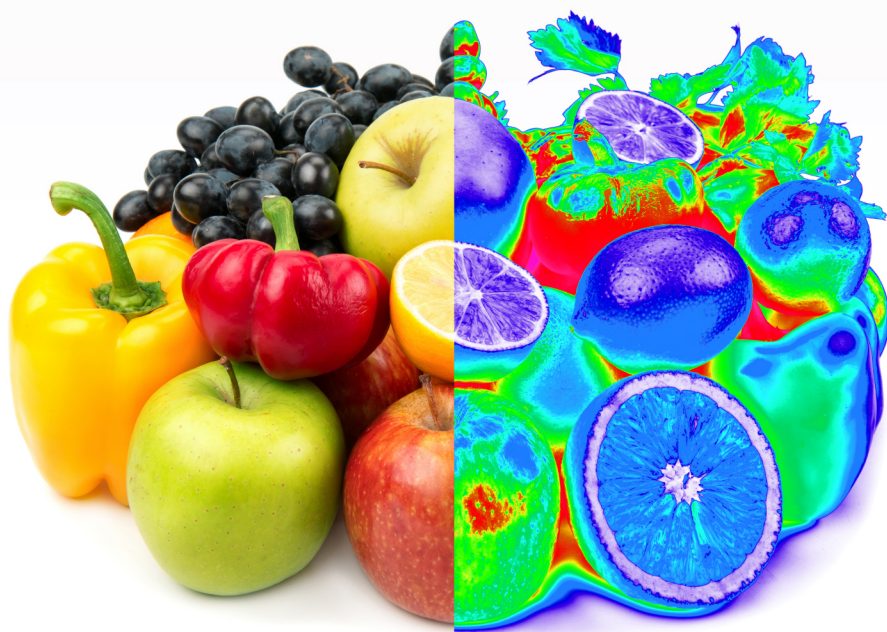


A Guide to
NEAR-INFRARED
SPECTROSCOPY



Juan Francisco García Martín

E D I T O R

NOVA

Analytical Chemistry and Microchemistry



No part of this digital document may be reproduced, stored in a retrieval system or transmitted in any form or by any means. The publisher has taken reasonable care in the preparation of this digital document, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained herein. This digital document is sold with the clear understanding that the publisher is not engaged in rendering legal, medical or any other professional services.

Analytical Chemistry and Microchemistry

Chemometrics: Advances in Applications and Research

Larry D. Crenshaw (Editor)

2022. ISBN: 979-8-88697-274-0 (Hardcover)

2022. ISBN: 979-8-88697-311-2 (eBook)

Functional Analytical Techniques in Pharmaceutical Chemistry

Malik Saadullah (Editor)

2022. ISBN: 979-8-88697-059-3 (Hardcover)

2022. ISBN: 979-8-88697-288-7 (eBook)

FTIR Spectroscopy: Advances in Research and Applications

Adriana S. França, PhD (Editor)

Leandro S. Oliveira, PhD (Editor)

2022. ISBN: 978-1-68507-571-2 (Softcover)

2022. ISBN: 978-1-68507-939-0 (eBook)

Trace Metals: Sources, Applications and Environmental Implications

Oscar M. Thygesen (Editor)

2022. ISBN: 978-1-68507-797-6 (Softcover)

2022. ISBN: 978-1-68507-827-0 (eBook)

Mass Spectrometry: Theory and Applications

William O. Nichols (Editor)

2021. ISBN: 978-1-53619-790-7 (Hardcover)

2021. ISBN: 978-1-53619-874-4 (eBook)

More information about this series can be found at

<https://novapublishers.com/product-category/series/analytical-chemistry-and-microchemistry/>

Juan Francisco García Martín

Editor

A Guide to Near-Infrared Spectroscopy



Copyright © 2023 by Nova Science Publishers, Inc.

DOI: <https://doi.org/10.52305/NBCM0163>.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means: electronic, electrostatic, magnetic, tape, mechanical photocopying, recording or otherwise without the written permission of the Publisher.

We have partnered with Copyright Clearance Center to make it easy for you to obtain permissions to reuse content from this publication. Please visit copyright.com and search by Title, ISBN, or ISSN.

For further questions about using the service on copyright.com, please contact:

	Copyright Clearance Center	
Phone: +1-(978) 750-8400	Fax: +1-(978) 750-4470	E-mail: info@copyright.com

NOTICE TO THE READER

The Publisher has taken reasonable care in the preparation of this book but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained in this book. The Publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance upon, this material. Any parts of this book based on government reports are so indicated and copyright is claimed for those parts to the extent applicable to compilations of such works.

