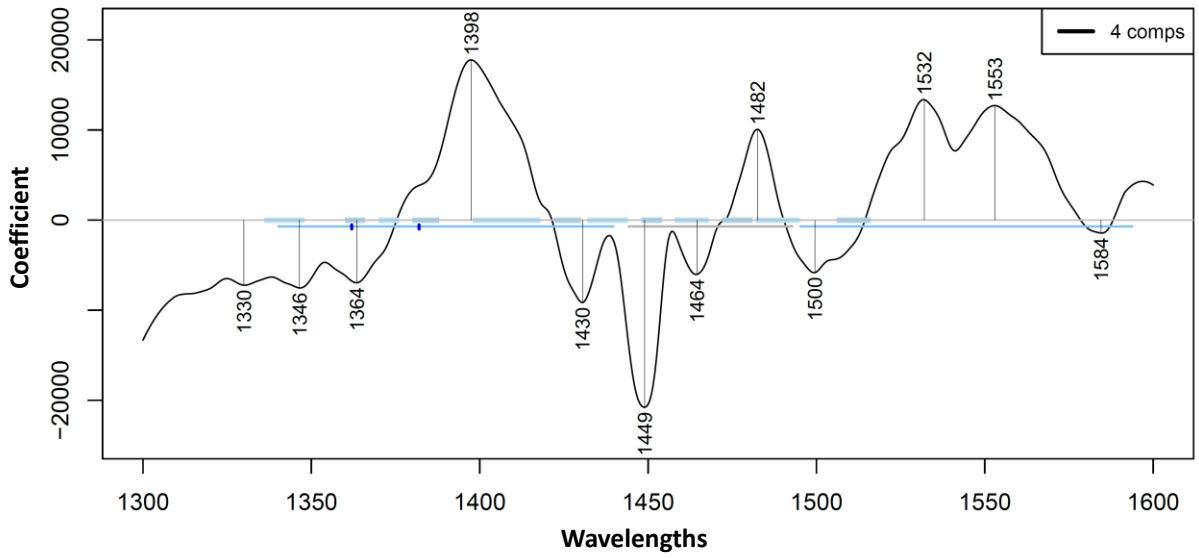

Figure 69: Regression vector of the predictive model of TAC in basil in the first overtone 1300–1600 nm range.

Amongst all the studied herbs, the peculiarity of basil resides in the fact that its total polyphenol content (TPC) decreased after being subjected to every one of the preservation methods. While

frozen samples were the least affected (295.52 mg GAE/g d.w), the decrease was most pronounced in case of oven drying at 60°C (83.46 mg GAE/g d.w). The intensity of the applied preservation method was decisive in determining the total polyphenols content. For instance, performing oven drying at a lower temperature of 40°C resulted in a higher TPC of 119.54 mg GAE/g d.w. For microwave drying, a higher intensity of 700W resulted in a higher TPC (252.03 mg GAE/g d.w as opposed to 169.51 mg GAE/g d.w when carrying out microwave drying at 250W).

Portrayed in Figure 70, the PLSR model corroborates the efficacy of predicting TPC values in basil aqueous extracts and validates the correlation between these values and those measured using the reference method. The R^2 of the cross-validated model was estimated at 0.984 and the RMSECV reached 11.98 mg GAE/g d.w.

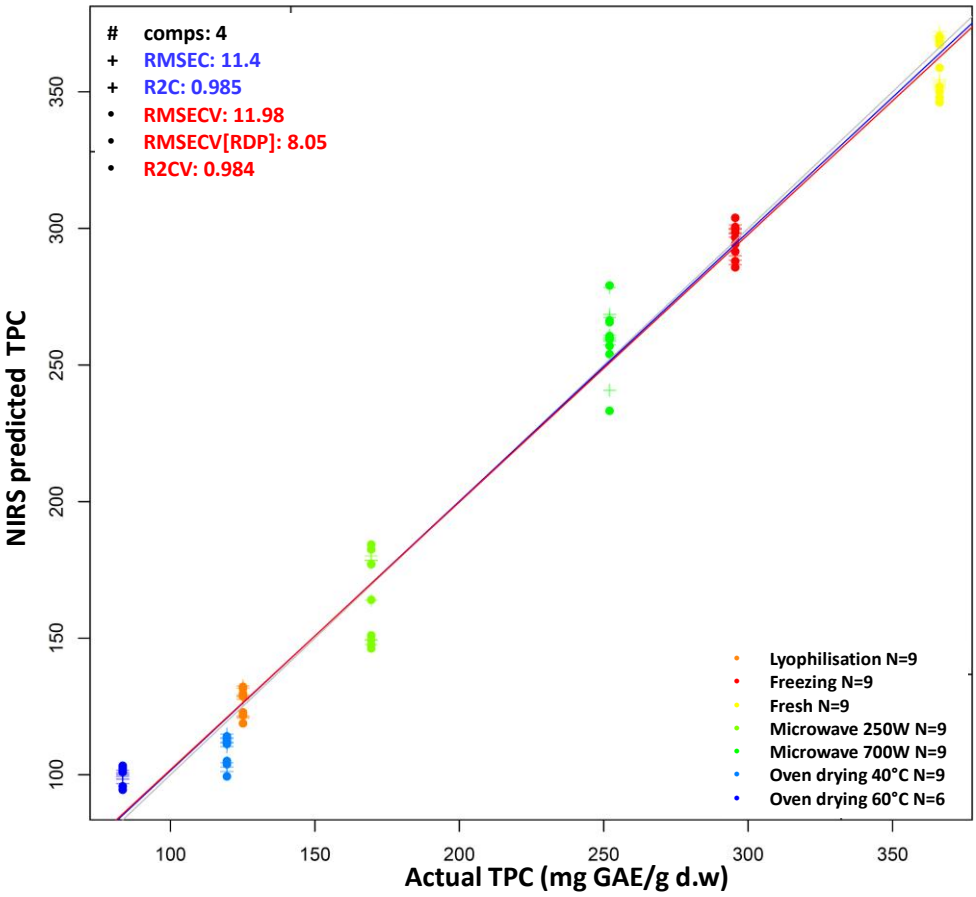

Figure 70: PLSR analysis on basil extracts for the prediction of the TPC (mg GAE/g d.w); N=60, LV=5, Spectral Preprocessing: Savitzky-Golay smoothing (window size 35).

The key wavelengths, contributing the most to the obtainment of the predictive model comprised mainly 1364nm, 1398nm, 1426nm, 1448nm, 1464nm, 1480nm, 1500nm, 1532nm, 1551nm (Figure 71). The first out of these absorbance bands has been assigned to solvation shell with 1, 2 or 4 H2O molecules and was proven to be associated with water evaporation in other studies

(Malegori *et al.*, 2022) while 1398nm was identified as being characteristic of free water and free OH- or trapped water (Iwamoto, Uozumi and Nishinari, 1987). The band at 1426nm falls within the C6 WAMAC assigned to water hydration, and H-OH bend (Esquerre et al., 2009). 1448nm has been attributed to $\nu 2 + \nu 3$, Water solvation shell and OH-$(H_2O)_{4,5}$ (Cattaneo *et al.*, 2009). As per wavebands at 1464nm and 1480nm, they were shown to be ascribed to water molecules with 2 or 3 hydrogen bonds, respectively (Ozaki, 2002).

