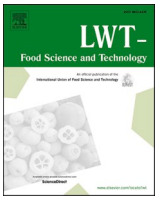
LWT - Food Science and Technology 212 (2024) 117000

Contents lists available at ScienceDirect

LWT

journal homepage: www.elsevier.com/locate/lwt



Benchtop NMR spectroscopy for quantitative determination of milk fat and qualitative determination of lactose: From calibration curve to deep learning

José Raúl Belmonte-Sánchez [a], Roberto Romero-González [a], Manuel Ángel Martínez Orosa [b], María Calvo Morata [b], Antonia Garrido Frenich [a,*]

[a] Department of Chemistry and Physics, Analytical Chemistry Area, Agrifood Campus of International Excellence ceiA3, University of Almería Research Centre for Agricultural Food Biotechnology (CIAIMBITAL), E-04120, Almería, Spain
[b] Tentamus Laboratorio de Control S.L.U., Edificio Las Rozas, 23, Ctra. de la Coruña, Km 23.200., Las Rozas, E- 28290, Madrid, Spain

ARTICLE INFO

Keywords:
Milk fat quantification
Benchtop NMR spectroscopy
Machine learning
Artificial neural networks
Partial least squares regression

ABSTRACT

This study compares three different methodologies for the quantification of the fat content of ultra-high temperature (UHT) milk using benchtop proton nuclear magnetic resonance (^{1}H NMR) spectroscopy, a flagship of green, accessible, and state-of-the-art technology suitable for modern laboratory environments. The evaluated approaches included traditional calibration curve and machine learning algorithms, with emphasis on partial least squares regression (PLS-R) and artificial neural networks (ANN), to estimate the fat content in skimmed, semi-skimmed and whole milk. Among these, ANN provided the most accurate results for all types of milk, particularly in skimmed milk, with a relative standard deviation (RSD) of 14.9% and an accuracy of −7.3%. The calibration curve showed higher variability, with an RSD of 34.1% and trueness of 25.3% for skimmed milk. PLS-R improved accuracy in relation to the calibration curve approach, reducing RSD to 18.9% and trueness to −17.7%. The developed method has been successfully applied to determine the fat content in 51 samples of UHT milk purchased in different Spanish supermarkets, providing adequate results for each of the three categories considered, including goat's milk, sheep's milk, and milk coffee. Furthermore, the application of machine learning has proven its validity by successfully distinguishing between lactose and lactose-free UHT milk.

1. Introduction

In the modern dairy industry, the assessment of fat content in food matrices, such as ultra-high-temperature (UHT) cow's milk, is of paramount importance. UHT milk, characterized by its sterilization process that involves heating to temperatures above 135 °C for a short period of time (typically 2–5 s), is specifically designed to eliminate microbial life, including spores. This process significantly extends the shelf life of milk, allowing it to be stored without refrigeration until it is opened (Tamime, 2008). The widespread popularity of UHT milk has led to its increased availability and consumption around the world, with a notable presence in countries such as Spain, where 95% of the milk consumed is UHT (Aquini & Gil, 2017). These factors highlight the importance of UHT milk in ensuring food safety, improving accessibility, and meeting consumer preferences.

Based on the importance of UHT milk, Regulation (EU) No. 1308/2013 sets strict EU standards for milk fat labelling. Therefore, whole milk must contain at least 3.25 g/100 mL milk fat, semi-skimmed between 1.5 g/100 mL and 1.8 g/100 mL, and skimmed milk less than 0.5 g/100 mL (Amaral et al., 2018). These rules are designed to maintain transparency, support consumer choice, and ensure fair competition.

For the analysis of the fat content in milk and dairy products, various analytical techniques are employed. The most widely adopted method involves gas chromatography coupled with flame ionization detection (GC-FID). This technique is particularly valued for its ability to separate and analyze the composition of fatty acids (FAs), although it requires extensive sample preparation (Danudol & Judprasong, 2022). Another variant of this determination involves the coupling of GC with mass spectrometry (GC-MS), which improves the specificity and sensitivity of the analysis, but it comes with higher complexity and cost (Chen et al., 2023). In addition to these chromatographic techniques, near-infrared spectroscopy (NIRS) (Evangelista et al., 2021) and Raman

* Corresponding author.
E-mail address: agarrido@ual.es (A. Garrido Frenich).

https://doi.org/10.1016/j.lwt.2024.117000
Received 4 June 2024; Received in revised form 25 October 2024; Accepted 5 November 2024
Available online 6 November 2024

spectroscopy (Reiner et al., 2020) are well established as non-destructive methods for the analysis of FA in milk and dairy products, although they can compromise on specificity and sensitivity due to fluorescence interferences. Each of these methods has its own set of applications, advantages, and limitations, which makes the choice of method dependent on the specific requirements of the analysis.

The introduction of nuclear magnetic resonance (NMR) spectroscopy, specifically proton NMR (^{1}H NMR), as an alternative to analyze the fat content of milk and dairy products, brings significant benefits. Previous studies have successfully applied high-resolution NMR techniques for quantifying milk fat using various strategies, such as two-dimensional NMR (Hu et al., 2007), time-domain (P. M. Santos et al., 2016) or regression models (Monakhova et al., 2012), proving the technique's efficacy. However, benchtop NMR technology is a recent advancement and its application in this context is currently under investigation.

Benchtop NMR instruments, with their reduced size, make this sophisticated technique more accessible to smaller laboratories and production facilities, facilitating more frequent and extensive testing (Draper & McCarney, 2023; Galvan et al., 2021; van Beek, 2021). These cost-effective versions of traditional high-resolution NMR systems offer a viable option for routine analyses. NMR spectroscopy enables the direct and non-destructive quantification of various components in complex mixtures like dairy products, eliminating the need for extensive sample preparation or chemical reagents (Burger et al., 2022). The use of benchtop NMR also aligns with the industry's move toward more rapid and efficient testing methodologies that can provide real-time data for quality control and assurance processes (Ezeanaka et al., 2019).

So far, some studies have been published on cow UHT milk using benchtop NMR. *Soyler* et al. differentiated milk samples based on properties such as glycerol, fat, and sugar content using NMR spectra combined with an artificial neural network (ANN) model (Soyler et al., 2021). In another study, a method focused on T_2 relaxation times (acquisition of the time domain) for determining the fat content (between 0.1 g/100 mL and 9 g/100 mL) in milk. This approach required a contrast solution consisting of a 6 g/100 mL NaCl solution containing 2 g/100 mL sodium ferric ethylenediaminetetraacetate (Fe-EDTA) to adapt the relaxation rate of water (Sørensen et al., 2022). It is important to clarify that the distinction between studies does not imply superiority to one over the other; rather, it highlights divergent methodological choices based on different research objectives and applications.

