University of Milan

Faculty of Agricultural and Food Sciences

Department of Food, Environmental and Nutritional Sciences

PHD in

Food Systems

XXXIII Cycle

PROCESS ANALYTICAL TECHNOLOGY APPROACHES FOR DAIRY INDUSTRY

Thesis of

Lorenzo Strani

R11978

Supervisor: Prof.ssa Ernestina CASIRAGHI

PhD Coordinator: Prof. Ella PAGLIARINI

Academic Year 2019-2020

Table of contents

1 – Abstract

This thesis work wants to answer the need of dairy industry to increase productivity while satisfying the consumers request for higher quality products. In order to do that, dairy companies need innovative methods to improve the understanding and the monitoring of production processes. Process Analytical Technology (PAT) approaches are the perfect tool for this purpose, as they use green, fast, non-invasive and non-destructive sensors that allow to perform measurements in real time. The most used techniques in this field are Near- and Mid-Infrared (NIR and MIR, respectively) spectroscopy, whose probes can be directly installed in critical points of the process providing both physical and chemical information of the product. However, these techniques have the drawback of providing results (spectra) difficult to be interpreted without proper statistical tools. In this context, Chemometric methods and algorithms allow the extraction of relevant information from spectroscopic data, providing a better understanding of the studied system.

The first part of the present work focused on the monitoring of the coagulation process, one of the most critical moments of cheesemaking. To this aim, an FT-NIR spectroscopy system was used, acquiring spectra along the rennet coagulation process. According to a Box-Behnken experimental design, several coagulation trials were carried out, changing crucial technological factors, such as temperature, fat content and pH. Through Multivariate Curve Resolution optimized by Alternating Least Squares (MCR-ALS) algorithm it was possible to both have a reliable description of the three different coagulation phases and, most importantly, to build Multivariate Statistical Process Control (MSPC) charts, able to detect failures from the first moment of the process. Moreover, ANOVA-Simultaneous Component Analysis (ASCA) method was applied on spectral data to obtain a better understanding of the process, highlighting in which way each physicochemical parameter affects the process.

In the second part of the work, FT-NIR spectroscopy was tested as a possible tool to replace the golden standards of coagulation ability, i.e. Formagraph. Coagulation trials were carried out using different milk powder samples. The use of MCR-ALS algorithm permitted the assessment of the best powder in terms of coagulation attitude and, in addition, it highlighted the non-significant effect on coagulation occurrence of $CaCl_2$ concentration and of heat treatment on reconstituted milk. Finally, experimental trials carried out with mixtures of skimmed milk and reconstituted milk showed a slower coagulation time when a higher reconstituted milk percentage was used.

The last part of the work regarded the use of MIR spectroscopy to monitor Galactooligosaccharides (GOS) production from cheese whey, in order to avoid the waste of this compound and to optimize the studied process. To do so, Partial Least Square (PLS) regression was used to predict the specific compounds resultant from the different enzymatic reaction studied.

In conclusion, the application of the proposed methods will implicate, with a modest environmental impact, an efficient control of the process, satisfying at the same time law requirements and consumers' needs. Furthermore, reliability of PAT approaches could be strengthened by future industrial applications.

Riassunto

Questo lavoro di tesi risponde al bisogno dell'industria lattiero-casearia di incrementare la produttività e allo stesso tempo soddisfare la richiesta dei consumatori di prodotti di elevata qualità. A tal fine, si possono proporre alle aziende lattiero-casearie nuovi metodi per migliorare la comprensione e il monitoraggio dei processi di produzione. La *Process Analytical Technology* (PAT) rappresenta uno strumento ideale per raggiungere questo scopo, grazie a sensori in grado di eseguire analisi rapide, *green*, non distruttive ed in tempo reale. Le tecniche più utilizzate in questo campo sono la spettroscopia del vicino e del medio infrarosso (NIR e MIR, rispettivamente), che forniscono informazioni chimico-fisiche sul prodotto grazie a sonde installate direttamente in punti critici del processo. Tuttavia, queste tecniche hanno lo svantaggio di fornire risultati (spettri) difficilmente interpretabili senza l'aiuto di un adeguato metodo statistico. In questo contesto, gli algoritmi Chemiometrici permettono l'estrazione di informazioni rilevanti dai dati spettroscopici, permettendo la comprensione del sistema studiato.

La prima parte del presente lavoro è focalizzata sul monitoraggio del processo di coagulazione, uno dei momenti più critici della caseificazione. A tal fine, è stata utilizzata una sonda FT-NIR per acquisire spettri durante il processo di coagulazione. variando alcuni fattori tecnologici cruciali, come temperatura, contenuto di grasso e pH, secondo un disegno Box-Behnken. Attraverso l'algoritmo *Multivariate Curve Resolution - Alternating Least Squares* (MCR-ALS) è stato possibile ottenere sia una efficiente descrizione delle tre differenti fasi del processo di coagulazione, sia lo sviluppo di carte di controllo multivariate (*Multivariate Statistical Process Control charts*, MSPC), capaci di individuare possibili non-conformità fin dai primi momenti del processo. Inoltre, il metodo *ANOVA-Simultaneous Component Analysis* (ASCA) è stato applicato ai dati spettrali al fine di ottenere una migliore comprensione della coagulazione, evidenziando in che modo ogni fattore sperimentale influenzi il processo.

Nella seconda parte del lavoro, la spettroscopia FT-NIR è stata studiata come possibile strumento per sostituire le tecniche standard, come il Formagraph, per valutare l'attitudine alla coagulazione. Le prove di coagulazione sono state effettuate usando differenti campioni di latte in polvere. L'utilizzo dell'algoritmo MCR-ALS ha permesso la valutazione della miglior polvere in termini di attitudine alla coagulazione e, inoltre, ha evidenziato la non significatività degli effetti della concentrazione di $CaCl_2$ e del trattamento termico del latte ricostituito sul tempo di coagulazione. Infine, le prove sperimentali eseguite con miscele di latte scremato e percentuali più elevate di latte ricostituito hanno mostrato una coagulazione più lenta.

L'ultima parte del lavoro ha riguardato l'utilizzo della spettroscopia MIR per monitorare la produzione di galattooligosaccaridi (GOS) dal siero di formaggio, allo scopo di valorizzare questo prodotto e di ottimizzare il processo. La regressione Partial Least Square (PLS) è stata utilizzata con l'obiettivo di predire le componenti specifiche derivanti dalle differenti reazioni enzimatiche studiate.