Independent verification should be sought for any data, advice or recommendations contained in this book. In addition, no responsibility is assumed by the Publisher for any injury and/or damage to persons or property arising from any methods, products, instructions, ideas or otherwise contained in this publication.

This publication is designed to provide accurate and authoritative information with regards to the subject matter covered herein. It is sold with the clear understanding that the Publisher is not engaged in rendering legal or any other professional services. If legal or any other expert assistance is required, the services of a competent person should be sought. FROM A DECLARATION OF PARTICIPANTS JOINTLY ADOPTED BY A COMMITTEE OF THE AMERICAN BAR ASSOCIATION AND A COMMITTEE OF PUBLISHERS.

Library of Congress Cataloging-in-Publication Data

ISBN: 979-8-88697-; 8: /: "gDqqm-

Published by Nova Science Publishers, Inc. † New York

Contents

Preface	vii
Chapter 1	Principles, Theories and Applications of Near-Infrared Spectroscopy for Food Quality and Safety Control	1
	Wei jie Lan, Yan Ge, Hui Ma, Songchao Chen, Kang Tu and Leiqing Pan	
Chapter 2	The Current and Potential Roles of Near-Infrared Spectroscopy in the Digital Food Era	33
	Daniel Cozzolino	
Chapter 3	The Determination of Olive Oil's Moisture by a Hand-Held Near-Infrared Spectrometer	53
	Noelia García-Criado and Juan F. García Martín	
Chapter 4	Near-Infrared Hyperspectral and Multispectral Imaging Principles and Applications in the Quality of Fruits and Vegetables	83
	Sneha Sharma, Sumesh K.C., Bim Prasad Shrestha and Panmanas Sirisomboon	
Chapter 5	The Prediction of Respiratory and Degradation Rates of Horticultural Crops by Near-Infrared Spectroscopy and Hyperspectral Imaging.....	117
	Yoshio Makino	

Chapter 6	Near-Infrared Spectroscopy, Hyperspectral, Multispectral Imaging Principles and Applications in Energy Properties of Biomass.....	133
	Jetsada Posom, Bijendra Shrestha, Kanvisit Maraphum, Lakkana Pitak, Khwantri Saengprachatanarug, Panmanas Sirisomboon and Bim Prasad Shrestha	
Chapter 7	The Determination of D-Xylose and Xylitol by Near-Infrared Spectroscopy Over the Fermentation of Olive Stone Hydrolysates.....	201
	Juan F. García Martín, Manuel Cuevas and Sebastián Sánchez	
Index	217
Editor's Contact Information	225

Preface

Near-infrared spectroscopy (NIRS) is a technique with several potential applications in many fields due to its numerous advantages. Some of these advantages are that minimal or no sample preparation is required, the non-destruction of the samples and that NIRS does not involve the use of any hazardous chemicals during the analysis. NIR spectroscopy is currently used for the quantification of parameters of interest in samples throughout the development of calibration models that relate the samples NIR spectra with the values of the property of interest, these last obtained by a reference method.

In Chapter 1, Lan et al. provide an overview of the principles, theory, instrumentation and data analysis procedures for the use of NIRS, following with a comprehensive discussion on the applications of NIRS for food quality and safety control during the last 10 years, critical issues, challenges and future trends of research, especially for minimizing instrumentation, improving data acquisition accuracy, and developing more robust mathematical modeling approaches.

Subsequently, Cozzolino explores in Chapter 2 the current and future roles of NIRS in the digital food era with particular emphasis in food functionality, safety, and security. With regard to food functionality, NIRS is expected not only to measure the proximate composition (e.g., protein, starch and dry matter contents), but also to monitor the functional characteristics or properties of the same food sample. As for food safety and security, NIRS can undertake a major role in monitoring the authenticity, origin, and provenance of both food ingredients and products along the supply and value chains.

In Chapter 3, García-Criado and García Martín demonstrate the feasibility of NIRS to determine the moisture in olive oils. These authors built two partial least squares regression models, using different internal validation methods (namely leverage and full cross validation), to relate the NIR absorbance of virgin olive oils to their moisture. The use of the model built with full cross validation in an external set of samples (validation set) not used for the model building led to the lowest standard error of prediction.

Hyperspectral (HSI) and multispectral (MSI) imaging technologies based on the features of the NIR region are being widely investigated and applied as a non-destructive, reliable, and accurate technique to monitor the quality and composition of agricultural products. Unlike NIRS, which provides spectral information at a point of the sample, multispectral and hyperspectral imaging systems provide spectral and spatial information of the whole sample.

In Chapter 4, Sharma et al. focus on the theory and principles (components, image acquisition and image processing) of HSI and MSI technologies along with the most recent applications of these imaging techniques for predicting physicochemical properties, antioxidants, chemical components, texture, defects, maturity classification, shape, and size of fruits and agricultural products.