Quantitative NMR, including both high and low resolution measurements, working with different regions of the spectra, has been applied to simultaneously measure different components in various dairy products (Hatzakis, 2019). To achieve this, one of the most commonly used methodologies involves the use of calibration curves, where known standards are plotted against their responses to determine concentrations, guided by current validation guidelines (Bharti & Roy, 2012). Machine learning, especially through regression models, offers a more specific approach to complex data that may not fit simple univariate linear relationships. Within machine learning, ANNs stand out by emulating human neural networks, allowing for pattern recognition within large datasets that might elude conventional statistical methods (da Costa et al., 2021; Joshi, 2023; M. C. Santos et al., 2019). Each of these quantification methods has its own set of advantages and applications, often chosen based on the specific requirements of the analysis, the complexity of the sample matrix, and the available instrumentation and computational resources.

In the case of regression models applied within machine learning contexts, particularly in chemometric studies using NMR, partial least squares regression (PLS-R) stands out as one of the most commonly utilized and robust methods, especially valued for its ability to handle highly collinear and multidimensional data typical of NMR studies. The popularity of PLS-R stems from its dual ability to perform dimensionality reduction and regression simultaneously (Dhaulaniya et al., 2020; Galvan et al., 2021; Huang et al., 2022). In addition to PLS-R, other models such as support vector machines (SVM) (Hou et al., 2019) or random forests (RF) (Zhao et al., 2019) have also been used. However, the choice of model is just the beginning, and adherence to specialized validation guidelines, such as those from EUROLAB (Schönberger et al., 2015), is essential for ensuring the model's reliability and generalizability.

For ANN, the model design process is even more sophisticated, given the complexity and "black box" nature of these models. In addition to generalization, optimization of ANNs involves optimizing the network architecture, including the number of layers and nodes, and tuning hyperparameters, such as the learning rate and regularization terms (Debik et al., 2022). To the best of our knowledge, there are no guidelines that address the optimization or validation of ANN based quantification methods. However, it is possible to cover this issue from the point of view of validation of chemometric methods (Benzaama et al., 2022; Taylor, 2006). Cross-validation techniques, such as k-fold cross-validation, are commonly employed to ensure that the model's performance is robust across different subsets of the data (Corsaro et al., 2022; Pomyen et al., 2020).

Attempting to address the differences between measurement methods and offering a rigorous approach applicable to contemporary analytical problems, this study compares three methods to quantify the content of UHT milk fat using desktop NMR: traditional calibration curves, and machine learning from PLS-R to ANN (deep learning). In addition to analyzing their benefits, limitations, and validation to identify the most accurate, efficient, and reliable approach to improve quality control in the dairy industry, an analysis of environmental sustainability was also carried out. This demonstrates the synergy between these types of studies and their alignment with green analytical chemistry principles, showcasing our commitment to promoting sustainable practices within analytical research.

2. Materials, reagents, and methods

2.1. Milk samples and sample preparation

A total of 55 milk and dairy product samples were collected from different Spanish supermarkets. **Table S-1** shows the samples randomly ordered as analyzed. The samples included different types of fat content: skimmed (14), semi-skimmed (23), and whole (18), with fat content ranging from less than 0.1 g/100 mL to 3.6 g/100 mL, except for condensed milk with g/100 mL fat, and a cream sample with 33 g/100 mL fat. Most of the samples were of cow origin (51), with a minor representation of goat and sheep milk (two from each), ensuring a comprehensive analysis of various types of milk. The sample set also included dairy derivatives such as coffee with milk (2), kefir (1), and yoghurt (1), providing a broad examination of milk and dairy products.

Additionally, for the calibration curve method, a set of UHT cow milk reference standard from QSE-GmbH (Wolnzach, Germany) was purchased, covering the range of fat content within the samples (0.06 g/100 mL, 1.52 g/100 mL, 3.48 g/100 mL and 4.33 g/100 mL).

Before NMR measurements, 320 μL of a 1 g/100 mL solution of 3-(trimethylsilyl)propionic-2,2,3,3-d4 acid sodium salt (TMSP) in deuterium oxide (D_2O), both obtained from Sigma-Aldrich (St Louis, MO, USA), serving as the internal standard and locking solvent, respectively, were added to 680 μL of the sample. The mixture was vortexed at room temperature for 5 min. The resulting solution was then transferred directly into the NMR tube, which was sealed and pre-warmed at 32 °C in an NMR tube heater for 1 min before analysis.

2.2. Instrumentation and software

NMR experiments were carried out on a 100 MHz benchtop NMR system (Nanalysis Corp., Alberta, Canada) equipped with a 24-position autosampler and sample heater. Spectroscopic analysis was performed using the presaturation technique (Presat) to suppress the water signal at an operating frequency of 102.5 MHz and a spectral width of 1500.20 Hz

(15.0 ppm). A pulse angle of 30° was used for excitation of the sample. To ensure optimal water suppression, a total pre-saturation time of 2 s was applied prior to the acquisition of the NMR signal. The bandwidth for the presaturation was set at 150 Hz to cover the water resonance without affecting the other signals of interest in the spectrum.

A total of eight scans were applied per analysis, except for samples labelled as skimmed, where 16 scans were used. While increasing the number of scans affects the signal intensity, this does not impact our calculations, as they are relative to the internal standard used; therefore, normalization was applied later to adjust the results. An interscan delay time of 30 s was used to allow complete relaxation of the nuclei, ensuring that each scan was acquired under fully relaxed conditions. Each milk acquisition lasted approximately 4 min under the conditions of analysis described above (8 min, approximately, for skimmed samples). Shimming was performed every three samples to ensure a uniform magnetic field throughout the series of measurements. All samples were measured at 32.0 ± 0.1 °C and under non spin conditions.

Each spectrum was manually calibrated against the TMSP standard at δ 0.0 ppm as a reference point. Prior to Fourier transformation, the ^{1}H NMR spectra received a line broadening application of 0.3 Hz to smooth the data. The spectra were automatically phased, baseline-corrected, normalized and integrated using MestReNova 10.0.2 software (Mestrelab Research SL, Santiago de Compostela, Spain).

Multivariate optimization was performed using MODDE™ Pro software version 13.0.2 (Sartorius AG, Göttingen, Germany), which facilitated a structured exploration of the experimental conditions (Nguyen et al., 2024; Sagmeister et al., 2020). Chemometric analyses (PLS-R) were performed with SIMCA 17 software (Umetrics, Umea, Sweden), while the ANN model was developed using Python (v3.11) and the Keras library (v2.2.4) with a TensorFlow backend (v1.13.1, CUDA 10.1).