In conclusione, l'applicazione dei metodi proposti permetterà un efficiente controllo del processo garantendo un modesto impatto ambientale e soddisfacendo allo stesso tempo requisiti di legge e esigenze dei consumatori. Infine, l'affidabilità degli approcci PAT può essere rafforzata da future applicazioni industriali.

2 – Introduction

2.1 – Process Analytical Technology *(Banga et al., 2003; Parfitt et al., 2010; van de Berg et al., 2013)*

The Process Analytical Technology (PAT) concept is based on the idea that "quality could not be tested into final products but it should be build-in or should be by design", especially by "analyzing and controlling the manufacturing processes by real-time measurements of critical quality parameters and performance attributes of raw materials and processes to assure acceptable end products quality", as stated by the United States Food and Drug Administration (FDA) on 2004. Although these applications were thought for pharmaceutical industry, they can easily fit to all kind of processes, including food production ones. Therefore, PAT-based approaches answer the need of industrial companies to have smart solutions for the demands of regulatory organizations, such as the increase of products safety and quality, as well as the reduction of foodstuff environmental footprint. In fact, monitoring processes and assessing quality of the products in real time along the food chain are some of the most significant concerns for the today's food industrial sector. Furthermore, companies want to find efficient, reliable, but at the same time cheap, green and fast technologies that allow to decrease the occurrence of malfunctions in the production lines and to ensure high quality final products. The Process Analyzers (PA), i.e. sensors that allow data collection, are placed in critical steps/positions all along the process, performing measurements and providing results in real time. These datapoint are immediately projected in a prediction model, previously created using measurements acquired at optimal operative conditions, and if a measured point does not fit properly in the model, operators can immediately understand and control the problem. Some of the most used PA are Infrared spectroscopic probes, thanks to their ability to give both chemical and physical characterization of the sample without preprocessing, being at the same time fast, non-destructive and non-invasive. Infrared (IR) and Near-

Infrared (NIR) spectroscopy fundamentals are described in the following chapter.

References

Banga, J. R., Balsa-Canto, E., Moles, C. G., Alonso, A. A., 2003. Improving food processing using modern optimization methods. Trends Food Sci. Tech., 14(4), 131-144.

van den Berg, F., Lyndgaard, C. B., Sørensen, K. M., Engelsen, S. B., 2013. Process analytical technology in the food industry. Trends Food Sci. Tech., 31(1), 27-35.

Parfitt, J., Barthel, M., Macnaughton, S., 2010. Food waste within food supply chains: quantification and potential for change to 2050. Philos. Trans. R. Soc. B., 365(1554), 3065-3081.

2.2 – Spectroscopy *(Miller, 2001; Sandorfy et al., 2007; Stuart, 2004; Wilson et al., 1980)*

The term "spectroscopy" indicates the science field that deals with the measurement and the interpretation of the light emitted or absorbed by a sample. In fact, the study of the interactions between radiations and matter allows to obtain physical and chemical information about the nature and the characteristics of a sample. The electromagnetic radiation can be described as a perturbation of both an electric and a magnetic oscillatory field, perpendicular to each other, as showed in Figure 1.

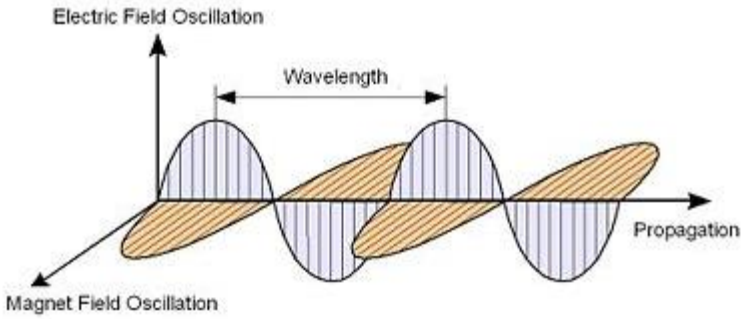


Figure 1: Representation of an electromagnetic radiation (reproduced from https://depts.washington.edu)

This radiation is described by two parameters: periodicity and amplitude. The periodicity can be expressed as its value of wavelength (λ), i.e. the distance between two wave points having the same phase, or frequency (ν), the number of wave cycles per unit distance. The wavenumber is the reciprocal of the wavelength, as reported in Equation 1, and it is expressed by Hertz (cm^{-1}).

Equation 1 $$\lambda = \frac{c}{\nu}$$

Where c is the speed of light in vacuum. So, being c a constant (2.9979×10^{10} cm/s), it is possible to calculate one parameter knowing the other one. Up to

now, the electromagnetic radiation has been described as a wave but, interacting with the matter, it reveals its particle nature. In fact, the energy transmission between radiation and matter is not continuous, but it is discrete, as a stream of particles (photons). When the interaction starts, an atom or a molecule can absorb or emit a certain amount of energy, changing between its ground state (*j*) and an excited state (*k*). This occurs only if the frequency of the radiation *v* matches the energy difference (Δ*E*) between the two states (resonance), as stated by Equation 2:

Equation 2 $$\mathbf{\Delta E} = \boldsymbol{E_j} - \boldsymbol{E_k} = \boldsymbol{h}\,\boldsymbol{\nu} = \boldsymbol{h}\,\frac{\boldsymbol{c}}{\boldsymbol{\lambda}}$$

where *h* is the Planck constant (6.62×10^{-34} J s^{-1}) and E is the energy (J) transported by a photon. So, the energy of a radiation is directly proportional to its frequency, and inversely proportional to its wavelength. Furthermore, the energy of each state is different for each atom or molecule, and may be affected by the environment.

The whole electromagnetic spectrum is conventionally partitioned into smaller regions according to wavelength/frequency values, as shown in Figure 2.

