LWT - Food Science and Technology 69 (2016) 348–357

Contents lists available at ScienceDirect

LWT - Food Science and Technology

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

On-line measure of donkey's milk properties by near infrared spectrometry

Giuseppe Altieri*, Francesco Genovese, Naouel Admane, Giovanni Carlo Di Renzo

SAFE - Scuola di Scienze Agrarie, Forestali, Alimentari e Ambientali, Università degli Studi della Basilicata, Viale dell'Ateneo Lucano, 10, 85100 Potenza, Italy

ARTICLE INFO

Article history:
Received 15 July 2015
Received in revised form
16 January 2016
Accepted 28 January 2016
Available online 30 January 2016

Keywords:
Regression methods
NIR spectroscopy
Discrete NIR LED device

ABSTRACT

The ranchers handle donkeys breeding in small livestock farms and there is a need to supply a suitable tool to assess milk characteristics.

To this aim, near infrared (NIR) spectroscopy was assessed through nine statistical methods: partial least squares, principal components, multivariate adaptive splines, regression trees using M5′ method, least squares support vector machine, artificial neural network, regression trees, ensemble regression trees and regularized least squares regression. These methods were implemented using a forward sequential feature selection algorithm; their predicting capability was validated with ten-fold cross validation; different types of spectra pre-processing methods were tested.

The trials confirmed that NIR analysis is an efficient method for quantitative analysis of protein, lactose and dry matter content of donkey's milk with a full-scale prediction error (FSPERR) of 3.0%, 4.4% and 4.5% respectively. Furthermore, a 16 NIR light emitting diodes (LED) device was investigated using simulation; in this case the prediction models of protein content showed the best FSPERR (3.1%), followed by those of dry matter content (7.0%) whereas those of the lactose content showed the worst (19.8%). These results could be used to design a low cost NIR LED device for the real-time control of donkey's milk in breeding farms.

1. Introduction

Donkey's milk is the best substitute of human milk for its content in lactose, proteins, minerals, and omega-3 fatty acids, furthermore, donkey's milk contains several antimicrobial factors, such as lactoferrin, lactoperoxidase and high amounts of lysozyme, suitable to prevent intestine infection to infants (Vincenzetti, Polidori, & Vita, 2007). Hypoallergenic alternatives to breast milk have been proposed but rarely these products solve the problem and often present poor palatability (Criscione et al., 2009). Moreover, the increasing number of children intolerant to cows' milk protein makes donkey's milk extremely important in the first years of life (Sampson, 2004).

Donkey's milk is a very expensive food for different reasons, particularly, (a) for the low productivity of the donkeys and (b) for the very long gestation stage (12–14 months) with a consequent long period of no-milk production, thus the animal continues to produce only if it has beside the foal and the share of milk removed by the milking operation is very low; moreover, (c) the period of nursing mothers is very short.

Farmers handle donkeys breeding mostly in small livestock farms with a maximum of 60 animals, the average milk production is of 12–72 kg per day being related to the donkey-specific milk production of 0.3–1.2 kg per day. Donkey's milk is generally sold as fresh pasteurized milk with market price between 12 € and 16 € per litre; furthermore, due to the seasonal variations in market demand, limited milk amount is retained for instant soluble powder production (Di Renzo, Altieri, & Genovese, 2011; 2013).

Despite the high price, the interest of consumers, looking for a natural product as a substitute for breast milk or cow's milk, has led to an increase in demand for donkey milk (Criscione et al., 2009).

Due to increasing donkeys breeding farms, the Italian Breeders Association has introduced greater controls on the donkey's milk production and the Italian law allows farmers the opportunity to sell their product as raw milk directly from farm (Criscione et al., 2009); therefore, there is a clear need for small farms to provide themselves with suitable tools allowing the assessment of qualitative and quantitative milk characteristics. To this aim spectrophotometric methods could be very attractive, considering that

* Corresponding author.
E-mail address: giuseppe.altieri@unibas.it (G. Altieri).

http://dx.doi.org/10.1016/j.lwt.2016.01.069
0023-6438/

they are valuable investigative tools to evaluate the composition of many foods. In particular, the near infrared (NIR) spectroscopy is a suitable alternative to traditional analytical techniques, being simple, fast, non-destructive and requiring minimal or no sample preparation; furthermore, it is inexpensive, non-polluting and safely, because it does not require chemical reagents and particularly trained staff (Huang, Yu, Xu, & Ying, 2008; Nicolai et al., 2007; Porep, Kammerer, & Carle, 2015; Ru & Glatz, 2000).

The key element for an efficient NIR analysis is the construction of a good chemometric model able to extract all the information present, and sometime hidden, in the spectrum; in order to decrease the prediction error level of the chemometric model it is necessary to reduce the number of variables; in addition, it is possible to process the spectrophotometric signal with mathematical transformations of the spectrum using various types of pre-treatments reducing and/or eliminating the effects unrelated to the property of interest (Huang et al., 2008; Porep et al., 2015).

To remove unwanted effects the pre-treatments include multiplicative scatter correction (MSC), standard normal variate (SNV) correction, direct orthogonal signal correction (DOSC) (Azzouz, Puigdoménech, Aragay, & Tauler, 2003; Westerhuis, de Jong, & Smilde, 2001; Wold, Antii, Lindgren, & Ohman, 1998), de-trending or baseline correction, spectral truncation, data standardization and use of first and second derivative (Savitzky & Golay, 1964).