2.3. Calibration curve method

The proposed ^{1}H NMR method to determine fat content was validated in terms of linearity (R^2), selectivity, precision, limit of quantification (LOQ), accuracy (recovery), and working range (0.06 g/100 mL, 1.52 g/100 mL, 3.48 g/100 mL and 4.33 g/100 mL) (Bachmann, 2023). Integration of the signals corresponding to fat was performed within the spectral range of 0.5–2.5 ppm.

Accuracy and precision were evaluated with fortified samples (from available samples at a concentration of 0.1 g/100 mL) at three concentration levels, skimmed (0.06 g/100 mL), semi-skimmed (1.5 g/100 mL) and whole (3.6 g/100 mL), with each of the three replicates. The LOQ was established at the lowest point of the working range. Importantly, determining LOQ required achieving a signal-to-noise ratio greater than 10, ensuring reliable quantification (Belmonte-Sánchez et al., 2019; FDA, 2019).

2.4. Partial least squares regression

For the investigation of UHT milk, an initial unsupervised statistical approach was applied to uncover patterns within the data set. Principal Component Analysis (PCA) was employed for exploratory purposes, allowing for the visualization of relationships between the milk samples. Hotelling's T^2 analysis and the Distance to Model in X-space (DModX) plot were employed to detect and remove outliers, ensuring the robustness of the statistical analysis.

Following this exploratory phase, making use of supervised techniques, a PLS-R model was developed and evaluated using an arbitrary data partitioning strategy, in which the data set was split into a training set comprising 80% of the data and a validation set comprising the remaining 20%. Within the training set, the model was tuned using a seven-fold cross-validation, ensuring that the model was iteratively optimized. During this process, the optimal number of components in the latent dimensions (NSCs) was determined by monitoring the R^2Y and Q^2Y values as the number of NSCs increased. The number of NSCs was considered optimal when these values stabilized, meaning that adding more components no longer significantly improved the model's ability to explain or predict the data.

The goodness of fit of the model was assessed by the determination coefficient (R^2X) for the training set and the explained variance (R^2Y), reflecting the model's ability to capture the variance in the data. Additionally, its predictive performance in the validation set was evaluated using the predictive ability parameter (Q^2Y).

Analysis of variance coupled with cross-validation (ANOVA-CV) provided a detailed understanding of the model's variance components, enhancing insight into its predictive power and statistical significance.

The root mean square error of cross-validation (RMSE-CV) was calculated to gauge the model's predictive accuracy by measuring the average deviation of predictions from actual values during cross-validation, offering a precise measure of performance on unseen data.

To further validate the precision and accuracy of the results, the developed model was tested against fortified samples (testing set) in water at three concentration levels: skimmed (0.06 g/100 mL), semi-skimmed (1.5 g/100 mL), and whole (3.6 g/100 mL), with each concentration tested in three replicates.

2.5. Artificial neural network (deep learning)

For the quantitative analysis of UHT processed cow's milk, a feed-forward ANN was utilized to address the nonlinear complexities of the data. The model was operated on an NVIDIA Titan RTX GPU for improved computational efficiency. The architecture featured an input layer of 21 neurons and a hidden layer of 32 neurons with Rectified Linear Unit (ReLU) activation, including dropout layers at a rate of 0.2 to mitigate overfitting (Fig. S-1). Compiled with the RMSprop optimizer and the mean squared error loss function, ANN supported regression analysis through a single-neuron output layer with linear activation.

Training involved 30 epochs with a batch size of 8, and model validation on unseen data assessed generalization capabilities. RMSE-CV and mean absolute error (MAE) measured predictive accuracy, while learning curve analysis identified optimal complexity and training duration to balance pattern recognition and avoid overfitting, ensuring robust evaluation and accurate predictions of the UHT milk fat content.

As an additional step to ensure the reliability of the analysis, the performance of the ANN model was assessed by applying them to the test set of fortified samples (see section 2.3) on a gradient of fat concentrations: skimmed (0.06 g/100 mL), semi-skimmed (1.5 g/100 mL), and whole (3.6 g/100 mL), performing this test in three separate measurements for each concentration level.

3. Results and discussion

3.1. Optimization of NMR acquisition conditions

Considering the numerous variables involved in this technique, such as relaxation times, signal suppression considerations, bandwidth for suppression, the ratio of sample to solvent or the number of scans, among others, a design of experiments approach has been employed for optimization (Peris-Díaz & Krężel, 2021). A central composite design (CCD) was used in the response surface methodology (RSM) framework to optimize the parameters of desktop NMR spectroscopy, focusing on minimizing the spectral distortion of the specific spectral region of the predominant water signal at δ 4.7 ppm.

The optimization process considered four key parameters: total presaturation time, bandwidth, pulse angle, and sample-solvent ratio, each within specified ranges. The total presaturation time was varied from 1 to 4 s, bandwidth from 30 to 200 Hz, pulse angle from 30 to 90°, and sample-solvent ratio from a minimum of 20:80 *v/v* to a maximum of 80:20 *v/v*. The primary objective was to identify conditions under which the integrated area of a designated spectral region would be as close to zero as possible, indicating minimal spectral distortion. This approach

emphasized that the significant variable was the integrated area quantitatively measured, rather than a characteristic visually assessed.

Given the number of variables and the aim of a minimum number of degrees of freedom of 5, with a restriction on the maximum number of runs set at 15, the MODDE™ software recommended a central composite face (CCF) design as the most suitable for this study. **Table S-2** details all experiments carried out within the framework of this study.

The optimal conditions identified through this systematic approach were a sample-solvent ratio of 32:68 *v/v*, a pulse angle of 30°, a total presaturation time of 2 s, and a bandwidth of 150 Hz. Fig. S-2 shows the remarkable differences obtained between different acquisition conditions, where 9 spectra have been represented to avoid saturating the image with spectral overlap. These settings were found to significantly reduce the spectral distortion, thereby enhancing the quality and reliability of the NMR spectra obtained. Remarkably, under these optimized conditions, no distortions were observed throughout the spectrum, with particular emphasis on the critical region of fats in milk, which ranged between δ 2.8 and δ 0.5 ppm (Soyler et al., 2021), and they remained undistorted.

3.2. Fat content determination

First, for the baseline correction, the Whittaker algorithm was utilized, effectively normalizing the baseline of the spectral signals to zero. This approach ensured that any negative distortions encountered during spectrum alignment did not detract from the analysis, allowing exclusive consideration of positive value peaks (Cobas, 2018).

Subsequently, selectivity was evaluated to identify potential interferences or impurities that arise from variations in the types of milk (cow, goat, or sheep) or between different brands. This involved comprehensive spectral profiling of all final spectra, using certified UHT cow milk reference standards for comparison. The investigation covered the entire chemical shift range relevant to the fat content of UHT milk, ensuring that no signal overlap or confuse with either the signals of interest or the reference standards. This process confirmed the high selectivity of the method across the range of selected whole milk products. Furthermore, the analysis was extended to dairy derivatives, such as coffee with milk, condensed milk, cream, kefir, and yoghurt. Due to their unique spectral profiles, condensed milk, cream, yoghurt, and kefir were excluded from further modelling to preserve the focus and integrity of the analysis, thus limiting the application of the present study to the rest of the 51 remaining samples (Fig. 1).

