NIR spectroscopy: optimisation of linear regression model to predict milk coagulation trends

C. Giardina[a], N. Estephan[b], T.M.P. Cattaneo[a], R. Giangiacomo and D.N. Rutledge[b]

[a]*Istituto Sperimentale Lattiero-Caseario, via A. Lombardo, 11 – 26900 Lodi - ITALY*

[b]*Institut National Agronomique Paris-Grignon, Laboratoire de Chimie Analytique, UMR INRA-INA PG, 16, rue Claude Bernard – 75005 Paris -FRANCE*

Introduction

The structural properties of most dairy products play an important role in their final quality. Milk coagulation is the primary step in the development of these properties: clotting and cutting times are the key parameters during the coagulation process. In previous works[1-2], we found three critical points (iCT, CT, GT), detected by NIRS at a single wavelength, confirming the presence of three steps in the early cheese-making phase.

The aim of this work was to optimise a linear model which could describe the coagulation process monitored by NIRS, in order to generalize the procedure under various technological conditions. For this reason, a Partial Least Squares (PLS) model was applied to full spectra collected from four different types of milk coagulation processes.

Materials and methods

Materials

Four different coagulation tests were carried out in duplicate at 35°C, using reconstituted skim milk powder at 10% of solid content (w/w) in Ca Cl_2 0.05% as substrate[1] and four different clotting agents: liquid calf rennet (LCR), lactic acid bacteria (LAB), rennet combined with LAB and glucono-δ-lactone (GDL).

LCR (containing 17% of bovine pepsin, strength 1:16000 Soxhlet units, Caglificio Clerici, Cadorago, CO, Italy) was diluted to 0.8% (v/v) with distilled water and added to milk at 2%. Lyofast starter 034 (Sacco, Cadorago, CO, Italy) was used as LAB and added to milk at 1% for direct use. For LCR+LAB, at first LAB (starter 034) and, then, LCR under the same conditions used for LCR test were added to milk. GDL (99%, Sigma, Milano, Italy) was added to milk using a concentration of 12 g/L. This concentration of GDL was chosen to obtain coagulation times comparable to those obtained by LAB coagulation.

Methods

NIR spectra were collected by a dispersive spectrometer InfraAlyzer 500 (Bran+Luebbe, Germany) in transflectance mode as *log (1/R)* in the range 1100-2500 nm at 4 nm intervals using Sesame software (Bran+Luebbe, Norderstedt, Germany). The instrument was equipped with a thermostatable liquid sample cell. Temperature was controlled by an external circulating bath (Haake, mod. F3-CH, Karlsbruhe, Germany). Milk samples were injected into the cell after the

Near Infrared Spectroscopy: Proceedings of the 11th International Conference.

addition of the clotting agent, and spectra were recorded in sequence every 71 (LCR, LCR+LAB) and 120 (LAB, GDL) seconds until curd formation.

Visual observation of flock formation was used as reference method.

Data processing

NIR data were processed by using a software package developed by Barros and Rutledge[3]. In order to minimise the effect of baseline shifts, the spectra were autoscaled and, then, normalised between 0 and 1.

Principal Component Analysis (PCA) was applied as exploratory analysis to all the NIR data. A Calibration model was built for the prediction of the different steps in each kind of coagulation process. This model was created by using a PLS regression between the NIR data sets and the actual time matrices. The number of Latent Variables (LV) was optimised combining results obtained by calculating the Durbin-Watson criterion for the B-coefficients vectors and by PLS Cross-Validation. Replicates of coagulation tests, carried out on different days, were used as independent prediction sets.

Results and discussion

An example of NIR spectral data collected during coagulation processes is shown in Figure 1. Apparently, only LCR spectra seem different from those of the other processes.

An exploratory PCA analysis applied to all four sets of NIR data together and discriminated clearly the different coagulation processes. The scores plot is shown in Figure 2: each set is separated from the others along PC3. This result is due to the different mechanisms of curd aggregation that occur during each kind of process.