Makino provides an example of the usefulness of the features of the infrared spectrum for the determination of the quality of horticultural crops, both by NIRS and HSI, in Chapter 5. This author uses NIRS combined with artificial neural networks to non-destructively estimate the O₂ uptake rate by NIRS based on the light absorption derived from the cytochrome *c* oxidase. Besides, he studies the local degreening rate of broccoli buds immediately after harvesting by measuring the spatial distribution of the spectral reflection spectrum by hyperspectral imaging.

Finally, NIRS can be also used for the non-destructive measurement of biomass quality. In Chapter 6, Posom et al. describe the applications of NIRS and HSI for analyzing energy parameters of biomass, such as heating value, proximate data, elemental composition, combustion index, pyrolysis characteristics, mechanical properties and so on. Besides, NIRS has proved its feasibility for the determination of the ethanol produced by fermentation with yeasts of lignocellulose biomass.

Notwithstanding, the determination of other high-added value products obtained during ethanolic fermentation of lignocellulose biomass is of high interest for the industry. Thus, García Martín et al. show in Chapter 7 the potential of NIRS for the real-time determination of xylose and xylitol over the fermentation of olive stones' hydrolysates with the yeast *Pachysolen tannophilus*.

Chapter 1

Principles, Theories and Applications of Near-Infrared Spectroscopy for Food Quality and Safety Control

Weijie Lan^{1,*}, PhD

Yan Ge²

Hui Ma¹

Songchao Chen³, PhD

Kang Tu¹, PhD

and Leiqing Pan¹, PhD

¹ College of Food Science and Technology,
Nanjing Agricultural University, Nanjing, Jiangsu, China

² College of Engineering,
Nanjing Agricultural University, Nanjing, Jiangsu, China

³ Hangzhou Global Scientific and Technological Innovation Center,
Zhejiang University, Hangzhou, China

Abstract

Since Karl Norris, “the father of near-infrared (NIR) spectroscopy,” demonstrated the NIR region was very useful for quantitative analysis of agricultural samples in the 1980s, it has played an important role in the safety inspection and quality analysis of agricultural products, with the advantages of rapid data acquisition, limited sample preparation requirements, and no chemical wastes. This chapter provides an overview of the principles and theory of NIR spectroscopy to measure

* Corresponding Author’s Email: weijie.lan@njau.edu.cn.

and characterize a large variability of food products. It then presents instrumentation and data analysis procedures for implementing different types of commercially available NIR systems.

Applications of these NIR techniques for food quality and safety control during the last 10 years, particularly in meat, fish, and fishery products, edible oils, milk and dairy products, grains and grain products, fruits, and vegetables, are then comprehensively summarized. Finally, discussions are given on critical issues, challenges and future research using NIR spectroscopy, especially for minimizing instrumentation, improving data acquisition accuracy, and developing more robust mathematical modeling approaches. This chapter aims to help researchers and practitioners gain a better understanding of NIR techniques and their advantages, and limitations, thus stimulating further research on NIR applications in the food and agricultural sector.

Keywords: near infrared spectroscopy, basic theory, commercial instrumentation, food quality and safety

Introduction

For many years, several vibrational spectroscopy techniques have been considered as rapid, non-destructive, and inexpensive tools for molecular structure and organic matters in food products (Abbas, Pissard, & Baeten, 2020; Nicolai et al., 2007). The development of advanced vibrational spectroscopic instrumentations makes them possible to be applied both in the laboratory basic researches (Dupuy, Duponchel, Huvenne, Sombret, & Legrand, 1996; Dupuy, Galtier, Ollivier, Vanloot, & Artaud, 2010; Galtier et al., 2011) and the online or at-line food industrial production and process (De Beer et al., 2009; Picouet, Gou, Hyypiö, & Castellari, 2018; Porep, Kammerer, & Carle, 2015).

Near infrared (NIR) spectroscopy has played an important role in the analysis of food products due to their advantages such as being non-destructive, quick and limited chemical wastes. Since William Herschel discovered the NIR spectral region in the 1800s, NIR technology appeared in the 1960s and reached a rapid and exciting development in the late 1980s. In recent years, it has become more accurate and efficient with the development of new instruments and advanced chemometric methods. NIR light is based on the absorption of the electromagnetic spectrum in the wavenumbers of 12500–4000 cm^{-1} (780–2500 nm). The basic principle of NIR spectroscopy is