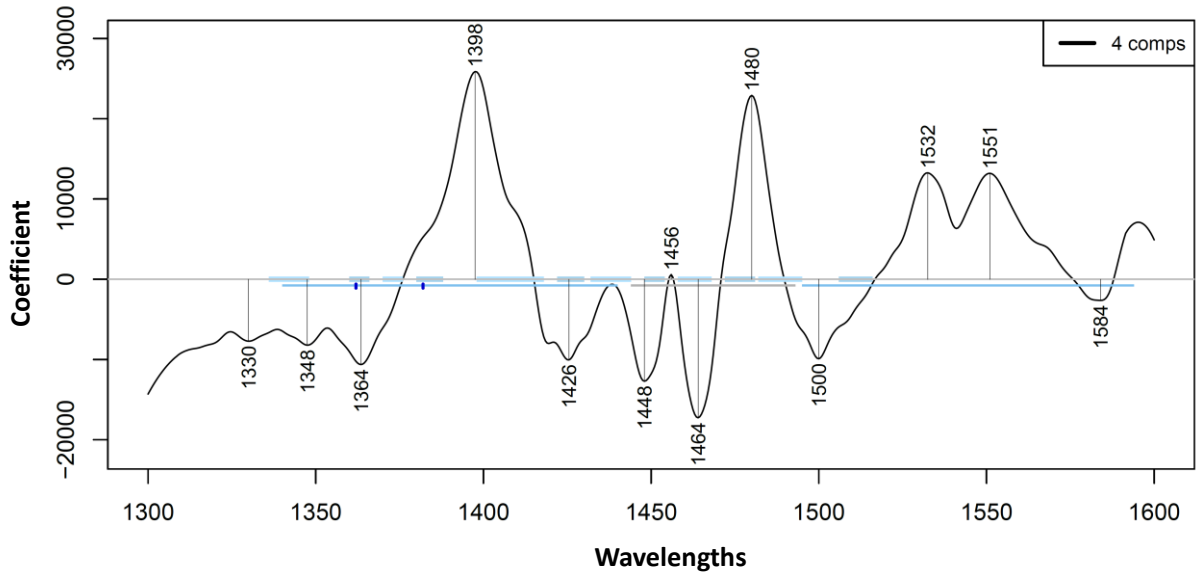

Figure 71: Regression vector of the predictive model of TPC in basil in the first overtone 1300–1600 nm range.

Overall, the presented results highlighted the implication of different water conformations and their decisive role in determining the content of bioactive compounds in herbal aqueous extracts post-preservation.

If anything, these findings not only ascertain the dynamic non-inert behavior of water and the role it plays in reflecting the state of preserved powdered herbs but also showcases the adequacy of aquaphotomics as novel breakthrough method that can serve as an alternative to other quality assessment tool for powdered herbs, when presented in their liquid form.

6. CONCLUSION AND RECOMMENDATIONS

Vibrational spectroscopy has been at the forefront of techniques adopted to tackle global and pressing challenges such as ensuring food integrity without compromising their quality. Yet, amidst the ever-growing urge to achieve economic gains, more sophisticated adulteration methods are emerging, requiring high throughput yet cost-effective authentication methods. Today, not only did the improvements in instrumentation and data processing facilitate instilling near infrared spectrometers into industrial applications but paradigm shifts pertaining to analysed matrices have been equally essential and paved the way for what were once considered uncharted applications (Otal, Iñón and Andrade, 2003). Among these enhancements, aquaphotomics provided a new take on the analysis of water based biosystems permitting a re-assessment of the prospect of NIR in terms of analysing samples, where water is ubiquitous — a prospect that was often avoided due to water's unfavourably strong absorptions in spectral-based analysis.

According to the discipline, water is constantly restructured in the vicinity of water-rich biosystems, in response to surrounding perturbations. Given their inherent richness in water, food matrices are no exception to such dynamism. In this regard, we explored the feasibility of the method in the qualitative assessment of powdered food matrices, when presented in their liquid format. The studied samples consisted of tomato powders, ground coffee and preserved herbs. Our research particularly addressed the adulteration of tomato powders with both colouring and bulking agents, the fraudulent mixing of Robusta and Arabica, as well as the preservation-induced alterations of selected herbs.

In the first study, by subjecting aqueous extracts of three different tomato powder varieties (Tytanium, Tuobodom and Navrongo) comprising 0.5%, 1%, 2%, 5%, 10%, 15% up to 20% w/w of single adulterants [bulking agent (corn flour CF) and coloring agents (food colorant FC and annatto powder AP)] or their combination [(CF+FC) or (AP+FC)] to spectral screening using the benchtop XDS rapid liquid analyzer we proved that PCA-LDA classification models can accurately distinguish the samples according to their adulteration level with average recognition and prediction accuracies ranging from [97.76%-100%] and [88.59%-100%], depending on the tomato variety and the studied adulterant. As per the PLSR predictive models, they gave good estimates of the extent of adulteration, with respective R2 and RMSECV values in the ranges [0.85-0.98] and [0.55 -1.59%w/w], depending on the adulterant (CF, FC, or AP) and adulteration type (single or dual). From aquaphotomics perspective, the results have shown that, even when average performances were achieved in PLSR, e.g., when attempting to predict the added corn,

the water spectral pattern as illustrated by the aquagrams highlighted the extent of adulteration, hence confirming the feasibility of relying on aquaphotomics as a complementary technique to conventional Near infrared spectroscopy (NIRS). Enhancing the reliability of the method and ensuring the potential implementation of this approach on an industrial scale requires further investigations with the inclusion of other prominent adulterants and tomato varieties as well as addressing the complexity of accounting for other ingredients in the tested food matrix.

Next, an assessment of the ground Arabica and Robusta from different geographical origin (Brazil, Columbia, Ethiopia, Vietnam, Uganda, India) by means of PCA-LDA modelling validated the efficiency of conventional NIR spectroscopy for the varietal and geographical discrimination of coffee with 100% correct accuracy. Using PLSR at the first overtone range (1300-1600nm), the Robusta-to-Arabica ratio in ground mixtures could be predicted with R2CV of 0.99 and an RMSECV of 2.4%. The efficacy of the implemented approach was further corroborated when marketed blends were examined, correctly discriminated, and their Robusta contents accurately estimated, the corresponding R2CV and RMSECV amounted to 0.97 and 3.93%.

The aquaphotomics based analysis of the aqueous Arabica and Robusta coffees proved that 100% authentic Robusta coffee extracts were majorly characterized by water molecules that are structured into water shells (908 nm), V1 and V2 bonded water while authentic Arabica has high hydrogen bonded water structures (1060 nm) and is rich in water clusters with two, three, and four hydrogen bonds (1018, 1036, and 1044 nm). Such results prove that a peculiar water spectral pattern (WASP) can be assigned to the pure aqueous coffee samples depending on the variety (Arabica or Robusta), with one having less hydrogen bonded water than the other. The aquagrams of the aqueous blends of Arabica and Robusta, obtained by adding Robusta to Arabica in concentrations of 0.5, 1, 2, 3, 5, 10, 20, 35% w/w demonstrated that the mixing ratio can be translated in the gradual increase of certain water conformations and the decrease of other water structures. These spectral features consisted of an abundance of free water structures in Robusta-rich coffee whereas blends containing higher arabica content were characterized by high hydrogen bonded water.

As per the quantitative prediction of the added robusta in the aqueous “Arabica-Robusta” blends by Partial least squares regression (PLSR), we proved that the NIR-based spectral analysis of these two different quality grades of coffee can accurately predict the added robusta ratio with R2 equal to 0.95 and RMSECV of 6.35% w/w. Once again, including the marketed blends resulted in a decrease in the accuracy of prediction to reach an R2CV of 0.90 and an RMSECV of 3.31%. Bands

corresponding to water molecules with 2 hydrogen bonds (S2), O-H stretch second overtone of sugars and oxalic acid were the ones contributing to the obtained predictive model.

This study offers new insights into the alterations caused by the incorporation of lower grade Robusta to higher grade Arabica, on a molecular level and can be considered an alternative to laborious analysis, namely when access to ground coffee is limited. Nevertheless, ensuring the practicality in food industries would require further refining of the applied method with the inclusion of other coffee mixtures prepared under different roasting and brewing processes.

Next, in our assessment of different herb preservation techniques (Sun and shade drying, slow and fast freezing, microwave drying at 250W and 700W and oven drying at 40°C and 60°C), we could prove that each of the applied methods had a peculiar effect on Thyme, Marjoram and Rosemary that could be traced back to their spectral features. This translated into correct assignment of the aqueous extracts of thyme, marjoram and rosemary to their corresponding groups according to the varying tested preservation methods. The PCA-LDA classification models yielded good recognition and prediction accuracies of [100%-98.77%], [100%-100%], and [100%-98.77%] respectively. The method was shown to be equally effective when testing a second set of herbs, consisting of peppermint, lovage, lemon balm, Greek oregano, tarragon, garden sage, summer savory and sweet basil and preserved by lyophilization, oven-drying (40°C and 60°C), microwave drying (250 W and 700 W), and slow freezing. The NIR based classification of the aqueous extracts of these herbs resulted in good recognition accuracies ranging from [80.16% -100%] while the prediction accuracies fluctuated between 74.62% and 100%, depending on the herb.