Particularly for skimmed milk, where signal resolution was crucial, the use of additional scans significantly improved signal clarity. For example, by doubling the number of scans from 8 to 16, a marked improvement in the quality of the spectral data was observed. This increase, although extending the analysis duration by approximately 4 min per sample, proved to be a valuable adjustment. This strategic increase in scans represented a practical approach to optimizing the analytical process, ensuring that even samples with a low-fat content can be accurately quantified.

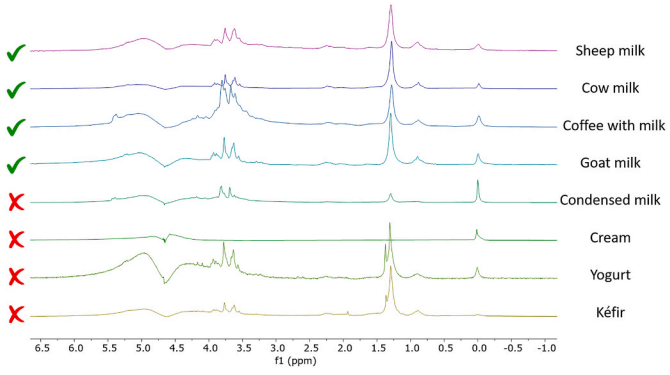


Fig. 1. Representation of spectral profiles obtained from different dairy derivatives by ^{1}H NMR, including whole sheep's milk, whole cow's milk, coffee with whole milk, whole goat's milk, condensed milk, kefir and yoghurt, showing differentiation of UHT milk and justification for its exclusion ('x' symbol) from further analysis.

3.2.1. Calibration curve

The working range of the calibration curve was established through triplicate measurements at four concentrations (0.06 g/100 mL, 1.52 g/100 mL, 3.48 g/100 mL and 4.33 g/100 mL) (Fig. S-3), ensuring it encompassed the typical levels of fat content found in these milk categories through the integration of signals corresponding to fat within the spectral range of 0.5–2.5 ppm (Soyler et al., 2021).

The linearity of the calibration curve was critically assessed using R^2, which revealed excellent linearity (R^2 = 0.993) within the defined range. For samples exceeding the upper limit of the working range, simple water dilution techniques could be employed, allowing for accurate analysis without compromising the method's integrity.

The precision and trueness results are shown in Table 1. For semi-skimmed and whole milk, precision and accuracy metrics (16.2% and 9.0% for semi-skimmed, 7.9% and 3.7% for whole milk, respectively) demonstrated good alignment with the stringent criteria set (Bachmann, 2023; FDA, 2019), supporting the method's ability to deliver consistent and reliable results for these milk fat concentrations. However, the values obtained for skimmed milk, with a precision of 34.1% and an accuracy of 25.3%, were significantly higher.

Furthermore, in anticipation of creating a prediction ensemble for multivariate models with samples, an analysis of these same samples was performed using the calibration curve for comparison purposes with samples of the models. Table 2 shows the results of the prediction set and the variance compared to the labelled values. It includes a mean percentage relative error (Diff. (%)), which was calculated as (observed value−true value)/true value × 100, of 5.2% for the calibration curve method.

3.2.2. Partial least squares regression

Before PLS-R analysis, the spectra acquired from the UHT milk samples were subjected to a bucketing process with a resolution of 0.05 ppm, resulting in a data set comprising 43 distinct variables. This specific bucket size was chosen on the basis of a thorough examination of the resulting spectral data to ensure that crucial information, such as signal multiplicity, was preserved. Furthermore, retrospective analyses conducted on a trial basis revealed that this bucketing resolution consistently yielded the most reliable and informative results.

In a preliminary phase, a PCA was carried out on milk samples to discern possible variations between the groups. Logarithmic transformation was used for all data sets, obtaining values in a smaller range without masking the effect of small values in the data set and allowing an optimal variable range for the proposed model (Lubes & Goodarzi, 2017). From 153 spectra, corresponding to three replicates each of 51 samples, 51 averaged observations were obtained for further analyses. Using Hotelling's T^2 and DModX analyses, 7 outliers (samples 5-Milk, 12-Milk, 9-Milk, 24-Milk, 36-Milk, 50-Milk, and 55-Milk) were identified and excluded to refine the statistical models. These 7 outliers corresponded to 21 observations, resulting in a total of 132 samples after their removal.

The PCA revealed significant variability in the fat content of milk and the high fit of the model, with R^2 exceeding 98% and a Q^2 of 91% using only four NSCs. The score plot confirmed the reproducibility of the analyses, bolstering confidence in the findings (Fig. S-4). The PLS-R analysis identified significant components to modelling the fat content in milk, leading to the creation of a prediction curve by plotting predicted sample concentrations against their known values (Fig. 2).

ANOVA-CV underscored the model's efficacy in explaining the variance of fat content, with a total sum of squares of 43 and a regression sum of squares of 39.2, highlighting the model's explanatory strength.

Table 1
Comparative analysis of the results of fat content quantification in three replicates of UHT milk fortified samples using calibration curve, machine learning and ANN methods.[a]

Milk type	Calibration curve		PLS-R		ANN	
	Precision (%)	Trueness (%)	Precision (%)	Trueness (%)	Precision (%)	Trueness (%)
Skimmed	34.1	25.3	18.9	−17.7	14.9	−7.3
Semi-skimmed	16.2	9.0	12.4	12.6	10.2	−0.3
Whole	7.9	3.7	4.8	−2.1	6.9	−2.1

[a] Abbreviations: ANN: Artificial neural network; PLS-R: Partial least squares regression.

Table 2
Fat content in the prediction set of commercial UHT milk samples obtained by calibration curve, PLS-R, and ANN, including the percentage relative error between the observed and true values (Diff.).[a]

ID	Labelled fat content	Calibration curve	Diff. (%)	PLS-R	Diff. (%)	ANN	Diff. (%)
1-Milk	3.6	3.5	−3.2	3.6	0.7	3.6	−0.3
6-Milk	1.6	1.7	6.4	1.6	−3.4	1.6	0.2
15-Milk	0.3	0.4	17.3	0.3	−11.7	0.3	2.8
18-Milk	3.6	3.6	0.6	3.5	−1.9	3.6	−1.3
31-Milk	0.3	0.3	24.0	0.2	−3.2	0.3	−0.9
41-Milk	1.6	1.6	−0.3	1.5	−7.3	1.6	0.1
42-Milk	0.3	0.3	−5.3	0.3	12.6	0.3	0.2
44-Milk	3.6	3.5	−1.7	3.5	−1.9	3.5	−1.7
52-Milk	1.6	1.7	8.8	1.5	−3.7	1.6	−0.1
Mean Diff. (%)		–	5.2	–	−2.2	–	0.1

[a] Abbreviations: ANN: Artificial neural network; PLS-R: Partial least squares regression.