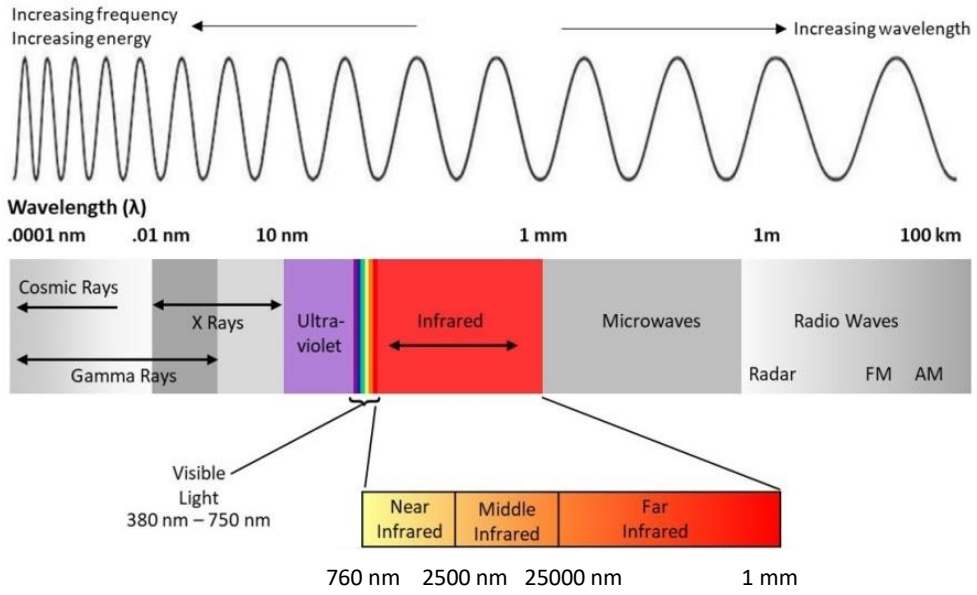


Figure 2: Description of the different region of the electromagnetic spectrum (reproduced from https://www.azooptics.com).

These regions are Gamma-rays, X-rays, Ultraviolet, Visible, Infrared, Microwaves and Radio waves. Each radiation from different spectral regions has a different kind of interaction with the matter, explaining different properties. The infrared region, whose spectral range goes from 12500 to 10 cm^{-1}, can be further divided into three subregions: near infrared (12500 to 4000 cm^{-1}), mid infrared (4000 to 400 cm^{-1}) and far infrared (400 to 10 cm^{-1}). In the next Chapter, the basic principles of Infrared spectroscopy are explained, as well as a focus on Near- and Mid-Infrared spectroscopies, the main techniques used in the present study.

2.2.1 – The basic principle of Vibrational Spectroscopy

As stated before, a photon can transfer its energy to a molecule only if this energy (frequency) is exactly equal to the one necessary to the molecule to pass from the ground to an excited state. When this occurs, one of the electrons in the equilibrium state gains the possibility to move to the higher level, and also the photon is absorbed by the molecule. Then, the electron

returns to its basic state emitting the photon having an amount of energy equal to the difference of energy between the two levels, as showed in Figure 3.

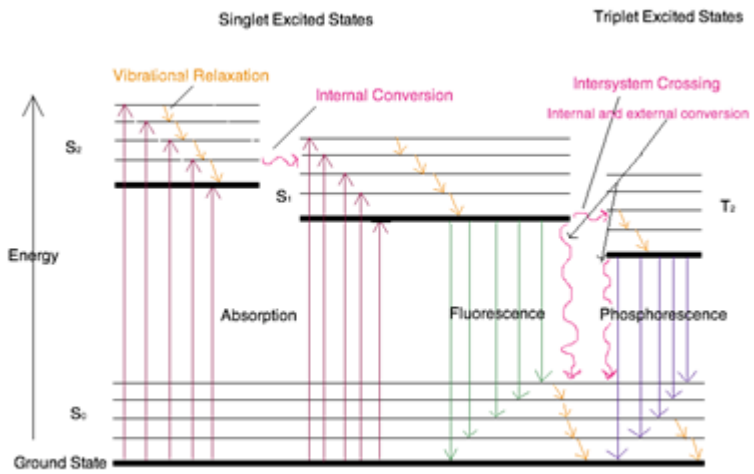


Figure 3: Schematic representation of the different molecular energy levels (reproduced from http://chemwiki.ucdavis.edu).

Obviously, protons with high energy can allow a higher jump of the electron, bringing it to upper levels of excitation. The Exclusion Principle of Pauli imposes the selection rules, that forbid more than two electrons to share the same orbit. Therefore, not all the transitions between the different levels are possible. An infrared radiation is not able to excite the molecule, but its energy is high enough to induce oscillatory motions of the electrons, also called vibrations. In general, the total amount of a molecule's energy can be obtained by adding the contributions of transitional, rotational, vibrational and electronic energies, as reported in Equation 3:

Equation 3

$$E_{tot} = E_{trans} + E_{rot} + E_{vib} + E_{el}$$

In the molecular spectra, for each electronic level are possible various vibrational levels, as well as several rotational states are possible in each vibrational state. Moreover, in the case of Near-Infrared, are present also

combinations of these level and overtones, i.e. summation and multiples of the fundamental vibrational frequencies of absorption, respectively. It is possible to explain the concept of vibrations through the harmonic oscillator model, where two balls are attached to a spring (Figure 4).

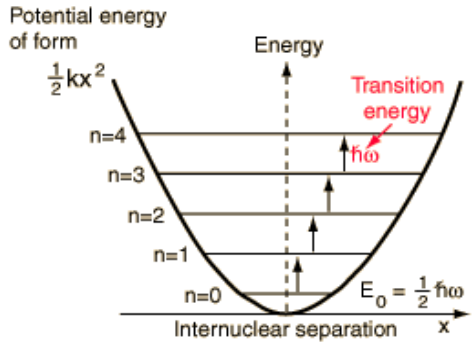


Figure 4: Schematic description of the harmonic oscillator model (reproduced from http://hyperphysics.phy-astr.gsu.edu).

In our case the balls represent the atoms and the spring represents their bond; this system it is described by the Hooke's Law (Equation 4):

Equation 4

$$E = \frac{hc}{2\pi} = \sqrt{\frac{k}{\mu}}$$

where *k* is the force constant of the bond and μ is the reduced mass (Equation 5).