Using XDS rapid liquid analyzer in water's 2nd overtone range, the correlation between NIR predicted total polyphenols content (TPC) and total antioxidant capacity (TAC) values and those measured using reference methods could be established based on the spectral fingerprints of aqueous rosemary, thyme, and marjoram extracts. The PLSR models validated using leave-one group-out cross-validation could predict TPC values with R2 CV of 0,94 and RMSECV of 23.12 mg GAE/g d.w and could predict TAC values with R2 CV of 0.85 and RMSECV of 32.63 mg AAE/g d.w. Such correlation could also be demonstrated in the case of the second set of herbs. The PLSR models could estimate TPC values with R2 CV and RMSECV in the ranges [0.93 - 0.98] and [2.76 -15.98] mg GAE/g d.w. and could predict TAC values with R2 CV and RMSECV values in the ranges [0.93-0.97] and [1.69 – 24.05] mg AAE/g d.w, respectively. The inspection of the wavelengths contributing most to the accurate prediction of TAC and TPC accentuated that those were linked to different water conformation, from less hydrogen bonded water to more ice-like clusters of highly hydrogen bonded water. These findings prove that aquaphotomics based

results align, to some extent, with the physicochemical results. Ascertaining the reliability of the method as a viable alternative to conventional methods would however necessitate expanding the range of herbs studied to include a more diverse selection that takes into consideration different varieties and harvesting seasons but also evaluating other herb extraction methods.

To conclude, given the substantial economic impact automated quality control could have on the industry, we presume that shifting from laboratory measurements to on-site measurements, market-based ones through incorporating aquaphotomics based analysis via the integration of handheld spectrometers in industrial settings would be valuable. The practical implications of the presented studies could be particularly promising in cases where the authentication and detection of the adulteration is not possible in powder form and only the liquid format is available (E.g., vending machines, the expansion of the Ready-to-Drink (RTD).

Tailoring future studies towards examining other matrices would substantially contribute to the enrichment of this field and help addressing prominent challenges of our days that meet the urgent needs for food integrity and quality assurance.

7. NEW SCIENTIFIC RESULTS

The new scientific findings, herein presented, take into consideration specific instrumental setups and well determined coffee, tomato, and herb varieties. Hence, the benchtop XDS Rapid liquid Analyzer refers to the XDS-RLA spectrometer (Metrohm, Herisau, Switzerland) used with a 1 mm pathlength quartz cuvette in transmission mode while the second benchtop spectrometer corresponds to MetriNIR spectrometer (MetriNIR Research, Development and Service Co., Budapest, Hungary) used with a circular cuvette of 0.4 mm layer thickness and operating in transflectance mode.

1. When subjecting aqueous Arabica coffee from Brazil, Columbia and Ethiopia and aqueous Robusta coffee from Vietnam, Uganda, and India to spectral analysis using the MetriNIR spectrometer I proved that a peculiar water spectral pattern (WASP) can be assigned to pure aqueous coffee samples depending on the variety. The Arabica coffee features highly hydrogen-bonded water structures (1060 nm) and is rich in water clusters with two, three, and four hydrogen bonds (1018, 1036, and 1044 nm), while Robusta coffee is primarily characterized by water molecules structured into water shells (908 nm) and V1 and V2 bonded water, having less hydrogen-bonded water than Arabica coffee. In case of blending the Arabica and Robusta coffee types, the mixing ratio is translated in the corresponding WASP with the abundance of free water structures in Robusta-rich coffee whereas blends containing higher arabica content were characterized by high hydrogen bonded water conformations.
2. In the aqueous coffee mixtures comprising 0.5, 1, 2, 3, 5, 10, 20, 35% w/w Robusta coffee from India mixed with Arabica coffee from Ethiopia, and subjected to spectral analysis using the MetriNIR spectrometer in the 1300-1600nm range, the Robusta-to-Arabica ratio could be predicted via PLSR modelling with an R^2CV of 0.95 and an RMSECV of 6.35% w/w.
3. Aqueous thyme (*Thymus vulgaris*), marjoram (*Origanum majorana*) and rosemary (*Rosmarinus officinalis*) extracts obtained by hot water extraction following preservation by sun drying, shade drying, slow freezing, fast freezing, microwave drying (250W or 700W) or oven drying (40°C or 60°C) could be classified using PCA-LDA method based on the NIR spectra acquired in the 1300-1600nm range using the XDS Rapid liquid Analyzer. Their respective recognition and prediction accuracies consisted of [100%-98.77%], [100%-100%] and [100%-98.77%], respectively.

4. Aqueous peppermint (*Mentha × piperita*), lovage (*Levisticum officinale*), lemon balm (*Melissa officinalis*), Greek oregano (*Origanum Vulgare*), tarragon (*Artemisia Dracunculus*), garden sage (*Salvia Officinalis*), summer savory (*Satureja hortensis*) and sweet basil (*Ocimum Basilicum*) extracts obtained by hot water extraction following their preservation by lyophilization, oven-drying (40°C and 60°C), microwave drying (250 W and 700 W) or slow freezing could be classified using PCA-LDA modelling of their NIR spectra acquired in the 1300-1600nm range using the XDS Rapid liquid Analyzer with average recognition and prediction rates in the ranges [80.16% -100%] and [74.62% -100%].
5. PLSR models built on the NIR spectra of aqueous extracts of peppermint (*Mentha × piperita*), lovage (*Levisticum officinale*), lemon balm (*Melissa officinalis*), Greek oregano (*Origanum Vulgare*), tarragon (*Artemisia Dracunculus*), garden sage (*Salvia Officinalis*), summer savory (*Satureja hortensis*) and sweet basil (*Ocimum Basilicum*) obtained by different preservation methods and collected by the XDS Rapid liquid Analyzer in the 1300-1600nm range could predict their total polyphenols content (TPC) with R^2 CV and RMSECV in the ranges [0.93-0.98] and [1.69-24.15] mg GAE/g d.w and could predict total antioxidant capacity (TAC) with R^2 CV and RMSECV values in the ranges [0.94-0.98] and [0.87-15.98] mg AAE/g d.w.
6. Subjecting aqueous extracts of three different tomato powder varieties (Tytanium, Tuobodom and Navrongo) comprising of 0.5%, 1%, 2%, 5%, 10%, 15% or 20% w/w of single adulterants [bulking agent (corn flour CF), coloring agents (food colorant FC or annatto powder AP)] or their combination [(CF+FC) or (AP+FC)] to spectral screening using the XDS rapid liquid analyzer in the 1300-1600nm range we proved that PCA-LDA models can accurately classify the samples according to their adulteration level with average recognition and prediction accuracies ranging from [97.76%-100%] and [88.59%-100%], depending on the tomato variety and the studied adulterant.
7. PLSR models built on the NIR spectra of aqueous extracts of three different tomato powder varieties (Tytanium, Tuobodom and Navrongo) comprising of 0.5%, 1%, 2%, 5%, 10%, 15% or 20% w/w of single or dual adulterants acquired by the XDS rapid liquid analyzer in the 1300-1600nm range could predict the added adulterant, with respective R^2 and RMSECV values in the ranges [0.85-0.98] and [0.55-1.59%w/w], depending on the adulterant (Corn flour, Food colorant, or Annatto powder) and adulteration type (single or dual).

8. SUMMARY

Water, ubiquitous across biosystems and within food matrices, confers a set of organoleptic and functional characteristics, that are inextricably linked to the overall state of such systems. Qualitative assessments have been, however, largely dependent on the combination of omics, each targeting specific criteria and indicators.

Aquaphotomics based research has shifted the focus away from such macroscale measurements onto a more integrative, an all-encompassing approach.

To tap into the potential of the method in the determination of both the quality and safety of powdered products, we resorted to their analysis in liquid form. The presented studies investigated the fraudulent blending of Robusta and Arabica, the adulteration of tomato powders with coloring and bulking agents, and the preservation-induced changes of selected herbs.