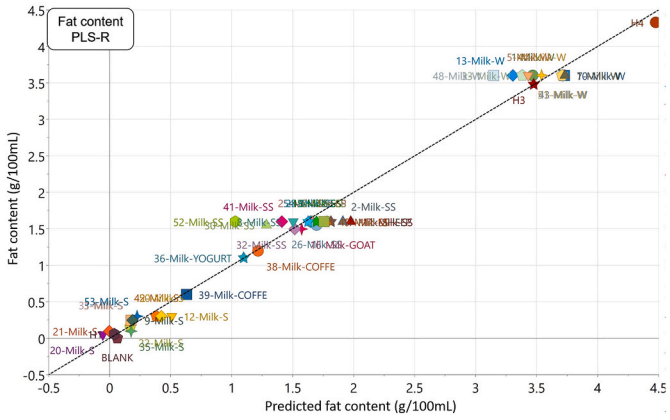


Fig. 2. Comparison of expected *vs.* actual fat content in UHT milk samples obtained by PLS-R.

An F-statistic of 216.6 and a significant p-value of 1.59e-22 emphasized the role of regression in explaining variance (**Table S-3**). The disparity in standard deviations between the regression and residual components further validated the precision of the model.

Collectively, these results underscore the robustness and precision of the PLS-R model in quantifying the dairy fat content, showcasing its significant analytical potential. With 44 averaged observations (35 for the training set and 9 for the validation set, as indicated in **Table S-1**), four NSCs (Fig. S-5), a R^2Y of 0.98, a Q^2 of 0.914, and a RMSE-CV of 0.39, the findings highlight the substantial capability of the model in the analysis of dairy products. Furthermore, the analysis of overfitting was reinforced by a permutation test, which yielded low values ($R^2 = 0.16$ and $Q^2 = -0.17$), further affirming the robustness of the model and its ability to generalize beyond the training data. The results obtained for the validation set showed a difference value with respect to the labelling value of −2.2%, improving the values previously obtained by quantification of the calibration curve (Table 2).

The application of machine learning PLS-R for the milk fat content in the test set at three concentration levels demonstrated precision and accuracy results that, while variable, represented a significant improvement over the traditional calibration curve method (Table 1). The method demonstrated a significant improvement in the quantification of fat content in relation to the calibration curve, with precision values of 18.9%, 12.4%, and 4.8% for skimmed (0.06 g/100 mL), semi-skimmed (1.5 g/100 mL), and whole (3.6 g/100 mL) milk, respectively. The accuracy across the tested concentrations was also notable, showing absolute values of 17.7%, 12.6%, and 2.1% for each type of milk, respectively.

3.2.3. *Artificial neural network*

A calibrated feedforward ANN was applied, using bucketing at 0.1 ppm after showing improved results compared to the previous bucketing at 0.05 ppm used in the PLS-R analysis, for NMR spectra analysis (21 variables) to quantify the fat content of UHT milk. The data set, consisting of 132 averaged samples after the extraction of outliers, was initially divided into a training set and a validation set with the remaining samples from the previous exploratory PCA analysis after extraction of the outlier (35 for the training set and 9 for the validation set). The process involved iterative parameter optimization, focusing on balancing complexity, computational efficiency, and performance. Using Python, a grid search method evaluated various configurations, identifying the optimal setup based on activation function, batch size, optimizer, and number of hidden layers (Jiang & Xu, 2022).

ReLU was selected for its efficiency in maintaining gradient flow and learning complex patterns without excessive computational demands (Sa'adah, 2023). The optimal batch size was determined to be 8, balancing computational load and learning stability. RMSprop was chosen as the optimizer for its adaptive learning rate, which improves the convergence speed. A single hidden layer proved sufficient for the model, adhering to the parsimony principle, while seven-fold cross-validation and strategic dropout layers combated overfitting, improving generalizability.

The chosen model configuration featured ReLU activation, an eight-unit batch size, an RMSprop optimizer, and one hidden layer, which combined theoretical and empirical efficacy for the quantification of fat content in UHT milk. Training ended within ten epochs, indicating rapid convergence of the model. Fig. S-6 showcases the model's pattern recognition capability and predictive accuracy, with visuals illustrating loss trends and a scatter plot of predicted versus actual fat content. The

RMSE-CV value of 0.25 offered insight into the expected error when the model was applied to novel data within this range. The results obtained for the validation set, together with their differences with respect to the labelling value, for each of the methods applied, are shown in Table 2. The table shows that the best overall results were obtained with the application of ANNs, with an average value of 0.1%.

Furthermore, the results obtained by applying the ANN approach to quantify fat content against the set of tests (fortified samples) at three concentration levels were particularly promising (Table 1). With precision values of 14.9% for skimmed, 10.2% for semi-skimmed, and 6.9% for whole milk, together with accuracy values of −7.3% for skimmed, −0.3% for semi-skimmed, and −2.1% for whole milk, this method demonstrated a significant improvement in both parameters compared to previous methodologies.

3.3. Qualitative determination of lactose content

Although the primary focus was the determination of the fat content in milk, spectral regions around δ 3.1 and δ 4.0 ppm, typically related to lactose (Soyler et al., 2021), played an unexpectedly important role in the determination of lactose. In particular, without specific optimization for these regions, a binary classification model that could differentiate between lactose and lactose-free milk was developed, with notable preliminary success, such as the PLS Discriminant Analysis (PLS-DA) model depicted in Fig. 3.

Despite the fact that the data set was not being specifically balanced for this purpose, with only 15 lactose-free samples out of a total of 50 (after removing outliers), the results clearly indicated a robust classification between lactose-containing and lactose-free milk. This was particularly evident in the distinction between lactose-free milk and both semi-skimmed and whole milk, as well as between lactose-free and skimmed milk. The model's validation process was carefully designed ensuring that all replicates for each sample were treated together in each fold during a seven-fold cross-validation, which minimized intra-sample variability and allowed for the detection of potential outliers within the replicates.

In addition to the classification accuracy, which reached 98.67% (**Table S-4**), the results were further validated with Fisher's probability test, yielding an of 2.7e-33, confirming the statistical significance of the classification. The validation results, including the permutation test to prevent overfitting, are summarized in **Table S-5**, further demonstrating the model's reliability.