Equation 5

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

In this equation, m_1 and m_2 represent the two atoms masses. The potential energy (*V*) of this kind of model is assumed to be the quadratic function of the dislocations of the considered atoms, according to Equation 6:

Equation 6

$$V = \frac{1}{2} k x^2$$

where *k* is the restoring force constant and *x* is the dislocation from the atoms' equilibrium position. However, a quantum theory model is required to assess

the possible energy levels for a specific vibration. This model, described by Equation 7, shows that the vibrational levels are a group of quantized energy levels:

Equation 7 $$E_v = \left(\omega + \frac{1}{2}\right) h\nu$$

where E_v indicates the energy of the quantum level of a specific vibration, ω is the vibrational quantum number and ν is the vibration frequency. In general, only vibrations that change the dipolar electric motion of the molecule can absorb the infrared radiation. In Figure 5 are shown the main types of bond vibration in a non-linear diatomic molecule, i.e. stretching (vibration along the plane) and bending (variation of the angle between two atoms). Stretching vibration can be in-phase, or symmetrical, and out-of-phase, or anti-symmetrical, whereas bending vibrations present four different typical patterns: two in the plane (rocking and scissoring) and two out of plane (wagging and twisting). All these vibrations have different frequencies.

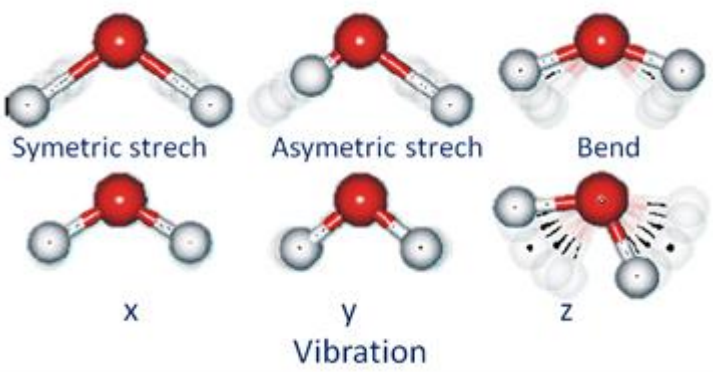


Figure 5: Graphical description of the different vibration patterns of a non-linear diatomic molecule.

Nonetheless, this model fails to efficiently describe all the possible energy transitions that take place in a molecule containing a significant number of atoms, especially if not organized symmetrically. In fact, this approximation works well just with diatomic molecules, but normally we deal with asymmetric

diatomic molecules. This cause the system to be far from the optimal condition described so far, due to the mechanical anharmonicity, that is the non-equidistance between the different energy levels, and anharmonicity of electricity, i.e. the change of the moment dipolar electric equation. In Figure 6 it is represented an anharmonic oscillatory model.

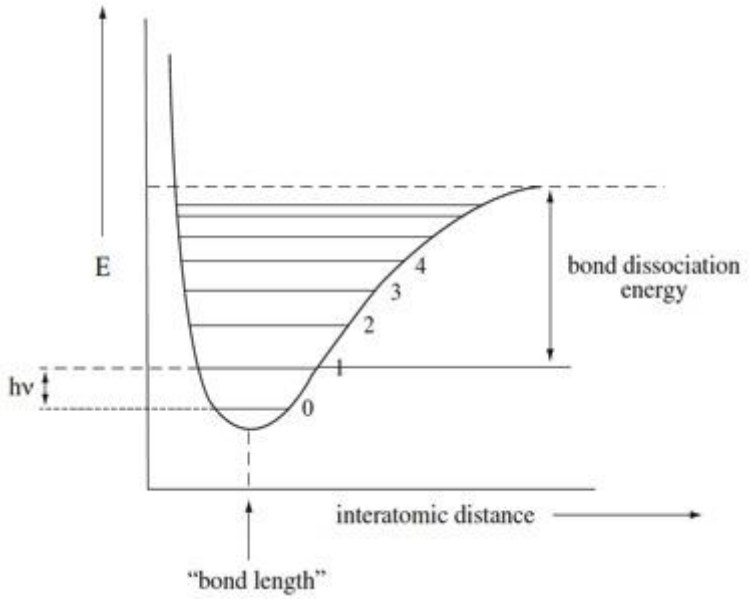


Figure 6: Representation of the anharmonic oscillatory model (reproduced from https://serc.carleton.edu).

As a result, these phenomena lead to anharmonic bands (overtones), whose frequency is not an integer multiple of the fundamental frequency with which oscillates the bond dipole of the molecule. Obviously, the higher are the mutual influences between atoms in a molecule, the more evident are these phenomena. When the difference between the masses of two atoms is high, the vibrations have a great amplitude and, consequently, a more intense absorption band. Different bonds have unique vibrational frequencies, thus the absorption frequency in an Infrared spectrum is used to detect the presence of different bonds and, therefore, permits the discovery of the compounds present in a sample.

Infrared spectroscopy can provide both qualitative and quantitative information through the ability of samples to absorb light in a specific region of the electromagnetic spectrum. NIR spectroscopy analyzes the absorption of radiation in the 12500 – 4000 cm^{-1} region, whereas MIR spectroscopy refers to the 4000 - 400 cm^{-1} region. The spectra obtained from these techniques allow the assessing of an organic compound, because of its individual fingerprint. Furthermore, an analyte can be almost always quantified even in a complex mixture, thanks to the high selectivity of these methods.

2.2.2 Near-Infrared Spectroscopy

Instrumentation

In this section are described the characteristics and the features of a Fourier-transform NIR (FT-NIR) spectrometer, the instrument used in all the NIR applications presented in this thesis.

The most important part of an FT-NIR spectrometer is the interferometer, composed by a beam splitter, a stationary mirror and a mirror moving back and forth, which forms a 90° angle with the first. The light coming from the source is divided in two beams by the beam splitter: one is reflected while the other one exceeds the splitter. Then, both the beams are reflected by the mirrors to the splitters and pass through it, as showed in Figure 7. The change of the mirrors distance from the splitter causes a continuous modification in the optical path difference between the two beams, creating interference in the radiation. From this procedure the interferogram is obtained, i.e. the intensity as a function of the mirror shift. All the spectral information contained in the interferogram is then recovered applying a Fourier transformation.