Using PCA-LDA modeling, conventional NIR spectroscopy was proven 100% efficient in terms of the varietal and geographical discrimination of ground Arabica and Robusta from different geographical origins. The prediction of the Robusta-to-Arabica ratio in ground mixtures yielded an R^2CV of 0.99 and an RMSECV of 2.4%. The approach was further validated with marketed blends, achieving an R^2CV of 0.97 and an RMSECV of 3.93%. Aquaphotomics-based analysis of the aqueous coffee revealed distinct water molecular structures per variety where authentic Robusta was characterized by less hydrogen bonded water while authentic Arabica had high hydrogen bonded water structures. The gradual addition of Robusta, in concentrations from 0% to 35% w/w, was reflected in the aquagrams of the respective aqueous blends. Robusta-rich blends showed an abundance of free water structures, while those with higher Arabica content exhibited more hydrogen bonded water.

PLSR models provided an accurate prediction of the added Robusta where an R^2 CV of 0.95 and an RMSECV of 6.35% w/w were achieved.

Having proved the adequacy of the method in pinpointing the blending of two coffee quality grades, the next study aimed at assessing herb preservation methods and whether or not changes in the bioactive compounds are correlated with those of water structures. A first set of herbs (rosemary, thyme and marjoram) was subjected to sun and shade drying, slow and fast freezing, microwave drying (250W and 700W) and oven drying (40°C and 60°C) whereas a second set of herbs (peppermint, lovage, lemon balm, Greek oregano, tarragon, garden sage, summer savory and sweet basil) was preserved by lyophilization, oven-drying (40°C and 60°C), microwave drying (250 W and 700 W), and slow freezing. In both experiments, we could demonstrate that the applied

methods had a peculiar effect on the herbs that can be reflected in their spectral features through an accurate assignment of the herbal extracts to their distinctive groups. Accuracies of the PCA-LDA models during recognition and prediction reached [100%-98.77%], [100%-100%], and [100%-98.77%] in case of thyme, marjoram, and rosemary. Classification models fluctuated from 80.16% to 100% during training and from 74.62% to 100% during prediction for the 2nd set of herbs.

The correlation between NIR-predicted total polyphenols content (TPC) and total antioxidant capacity (TAC) and those measured using reference methods could be established for both herbal extract sets with an R^2CV and RMSECV in the ranges of [0.93 - 0.98] and [2.76 – 23.12] mg GAE/g d.w. for TPC and R^2CV and RMSECV in the ranges of [0.85 - 0.97] and [1.69 – 32.63] mg AAE/g d.w in case of TAC. The predominant wavelengths contributing to the precise prediction of TAC and TPC underscored their association with varying water conformations, ranging from less hydrogen-bonded water to more ice-like clusters of highly hydrogen-bonded water.

In the third study, we analyzed aqueous extracts of three tomato powder varieties (Tytanium, Tuobodom, and Navrongo) containing single adulterants (corn flour CF, food colorant FC, annatto powder AP) or their combinations [(CF+FC) or (AP+FC)] in concentrations of (0.5% to 20% w/w). PCA-LDA classification accurately distinguished samples based on adulteration levels, achieving recognition and prediction accuracies of [97.76%-100%] and [88.59%-100%], depending on the tomato variety and adulterant type. PLSR models provided good estimates of adulteration extent, with R^2 and RMSECV values ranging from [0.85-0.98] and [0.55-1.59%w/w], respectively, depending on the adulterant and adulteration type. Aquagrams revealed the extent of adulteration through the water spectral pattern.

Across the conducted experiments, the wavelengths contributing the most to the obtainment of accurate qualitative and quantitative models were proven to fall within the overtone ranges of water, hence accentuating the feasibility of relying on the water spectral pattern as a holistic biomarker of different food matrices as opposed to the overwhelming multiplicity of food safety and quality assessment markers. The method proves to be of particular interest with regards to eliminating sources of heterogenicity when analyzing powdered samples, which were conventionally proven to impact the spectral data and influence classification and prediction accuracies (Faqeerzada *et al.*, 2020).

9. LIST OF PUBLICATIONS IN THE FIELD OF STUDY

9.1. Journal articles

Bodor, Z., Benedek, C., **Aouadi, B.**, Zsom-Muha, V., & Kovacs, Z. (2022). Revealing the Effect of Heat Treatment on the Spectral Pattern of Unifloral Honeys Using Aquaphotomics. Molecules, https://doi.org/10.3390/molecules27030780. Q1 - IF 4.927

Aouadi, B.; Zaukuu, J.-L.Z.; Vitális, F.; Bodor, Z.; Fehér, O.; Gillay, Z.; Bazar, G.; Kovacs, Z. (2020). Historical Evolution and Food Control Achievements of Near Infrared Spectroscopy, Electronic Nose, and Electronic Tongue—Critical Overview. *Sensors*, https://doi.org/10.3390/s20195479. Q1 - IF 3.847

Aouadi, B.; Vitalis, F.; Bodor, Z.; Zinia Zaukuu, J.-L.; Kertesz, I.; Kovacs, Z (2022). NIRS and Aquaphotomics Trace Robusta-to-Arabica Ratio in Liquid Coffee Blends. Molecules, https://doi.org/10.3390/molecules27020388. Q1 - IF 4.927

Aouadi, B.; Laryea, D.; Aguinaga Bósquez, J.P.; Majadi, M.; Kertész, I.; Bodor, Z., Zaukuu, J.-L.Z.; Kovacs, Z.(2024). Aquaphotomics based screening of tomato powder extracts reveals susceptibility to bulking and coloring agents. Food Control, https://doi.org/10.1016/j.foodcont.2023.110163. Q1 - IF 6

9.2. Book chapters

Muncan, J; **Aouadi, B**; Tsenkova, R. (2024). Potential of Aquaphotomics for Determination of Quality of Natural Foods, In Megh R. Goyal, Arijit Nath, & Zoltán Kovács, (Eds.), Sustainable and Functional Foods from Plants (Chapter 12, pp.371-420). Publisher: Apple Academic Press, Inc. Co-published with CRC Press (Taylor & Francis)

Muncan, J; **Aouadi, B**; Vitalis, F; Kovacs, Z.; Tsenkova, R. (2020). Soil Aquaphotomics for Understanding Soil–Health Relation through Water–Light Interaction, In: Lal, Rattan (ed.) The Soil–Human Health Nexus (Chapter 10, pp. 197-221). Publisher: CRC Press

10. APPENDICES

References

Abascal, K., Ganora, L. and Yarnell, E. (2005) 'The effect of freeze-drying and its implications for botanical medicine: A review', *Phytotherapy Research*, 19(8), pp. 655–660. Available at: https://doi.org/10.1002/ptr.1651.

Abul-Fadl, M.M. and Ghanem, T.H. (2011) 'Effect of Refractance-window (RW) drying method on quality criteria of produced tomato powder as compared to the convection drying method', *World Applied Sciences Journal*, 15(7), pp. 953–965.

Adnan, A. *et al.* (2020) 'Reliable Discrimination of Green Coffee Beans Species: A Comparison of UV-Vis-Based Determination of Caffeine and Chlorogenic Acid with Non-Targeted Near-Infrared Spectroscopy', *Foods*. Available at: https://doi.org/10.3390/foods9060788.

Aires, P. *et al.* (2018) 'Near infrared hyperspectral images and pattern recognition techniques used to identify etiological agents of cotton anthracnose and ramulosis', *Journal of Spectral Imaging*, 1. Available at: https://doi.org/10.1255/jsi.2018.a8.

Alwafa, R.A. *et al.* (2022) 'Comparison between Quality Traits of Solar-Dried and Freeze-Dried Origanum syriacum L. (Za'atar)', *Plants*, 11(9). Available at: https://doi.org/10.3390/plants11091110.

Arvanitoyannis, I.S. and Vaitsi, O.B. (2007) 'A review on tomato authenticity: Quality control methods in conjunction with multivariate analysis (chemometrics)', *Critical Reviews in Food Science and Nutrition*, 47(7), pp. 675–699. Available at: https://doi.org/10.1080/10408390600948568.

Assis, C., Oliveira, L.S. and Sena, M.M. (2018) 'Variable Selection Applied to the Development of a Robust Method for the Quantification of Coffee Blends Using Mid Infrared Spectroscopy', *Food Analytical Methods*, 11(2), pp. 578–588. Available at: https://doi.org/10.1007/s12161-017-1027-7.