The model performed robustly, and the accuracy of the classification underscores the strength of the approach, suggesting that the differentiation observed between lactose-containing and lactose-free milk is reliable and significant. This result opens the door to further investigations into the underlying spectral differences that drive this classification.

3.4. Environmental sustainability analysis

To further emphasize the commitment to sustainable practices, an environmental sustainability analysis was incorporated into the study, compared to other established methods. This analysis was performed using the Analytical Greenness (AGREE) calculator, a tool designed to evaluate environmental and occupational hazards associated with analytical procedures. AGREE evaluates the "greenness" of analytical methods based on the 12 principles of green analytical chemistry, providing a comprehensive overview of the environmental impact of the present research approach (Pena-Pereira et al., 2020).

The GC-MS and GC-FID methods were chosen for their comparison with the method proposed in this study in terms of analytical greenness (Chen et al., 2023; Danudol & Judprasong, 2022). The main characteristics of the three methods compared are shown in **Table S-6**. In addition, the pictograms obtained by using the AGREE software to evaluate them are represented in Fig. 4.

Environmental sustainability revealed significantly favorable results for the benchtop NMR method, with a score of 0.73, compared to the lower scores achieved by GC-MS (0.36) and GC-FID (0.33). This superiority of benchtop NMR was particularly evident in critical aspects, such as sample preparation stages, automation or miniaturization, and notably in the absence of derivatization and the use of toxic agents in sample preparation. The latter point highlighted a significant reduction in waste generation, demonstrating a considerably "greener" environmental profile. However, the AGREE analysis highlighted areas for improvement in all the techniques analyzed, such as the need for off-line measurements due to the inability to conduct in-situ analysis, and the energy demands. It was important to mention that AGREE calculator did not distinguish between benchtop and conventional cryogen-cooled superconducting electromagnet NMR systems. As such, we considered that the actual energy consumption of benchtop systems would have been lower due to their use of permanent magnets. Nevertheless, even in the worst-case scenario, where the higher energy consumption of conventional systems was assumed, the benchtop NMR method still demonstrated a highly favorable environmental profile.

4. Conclusions

This study presents, for the first time, the simultaneous use of three different quantification models to determine fat content in UHT milk samples using benchtop NMR. This comparison provided a better understanding of the differences between traditional calibration curve applications, and PLS-R and ANNs based machine learning algorithms for estimating fat content in skimmed, semi-skimmed and whole milk. In analyzing the most challenging cases, such as skimmed milk, the ANN method exhibited better precision and accuracy (−14.9%, 7.3%) compared to both the calibration curve method (34.1%, 25.3%) and the PLS-R approaches (18.9%, −17.7%). This trend in fat quantification was again observed, with ANN consistently achieving better results in semi-skimmed (precision 10.2% and accuracy −0.3%) and whole milk

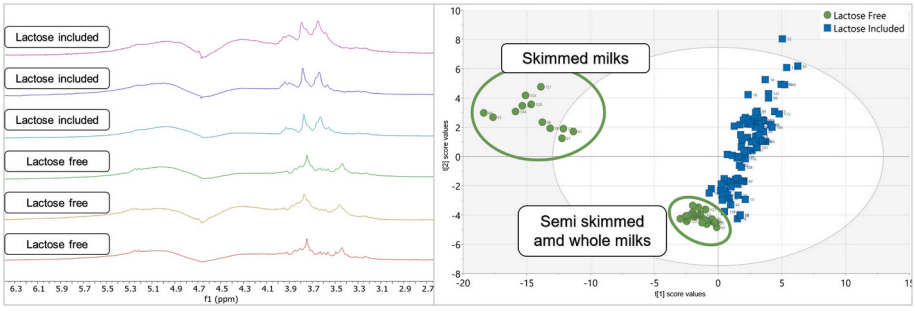


Fig. 3. Example of the spectral analysis obtained by benchtop NMR in the target region (left) and the binary classification PLS-DA (right) of lactose and lactose-free milk.

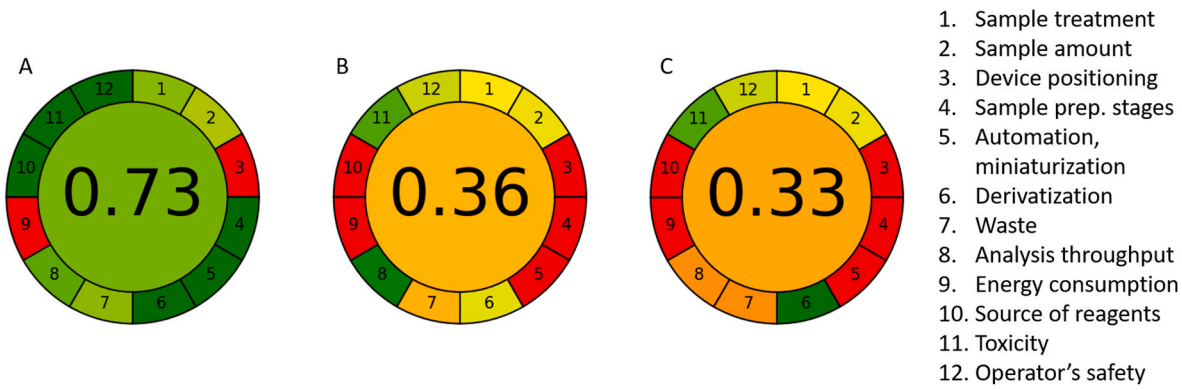


Fig. 4. Results of the AGREE analysis for environmental sustainability of the benchtop NMR method (A) compared to the GC-MS (B) and GC-FID (C) methods.

(precision 6.9% and accuracy −2.1%). All proposed methods have been adequately validated, demonstrating suitable values for selectivity, linearity within the working range, precision and accuracy.

Regarding sample analysis, the comparative analysis of the methods demonstrated a commendable level of accuracy for skimmed, semi-skimmed, and whole milk, with the overall average differences from the labelled values being 5.2%, −2.2%, and 0.1%, respectively. This accuracy underscored the potential of these methodologies to meet the rigorous demands of quality control within the dairy industry. The extension of the study to goat, sheep and coffee-flavoured milk further attests to the versatility and reliability of the methods in accurately determining the fat content, corresponding closely with the product labels.

Additionally, the application of machine learning to full spectra, rather than specific regions, has shown potential for binary classification models such as PLS-DA, successfully distinguishing between lactose and lactose-free milk, suggesting broader applications for future research.

In light of comparative analysis with established techniques such as GC-MS and GC-FID for quantifying fat content in milk samples, the AGREE analysis highlighted the exceptional greenness score of 0.73 for the benchtop NMR method, showcasing a robust commitment to environmentally friendly research practices. This comparison showed the significant sustainability advantages of benchtop NMR over these traditional techniques, with benefits including reduced sample preparation stages, enhanced automation and miniaturization, and the elimination of toxic reagents and derivatization processes. By markedly outperforming established techniques in terms of environmental impact, benchtop NMR cements its leading position in promoting a greener approach to quantifying fat content in milk.