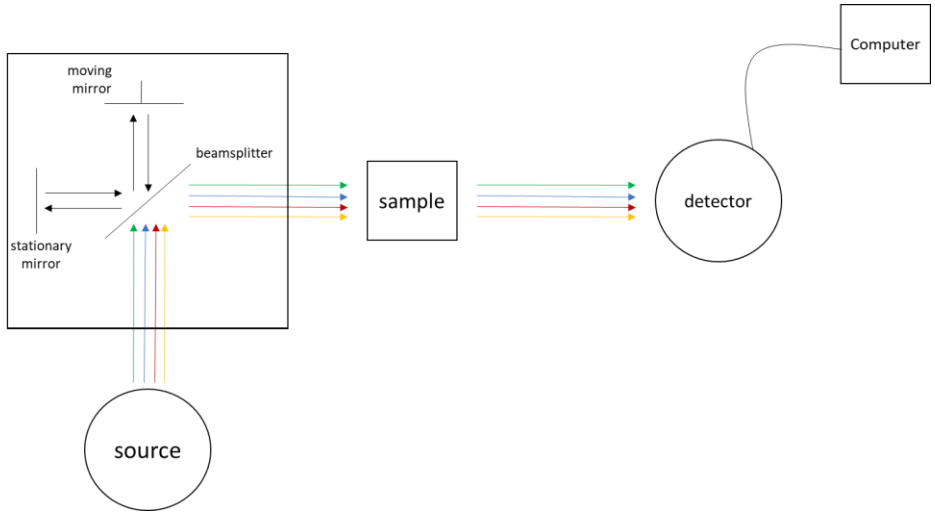


Figure 7. Scheme of an FT-NIR spectrometer.

Sample interface and measurements modes

The approaches to perform a measurement using an NIR instrument depends on the nature of the sample. Transmission mode (Figure 8, left) is often used to measure liquid and gasses, even if can be also used for reflecting and scattering samples. Using this mode, the NIR radiation is directed to the sample with a parallel or focused beam. A portion of the radiation is absorbed by the sample, while the rest is transmitted to the detector, as shown in the left part of Figure 8.

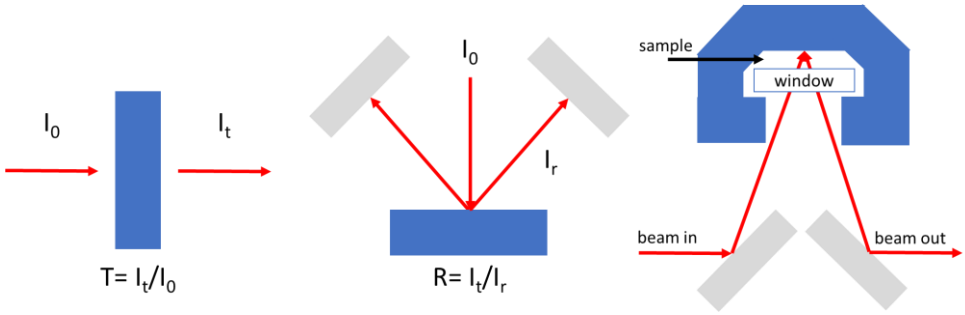


Figure 8. Schematic representation of transmission, reflection and transflection modes.

Reflection mode is used for solid surfaces, powders, pallets and granulates. When the incident light hits the sample surface, it is reflected with a different angle (generally between 5 and 85 degrees) than the incident one (middle part of Figure 8). The third approach is called transflection and it is a combination of transimission and reflection modes, often used for gels, turbid liquids and emulsions. A mirror is placed in the light path, the NIR radiation transmitted through the sample is reflected by the mirror at the same or almost the same angle as the incident beam and it goes back to the detector (right part of Figure 8).

To achieve an efficient process control it is fundamental to have an optimal timing of process measurements. NIR analysis can be performed off-line, at-line, on-line and in-line. The off-line measurements are made in the laboratory, taking the sample from the process line and for analysing it in a second moment. However, if the instrument is placed close to the process line, it is possible to perform an at-line analysis. These two strategies of analysis requires a manual sampling by the operators. An on-line analysis allows to have an automatic sampling, as a NIR probe is placed directly in the production line (in-line analysis) or in a parallel line, where the samples are deviated, analysed and reintroduced in the principal line (on-line analysis).

Spectra interpretation

Combinations and overtones of fundamental vibrations of molecules containing -CH, -OH, -NH groups are the elements that characterize a near infrared spectrum. These groups are contained in molecules of food products as constituent of protein, sugar, fat and moisture. A schematic representation of combination bands and overtones absorptions in the NIR region are showed in Figure 9. Thus, changes in bands position and intensity could be associated to changes in nature and quantity of food constituents.

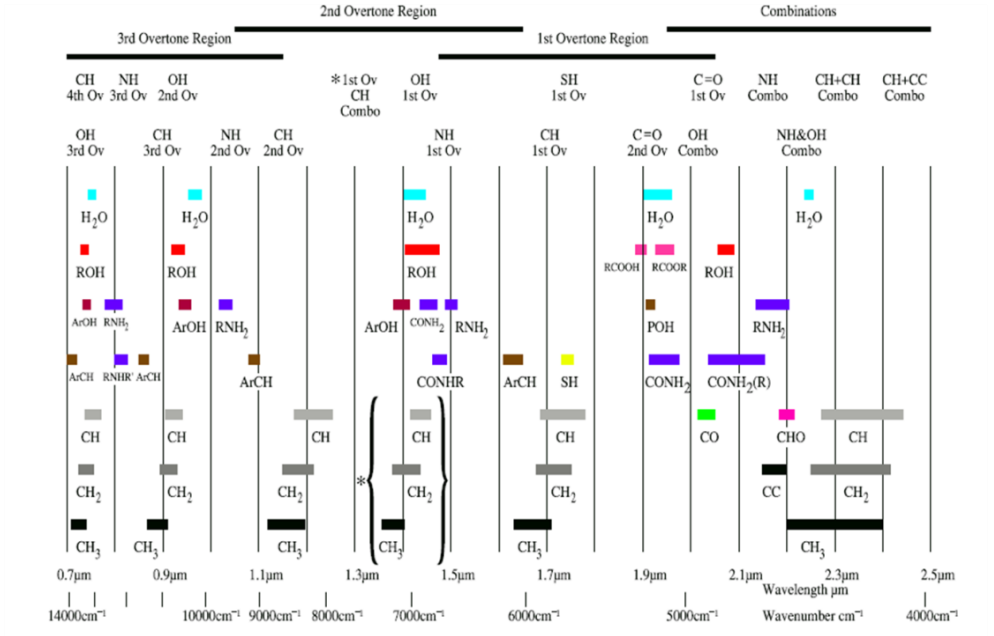


Figure 9. Schematic representation of overtone and combination bands absorptions in the NIR region (reproduced from Harris, P. J., & Altaner, C. M., 2013).