Atanassova, S. *et al.* (2016) 'Near infrared spectroscopy and aquaphotomics for monitoring changes during yellow cheese ripening', *Agric. Science Technol.*, 3(4), p. 390. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84886778279&partnerID=tZOtx3y1.

Baishya, N. *et al.* (2020) 'Near Infrared and Aquaphotomic analysis of water absorption in lactate containing media', *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020-July(July), pp. 4381–4384. Available at: https://doi.org/10.1109/EMBC44109.2020.9176675.

Bázár, G. *et al.* (2015) 'Water revealed as molecular mirror when measuring low concentrations of sugar with near infrared light', *Analytica Chimica Acta*, 896, pp. 52–62. Available at:

https://doi.org/10.1016/j.aca.2015.09.014.

Bázár, G. *et al.* (2016) 'NIR detection of honey adulteration reveals differences in water spectral pattern', *Food Chemistry*, 194, pp. 873–880. Available at: https://doi.org/10.1016/j.foodchem.2015.08.092.

Belović, M. *et al.* (2018) 'Tomato pomace powder as a raw material for ketchup production', *Food Bioscience*, 26(October), pp. 193–199. Available at: https://doi.org/10.1016/j.fbio.2018.10.013.

Benzie, I.F.F. and Strain, J.J. (1996) 'The Ferric Reducing Ability of Plasma (FRAP) as a Measure of "Antioxidant Power": The FRAP Assay', *Analytical Biochemistry*, 239(1), pp. 70–76. Available at: https://doi.org/https://doi.org/10.1006/abio.1996.0292.

Berrar, D. (2018) 'Cross-validation', *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3, pp. 542–545. Available at: https://doi.org/10.1016/B978-0-12-809633-8.20349-X.

Bilge, G. (2020) 'Investigating the effects of geographical origin, roasting degree, particle size and brewing method on the physicochemical and spectral properties of Arabica coffee by PCA analysis', *Journal of Food Science and Technology*, 57(9), pp. 3345–3354. Available at: https://doi.org/10.1007/s13197-020-04367-9.

Bodor, Z. *et al.* (2019) 'Application of NIRS and Aquaphotomics for the detection of adulteration of honey, paprika and tomato Paste', in *SZIENtific meeting for young researchers - Ifjú Tehetségek Találkozója (ITT)*, pp. 76–91.

Borba, K.R. *et al.* (2021) 'Portable near Infrared Spectroscopy as a Tool for Fresh Tomato Quality Control Analysis in the Field', *Applied Sciences*, 11(7), p. 3209. Available at: https://doi.org/10.3390/app11073209.

Brudzewski, K., Osowski, S. and Dwulit, A. (2012) 'Recognition of Coffee Using Differential Electronic Nose', *IEEE Transactions on Instrumentation and Measurement*, 61(6), pp. 1803–1810. Available at: https://doi.org/10.1109/TIM.2012.2184011.

Bu, D. (2007) 'Chemometric Analysis for Spectroscopy', *Tech Flash* [Preprint].

Burrows, A. (2009) 'Palette of our palates: A brief history of food coloring and its regulation', *Comprehensive Reviews in Food Science and Food Safety*, 8(4), pp. 394–408. Available at: https://doi.org/10.1111/j.1541-4337.2009.00089.x.

Cai, W. *et al.* (no date) 'Observing the complexity of water structures from near infrared spectra and simulation', pp. 0–1.

Calín-Sánchez, Á. *et al.* (2020) 'Comparison of Traditional and Novel Drying Techniques and Its Effect on Quality of Fruits, Vegetables and Aromatic Herbs', *Foods*. Available at: https://doi.org/10.3390/foods9091261.

Cattaneo, T.M.P. *et al.* (2009) 'Contribution of Light Scattering to near Infrared Absorption in

Milk', *Journal of Near Infrared Spectroscopy*, 17(6), pp. 337–343. Available at: https://doi.org/10.1255/jnirs.867.

Cattaneo, T.M.P. *et al.* (2016) 'The aquaphotomics approach as a tool for studying the influence of food coating materials on cheese and winter melon samples', *Journal of Near Infrared Spectroscopy*, 24(4), pp. 381–390. Available at: https://doi.org/10.1255/jnirs.1238.

Chatani, E. *et al.* (2014) 'Water molecular system dynamics associated with amyloidogenic nucleation as revealed by real time near infrared spectroscopy and aquaphotomics', *PLoS ONE*, 9(7), pp. 1–10. Available at: https://doi.org/10.1371/journal.pone.0101997.

Chaves, R.P. *et al.* (2023) 'Convective Drying of Purple Basil (Ocimum basilicum L.) Leaves and Stability of Chlorophyll and Phenolic Compounds during the Process', *Plants*. Available at: https://doi.org/10.3390/plants12010127.

Chen, Y. *et al.* (2017) 'Using near infrared spectroscopy to determine moisture and starch content of corn processing products', *Journal of Near Infrared Spectroscopy*, 25(5), pp. 348–359. Available at: https://doi.org/10.1177/0967033517728146.

Čurlej, J. *et al.* (2021) 'Sights To Authentication and Adulteration of the Coffee in Global Aspect', *Journal of microbiology, biotechnology and food sciences*, 10(6), p. e4793. Available at: https://doi.org/10.15414/jmbfs.4793.

Czarnecki, M.A. *et al.* (2015) 'Advances in Molecular Structure and Interaction Studies Using Near-Infrared Spectroscopy', *Chemical Reviews*, 115(18), pp. 9707–9744. Available at: https://doi.org/10.1021/cr500013u.

Danezis, G.P. *et al.* (2016) 'Food authentication: Techniques, trends & emerging approaches', *TrAC - Trends in Analytical Chemistry*, 85, pp. 123–132. Available at: https://doi.org/10.1016/j.trac.2016.02.026.

Daniel, D. *et al.* (2018) 'Detection of coffee adulteration with soybean and corn by capillary electrophoresis-tandem mass spectrometry', *Food Chemistry*, 243, pp. 305–310. Available at: https://doi.org/https://doi.org/10.1016/j.foodchem.2017.09.140.

Downham, A. and Collins, P. (2000) 'Colouring our foods in the last and next millennium', *International Journal of Food Science and Technology*, 35(1), pp. 5–22. Available at: https://doi.org/10.1046/j.1365-2621.2000.00373.x.

Dumitriu, R.P. *et al.* (2015) 'Sulfadiazine—Chitosan Conjugates and Their Polyelectrolyte Complexes with Hyaluronate Destined to the Management of Burn Wounds', *Materials*, pp. 317–338. Available at: https://doi.org/10.3390/ma8010317.

ElNaker, N.A. *et al.* (2021) 'A metabolomics approach to evaluate the effect of lyophilization versus oven drying on the chemical composition of plant extracts', *Scientific Reports*, 11(1), pp. 1–11. Available at: https://doi.org/10.1038/s41598-021-02158-6.

Esquerre, C. *et al.* (2009) ‘Water absorbance pattern of physically-damaged mushrooms stored at ambient conditions’, *Journal of Near Infrared Spectroscopy*, 17, pp. 353–361.

Esslinger, S., Riedl, J. and Fauhl-Hassek, C. (2014) ‘Potential and limitations of non-targeted fingerprinting for authentication of food in official control’, *Food Research International*, 60, pp. 189–204. Available at: https://doi.org/10.1016/j.foodres.2013.10.015.

Esslinger, S, Riedl, J. and Fauhl-Hassek, C. (2014) ‘Potential and limitations of non-targeted fingerprinting for authentication of food in official control’, *Food Research International*, 60, pp. 189–204. Available at: https://doi.org/https://doi.org/10.1016/j.foodres.2013.10.015.

Esteban-Díez, I., González-Sáiz, J.M. and Pizarro, C. (2004) ‘Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy’, *Analytica Chimica Acta*, 525(2), pp. 171–182. Available at: https://doi.org/10.1016/j.aca.2004.08.057.

Eyiler, E. and Oztan, A. (2011) ‘Production of frankfurters with tomato powder as a natural additive’, *Lwt*, 44(1), pp. 307–311. Available at: https://doi.org/10.1016/j.lwt.2010.07.004.