CRediT authorship contribution statement

José Raúl Belmonte-Sánchez: Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Roberto Romero-González:** Writing – review & editing, Visualization, Supervision, Conceptualization. **Manuel Ángel Martínez Orosa:** Writing – review & editing, Formal analysis. **María Calvo Morata:** Writing – review & editing, Resources, Project administration, Funding acquisition. **Antonia Garrido Frenich:** Writing – review & editing, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Grant project ref. CPP2021-008672 funded by MICIU/AEI/10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR".

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lwt.2024.117000.

Data availability

The data that has been used is confidential.

References

Amaral, J. S., Mafra, I., Pissard, A., Fernández Pierna, J. A., & Baeten, V. (2018). Milk and milk products. In *FoodIntegrity handbook* (pp. 1–23). Eurofins Analytics France. https://doi.org/10.32741/fihb.1.milk.

Aquini, S., & Gil, J. M. (2017). Innovation and brand effects on the consumers' demand for fresh milk in Spain. *Agricultural and Applied Economics Association, 258131*, 1–30. https://doi.org/10.22004/ag.econ.258131

Bachmann, R. (2023). Guide to NMR method development and validation-Part I: Identification and quantification (update 2023). https://doi.org/10.13140/RG.2.2.30200.83208.

Belmonte-Sánchez, J. R., Aguilera-Sáez, L. M., Romero-González, R., Martínez Vidal, J. L., Arrebola, F. J., & Garrido Frenich, A. (2019). Determination of etidronic acid in vegetable-washing water by a simple and validated quantitative 31P nuclear magnetic resonance method. *Microchemical Journal, 150*, Article 104083. https://doi.org/10.1016/j.microc.2019.104083

Benzaama, M. H., Menhoudj, S., Mokhtari, A. M., & Lachi, M. (2022). Comparative study of the thermal performance of an earth air heat exchanger and seasonal storage systems: Experimental validation of Artificial Neural Networks model. *Journal of Energy Storage, 53*, Article 105177. https://doi.org/10.1016/j.est.2022.105177

Bharti, S. K., & Roy, R. (2012). Quantitative ^{1}H NMR spectroscopy. *TrAC, Trends in Analytical Chemistry, 35*, 5–26. https://doi.org/10.1016/j.trac.2012.02.007

Burger, R., Lindner, S., Rumpf, J., Do, X. T., Diehl, B. W. K., Rehahn, M., Monakhova, Y. B., & Schulze, M. (2022). Benchtop versus high field NMR: Comparable performance found for the molecular weight determination of lignin. *Journal of Pharmaceutical and Biomedical Analysis, 212*, Article 114649. https://doi.org/10.1016/j.jpba.2022.114649

Chen, M., Wang, F., Wu, X., Si, B., Pan, J., Zheng, N., Zhang, Y., & Wang, J. (2023). Updating the fatty acid profiles of retail bovine milk in China based on an improved GC-MS method: Implications for nutrition. *Frontiers in Nutrition, 10*. https://doi.org/10.3389/fnut.2023.1204005

Cobas, C. (2018). Applications of the Whittaker smoother in NMR spectroscopy. *Magnetic Resonance in Chemistry, 56*(12), 1140–1148. https://doi.org/10.1002/mrc.4747. John Wiley and Sons Ltd.

Corsaro, C., Vasi, S., Neri, F., Mezzasalma, A. M., Neri, G., & Fazio, E. (2022). NMR in Metabolomics: From conventional statistics to machine learning and neural network approaches. *Applied Sciences, 12*(6), 2824. https://doi.org/10.3390/app12062824

da Costa, N. L., da Costa, M. S., & Barbosa, R. (2021). A review on the application of chemometrics and machine learning algorithms to evaluate beer authentication. *Food Analytical Methods, 14*(1), 136–155. https://doi.org/10.1007/s12161-020-01864-7

Danudol, A., & Judprasong, K. (2022). Development and method validation of butyric acid and milk fat analysis in butter blends and blended milk products by GC-FID. *Foods, 11*(22), 3606. https://doi.org/10.3390/foods11223606

Debik, J., Sangermani, M., Wang, F., Madssen, T. S., & Giskeødegård, G. F. (2022). Multivariate analysis of NMR-based metabolomic data. *NMR in Biomedicine, 35*(2). https://doi.org/10.1002/nbm.4638. John Wiley and Sons Ltd.

Dhaulaniya, A. S., Balan, B., Sodhi, K. K., Kelly, S., Cannavan, A., & Singh, D. K. (2020). Qualitative and quantitative evaluation of corn syrup as a potential added sweetener in apple fruit juices using mid-infrared spectroscopy assisted chemometric modeling. *LWT, 131*, Article 109749. https://doi.org/10.1016/j.lwt.2020.109749

Draper, S. L., & McCarney, E. R. (2023). Benchtop nuclear magnetic resonance spectroscopy in forensic chemistry. *Magnetic Resonance in Chemistry, 61*(2), 106–129. https://doi.org/10.1002/mrc.5197

Evangelista, C., Basiricò, L., & Bernabucci, U. (2021). An overview on the use of near infrared spectroscopy (nirs) on farms for the management of dairy cows. *Agriculture, 11*(4), 296. https://doi.org/10.3390/agriculture11040296

Ezeanaka, M. C., Nsor-Atindana, J., & Zhang, M. (2019). Online low-field nuclear magnetic resonance (LF-NMR) and magnetic resonance imaging (MRI) for food quality optimization in food processing. *Food and Bioprocess Technology, 12*(9), 1435–1451. https://doi.org/10.1007/s11947-019-02296-w. Springer New York LLC.

FDA. (2019). Guidelines for the validation of chemical methods in food, feed, cosmetics, and veterinary products. *Guidelines for the Validation of Chemical Methods in Food, Feed, Cosmetics, and Veterinary Products, 23*. https://fda.report/media/121751/ValidationNucleicAcidSequenceBasedAnalysisFoodFeedCosmeticsVeterinary.pdf.