the absorbance of radiation at molecular vibrational frequencies occurs for the O–H, N–H, and C–H groups and for the C–C, C–O, C–N, and N–O groups, which are the main chemical bonds of organic compounds. In addition, NIR detection conforms to Lambert-Beer's law, thus, the specific chemical bond determines the wavelength and quantity of absorbed light. Therefore, the quality attributes of food products can be assessed through the analysis of their NIR spectrum. However, there are some inherent factors influenced by conventional NIR techniques, such as type of instrument, sensing mode, and light source/detecting probe setup. A varied selection of spectroscopic instruments is accessible and there are around sixty NIR spectrometer manufacturers around the world. These instruments can be divided into three groups: laboratory devices, sorting and grading, and portable devices. Based on these variable NIR spectrometers, they have been widely used for the quality evaluation of physicochemical properties, specific nutrient components, authentication and geographical origin traceability of foods. Meanwhile, the applicability of NIR spectroscopy for the off-line and in-line monitoring of quality in foods also has shown great potential.

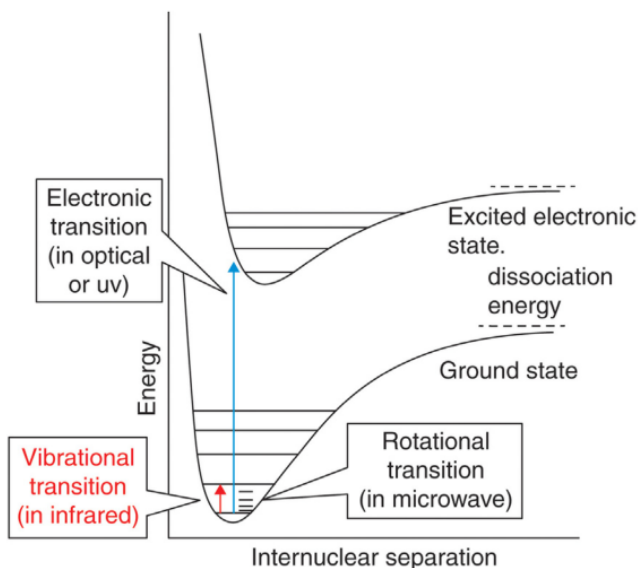
In this chapter, we introduced the basic knowledge of NIR technique and provided an overview of the principles and theory of NIR spectroscopy and presents instrumentation and data analysis procedures for implementing different types of commercially available NIR systems, then comprehensively summarized the applications of these NIR techniques for food quality and safety control during the last 10 years. Finally, discussions are given on critical issues, challenges and future research using NIR spectroscopy.

Theory

When the IR light irradiates a molecule, the interaction of electromagnetic radiation with a molecule might result in the light absorption, which induces an electronic, vibrational, or rotational transition, according to the energetic level of this radiation (Figure 1). The energy of the absorbed IR light ($h\nu$) is equal to an energy difference between a certain energy level of vibration of the molecule (having an energy E_m) and another energy level of vibration of a molecule (having an energy E_n). In the form of an equation (1) as follows:

$$h\nu = E_n - E_m \quad (1)$$

This equation is named as Bohr frequency condition, which indicates the absorption of IR light takes place depends on a transition between energy levels of a molecular vibration. Therefore, an IR absorption spectrum is a vibrational spectrum of a molecule.



Source: <http://hyperphysics.phy-astr.gsu.edu/Hbase/molecule/imgmol/molwel.gif>.

Figure 1. Molecular spectra involving electronic, vibrational, and rotational transitions.

Vibrational energy, like other molecular energies, can be quantized. Molecules can occupy discrete energy levels defined by whole numbers (0, 1, 2, etc.). Normally, molecules occupy the lowest energy level 0. A transition from level 0 to 1 leads to a fundamental transition in the molecule, whereas transitions from energy level 0 to 2 or 3 are defined as first and second overtones. Overtone frequencies are characteristic of the NIR spectral region (Abbas et al., 2020).

NIR absorption frequencies are expressed as wavelengths (λ) in nm from 780 nm to 2500 nm. NIR spectra consist of broad bands produced by overlapping absorption and correspond mainly to overtones and combinations of vibrational modes of C-H, N-H, and O-H chemical bonds. Harmonics correspond to energy transitions higher than those of fundamental vibrations do. The frequencies of the first and second harmonics correspond to about

twice or three times those of the fundamental vibrations. Combination bands result from transitions in which two or more different vibrational modes of a functional group occur simultaneously; the frequency of a combination band is the sum or multiple of the corresponding frequencies. Absorption intensity decreases as the overtone level increases (Osborne, 2000).

Instrumentation

There are three main types of NIR spectrometers on the market, including laboratory stationary NIR spectrometers, portable spectrometers, and industrial on-line spectrometers.