Faqeerzada, M.A. *et al.* (2020) ‘Non-Targeted Detection of Adulterants in Almond Powder Using Spectroscopic Techniques Combined with Chemometrics’, *foods* [Preprint].

Forchetti, D.A.P. and Poppi, R.J. (2020) ‘Detection and Quantification of Adulterants in Roasted and Ground Coffee by NIR Hyperspectral Imaging and Multivariate Curve Resolution’, pp. 44–49.

Giraudo, A. *et al.* (2019) ‘Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis’, *Food Control*, 99, pp. 137–145. Available at: https://doi.org/https://doi.org/10.1016/j.foodcont.2018.12.033.

Gloess, A.N. *et al.* (2013) ‘Comparison of nine common coffee extraction methods: instrumental and sensory analysis’, *European Food Research and Technology*, 236(4), pp. 607–627. Available at: https://doi.org/10.1007/s00217-013-1917-x.

Gowen, A.A. *et al.* (2012) ‘Vibrational spectroscopy for analysis of water for human use and in aquatic ecosystems’, *Critical Reviews in Environmental Science and Technology*, 42(23), pp. 2546–2573. Available at: https://doi.org/10.1080/10643389.2011.592758.

Gowen, A.A. *et al.* (2015) ‘On the feasibility of near infrared spectroscopy to detect contaminants in water using single salt solutions as model systems’, *Talanta*, 131, pp. 609–618. Available at: https://doi.org/10.1016/j.talanta.2014.08.049.

Gowen, A.A., Amigo, J.M. and Tsenkova, R. (2013) ‘Characterisation of hydrogen bond perturbations in aqueous systems using aquaphotomics and multivariate curve resolution-alternating least squares’, *Analytica Chimica Acta*, 759, pp. 8–20. Available at: https://doi.org/10.1016/J.ACA.2012.10.007.

Grassi, S. *et al.* (2024) ‘Use of ultrasound and NIRs as tools for monitoring ice formation in

superchilled meat', *Journal of Food Engineering*, 369, p. 111957. Available at: https://doi.org/https://doi.org/10.1016/j.jfoodeng.2024.111957.

Hagos, M. *et al.* (2022) 'Development of Analytical Methods for Determination of β -Carotene in Pumpkin (Cucurbita maxima) Flesh , Peel , and Seed Powder Samples', *Journal of analytical chemistry*, 9363692. Available at: https://doi.org/10.1155/2022/9363692.

Hayashi, T. (2021) 'Water at biointerfaces: what makes surfaces bioinert?', in *PROCEEDING OF The 4th Aquaphotomics International Conference*, pp. 46–47.

Hazarika, U. and Gosztola, B. (2020) 'Lyophilization and its Effects on the Essential Oil Content and Composition of Herbs and Spices - A Review', *Acta Scientiarum Polonorum, Technologia Alimentaria*, 19(4), pp. 467–473. Available at: https://doi.org/10.17306/J.AFS.2020.0853.

Headrick, J.M. *et al.* (2005) 'Spectral signatures of hydrated proton vibrations in water clusters', *Science*, 308(5729), pp. 1765–1769. Available at: https://doi.org/10.1126/science.1113094.

ICO, I.C.O. (2022) 'Coffee market report', (July), pp. 1–10. Available at: https://www.ico.org/documents/cy2021-22/cmr-0722-e.pdf.

Iwamoto, M., Uozumi, J. and Nishinari, K. (1987) 'Preliminary investigation of the state of water in foods by near infrared spectroscopy', in J. Hello, K. Kaffka, and J. Gonczy (eds) *Proceedings of International Near Infrared Spectroscopy / Near Infrared Technology Conference*. Akademiai Kiado, pp. 3–12.

Izutsu, K. *et al.* (2006) 'Near-Infrared Analysis of Protein Secondary Structure in Aqueous Solutions and Freeze-Dried Solids', *Journal of pharmaceutical sciences*, 95(4), pp. 781–789. Available at: https://doi.org/10.1002/jps.

Jollife, I.T. and Cadima, J. (2016) 'Principal component analysis: A review and recent developments', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065). Available at: https://doi.org/10.1098/rsta.2015.0202.

Kakuta, N. *et al.* (2018) 'Near-infrared temperature measurement technique for water surrounding an induction-heated small magnetic sphere', *Journal of Visualized Experiments*, 2018(134), pp. 1–10. Available at: https://doi.org/10.3791/57407.

Kamiloglu, S. (2019) 'Authenticity and traceability in beverages', *Food Chemistry*, 277, pp. 12–24. Available at: https://doi.org/https://doi.org/10.1016/j.foodchem.2018.10.091.

Kaur, H., Künnemeyer, R. and McGlone, A. (2020) 'Investigating aquaphotomics for temperature-independent prediction of soluble solids content of pure apple juice', *J. Near Infrared Spectroscopy*, 28(2), pp. 103–112. Available at: https://doi.org/10.1177/0967033519898891.

Kovacs, Z. *et al.* (2019) 'Rapid bacteria selection using Aquaphotomics and near infrared spectroscopy', *Proceedings of the 18th International Conference on Near Infrared Spectroscopy*, pp. 65–69. Available at: https://doi.org/10.1255/nir2017.065.

van de Kraats, E.B., Munćan, J. and Tsenkova, R.N. (2019) 'Aquaphotomics – Origin, concept, applications and future perspectives', *Substantia*, 3, pp. 13–28. Available at: https://doi.org/10.13128/substantia-702.

Kraats, E.B. Van De, Munćan, J.S. and Tsenkova, R.N. (2019) 'Aquaphotomics Origin , concept , applications and future perspectives', *Substantia*, 3, pp. 13–28. Available at: https://doi.org/10.13128/Substantia-702.

Kuroki, S. *et al.* (2019) 'Water molecular structure underpins extreme desiccation tolerance of the resurrection plant Haberlea rhodopensis', *Scientific Reports*, (July 2018), pp. 1–12. Available at: https://doi.org/10.1038/s41598-019-39443-4.

Li, D. *et al.* (2019) 'A feasibility study on quantitative analysis of low concentration methanol by FT-NIR spectroscopy and aquaphotomics', *Journal of Molecular Structure*, 1182, pp. 197–203. Available at: https://doi.org/10.1016/j.molstruc.2019.01.056.

Li, S. *et al.* (2020) 'Rapid detection of saffron (Crocus sativus L.) Adulterated with lotus stamens and corn stigmas by near-infrared spectroscopy and chemometrics', *Industrial Crops and Products*, 152(April), p. 112539. Available at: https://doi.org/10.1016/j.indcrop.2020.112539.

Lohumi, S. *et al.* (2015) 'A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration', *Trends in Food Science and Technology*, 46(1), pp. 85–98. Available at: https://doi.org/10.1016/j.tifs.2015.08.003.

Malegori, C. *et al.* (2022) 'Analysing the water spectral pattern by near-infrared spectroscopy and chemometrics as a dynamic multidimensional biomarker in preservation: rice germ storage monitoring', *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 265, p. 120396. Available at: https://doi.org/10.1016/j.saa.2021.120396.

Mamani-linares, L.W., Gallo, C. and Alomar, D. (2012) 'Identification of cattle, llama and horse meat by near infrared reflectance or transflectance spectroscopy', *Meat Science*, 90, pp. 378–385. Available at: https://doi.org/10.1016/j.meatsci.2011.08.002.

Martínez-Valdivieso, D. *et al.* (2014) 'Application of near-infrared reflectance spectroscopy for predicting carotenoid content in summer squash fruit', *Computers and electronics in agriculture*, 108, pp. 71–79.

Matija, L.R. *et al.* (2012) 'Aquagrams: Water spectral pattern as characterization of hydrogenated nanomaterial', *FME Transactions*, 40(2), pp. 51–56.

McGlone, V.A. and Kawano, S. (1998) 'Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy', *Postharvest Biology and Technology*, 13(2), pp. 131–141. Available at: https://doi.org/10.1016/S0925-5214(98)00007-6.

