Doctoral Dissertation
Doctoral Program in Chemical Engineering (37th Cycle)

Chemometrics applied to food industry: improving quality, safety and sustainability using multivariate data analysis

Mattia Sozzi
* * * * * *

Supervisors
Prof. Francesco Savorani, Supervisor
Prof. Francesco Geobaldo, Co-Supervisor
Dr. Nicola Cavallini, Co-Supervisor

Politecnico di Torino
March 28th, 2025

Summary

This PhD project investigates the role of chemometrics in the food industry and its applications in chemical engineering, with a particular focus on improving food quality, safety, and sustainability using multivariate data analysis. The study integrates analytical techniques, like Nuclear Magnetic Resonance (NMR) spectroscopy and Field Emission Scanning Electron Microscopy (FESEM), with chemometric methods to address specific challenges in the food sector.

The thesis can be divided in two main blocks: the first one corresponds to Chapter 2 and provides an overview of the materials and methods employed, including a discussion of metabolomics and foodomics, as well as the analytical techniques used, and a comprehensive review of the chemometric methods applied. The second block is composed by Chapter 3-6, that describe four different case studies explored within the PhD project, each addressing a distinct aspect of chemometric applications in food industry.

Chapter 3 presents the first case study, which focuses on the optimization of industrial food processing treatments. The study aims to explore and optimize sustainable technologies for food production. Experimental design principles were applied to evaluate process parameters, while Principal Component Analysis (PCA) was employed to explore the data and identify optimal conditions, demonstrating improvements in resource efficiency (e.g., reductions in time and energy consumption). In brief, PCA was applied to lentil flour samples treated with different enzymes, stirring rates and durations. The evaluation of clusters and time trends found in the PCA scores plot revealed that conditions of 120 minutes at 60 RPM were more advantageous than the initially hypothesized optimal conditions, leading to an optimization of this particular industrial food treatment.

Chapter 4 explores the application of chemometric techniques in food traceability and authentication, specifically in the context of hazelnuts. The study combines multiple analytical techniques, including ^{1}H-NMR and Liquid Chromatography coupled with Mass Spectrometry (LC-MS), with multivariate data analysis to determine the geographical origin and cultivar of hazelnuts. A dataset comprising 54 hazelnut samples from diverse origins and cultivars, collected over a two-year period, was analysed. The datasets coming from the two techniques were firstly explored independently, then, a data fusion approach was employed to integrate NMR and LC-MS data. This new fused dataset was firstly explored using PCA and then different classification models were developed using Partial Least Squares Discriminant Analysis (PLS-DA) to differentiate among origins and cultivars. In addition, the identification of metabolites, both from ^{1}H-NMR and LC-MS data, further reinforced the authentication findings, hypothesizing new potential biomarkers for the Piedmont TGT hazelnuts (an Italian product known worldwide for its excellence). This case study underscores the potential of data fusion approaches applied to spectroscopic techniques in combination with chemometric methods with the aim of verifying the authenticity and the origin of food products.

Chapter 5 explore analytical challenges associated with the application of ^{1}H-NMR for wine analysis, with a particular focus on the ethanol quantification in alcoholic beverages, where high ethanol concentrations complicate accurate quantification by NMR. This study systematically examined sample preparation procedures, including the choice of deuterated solvent, experimental parameters, like NMR pulse sequences, and different quantification strategies (e.g., Multivariate Curve Resolution (MCR), quantification with internal or external standard, application of regression methods like Partial Least Squares (PLS) regression) to enhance the accuracy and precision of ethanol quantification. The application of water suppression pulse sequences and D_2O as a solvent was confirmed as a valuable approach. However, challenges were identified with the use of TSP as an internal standard across varying ethanol concentrations, leading to the exploration of external standard methods to improve ethanol quantification. The findings highlight the necessity of rigorous methodological optimization and clear standard operating procedures for reliable quantitative NMR analysis in complex food matrices.

Chapter 6 investigates the application of image analysis in the agrifood sector, with a focus on rice. This study examines the Glycaemic Index (GI) and the internal structure of 25 Italian rice genotypes and 18 non-Italian varieties. An algorithm was *ad hoc* developed in MATLAB to automatically perform image analysis to extract morphological features of starch granules and endosperm porosity starting from images acquired using Field Emission Scanning Electron Microscopy (FESEM). GI values and other biochemical properties were determined in collaboration with University of Pavia. This analysis revealed substantial variability in GI among the examined genotypes, with five varieties exhibiting particularly low GI values. PCA conducted on starch granule features indicated a correlation between morphological features and GI values, suggesting that rice varieties with larger starch granules and lower porosity tend to have a lower GI. The findings contribute to a deeper understanding of the relationship between rice grain structure and its nutritional properties, potentially facilitating the development of rice varieties tailored to specific dietary needs. Additionally, a Design of Experiment (DoE) approach was used to evaluate and optimize the developed image analysis algorithm. By systematically testing different input parameters, the effect of those inputs on the algorithm result and on the processing time was evaluated and optimized using chemometric methods such as ANOVA Simultaneous Component Analysis (ASCA) and Multi Linear Regression (MLR).

Finally, the general conclusions summarize the key findings across all case studies and emphasize the wide range of applicability of chemometrics in addressing diverse challenges within the food industry. These include process optimization, food authentication, and quality assessment based on compositional and structural analysis. The PhD project highlights the power of integrating different analytical techniques with multivariate data analysis to gain deeper insights into food and chemical systems, contributing to more sustainable, safer, and higher-quality food production. More detailed challenges and future research directions for each case study are also discussed within their respective chapters.