Laboratory Stationary (i.e., Benchtop) NIR Spectrometer

Benchtop NIR spectrometers are nowadays fully mature instruments that follow a fairly standard design scheme using either a Michelson or a polarization interferometer and recording the spectrum according to the Fourier transform (FT) principle (Tamburini, Marchetti, & Pedrini, 2014). The design scheme of a benchtop NIR spectrometer is not significantly different from that of a general optical absorption spectroscopy instrument. It includes a light source, a wavelength selector, and a detector as the main building blocks/functional blocks connected by optics for beam propagation. While NIR spectrometers can be readily configured for transmission measurements, usually with a sample chamber integrated into the spectrometer housing, the acquisition of spectra by diffuse reflection is much more popular for analytical applications. This advantage arises from the physical principles of NIR spectroscopy (i.e., the high transmittance of typical organic materials to NIR wavelengths), which make it possible to obtain good quality reflectance spectra from samples without prior pretreatment. Differently, NIR spectrometers can be easily equipped with fiber probes for remote sensing in diffuse reflectance mode, further increasing the versatility and practical utility of NIR spectroscopy.

Handheld and Micro NIR Spectrometers

Recent advances have been made in portable and microscopic NIR instruments thanks to rapid advances in sensor technologies such as linear variable filters (LVF) or microelectromechanical systems (MEMS) and their integration with microoptics (MOEMS). The precise dimensions and alignment of MEMS devices, combined with the mechanical stability that comes with miniaturization, make MEMS optical sensors suitable for a variety of challenging measurements. Depending on the type of detectors, portable spectrometers can be divided into two categories (array detectors and single detectors). In miniaturized NIR spectrometers, cost and power consumption are the most important factors. Therefore, single element detectors are preferred in miniaturized NIR spectrometers. However, the disadvantage of this type of detector is that the spectra obtained are noisier than those from standard InGaAs detectors (with a cut-off of 1700 nm) and the detector needs to be cooled (Solgaard, Godil, Howe, Lee, Peter, & Zappe, 2014; Wiedemair, Langore, Garsleitner, Dillinger, & Huck, 2019). However, the challenge with miniaturized NIR spectroscopy is the lower resolution compared to benchtop instruments.

Online NIR Spectrometer

Recent rapid advances in NIR instrumentation, especially in fiber optic techniques, and software have activated the applications of NIR spectroscopy for in-process and on-line sample monitoring. The monochromator/detector principles in scanning NIR spectroscopy are variable. In addition to various types of spectrometers with moving parts, such as grating instruments and FT spectrometers with Michelson or polarization interferometers, there are also fast scanning instruments without moving parts, such as diode array, acousto-optic tunable filters, and light-emitting diode systems (Lin, A Rasco, G Cavinato, & Al-Holy, 2009).

Sample Presentation

NIR spectrometers has the potential to take measurements on a great variety of food samples in various forms, including all kinds of solids, liquids, purees,

powders, etc., for example, the Figure 2 displayed the NIR detections on solid (apple) and liquid (puree) samples.

The interaction between the NIR source and characterized matters can be analyzed through the transmittance, transreflectance and diffuse reflectance modes:

- The transmission mode is mainly adapted to the transparent matters with the constant sample thickness, quartz cuvette and flow cell. In these conditions, the absorption depends only on the concentration of the absorbing component.
- The diffuse transmission/ reflection mode is suitable for the opaque liquids, solid and granular samples, in this case part of light is scattered. The diffuse reflectance measurements can be adapted both on apples and cooked purees.
- The transfection mode, with the combination of reflections and transmission, is suitable for characterizing turbid or clear liquids.



Figure 2. An example of the measurements on (a) apple and (b) purees using the NIR spectrometer.

Chemometrics

The NIR prediction requires an initial calibration step, which is time consuming. Indeed, on a set of samples, representative of the expected variability, both NIR spectra and their corresponding reference data are needed to establish predictive models using multivariate statistical and

mathematical data analyses. Several parameters can thus be evaluated from a single spectrum, with varying precision. The application of spectroscopic techniques has been greatly reinforced by the development of chemometrics methods and the calculation speed of computers in parallel. The chemometrics field is very large and some of the most common mathematical and statistical methods have been detailed in reviews (Mishra, Roger, et al., 2021; Oliveri, Malegori, & Casale, 2020; Roger & Boulet, 2018). The applications of chemometrics tools to detect food quality can be mainly divided into two groups, including: i) classification modelling to discriminate the sample variability and regression modelling to predict quantitative variables.

Spectral Data Pre-Processing

Spectral data acquired on samples with infrared spectrometer are associated with background information and noises. Before using chemometric tools, it is often necessary to apply a pre-treatment on spectral dataset. There are some common pre-processing methods, such as baseline correction, smoothing, derivatives, standard normal variate transformation (SNV) etc.