McGrath, T.F. *et al.* (2018) 'What are the scientific challenges in moving from targeted to non-targeted methods for food fraud testing and how can they be addressed? – Spectroscopy case

study', *Trends in Food Science and Technology*, 76(March), pp. 38–55. Available at: https://doi.org/10.1016/j.tifs.2018.04.001.

Medina, S. *et al.* (2019) 'Trends in Food Science & Technology Current trends and recent advances on food authenticity technologies and chemometric approaches', *Trends in Food Science & Technology*, 85(December 2018), pp. 163–176. Available at: https://doi.org/10.1016/j.tifs.2019.01.017.

Moore, J.C., Spink, J. and Lipp, M. (2012) 'Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010', *Journal of Food Science*, 77(4). Available at: https://doi.org/10.1111/j.1750-3841.2012.02657.x.

Muncan, J. *et al.* (2021) 'Near infrared aquaphotomics study on common dietary fatty acids in cow's liquid, thawed milk', *Food Control*, 122(June 2020), p. 107805. Available at: https://doi.org/10.1016/j.foodcont.2020.107805.

Muncan, J. *et al.* (2022) 'Aquaphotomic Study of Effects of Different Mixing Waters on the Properties of Cement Mortar', *Molecules*, 27(22). Available at: https://doi.org/10.3390/molecules27227885.

Muncan, J., Tei, K. and Tsenkova, R. (2021) 'Real-time monitoring of yogurt fermentation process by aquaphotomics near-infrared spectroscopy', *Sensors (Switzerland)*, 21(1), pp. 1–18. Available at: https://doi.org/10.3390/s21010177.

Naes, T. *et al.* (2002) 'A user-friendly guide to multivariate calibration and classification', *Journal of chemometrics*, 17, pp. 571–572.

Ni, D. *et al.* (2024) 'Application of near-infrared spectroscopy and chemometrics for the rapid detection of insect protein adulteration from a simulated matrix', *Food Control*, 159(July 2023). Available at: https://doi.org/10.1016/j.foodcont.2023.110268.

Noah, L. *et al.* (1997) 'Near-Infrared Spectroscopy As Applied to Starch Analysis of Digestive Contents', *Journal of Agricultural and Food Chemistry*, 45(7), pp. 2593–2597. Available at: https://doi.org/10.1021/jf960891n.

Núñez, N. *et al.* (2020) 'Authentication of the origin, variety and roasting degree of coffee samples by non-targeted HPLC-UV fingerprinting and chemometrics. Application to the detection and quantitation of adulterated coffee samples', *Foods*, 9(3). Available at: https://doi.org/10.3390/foods9030378.

Nurhaslina, C.R., Andi Bacho, S. and Mustapa, A.N. (2022) 'Review on drying methods for herbal plants', *Materials Today: Proceedings*, 63, pp. S122–S139. Available at: https://doi.org/10.1016/j.matpr.2022.02.052.

Obadina, A., Ibrahim, J. and Adekoya, I. (2018) 'Influence of drying temperature and storage period on the quality of cherry and plum tomato powder', *Food Science and Nutrition*, 6(4), pp.

1146–1153. Available at: https://doi.org/10.1002/fsn3.658.

Oliveira, M.M. *et al.* (2020) 'Portable near-infrared spectroscopy for rapid authentication of adulterated paprika powder', *Journal of Food Composition and Analysis*, 87(September 2019). Available at: https://doi.org/10.1016/j.jfca.2019.103403.

Oliveira, M.M., Cruz-Tirado, J.P. and Barbin, D.F. (2019) 'Nontargeted Analytical Methods as a Powerful Tool for the Authentication of Spices and Herbs: A Review', *Comprehensive Reviews in Food Science and Food Safety*, 18(3), pp. 670–689. Available at: https://doi.org/10.1111/1541-4337.12436.

Otal, E.H., Iñón, F.A. and Andrade, F.J. (2003) 'Monitoring the temperature of dilute aqueous solutions using near-infrared water absorption', *Applied Spectroscopy*, 57(6), pp. 661–666. Available at: https://doi.org/10.1366/000370203322005355.

Ozaki, Y. (2002) 'Applications in chemistry in near-infrared spectroscopy: Principles, instruments and applications', in H. Siesler et al. (eds) *Near-infrared spectroscopy: principles, instruments, applications*. Chichester, UK: Wiley, pp. 179–211.

Pachura, N. *et al.* (2022) 'Chemical investigation on Salvia officinalis L. Affected by multiple drying techniques – The comprehensive analytical approach (HS-SPME, GC–MS, LC-MS/MS, GC-O and NMR)', *Food Chemistry*, 397, p. 133802. Available at: https://doi.org/https://doi.org/10.1016/j.foodchem.2022.133802.

Pasikatan, M.C. *et al.* (2001) 'Near Infrared Reflectance Spectroscopy for Online Particle Size Analysis of Powders and Ground Materials', *Journal of Near Infrared Spectroscopy*, 9(3), pp. 153–164. Available at: https://doi.org/10.1255/jnirs.303.

Pellegrini, N. *et al.* (2003) 'Total antioxidant capacity of plant foods, beverages and oils consumed in Italy assessed by three different in vitro assays', *Journal of Nutrition*, 133(9), pp. 2812–2819. Available at: https://doi.org/10.1093/jn/133.9.2812.

Pizarro, C., Esteban-Díez, I. and González-Sáiz, J.M. (2007) 'Mixture resolution according to the percentage of robusta variety in order to detect adulteration in roasted coffee by near infrared spectroscopy', *Analytica Chimica Acta*, 585(2), pp. 266–276. Available at: https://doi.org/https://doi.org/10.1016/j.aca.2006.12.057.

Poonia, A. *et al.* (2017) 'Detection of adulteration in milk: A review', *International Journal of Dairy Technology*, 70(1), pp. 23–42. Available at: https://doi.org/10.1111/1471-0307.12274.

Rahimmalek, M. and Goli, S.A.H. (2013) 'Evaluation of six drying treatments with respect to essential oil yield, composition and color characteristics of Thymys daenensis subsp. daenensis. Celak leaves', *Industrial Crops and Products*, 42(1), pp. 613–619. Available at: https://doi.org/10.1016/j.indcrop.2012.06.012.

Ribeiro, J.S., Ferreira, M.M.C. and Salva, T.J.G. (2011) 'Chemometric models for the quantitative

descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy', *Talanta*, 83(5), pp. 1352–1358. Available at: https://doi.org/10.1016/j.talanta.2010.11.001.

Schievano, E. *et al.* (2014) 'Rapid Authentication of Coffee Blends and Quantification of 16-O-Methylcafestol in Roasted Coffee Beans by Nuclear Magnetic Resonance', *Journal of Agricultural and Food Chemistry*, 62(51), pp. 12309–12314. Available at: https://doi.org/10.1021/jf505013d.

Sciuto, S. *et al.* (2017) 'Rapid screening technique to identify Sudan dyes (i to IV) in adulterated tomato sauce, chilli powder, and palm oil by innovatIVe high-resolution mass spectrometry', *Journal of Food Protection*, 80(4), pp. 640–644. Available at: https://doi.org/10.4315/0362-028X.JFP-16-313.

Segtnan, V.H., Isaksson, T. and Ozaki, Y. (2001) 'Studies on the Structure of Water Using Spectroscopy and Principal Component Analysis', *Analytical Chemistry*, 73(13), pp. 3153–3161. Available at: https://doi.org/10.1021/ac010102n.

Sheelarani, B. *et al.* (2022) 'Comparative interaction of an anionic dye, ponceau 4R with triple viz., anionic, non-ionic and cationic micellar systems: Spectral and conductometric analysis', *Food Chemistry Advances*, 1(July), p. 100113. Available at: https://doi.org/10.1016/j.focha.2022.100113.

Singleton, V.L. and Joseph A. Rossi, J. (1965) 'Colorimetry of Total Phenolics with Phosphomolybdic-Phosphotungstic Acid Reagents', *American Journal of Enology and Viticulture*, 16(3), pp. 144 LP – 158. Available at: https://doi.org/10.5344/ajev.1965.16.3.144.

Siriamornpun, S., Kaisoon, O. and Meeso, N. (2012) 'Changes in colour, antioxidant activities and carotenoids (lycopene, β-carotene, lutein) of marigold flower (Tagetes erecta L.) resulting from different drying processes', *Journal of Functional Foods*, 4(4), pp. 757–766. Available at: https://doi.org/10.1016/j.jff.2012.05.002.