The NIR spectra not only provides a chemical characterization, but can also extract information about physical properties of the sample. Thanks to the scattering variation, it is possible to obtain information about particle size and shape, as well as sample surface. The scattering effect plays a crucial role in the coagulation process, as showed in Papers I, II and III, where occurs the increase of the number and the size of casein micelles. Mainly what is observed in the spectra is a change in the baseline, and in co-linearity phenomena among the wavenumbers

2.2.3 Mid-Infrared Spectroscopy

Instrumentation

The spectrometer used in Paper IV was a FT-IR spectrometer, whose structure was already described in section 2.2.2.1. Briefly, FT-IR spectroscopy has a higher sensitivity in respect of FT-NIR spectroscopy,

because bands derive mainly from fundamental vibrations and they are less overlapped in respect to NIR spectroscopy.

Sample interface and measurement modes

IR spectroscopy is very sensible to how the sample is presented to the instrument. This is one of the most critical disadvantages of IR spectroscopy with respect to NIR spectroscopy.

Transmission mode, described previously, can be used for the analysis of different kind of samples, such as gasses, powders, liquids, and pastes. Due to the high IR absorptivity of water the optical path length, and consequently the sample, must be very thin. For this reason, transmission today is mostly used to analyse gas samples, where the path length can be higher. These preparations will not be reported in detailed as all the problems related to transmission windows have been solved by the implementation of attenuated total reflectance (ATR) cells. For solid and liquid samples are commonly used the ATR cells, where the light is guided in a transparent IR crystal by total reflection. The IR beam enters the ATR crystal and it is reflected at the sample-element interface, and multiple reflections occurs thanks to the angle of incidence of the accessory, as shown in Figure 10.

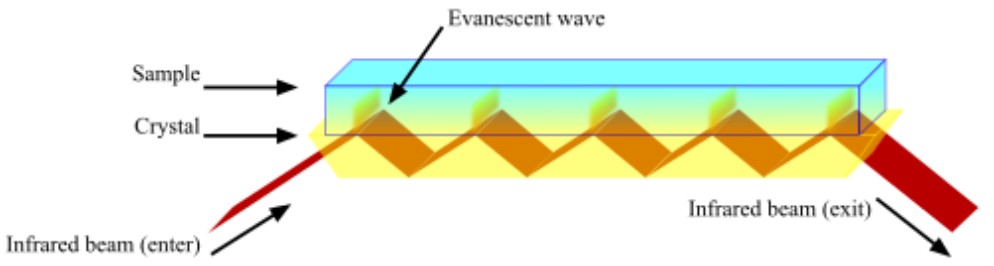


Figure 10. Schematic representation of an ATR cell (reproduced from http://sites.science.oregonstate.edu).

During each reflection, radiation slightly penetrates the sample and it decays logarithmically as a wave into the sample medium. In its way, the energy is reduced by the part absorbed by the sample.

Spectra interpretation

The mid infrared spectrum is characterized by the absorption owing to fundamental rotational molecular vibrations, i.e. bending and stretching. The fingerprint region (1500 – 1000 cm^{-1}) is crucial to delineate the characteristics of the analysed sample. IR spectroscopy is generally used for the identification of molecular structures, as IR bands can be ascribed to functional groups in terms of both intensity and frequency, as described in Figure 11. The mass and the bond strength are directly related to the vibrational energy, allowing the identification of specific chemical species.

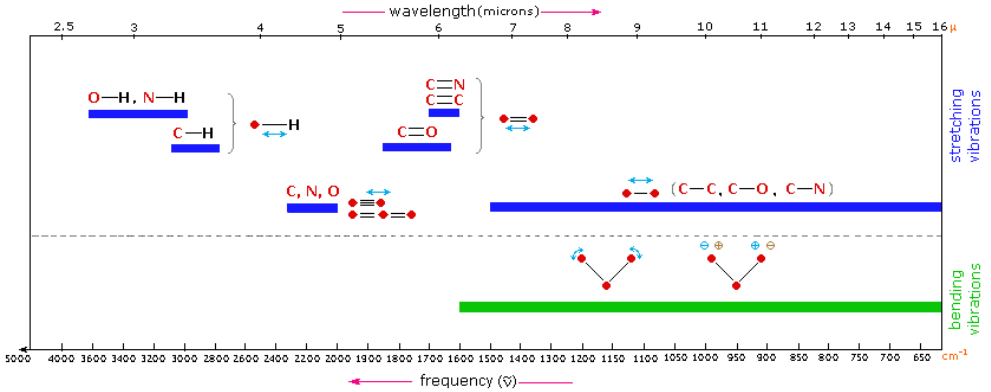


Figure 11. Schematic representation of vibration bands absorptions in the mid IR region (reproduced from https://www2.chemistry.msu.edu).

In these chapters a detailed description of both NIR and MIR spectroscopies highlighted the benefits but also the limits of these techniques. In fact, the obtained spectra can be difficult to interpret without a proper approach. Therefore, the next chapter introduces the concept of Chemometrics, a set of techniques that allows to extract the relevant information from this kind of data, minimizing the effects of instrumental noise and spectral artifacts.

References

Harris, P. J., Altaner, C. M., 2013. Workshop on commercial application of IR spectroscopies to solid wood.

Miller, C. E., 2001. Chemical principles of near infrared technology. In: Near-infrared technology in the agricultural and food industries, 2 .American Association of Cereal Chemists. Minnesota, USA.

Sandofy, C., Buchet, R., Lachenal, G., 2006 In: Ozaki, Y., McClure, W. F., & Christy, A. A. (Eds.). Near-infrared spectroscopy in food science and technology. Wiley.

Stuart, B. H., 2004. Infrared spectroscopy: fundamentals and applications. Wiley. Chichester, UK.

Wilson, E. B., Decius J.C., Cross P.C., 1980. Molecular vibrations: the theory of infrared and Raman vibrational spectra. Dover Publications Inc.

