Food Chemistry 134 (2012) 412–418

Contents lists available at SciVerse ScienceDirect

Food Chemistry

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

Analytical Methods

Quantitative determination of fat and total protein in milk based on visible light scatter

Andrey Bogomolov [a,*], Stefan Dietrich [a], Barbara Boldrini [b], Rudolf W. Kessler [b]

[a] J&M Analytik AG, 8 Willy-Messerschmitt-Strasse, 73457 Essingen, Germany
[b] Reutlingen University, Alteburgstrasse 150, 72762 Reutlingen, Germany

ARTICLE INFO

Article history:
Received 6 May 2011
Received in revised form 19 October 2011
Accepted 14 February 2012
Available online 22 February 2012

Keywords:
Milk analysis
Fat content
Total protein content
Visible spectroscopy
Near infrared spectroscopy
Light scatter
Mie theory
Design of experiment

ABSTRACT

A new optical spectroscopic method for milk fat and total protein analysis has been developed. In contrast to the conventional approach that generally relies on the components' absorption, the suggested method is based on the phenomenon of light scatter by fat and protein particles. This fundamental distinction enables shifting the measurement to the cost-effective visible and adjacent near infrared region (below 1000 nm), where the scatter strongly dominates.

Partial Least-Squares regression modelling on a designed set of training and validation milk samples resulted in root mean-square prediction errors of 0.05% and 0.03% for fat and protein content, respectively, which is close to the accuracy of reference analysis. It has been shown that multivariate data analysis is capable of distinguishing individual scatter spectra of fat and protein. This conclusion has been supported by Mie-theory calculations.

The method is suitable for routine laboratory analysis or in-line quality monitoring in the dairy production.

1. Introduction

Fat and protein are two main nutrients of milk, and their concentrations are routinely monitored during the commercial production and in the final product. Growing quality requirements in the food and pharmaceutical industries promote the dissemination of spectroscopic analysis enabling real-time monitoring of processes and products. Optical spectroscopy is one of the most favored techniques for the industrial quality monitoring due to its wide availability, undemanding sampling, quick measurement and the applicability of fibre optics and remote probes on its basis.

Spectroscopic analysis of homogenized and raw natural milk has been intensively investigated by infrared (IR) (Iñón, Garrigues, & De la Guardia, 2004; Luginbühl, 2002) or near IR (NIR) spectroscopy in both "silicon" detector range (700–1100 nm) (Brennan, Alderman, Sattler, O'Connor, & O'Mathuna, 2003; Chen, Iyo, Kawano, & Terada, 1999; Kalinin et al., 2008; Šašić & Ozaki, 2001; Tsenkova, Atanassova, Ozaki, Toyoda, & Itoh, 2001; Tsenkova et al., 1999; Wu, He, & Feng, 2008) and far-NIR (1100–2500 nm) (Kalinin, Krasheninnikov, & Potapov, 2009; Kalinin et al., 2008; Purnomoadi, Batajoo, Ueda, & Terada, 1999; Tsenkova, Atanassova, Itoh, Ozaki, & Toyoda, 2000; Tsenkova et al., 1999). Reported prediction errors for both fat and protein content typically lie between 0.07% and 0.10% or above.

Milk has pronounced light-scattering properties due to the presence of emulgated fat globules and suspended protein micelles (Walstra, Wouters, & Geurts, 2006, chap. 1). Applying the spectroscopic analysis, researches typically attempt to minimize the scatter and thus emphasize the absorption. The scatter can be reduced during the measurement or through a mathematical post-processing of the spectra, e.g. multiplicative scatter correction (MSC) by Geladi, MacDougall, and Martens (1985), standard normal variate (SNV) or numerical derivatives. Multivariate regression analysis is then commonly used to obtain the models capable of predicting component concentrations (Rodriguez-Otero, Hermida, & Centeno, 1997).

A complex multimodal size distribution of scattering particles significantly complicates the spectroscopic analysis of milk, specifically, in the economically very attractive region of visible (Vis) light (400–700 nm), where the scatter is essentially stronger than in NIR. For this reason, visible wavelengths are usually ignored as containing no intense characteristic absorption bands of the milk components. There are only a few published works exploiting the Vis region for milk analysis (Crofcheck, Payne, Hicks, Menguc, & Nokes, 2000; Muñiz et al., 2009).

The scatter is generally considered as a "parasitic" phenomenon, complicating the spectroscopic analysis of opaque media. At the same time, the scatter may deliver quantitative information. The detected intensities of scattered light at different wavelengths depend on the number and sizes of colloidal particles, and

* Corresponding author. Fax: +49 7361 92812.
E-mail address: bogomolov@j-m.de (A. Bogomolov).

0308-8146/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2012.02.077

consequently, on the respective component content. Hence, for systems with a stable composition (or having predictably narrow variations) like milk the scatter can potentially be used for quantitative analysis. This statement can be supported by an example from powder technology, where the application of diffuse reflectance NIR spectroscopy to the analysis of particle size distribution has been proved by Nieuwmeyer et al. (2007). Nevertheless, the feasibility of quantitative analysis of multi-component mixtures from the scatter spectra still stays poorly studied.