Smith, J.D. *et al.* (2005) 'Unified description of temperature-dependent hydrogen-bond rearrangements in liquid water', *Proceedings of the National Academy of Sciences (PNAS)*, 102(40), p. 14171.

Song, H.Y. *et al.* (2019) 'Analytical method to detect adulteration of ground roasted coffee', *International Journal of Food Science and Technology*, 54(1), pp. 256–262. Available at: https://doi.org/10.1111/ijfs.13942.

Spaniolas, S. *et al.* (2006) 'Authentication of Coffee by Means of PCR-RFLP Analysis and Lab-on-a-Chip Capillary Electrophoresis', *Journal of Agricultural and Food Chemistry*, 54(20), pp. 7466–7470. Available at: https://doi.org/10.1021/jf061164n.

Speer, K. and Kölling-Speer, I. (2006) 'The lipid fraction of the coffee bean', *Brazilian Journal of Plant Physiology*, 18(1), pp. 201–216. Available at: https://doi.org/10.1590/S1677-04202006000100014.

Spence, C. (2015) ‘On the psychological impact of food colour’, *Flavour*, 4(1), pp. 1–16. Available at: https://doi.org/10.1186/s13411-015-0031-3.

Stenberg, B. (2010) ‘Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon’, *Geoderma*, 158(1–2), pp. 15–22. Available at: https://doi.org/10.1016/j.geoderma.2010.04.008.

Suhandy, D. and Yulia, M. (2017) ‘The Quantification of Adulteration in Arabica Coffee using UV-Visible Spectroscopy in Combination with Two Different PLS Regressions’, *Aceh International Journal of Science and Technology*, 6(2), pp. 59–67. Available at: https://doi.org/10.13170/aijst.6.2.8457.

Tavares, K.M. *et al.* (2016) ‘Free tocopherols as chemical markers for Arabica coffee adulteration with maize and coffee by-products’, *Food Control*, 70, pp. 318–324. Available at: https://doi.org/10.1016/j.foodcont.2016.06.011.

Teran, E. (2024) ‘Enhancement of Coffee Quality Attributes by Combining Processing Methods and Varieties’, *Beverages*. Available at: https://doi.org/10.3390/beverages10010010.

Thamkaew, G., Sjöholm, I. and Galindo, F.G. (2021) ‘A review of drying methods for improving the quality of dried herbs’, *Critical Reviews in Food Science and Nutrition*, 61(11), pp. 1763–1786. Available at: https://doi.org/10.1080/10408398.2020.1765309.

Toci, A.T. *et al.* (2016) ‘Coffee Adulteration: More than Two Decades of Research.’, *Critical reviews in analytical chemistry*, 46(2), pp. 83–92. Available at: https://doi.org/10.1080/10408347.2014.966185.

Trygg, J. and Holmes, E. (2007) ‘Journal of Proteome Research 2007, 6, 469-479.pdf’, pp. 469–479.

Tsenkova, R. (2007) ‘Aquaphotomics: extended water mirror approach reveals peculiarities of prion protein alloforms’, 18(6).

Tsenkova, R. (2008a) ‘Aquaphotomics: The Extended Water Mirror Effect Explains Why Small Concentrations of Protein in Solution Can Be Measured with near Infrared Light’, *NIR news*, 19(4), pp. 13–14. Available at: https://doi.org/10.1255/nirn.1079.

Tsenkova, R. (2008b) ‘Aquaphotomics: The Extended Water Mirror Effect Explains Why Small Concentrations of Protein in Solution Can Be Measured with near Infrared Light’, *NIR news*, 19(4), pp. 13–14. Available at: https://doi.org/10.1255/nirn.1079.

Tsenkova, R. (2009) ‘Aquaphotomics: Dynamic spectroscopy of aqueous and biological systems describes peculiarities of water’, *Journal of Near Infrared Spectroscopy*, 17(6), pp. 303–313. Available at: https://doi.org/10.1255/jnirs.869.

Tsenkova, Roumiana, Muncan, J., *et al.* (2018a) ‘Essentials of aquaphotomics and its chemometrics approaches’, *Frontiers in Chemistry*, 6(AUG), pp. 1–25. Available at:

https://doi.org/10.3389/fchem.2018.00363.

Tsenkova, R. *et al.* (2018) 'Essentials of aquaphotomics and its chemometrics approaches', *Frontiers in Chemistry*, 6, p. 363. Available at: https://doi.org/10.3389/fchem.2018.00363.

Tsenkova, Roumiana, Munćan, J., *et al.* (2018) 'Essentials of Aquaphotomics and Its Chemometrics Approaches', *Frontiers in Chemistry*, 6(August), pp. 1–25. Available at: https://doi.org/10.3389/fchem.2018.00363.

Tsenkova, Roumiana, Muncan, J., *et al.* (2018b) 'Essentials of Aquaphotomics and Its Chemometrics Approaches', 6(August), pp. 1–25. Available at: https://doi.org/10.3389/fchem.2018.00363.

Vanoli, M. *et al.* (2018) 'Water spectral pattern as a marker for studying apple sensory texture', *Advances in Horticultural Science*, 32(3), pp. 343–351. Available at: https://doi.org/10.13128/ahs-22380.

Wermelinger, T. *et al.* (2011) 'Quantification of the Robusta Fraction in a Coffee Blend via Raman Spectroscopy: Proof of Principle', *Journal of Agricultural and Food Chemistry*, 59(17), pp. 9074–9079. Available at: https://doi.org/10.1021/jf201918a.

Wold, S. *et al.* (1998) 'Orthogonal signal correction of near-infrared spectra', *Chemometrics and Intelligent Laboratory Systems*, 44(1–2), pp. 175–185. Available at: https://doi.org/10.1016/S0169-7439(98)00109-9.

Wold, S., Sjöström, M. and Eriksson, L. (2001) 'PLS-regression: a basic tool of chemometrics', *Chemometrics and Intelligent Laboratory Systems*, 58(2), pp. 109–130. Available at: https://doi.org/https://doi.org/10.1016/S0169-7439(01)00155-1.

Wu, D., He, Y. and Feng, S. (2008) 'Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment', *Analytica Chimica Acta*, 610(2), pp. 232–242. Available at: https://doi.org/https://doi.org/10.1016/j.aca.2008.01.056.

Xantheas, S.S. (1995) 'Ab initio studies of cyclic water clusters (H2O)n, n=1–6. III. Comparison of density functional with MP2 results', *The Journal of Chemical Physics*, 102(11), p. 4505. Available at: https://doi.org/10.1063/1.469499.

Yakes, B.J. *et al.* (2017) 'Effects of Wet-Blending on Detection of Melamine in Spray-Dried Lactose', *Journal of Agricultural and Food Chemistry*, 65(28), pp. 5789–5798. Available at: https://doi.org/10.1021/acs.jafc.7b00834.

Yang, X. *et al.* (2020) 'Manuka honey adulteration detection based on near-infrared spectroscopy combined with aquaphotomics', *Lwt*, 132(June), p. 109837. Available at: https://doi.org/10.1016/j.lwt.2020.109837.

Zaukuu, J. *et al.* (2019) 'Spectroscopy as a rapid detecting paprika powder adulteration', *Hungarian Agricultural Engineering*, 36, pp. 38–43. Available at:

https://doi.org/10.17676/hae.2019.36.38.

Zhang, L. *et al.* (2022) ‘Effects of Different Processing Methods Based on Different Drying Conditions on the Active Ingredients of Salvia miltiorrhiza Bunge’, *Molecules*. Available at: https://doi.org/10.3390/molecules27154860.

11. ACKNOWLEDGMENT

| This is dedicated to the loving memory of my father, to my mother, sister, and brother |

Thank you |شكرا | Köszönöm szépen

To

Tempus Public Foundation and Stipendium Hungaricum Program

Family and Friends

Supervisor, Professors, Colleagues

at

Department of food measurements and Process Control, MATE

Department of Medicinal and Aromatic Plants, MATE

Department of Food Science and Technology, Kwame Nkrumah University of Science and Technology (KNUST)