Baseline correction is an important pre-processing technique used to separate true spectroscopic signals from interferences or to remove background effects (Liland, Rukke, Olsen, & Isaksson, 2011). It has been reported to improve the NIR prediction of SSC in apples (Vincent, Wang, Nibouche, & Maguire, 2018; Xia, Fan, Li, Tian, Huang, & Chen, 2020).

Spectral smoothing mainly includes moving smoothing and, Savitzky-Golay (SG) smoothing is one of the methods used to eliminate noises (Gorry, 1990). Optimizing the smoothing window width of spectral matrix is crucial to increase the Signal-to-Noise Ratio (SNR). It has been widely applied to the NIR prediction of SSC in apples (Sun, Zhang, Pan, & Liu, 2009; Tian, Fan, Li, Xia, Huang, & Zhao, 2019) and the detection of harvesting optimal picking dates (Peirs, Lammertyn, Ooms, & Nicolai, 2001). Moreover, it is often combined with several other pre-processing methods, such as SNV, multiplicative scatter correction (MSC) etc.

Regarding the elimination or minimization of unwanted systematic effects (the multiplicative interferences of scatter, particle size, and the change of light distance etc.) (Barnes, Dhanoa, & Lister, 1989), SNV as a typical mathematical method is useful to correct both baseline shifts and global intensity variations (Barnes et al., 1989). The SNV pre-treated NIR (Dong &

Guo, 2015; Qing, Ji, & Zude, 2007) and MIR spectra (Bureau et al., 2009; Labaky et al., 2021) give a good prediction accuracy of fruit quality traits.

Derivatives including first and second derivatives are used to remove background and increase spectral resolution. There are two algorithms, direct differentiation, and SG. The most used method is the SG, where the data within a moving window are fitted by a polynomial of a given degree to generate a differential of a chosen degree (Gorry, 1990). The selection of the proper width of the moving window is very important in this function. Many calibration models using derivatives of NIR spectra, give good prediction of apple quality traits (Beghi, Giovenzana, Civelli, & Guidetti, 2016; Giovanelli, Sinelli, Beghi, Guidetti, & Casiraghi, 2014; Lammertyn, Nicolaï, Ooms, De Smedt, & De Baerdemaeker, 1998).

Recently, Mishra, Biancolillo, Roger, Marini, and Rutledge (2020) demonstrated the benefits of using the multi-block approach for spectral data preprocessing, where several pre-treatments were combined using sequential and orthogonalized partial least squares (SO-PLS), thus leading to a boosting procedure. They proposed a chemometric strategy named ‘SPORT’, both allowing both the fusion of multiple preprocessings and the identification of the best preprocessing techniques and their combinations.

Classification Modelling

Principal Component Analysis (PCA)

PCA, which was proposed by Pearson (1901), is one of the most widely used basic tools for discriminating the samples based on their spectra and/or reference dataset. It is a data-reduction technique, which reduces the dimensions of a dataset into its principal components (PC) with minimal loss of information. The PCs are uncorrelated, and the former retain most of the variation in all the original dataset (Jolliffe & Cadima, 2016), indicating the directions where there is the most variance. Although some precision is lost, data analysis is facilitated since the number of variables is reduced and can be visualized graphically. This is an unsupervised method, which can be used to identify informative spectral features and reduce noise.

PCA depends on the assumption that a large variability (i.e., a high variance value) is synonymous with a high amount of information (Oliveri et al., 2020). For this reason, PCA algorithms search for the maximum variance direction, in the multidimensional space of the original dataset (Figure 3). The maximum variance direction represents the first principal component (PC1).

The second PC (PC2) keeps the maximum variance among all directions orthogonal to (= not correlated to) the PC1. These two new variables (PC1 and PC2) are thus not intercorrelated (Jolliffe et al., 2016; Oliveri et al., 2020). All subsequent principal components are calculated iteratively and are by construction not correlated with one another.

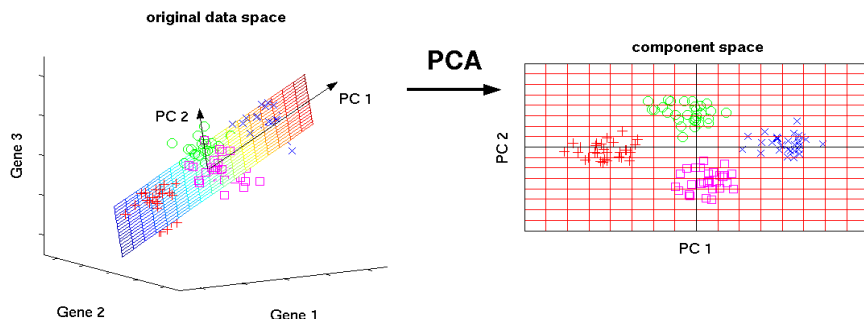


Figure 3. A brief graph to explain PCA (Figure adapted from Ghosh, 2019).