The linear regression methods commonly used to develop chemometric models are: “principal components regression” (PCR) and “partial least squares regression” (PLS) developed in 1975 by Herman Wold and Svante Wold (Wold, 2001). In PLS the decomposition is done by using both spectral data and data obtained from conventional analysis (Cavalcanti Inácio, De Fátima Vitória de Moura & Gomes de Lima, 2011). A widespread method is the “support vector machines” (SVM) that originates as a binary classification tool. The “least squares support vector machine” (LSSVM), as improved form of SVM, it is considered able to deal with linear and nonlinear multivariate calibration (Balabin & Lomakina, 2011; Balabin & Smirnov, 2011). Another largely used technique is the “artificial neural network” (ANN) being a non-linear method inspired by the nervous system (Nawi, Chen, Jensen, & Mehdizadeh, 2013).

The major problem of regression methods is represented by the outliers. Their detection is extremely important before data analysis because a single outlier can have a significantly greater and detrimental effect on the model than normal samples, creating misleading statistics (Moller et al., 2005).

The chemometric model performance is evaluated by the coefficient of determination (R2C for calibration and R2V for validation/prediction) and the root mean square error (RMSEC for calibration and RMSEV for validation/prediction); in addition, a further parameter has to be considered to evaluate the model fit goodness, namely the “ratio of standard error of performance to standard deviation” (RPD), firstly introduced by Williams and Norris (Kays, Archibald, & Sohn, 2005; Porep et al., 2015). The RPD is calculated as the ratio of the standard deviation of reference data for the prediction samples to standard error of prediction.

An accurate and robust model should have high R2C and R2V, high RPD, low RMSEC and RMSEV values (Wu et al., 2012). The recommended RDP value for the use of NIR in quality control is 5.0 or more; values higher than 8.1 represent excellent results and recommend the use of NIR chemometric model in whatever application (Kays et al., 2005).

The good model prediction capability is statistically assessed using k-fold cross-validation (CV) or leave-one-out CV depending on samples size: these involve the random partitioning of data samples into complementary subset, this operation is subsequently executed k times to reduce the variability of the results using different random partitions of the data set (Picard & Cook, 1984).

Furthermore, a sequential feature selection algorithm can be used to select a subset of features (wavelengths) that best predict the samples properties by sequentially selecting the feature of interest until there is no further improvement in prediction (Liu & Yu, 2005).

Currently, researchers have applied NIR techniques to evaluate and monitor food production (Nicolai et al., 2007; Porep et al., 2015; Triebold, 2000) with good results. NIR techniques have been applied to raw cow's milk (Jankowska & Sustova, 2003; Kawasaki et al., 2008; Melfsen, Hartung, & Haeussermann, 2013, 2012; Tsenkova, Atanassova, Itoh, Ozaki, & Toyoda, 2000) and to raw goat's milk (Albanell et al., 2003; Núñez-Sánchez et al., 2016); in addition, some researchers explored the possibility to apply these techniques to donkey's milk with satisfactory results (Altieri, Di Renzo, & Genovese, 2011; Zheng et al., 2007).

The aim of the this work is: (a) to plan and test improved models of calibration, (b) to compare the obtained results and validate them with respect to model prediction capability, (c) to design and test a low-cost on-line sensor in the NIR spectrum to monitor, in real-time, protein, lactose and dry matter content in donkey's milk.

2. Materials and methods

Donkey's milk samples (178) selected from three different livestock were collected at different periods from calving, and transported by keeping their temperature at 4 °C until the arrival to the laboratory, where they were subsequently frozen at −20 °C before the analysis.

Once defrosted, milk samples were analyzed for total protein content, lactose content and dry matter. Total protein content was measured using Biuret method (Sapan, Lundblad, & Price, 1999), lactose was measured using Fehling's solution method (Triebold, 2000), while the dry matter content was measured using gravimetric method (AOAC, 1990). The sample was scanned using a NIR spectrophotometer (Avaspec NIR256-2.5, Avantes) in transmission mode, and an overall number of 178 spectra were collected, constituted of 256 wavelengths in the region between 900 and 2500 nm. For the spectra collection was used the dedicated software Avasoft Basic ver. 7.3 from Avantes.

In order to use the acquired spectra in the subsequent phases and to improve the prediction performance of the different models, different types of spectra pre-processing methods were carried out on the spectral data. These pre-treatments included: SNV and OSC correction, data standardization (NORM), denoising using wavelet thresholding (WAV) (Cohen, 2012) and Savitzky-Golay first and second derivatives and normalization in order to remove unwanted effects.

Data loading, pre-processing, variable method selection (forward features selection), chemometric regression model construction and validation were performed in MATLAB. The model validation was assessed using the k-fold CV with $k = 10$.

The outliers detection was planned by removing samples with extreme values which exhibited increased influence on the model. Rousseeuw and Van Zomeren (1987, 1990, and 1992) suggest the use of “least median of squares” (LMS) estimator to detect regression outliers; this algorithm was implemented in MATLAB.

As further calibration methods have been investigated: the “adaptive regression splines” (ARES) (Jekabson, 2011) that works with piecewise-linear and piecewise-cubic regression models built using the “Multivariate Adaptive Regression Splines” technique (also known as MARS); the “regression trees and model trees” using M5′ method (M5TREE) (Jekabson, 2010); the “least-squares support vector machine” (LSSVM) (De Brabanter et al., 2010); the “regression trees” (REGTREE); the “ensemble modelling of regression trees