2.3 - Chemometrics

Chemometrics is a chemistry sector that studies the applications of mathematical and statistical methods to chemical data. The official definition of the International Chemometrics Society (ICS) is "a chemical discipline that uses mathematical and statistical methods to: design/select optimal procedures and experiments, provide maximum chemical information by analyzing data, give a graphical representation of this information, in other words information aspects of chemistry". The aims of chemometrics focus on the design and optimization of procedures and experiments, on extracting relevant information from the studied systems and on providing a graphic and more clear visualization of the results. To this purpose, chemometric methods represent the perfect tool to handle spectroscopic data, extracting relevant information from NIR and MIR spectra. In order to perform an exploratory data analysis, it is commonly used Principal Component Analysis (PCA), one of the most “famous” chemometric tools. For the compounds quantification, chemometrics utilizes multivariate regression methods, such as Principal Component Regression (PCR) and Partial Least Square (PLS) regression. Another big sector of chemometrics regards classification, that comprehends both discriminant and modelling methods, such as Partial Least Square – Discriminant Analysis (PLS-DA) and Soft Independent Modelling of Class Analogies (SIMCA). All these methods require a priori information, numerical in case of regression and categorical in case of classification. For the study and the modelling of complex mixtures one of the most used methods is Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS), whereas ANOVA-Simultaneous Component Analysis (ASCA) method is often used for the assessment of the influence of process factors in a multivariate way.

In this chapter are described in detail the chemometric methods that have been applied for the achievement of this thesis aims.

2.3.1 – Principal Component Analysis *(Jackson, 1981; Wold et al., 1987)*

Principal Component Analysis (PCA) is a multivariate tool that allows to detect and evaluate similarities, differences, outliers and cluster tendency on data matrices. It transforms the correlated variables of the data into a fewer number of independent and uncorrelated variables called Principal Components (PCs), orthogonal to each other. The first PC is constructed through detecting among the infinite directions of the *n*-dimensional space (where *n* is the number of variables), the one that involves the largest variability in the data. The second PC is constructed orthogonally to the first, incorporating the largest source of variability not expressed from the first one, and so on. Technically, the maximum PCs amount is equal to *n*, but practically, only the first components include the useful information, as variables co-vary, especially in spectroscopic data, and further PCs only contain random noise. Using this approach it is possible to create a new projection space, smaller in size, whose coordinates are represented by the PCs. In particular, each PC is composed by the product of two vectors: the scores vector (t_i), and the loadings vector (p_i), where *i* represents the number of components (Equation 8):

Equation 8 $$\boldsymbol{X = t_1 p_1^T + t_2 p_2^T + \cdots + t_i p_i^T + E = TP^T + \mathbf{E}}$$

X is a data matrix composed by *m* rows (observations) and *n* columns (variables), whereas **E** is the residual matrix, that contains the unmodelled variation and has the same dimensions of **X**. The scores matrix **T** contains information on how each sample relates to each other, whereas the loading matrix **P** expresses the influence of the measured variables on the scores. PCs number *i* defines the amount of variation in the data, i.e. the independent phenomena. In Figure 12 it is shown the PCA decomposition scheme.

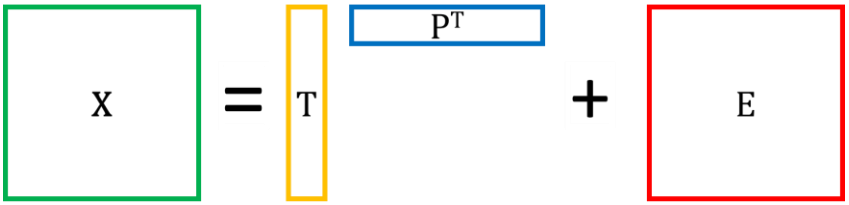


Figure 12: PCA decomposition of **X.**

Mathematically, components are calculated by an eigenvalue decomposition of the covariance matrix (Equation 9 and Equation 10):

Equation 9 $$cov(X) = \frac{X^T X}{m-1}$$

Equation 10 $$cov(X)p_i = p_i\lambda_i$$

where λ_i are the eigenvalues related to the eigenvectors p_i. The scores vectors t_i are the projections of the data matrix **X** onto p_i. T^2 and Q parameters are used to evaluate PCA models. T^2 statistic represents the distance of a sample in the model space (i.e. in the space of significant PCs), whereas Q statistic represents the distance of a sample from the model space, meaning that the PCA model cannot describe efficiently their variability (Figure 13). These two parameters are extremely useful for outliers detection.

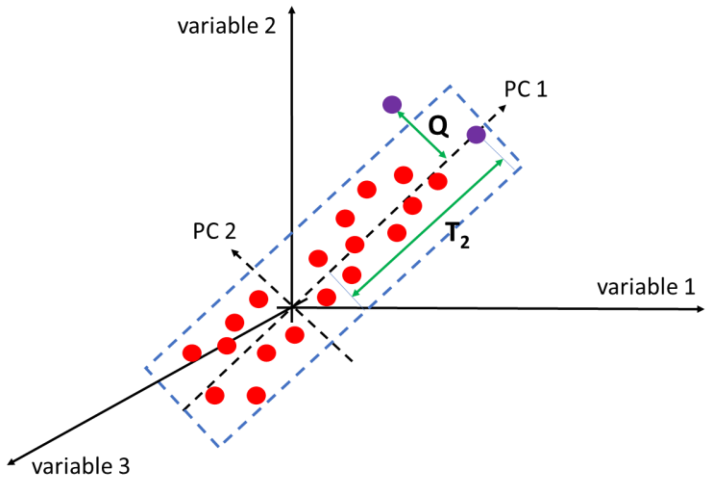


Figure 13: Graphical representation of T^2 and Q statistics.

2.3.2 – PCA-based Multivariate Statistical Process Control Charts *(Chen et al., 2004; Kourti, 2006)*

The PCA-based Multivariate Statistical Control (MSPC) charts are a powerful tool for the real time monitoring of production processes. These charts simplify the operators work, thanks to the PCA ability of decreasing the dimensionality of a dataset, reducing the number of variables and consequently the number of charts an operator should control. To build MSPC charts, a model is computed by using reference samples, i.e. samples that represent the optimal/normal operative conditions; then, new samples are projected in the model to assess how they fit in it. However, using the raw PCs, plotting them one against the other, could be confusing, as the number of charts would still be high, even if the visualization of the process behaviour would be easier. For this reason, MSPC charts are rendered using T^2 and Q statistics, introduced in the previous section. Each measurement, acquired from the process analyzers at each moment, is transferred into the two charts: these two points describe the process at that precise instant, providing a picture of the system wellness. If the points are below the confident limit of each charts, everything is in control; instead, if a point falls outside these confidence limits, an abnormality occurs, indicating that the process is not in its optimal conditions anymore (Figure 14). T^2 and Q can be calculated according to the following equations:

Equation 11
$$\boldsymbol{T_k^2 = \sum_{i=1}^{N} \frac{t_{ki}^2}{\sigma_i}}$$

Equation 12
$$\boldsymbol{Q_k = \sum_{j=1}^{N} t_{kj}^2}$$

where $\mathbf{T^2_k}$ and $\mathbf{Q_k}$ are values of T^2 and Q statistics related to samples collected at time instance **k**, t_{ki} and t_{kj} represent the values of i^{th} and j^{th} score variables at time instance **k**, respectively, σ_i is the estimated variance of the i^{th} score variable and **N** is the number of PCs.