LCR system is already different at the beginning of the process after the rennet addition. Rennet hydrolyses a specific peptide bond of k-casein in such a way that caseinomacropeptide (CMP) is released and dissolves in the whey, and para-k-casein remains in the micelles of casein. The altered casein is referred to as paracasein and, in this way, it cannot be dissolved in soluble fraction. Because of this, the paracasein micelles in the milk coagulate.[4]

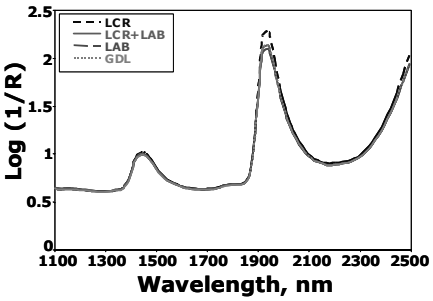

Figure 1. Examples of NIR spectra collected for each kind of coagulation in the range 1100-2500 nm.

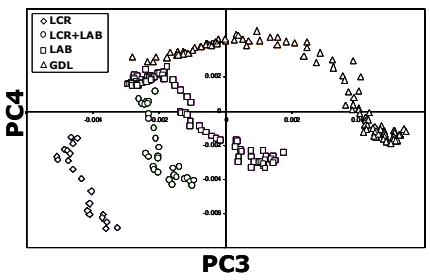

Figure 2. PCA applied to all the sets of NIR data.

In LAB system, milk clots by lowering the pH owing to LAB action. The casein micelles aggregate at the pH values near their isoelectric point without any breakdown of casein. This can explain why the LCR+LAB scores were located between LAB and LCR scores, but nearer to LAB than to LCR scores.

GDL system is in general used to simulate LAB experimental systems[5] because the milk achieves low pH thanks to the hydrolysis of GDL to lactic acid. At the beginning ,GDL system is similar to LAB system, as shown by its scores near to those of LAB. As the process evolves, the trend of the GDL scores becomes characteristic, placing itself -in the opposite zone of the plane related to LCR scores.

For each type of coagulation, a linear regression model was optimised in order to predict the different milk coagulation trends. To obtain the PLS Regression models, it was necessary to build a calibration set for each kind of coagulation and this goal was achieved by applying PCA to the four data sets one by one. The samples of the calibration sets were chosen observing the PC scores distribution as shown in Figure 3 for LCR.

By using the whole spectral range, PC2 behaviour confirms the presence of three critical points (iCT, CT, GT) detected by NIRS during milk clotting.[6] Furthermore, measurements of flock formation by visual observation were found to be able to detect process changes later than NIR indices.

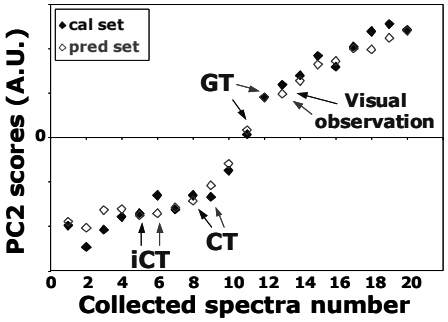

Figure 3. PC2 scores distribution for LCR process.

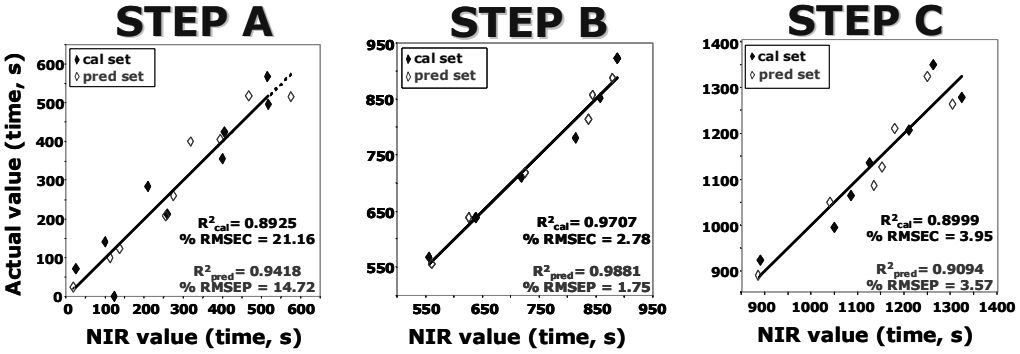

Figure 4. Calibration curves and PLRS results for LCR coagulation.

