Comparison between chemometric analysis and machine learning for the prediction of macronutrients in fresh cheeses from the Netherlands

1. **Introduction**

* What is Cheese?
* Why is it important to measure fat and protein contents?

Extensive research has been conducted in the past to explore the utilization of near-infrared (NIR) spectroscopy (NIRS) in forecasting cheese composition, employing various methodologies, instruments, and chemometric techniques.

Cristina Malegori et al, employed near-infrared hyperspectral imaging (NIR-HSI) to discern and predict dehydration, proteolysis, and lipolysis during the maturation period, with a specific focus on a particular cheese type: Formaggetta, a semi-hard cheese typical of a region in Italy. (1)

Furthermore, Calvini et al, used HSI to quantify the rind percentage in grated Parmigiano Reggiano cheese samples. These authors converted Each hyperspectral image into a one-dimensional signal referred to as a hyperspectrogram, which includes the information present in the image.(2)

* With regard to the chemometrics algorithms used to calibrate models for determining moisture and fat content, Adams et al in 1999 concluded that orthogonal models built by selected wavelengths provided highly accurate results. (3)
* Chemometrics used in the past
* In reference to the equipments that have been used

While most calibrations of NIR or HSI models conducted on cheese samples were specific to a single cheese type, da Costa Filho et al chose a more generalized approach, using five types of cheese in the same calibration.(4) This approach, where more than one type of sample is incorporated into the model, can be referred to as a 'broad-based approach.'

In 2019, Stocco et al. collected a total of 1,050 diverse cheeses, categorizing them into 37 groups. They calibrated eight models using NIR data obtained from three different instruments to assess and compare their performances.(5)

In the present study, XXXX different types of samples of cheeses from The Netherlands were utilized to calibrate and validate NIR models for predicting macronutrient content. To the best of our knowledge, such an extensive range of diverse cheese varieties has never been employed in a single model, representing an 'extremely broad-based approach.'

The objectives of this study were 1) to compare the performance of machine learning techniques and chemometrics to predict the %Fat and % Protein contents in 80 cheeses from the Netherlands; 2)to evaluate the performance of a cheese model using an extremely broad-based approach, where numerous varieties are included in the same model and 3) to compare different spectral regions, highlighting the most important wavelengths to predict both macronutrients.

1. Materials and methods

2.1. Samples

In this work, 80 samples of cheese purchased in XXXX the Netherlands were used. The description of these samples can be seen in **Table 1.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cheese name | Type | Origin | Made from |  |
|  | Blue-veined Rich and creamy | Germany (Bavarian Alps) | Cow’s milk |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

2.2. Data acquisition

The camera used were the Specim FX10 (spectral range of 400 – 780 nm, high spatial resolution of 1024 pixels) and the Specim FX17 (spectral range of 941.1 nm to 1723.94 nm, high spatial resolution of 640 pixels, spectral resolution of 7–8 nm spectral steps) (Specim, Spectral Imaging Ltd., Finland).(6) The setup and the illumination system is depicted and explained in Qing Han et al. 2023, and shown in **Figure 1**.(7)



**Figure 1:** Setup system with Specim Camera FX17 used in this study.



**Figure 2:** Sample presentation to the instrument.

In each measurement, four slices of cheese taken from the same package were placed on the black platform of the system, separated as shown in **Figure 2** XXXX this Figure should change for un unpolluted one

A white panel was initially captured and used as a white reference during the image acquisition in the same way as Qing et al. 2021.(7) Images were acquired using 12 minutes exposure, Frame Rate=30,00 Hz, Spectral binning=2, Spatial binning=2, and Trigger mode: Internal. For this experiment we only used the FX17 camera, because XXXXX to complete

Each sample was scanned in duplicate and saved for later analysis.

2.3. Data analysis

2.3.1. Unfolding of images. Images were converted to raw spectra, were each row corresponded to one pixel and one spectrum.

2.3.2. Unfolded spectra were analyzed using both chemometrics and machine learning approaches.

2.4. Chemometrics.

1. Cristina Malegori PO, Eleonora Mustorgi, Maria Alessandra Boggiani, Graziella Pastorini, Monica Casale An in-depth study of cheese ripening by means of NIR hyperspectral imaging: Spatial mapping of dehydration, proteolysis and lipolysis. Food Chemistry. 2021;343(128547).

2. Calvini RM, Sara & Pizzamiglio, Valentina & Foca, Giorgia & Ulrici, Alessandro. Exploring the potential of NIR hyperspectral imaging for automated quantification of rind amount in

grated Parmigiano Reggiano cheese. Food Control 2020;112(107111).

3. Michael J Adams KL, Neil W Barnett, Allan J Poynton. Calibration models for determining moisture and fat content of processed cheese using near-infrared spectrometry. Journal of the Science of Food and Agriculture. 1999.

4. Paulo Augusto da Costa Filho PV. Broad-based versus specific NIRS calibration: Determination of total solids in fresh cheese. Analytica Chimica Acta. 2005;544(1-2):82-8.

5. Giorgia Stocco CC-G, Alessandro Ferragina, Paolo Berzaghi, Giovanni Bittante Accuracy and biases in predicting the chemical and physical traits of many types of cheeses using different visible and near-infrared spectroscopic techniques and spectrum intervals. Journal of Dairy Science. 2019;102(11):9622-38.

6. Specim FX17 [Available from: <https://www.specim.com/products/specim-fx17/>.

7. Qing Han JP, Sara W. Erasmus, Christopher T. Elliott, Saskia M. van Ruth. Interpreting the variation in particle size of ground spice by high-resolution visual and spectral imaging: A ginger case study. Food Research International. 2023;170( 113023).