There is a number of earlier works establishing the correlation between the visible light scatter and milk fat content, the so-called turbidimetric methods (Ashworth, 1969; Walstra, 1967). None of them however resulted in a workable analytical technique having a practically acceptable accuracy. An essential fault of these early works was the univariate approach. Although full-spectrum data were available, the analysis still stuck to individual wavelengths, one at a time. In this way, the globule size differences and the presence of smaller protein particles became the factors seriously limiting the method accuracy. The method by Ashworth (1969) required a sample pre-treatment to disperse the protein micelles, but the univariate model still depended on the homogenization efficiency and could not be transferred from one dairy to another.

To our knowledge, there is no scatter-based method for quantitative determination of total protein in milk. The task of physically based rendering of the milk appearance has been successfully resolved by Frisvad, Christensen, and Jensen (2007). The authors have also stated the feasibility of a reverse solution, i.e. resolving the constituent scatter properties from an image. This conclusion is important as an independent theoretical confirmation of the formal modelling technique we suggest in the manuscript.

In the present work, the spectroscopy in the predominant Vis and adjacent short-wave (SW) edge of the NIR region (up to 1000 nm) has been successfully applied to the quantitative analysis of fat and total protein in bovine milk. Partial Least-Squares (PLS) regression analysis of milk spectra has been performed using a series of samples closely reproducing the natural milk composition. Fat and total protein concentrations in the training and validation sample sets were systematically varied in accordance with the authors' original design scheme. The main purpose of the research was to investigate the feasibility of accurate prediction of the milk composition from Vis/SW-NIR spectra. Comparative analysis of PLS diagnostic plots (scores and loadings) and mathematical scatter simulations has led to important conclusions on the role of milk components' individual scatter spectra in the multivariate regression modelling.

2. Materials and methods

2.1. Milk sample design and preparation

For a robust modelling the milk samples should be compositionally close to the analyzed production milk. The sample set should provide a systematic coverage of the whole design space, including the intervals of possible content variation. To achieve this, the samples should be prepared by a controlled dosing of the main constituents.

To prepare milk samples with a desired composition, the following technique was used (Kalinin et al., 2008, 2009). The samples were mixed of four source components. Two of them were cream (fat source) and skim milk (protein source), respectively containing (w/w): 34.20% and 0.06% of fat, 2.00% and 3.31% of protein and 2.97% and 4.83% of lactose. These were the intermediates, taken directly from the normalized milk production line of a large dairy (Schwarzwaldmilch GmbH, Freiburg, Germany). Lactose concentration was kept constant at 4.70%. This was achieved by additions of aqueous 10%-solution of lactose and pure water (sterile distilled water for injections was used in both cases).

Each sample was prepared by mixing calculated weights of the above four constituents and treating them with ultrasound in SONOREX TK30 50 kHz bath (Bandelin electronic, Berlin, Germany) for 5 min. It was initially supposed that ultrasound treatment would lead to the milk homogenization as reported by Ertugay, Şengül, and Şengül (2004). However, subsequent microscopic observation has shown that the samples still contained an essential share of fat particles with sizes above 5 μm. Hence, it should be concluded that the homogenization was still far from completeness.

The samples were composed in accordance with a scheme, presented in Fig. 1. Possibly wide intervals of fat and protein content, given the source material composition, were selected. A custom design, suggested by the authors, was put into the basis of the experiment. The main principle of this design is to cover possibly more concentration levels of the two components with the minimal number of samples. In accordance to it, fat and protein contents are varied in 11 levels each using only 11 samples. Several additional samples were added to update the existing scheme to the central composite circumscribed (CCC) design (Eriksson, Johansson, Kettaneh-Wold, Wikström, & Wold, 2008, chap. 6.4) with five replicates in the central point. The CCC design is often used to study the effects of multiple factors on a response function. In Section 3, dependences of Vis/SW-NIR spectra on the fat and protein content will be illustrated using the CCC samples. In total 21 samples were prepared.

The reference analysis of milk fat and total protein content was performed using MilkoScan™ FT120 (FOSS, Denmark). The actual concentrations of fat and protein in analyzed samples could slightly deviate from the design (Fig. 1), for example, because of component dosing inaccuracy or fat sedimentation on the walls of sample-containing vessels.

The validation samples (see also Section 2.3) were selected starting from 2 to 5 (central point replicates) and adding subsequent samples so that they never occupy any two adjacent levels of either component. Also, the validation set should not include any marginal samples, i.e. those forming the outer border of the design space (the numbers 12, 13, 16–18, 19–21). This algorithm

% Protein \ %Fat	3.02	3.10	3.18	3.26	3.34	3.42	3.50	3.58	3.66	3.74	3.82	3.90	3.98
3.20							19						
3.15												15	
3.10			12								**20**		
3.05										10			
3.00					**9**								
2.95								7					
2.90	18						1 **2-5**						21
2.85						6							
2.80									**8**				
2.75				11									
2.70			**17**								13		
2.65		14											
2.60							16						

Fig. 1. Experimental design of milk sample composition for multivariate regression analysis. Validation samples are selected in bold. Samples added to update to the CCC design (gray fields) are underlined.