NIR spectroscopic techniques coupled with PCA show a good ability to detect a large apple variability due to varieties (Cortés, Cubero, Blasco, Aleixos, & Talens, 2019; Eisenstecken, Stürz, Robatscher, Lozano, Zanella, & Oberhuber, 2019; He, Li, & Shao, 2005, 2007) and geographical origins (Eisenstecken et al., 2019; Schmutzler & Huck).

Discriminant Analysis (DA)

Discriminant analysis (DA) separates samples into classes, minimizing the variance within the class and maximizing the variance between classes, and finding the linear combination of the original variables (directions). It is a supervised method requiring the knowledge of group memberships for each sample. Thus, it is usually applied to the same sample types as PCA, where the latter technique can be used to reduce the number of variables in the data set and the resultant principal components (PCs) are then used in DA to define and predict classes (Mendlein, Szkudlarek, & Goodpaster, 2013).

DA is a form of supervised pattern recognition, such as linear discriminant analysis (LDA) or factorial discriminant analysis (FDA), well known as chemometric approaches for solving classification problems in chemistry. Different from PCA establishing the directions of maximal variance, LDA and FDA aim to separate the known classes by creating a new linear axis and projecting data on that axis (Figure 4). All the variance between classes is

calculated and defined as the distance between the mean of different classes to maximize the distance between classes. Besides, the variances of each class are also calculated and is defined as the distance between the mean and the sample of every class. Finally, the Fisher's criterion tries to find the lower-dimensional space that maximizes the variance between classes and minimizes the variance within classes. FDA has been applied on the NIR spectra of apples to successfully discriminate the different varieties and post-harvest storage conditions (Camps, Guillermin, Mauget, & Bertrand, 2007).

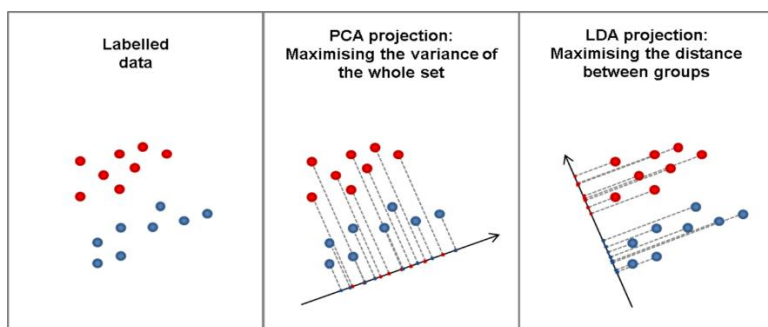


Figure adapted from Pelliccia, 2018.

Figure 4. The differences between PCA and LDA methods.

Regression Modelling

Partial Least Square Regression (PLS)

PLS is one of the most widely used multivariate linear regression methods to quantify the quality of F&V. In short, PLS regression combines principal component analysis and canonical correlation analysis. PLS models maximize the covariance between Y-matrix (references datasets of F&V) and X-matrix (spectral dataset of F&V) in a way that it has better prediction for Y-matrix by maximizing the variance of X-matrix.

The latent variables (LVs) of PLS models are directions in the space of the predictors. Particularly, the maximum covariance with the selected response variable is calculated as the first latent variable, subtracting from both the original predictors and the response. The second latent variable is orthogonal to the first one, being the direction of maximum covariance between the residuals of the predictors and the residuals of the response. This approach continues for the subsequent LVs.

The optimal PLS model is chosen according to the most appropriate number of latent variables, which are determined by a proper validation strategy, determination coefficients and prediction errors. The prediction ability of PLS models is usually presented by several parameters have been described (Abdi, 2003; Nicolai et al., 2007), such as:

- Determination coefficient of calibration (R_c^2) and/or validation (R_v^2) models, which determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, the coefficient of determination tells one how well the data fits the model.
- Root mean square error of calibration ($RMSE_c$) and/or validation ($RMSE_v$) sets, presents the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you the differences between predicted values by developed model and the observed values.
- Residual Predictive Deviation (RPD), which is defined as the standard deviation of observed values divided by the Root Mean Square Error (RMSE). It indicates the precision behavior of the prediction in comparison with the average composition of all the samples.

Machine Learning Regression

Machine learning (ML) methodologies aim at learning from training data to perform a task, including classification, regression, clustering, and dimensionality reduction models (Liakos, Busato, Moshou, Pearson, & Bochtis, 2018). Usually, each feature of an individual example can be nominal (enumeration), binary (i.e., 0 or 1), ordinal (e.g., A+ or B-), or numeric (integer, real number, etc.). The ML regression models may be linear (Ridge regression, Lasso regression etc.) and non-linear (random forest, support vector machine, decision tree, K Nearest Neighbors, Cubist etc.).