The models obtained for each coagulation were able to identify three steps: A, B, C. In Figure 4 calibration curves are shown only for LCR, while for LCR+LAB, LAB and GDL the PLSR results are reported in Table 1. Even if there were low structural modifications at the beginning of the process, in general step A was characterised by high %RMSEP values, from 14.72 to 30.23, due to

the important influence of temperature control on NIR data used in calibration. Step B was able to predict the coagulation times, regarding each kind of coagulation process, showing %RMSEC < 5.27, and %RMSEP < 4.35. Times of the step C were predicted with satisfactory accuracy in both calibration and prediction sets (%RMSEC < 6.98, %RMSEP < 7.14).

Table 1. PLSR results obtained for LCR+LAB, LAB and GDL processes.

TYPE OF COAGUALTION	STEP A				
	n*	R^2_{cal}	% RMSEC	R^2_{pred}	% RMSEP
LCR+LAB	7	0.8560	25.30	0.8148	26.55
LAB	46	0.6879	32.96	0.6183	30.23
GDL	25	0.9610	11.87	0.9172	14.22
	STEP B				
	n*	R^2_{cal}	% RMSEC	R^2_{pred}	% RMSEP
LCR+LAB	7	0.9722	3.71	0.9762	3.38
LAB	14	0.9584	1.60	0.9123	2.27
GDL	36	0.9556	5.27	0.9684	4.35
	STEP C				
	n*	R^2_{cal}	% RMSEC	R^2_{pred}	% RMSEP
LCR+LAB	18	0.9784	4.12	0.9182	7.14
LAB	32	0.8876	4.21	0.9438	2.81
GDL	41	0.7831	6.98	0.8172	5.67

n* = sample number

Within each step, the values of %RMSEC and %RMSEP were different from one type of process to the another. In step A, the lowest values were obtained for GDL and LCR. This result could be due to the way in which these processes evolved.

For all the systems, step B was characterised by the lowest values of %RMSEC and %RMSEP. These results indicated that PLS model for step B was more linear than those for the other two steps (A, C). In this middle step, all the curds began to gain substance up to the forming of gel. The best results were obtained for LCR and LAB processes.

Experimental %RMSEC and %RMSEP values for step C showed how the regression model for this step was sufficiently linear even if less so than step B. This behaviour could be due to the presence of other important system rearrangements as the early cheese-making phase progresses. As in step B, the best results were obtained for LCR and LAB.

On the basis of the PLS regression results obtained in all three steps, LCR was the best coagulation process described. Nevertheless, a satisfactory optimisation of the linear models was also obtained for the other three processes, with decreasing accuracy from LAB to GDL.

Conclusions

Experimental data suggested that the linear regression model of B and C steps could be used as a useful tool in monitoring clotting processes.

In all four cases, the greatest modifications occurred in step B, while step C may give additional information related to final characteristics of curd, provided further investigations are made.

This study shows that NIR spectroscopy together with chemometric tools such as PLS regression, could be applied to cheese-making processes, even if performed under different technological conditions. Furthermore, this approach could be applied in the study of other gelation and aggregation processes used in food industry.

References

1. R. Giangiacomo, R. Lizzano, S. Barzaghi and T.M.P. Cattaneo, *J.Near Infrared Spectrosc.* **6,** 205 (1998).
2. R. Lizzano, S. Barzaghi, T.M.P. Cattaneo and R. Giangiacomo, in *Near Infrared Spectroscopy: Proceedings of the 9th International Conference*, Eds. T. Davies and R. Giangiacomo, NIR Publications, Chichester, UK, p. 339 (2000).
3. A:S. Barros, in *Contribution à la selection et la comparison de variables caracteristiques. PhD thesis*, Institut National Agronomique, Paris-Grignon, chapter 1 (1999).
4. P. Walstra, T.J. Geurts, A. Noomen, A. Jellema and M.A.J.S. van Boekel, in *Dairy Technology: Principles of Milk Properties and Processes*, Ed by O.R. Fennema, M. Karel, G.W. Sanderson, S.R. Tannenbaum, P. Walstra, J.R. Whitaker. Marcel Dekker, Inc., N.Y., USA, p. 541 (1999).
5. J.A. Lucey, T van Liet, K. Grolle, T. Geurts and P. Walstra, *Int. Dairy Journal* **7**, 381 (1997)
6. R. Giangiacomo, in *CD-ROM Proceedings of the 26th IDF World Dairy Congress, Session "Tracers and Sensors"*, Paris, 24-27 September 2002.