Galvan, D., Tanamati, A. A. C., Casanova, F., Danieli, E., Bona, E., & Killner, M. H. M. (2021). Compact low-field NMR spectroscopy and chemometrics applied to the analysis of edible oils. *Food Chemistry, 365*, Article 130476. https://doi.org/10.1016/j.foodchem.2021.130476

Hatzakis, E. (2019). Nuclear magnetic resonance (NMR) spectroscopy in food science: A comprehensive review. *Comprehensive Reviews in Food Science and Food Safety, 18*(1), 189–220. https://doi.org/10.1111/1541-4337.12408

Hou, X., Wang, G., Su, G., Wang, X., & Nie, S. (2019). Rapid identification of edible oil species using supervised support vector machine based on low-field nuclear magnetic resonance relaxation features. *Food Chemistry, 280*, 139–145. https://doi.org/10.1016/j.foodchem.2018.12.031

Hu, F., Furihata, K., Kato, Y., & Tanokura, M. (2007). Nondestructive quantification of organic compounds in whole milk without pretreatment by two-dimensional nmr spectroscopy. *Journal of Agricultural and Food Chemistry, 55*(11), 4307–4311. https://doi.org/10.1021/jf062803x

Huang, W., Fan, D., Li, W., Meng, Y., & Liu, T. C. (2022). Rapid evaluation of milk acidity and identification of milk adulteration by Raman spectroscopy combined with chemometrics analysis. *Vibrational Spectroscopy, 123*, Article 103440. https://doi.org/10.1016/j.vibspec.2022.103440

Jiang, X., & Xu, C. (2022). Deep learning and machine learning with grid search to predict later occurrence of breast cancer metastasis using clinical data. *Journal of Clinical Medicine, 11*(19), 5772. https://doi.org/10.3390/jcm11195772

Joshi, P. B. (2023). Navigating with chemometrics and machine learning in chemistry. *Artificial Intelligence Review, 56*(9), 9089–9114. https://doi.org/10.1007/s10462-023-10391-w

Lubes, G., & Goodarzi, M. (2017). Analysis of volatile compounds by advanced analytical techniques and multivariate chemometrics. *Chemical Reviews, 117*(9), 6399–6422. https://doi.org/10.1021/acs.chemrev.6b00698

Monakhova, Y. B., Kuballa, T., Leitz, J., Andlauer, C., & Lachenmeier, D. W. (2012). NMR spectroscopy as a screening tool to validate nutrition labeling of milk, lactose-free milk, and milk substitutes based on soy and grains. *Dairy Science & Technology, 92*(2), 109–120. https://doi.org/10.1007/s13594-011-0050-5

Nguyen, H.-M., Duong, T.-K., Nguyen, V.-K., Nguyen, T.-K.-L., Dong, T.-H.-Y., Nguyen, C.-H., & Tung, N.-T. (2024). A two-step design of experiments approach to investigate the simultaneous effects of ion-pairing and chemical enhancers to improve the permeability of lornoxicam in a topical hydrogel patch. *Journal of Pharmaceutical Investigation, 54*(2), 229–247. https://doi.org/10.1007/s40005-023-00660-9

Pena-Pereira, F., Wojnowski, W., & Tobiszewski, M. (2020). Agree - analytical GREEnness metric approach and software. *Analytical Chemistry, 92*(14), 10076–10082. https://doi.org/10.1021/acs.analchem.0c01887

Peris-Díaz, M. D., & Krężel, A. (2021). A guide to good practice in chemometric methods for vibrational spectroscopy, electrochemistry, and hyphenated mass spectrometry. *TrAC, Trends in Analytical Chemistry, 135*, Article 116157. https://doi.org/10.1016/j.trac.2020.116157

Pomyen, Y., Wanichthanarak, K., Poungsombat, P., Fahrmann, J., Grapov, D., & Khoomrung, S. (2020). Deep metabolome: Applications of deep learning in metabolomics. *Computational and Structural Biotechnology Journal, 18*, 2818–2825. https://doi.org/10.1016/j.csbj.2020.09.033

Reiner, J., Protte, K., & Hinrichs, J. (2020). Investigation of the applicability of Raman spectroscopy as online process control during consumer milk production. *ChemEngineering, 4*(3), 45. https://doi.org/10.3390/chemengineering4030045

Sa'adah, A. (2023). Artificial intelligence with python. *Technometrics, 65*(3), 451–452. https://doi.org/10.1080/00401706.2023.2237826

Sagmeister, P., Poms, J., Williams, J. D., & Kappe, C. O. (2020). Multivariate analysis of inline benchtop NMR data enables rapid optimization of a complex nitration in flow. *Reaction Chemistry & Engineering, 5*(4), 677–684. https://doi.org/10.1039/d0re00048e

Santos, M. C., Nascimento, P. A. M., Guedes, W. N., Pereira-Filho, E. R., Filletti, É. R., & Pereira, F. M. V. (2019). Chemometrics in analytical chemistry - an overview of applications from 2014 to 2018. *Ecletica Quimica, 44*(2), 11–25. https://doi.org/10.26850/1678-4618eqj.v44.2.11-25. Atlantis Livros Ltda.

Santos, P. M., Pereira-Filho, E. R., & Colnago, L. A. (2016). Detection and quantification of milk adulteration using time domain nuclear magnetic resonance (TD-NMR). *Microchemical Journal, 124*, 15–19. https://doi.org/10.1016/j.microc.2015.07.013

Schönberger, T., Monakhova, Y. B., Lachenmeier, D. W., Walch, S., Kuballa, T., Team, N.-P. E., & Germany, (NEXT) -NMR working group. (2015). *EUROLAB Technical Report 01/2015 -Guide to NMR method development and validation-Part II: Multivariate data analysi. 01, 1–20*. http://www.eurolab.org/documents/NMRValGuidelineIIV6.pdf.

Sørensen, M. K., Balsgart, N. M., Beyer, M., Jensen, O. N., & Nielsen, N. C. (2022). On-Site measurement of fat and protein contents in milk using mobile NMR technology. *Molecules, 27*(3), 583. https://doi.org/10.3390/molecules27030583

Soyler, A., Cikrikci, S., Cavdaroglu, C., Bouillaud, D., Farjon, J., Giraudeau, P., & Oztop, M. H. (2021). Multi-scale benchtop 1H NMR spectroscopy for milk analysis. *LWT, 139*, Article 110557. https://doi.org/10.1016/j.lwt.2020.110557

Tamime, A. Y. (Ed.). (2008). *Milk processing and quality management*. Wiley. https://doi.org/10.1002/9781444301649.

Taylor, B. J. (2006). *Methods and procedures for the verification and validation of artificial neural networks*. Kluwer Academic Publishers. https://doi.org/10.1007/0-387-29485-6

van Beek, T. A. (2021). Low-field benchtop NMR spectroscopy: Status and prospects in natural product analysis. *Phytochemical Analysis, 32*(1), 24–37. https://doi.org/10.1002/pca.2921

Zhao, L.-L., Qiu, X.-J., Wang, W.-B., Li, R.-M., & Wang, D.-S. (2019). NMR metabolomics and random forests models to identify potential plasma biomarkers of blood stasis syndrome with coronary heart disease patients. *Frontiers in Physiology, 10*. https://doi.org/10.3389/fphys.2019.01109