- RF is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time (Ho, 1995). For classification tasks, the output of the RF is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. This method

frequently regards as "black-box" models, as it generates reasonable predictions across a wide range of data while requiring little configuration.

- Support vector machine (SVM) has been introduced for predicting numerical property values. SVM regression models can resolve nonlinear relationships in original feature spaces through dimensionality extension (Noble, 2006). However, it has black box characters, meaning that the predictions cannot be directly interpreted in chemical terms. Hence, it is generally difficult to rationalize model performance.

More specific explanations of different ML regression methods have been described in several outstanding publications (Jordan & Mitchell, 2015; Mitchell, 1997; Shalev-Shwartz & Ben-David, 2014).

Applications

Meat, Fish, and Fishery Products

Different types of meat, species of fish, and their products contain organic components containing hydrogen groups such as polysaccharides, proteins, and organic acids, and their NIR spectra contain multiple and combined frequency peaks of many groups, which makes NIR expected to be used as an effective means to detect (qualitative and quantitative) sensory indicators, composition, origin, and variety identification of meat, fish, and fishery products (Table 1).

Sensory indicators, including freshness, tenderness, and color, are important attributes influencing consumer opinion about the eating quality of meat, fish, and fishery products. Various studies have been conducted to reliably detect the freshness of meat and fish products based on NIR spectroscopy combined with chemometric methods, such as intact pork (Wang, Peng, Sun, & Long, 2017) and Rohu fish (*Labeo rohita*) (Dei, Naskar, Dasgupta, Nag, & Bandyopadhyay, 2020). In tenderness detection, NIR technology has been applied to porcine *Longissimus dorsi* muscles (Barbin, Valous, & Sun, 2013) and lamb (Knight et al., 2019) analysis. In addition, NIR spectroscopy-based detection and prediction of some indexes closely related to freshness and tenderness of meats has also been conducted, including pH

(Knight et al., 2019; Wang, Tang, Peng, Fan, Dong, & Li, 2016), color, and drip loss (Jia, 2017).

Table 1. Parameter evaluation of meat, fish, and fishery products based on visible (VIS) and NIR spectroscopy during the past 10 years

Types of products	Spectroscopy technique	Parameters	References
Intact pork	NIR	Freshness	(Wang et al., 2017)
Fish (Rohu, <i>Labeo rohita</i>)	NIR	Freshness	(Dei et al., 2020)
Porcine <i>Longissimus dorsi</i> muscles	NIR hyperspectral imaging	Tenderness	(Barbin et al., 2013)
Lamb	Visible (VIS)-NIR	Tenderness	(Knight et al., 2019)
Lamb	VIS-NIR	pH	(Knight et al., 2019)
Tan-sheep	NIRS hyperspectral imaging	pH	(Wang et al., 2014)
Chicken	NIR	pH	(Wang et al., 2016)
Chicken breast fillets	VIS-NIR	Color (L*, a*, b*), pH, and drip loss	(Jia, 2017)
Meatballs (beef, pork, chicken)	NIR	Starch content	(Vichasilp & Kawano, 2015)
Lamb	VIS-NIR	Intramuscular fat content	(Knight et al., 2019)
Lamb	Hand-held NIR spectroscopic device	Intramuscular fat content	(Fowler, Morris, & Hopkins, 2020)
Minced pork	NIR	Moisture, fat and protein content	(Mishra, Verkleij, & Klont, 2021)
Atlantic salmon (<i>Salmo salar</i>)	VIS-NIR reflectance spectroscopy	Fat content	(Brown, Kube, Taylor, & Elliott, 2012)
<i>Salmo salar</i> (minced)	VIS-NIR reflectance spectroscopy	Fatty acid constituents	(Brown et al., 2012)
Minced beef (fresh, frozen-thawed, and cooked)	FT-NIR	Adulteration	(Alamprese, Amigo, Casiraghi, & Engelsens, 2016)
Minced beef	NIR	Adulteration	(López-Maestresalas et al., 2018)
Minced lamb	NIR	Adulteration	(López-Maestresalas et al., 2018)
Zebrafish	NIR Fluorescent Probe	Hydrogen sulfide	(Yang et al., 2022)

It is important to assess and monitor the chemical composition and quality of various meat and fish products. Vichasilp et al. (2015) created a universal model by combining the NIR spectra of beef, pork, and chicken meatballs for predicting starch contents in these meatballs. The universal model had a coefficient of multiple determination of calibration set of 0.98, standard error