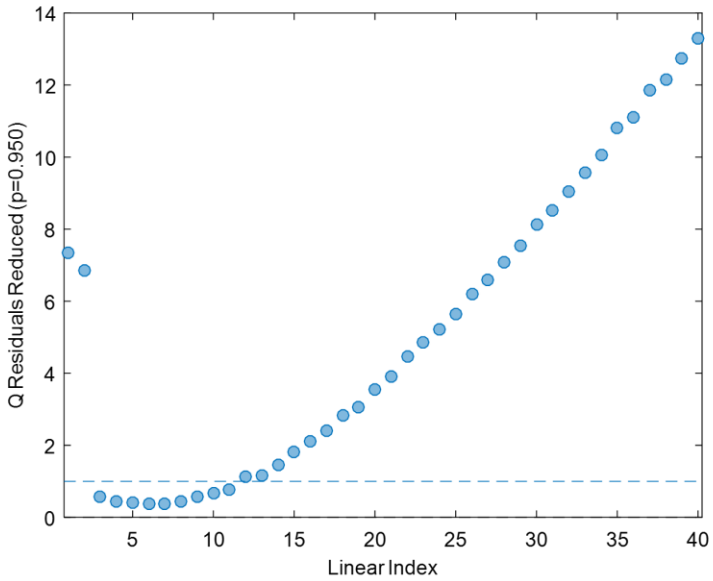


Figure 14: example of a Q statistic control chart. The first two samples and samples after the eleventh are above the confidence limit, suggesting that the process is not stable.

2.3.3 – Partial Least Square Regression *(Geladi & Kowalski, 1986; Wold et al., 1983)*

PLS method is one of the most used chemometric tools to evaluate the correlation between two data matrices performing a multivariate linear regression. The advantage of using this method comes from its ability to deal with data that have many more variables than samples, especially when these variables co-vary (e.g. spectral data). Furthermore, it is able to handle data matrices that are incomplete and with many noisy variables present in each of the **X** (descriptors matrix) and **Y** (response matrix) blocks. PLS operates a simultaneous decomposition of both **X** and **Y** matrices in order to explain as much as possible of the variability of **X** and to find the best correlation with **Y**. Its aim is to maximize the covariance between the two matrices, creating at the same time latent variables that describe the maximum variability of **X**. The

algorithm performs the decomposition of **X** and **Y** as in the case of PCA (Equation 13 and Equation 14:

Equation 13 $$\boldsymbol{X = TP^T + E}$$

Equation 14 $$\boldsymbol{Y = UQ^T + R}$$

The maximum covariance criterion is imposed through a regression model for each component between the scores of **X** (t_i) and **Y** (u_i), obtaining their inner relation (Equation 15 and Equation 16):

Equation 15 $$\boldsymbol{u_i = b_i t_i}$$

Equation 16 $$\boldsymbol{b_i = \frac{u_i^T t_i}{t_i^T t_i}}$$

where *b* is the regression coefficient related to the i^{th} component. However, this is not the best possible strategy, as the components are calculated separately for each block, resulting in a weak relation between them. For this reason, inner relation is improved by rotating the components, which means making *t* and *u* switch places in the NIPALS algorithm. As final step, scores are orthogonalized by introducing loading weights, **W**, which are orthonormal, calculated according to Equation 17:

Equation 17 $$\boldsymbol{W = T^T X}$$

Scores and loadings have the same properties described for PCA, but in PLS the latent variables (LV) explain the variability of **X** that most influences the responses predictions in **Y**. The choice of LVs number is critical: on the one hand is essential to include all the useful information to obtain better results, on the other hand LVs that contain only noise and irrelevant information should not be included. In fact, choosing to many LVs leads to the creation of a model that is perfectly suited to explain the variability of the data used to create it, even managing to represent its noise, but it would be hardly adaptable for the prediction of unknown samples, thus resulting not functional. The most used method for the determination of the optimal number

of LVs is the cross-validation. Once the model is created, it must be validated, i.e. subjected to a test that demonstrate its reliability. For this purpose, it is strongly recommended to use samples that were not used in the calculation of the model itself, in order to avoid an overly favorable estimate of the error. The previously generated prediction algorithm is applied on this new set of samples (test set), whose characteristics are known, in order to evaluate the predictive capability in terms of error on the real data.

2.3.4 – Multivariate Curve Resolution – Alternating Least Squares *(De Juan & Tauler, 2006; Tauler et al., 1995)*

Multivariate Curve Resolution is a mathematical method used to resolve complex mixtures systems into their pure components, without the assumption of any previous empirical model. As PCA, MCR aims to decompose the data matrix into a bilinear model, but in addition it constrains the profiles of the components to guarantee that the solution makes sense from a chemical point of view, not only statistically. In fact, MCR is based on the validity of the multicomponent version of the Lambert-Beer law, so that the profiles of the pure components acquire a chemical meaning. According to Equation 18, the algorithm decomposes a data matrix **D** (M x N) into a concentration profile matrix **C** (M x F) and a spectral profile matrix $\mathbf{S}^{\mathbf{T}}$ (F x N), as well as a matrix error **E** (M x N).

Equation 18 $$\boldsymbol{D} = \boldsymbol{C}\boldsymbol{S}^{\boldsymbol{T}} + \boldsymbol{E}$$

M is the number of samples or datapoint over time in the dataset, N the number of variables (i.e. wavenumbers) and F the number of components found relevant by the algorithm. In general, matrix **C** defines the variation of the F components influencing the spectral signals in the M different datapoints over time, whereas $\mathbf{S}^{\mathbf{T}}$ describes the response variation related to the different wavenumbers. In Figure 15 it is represented an example of a graphical MCR decomposition of a NIR spectra dataset.

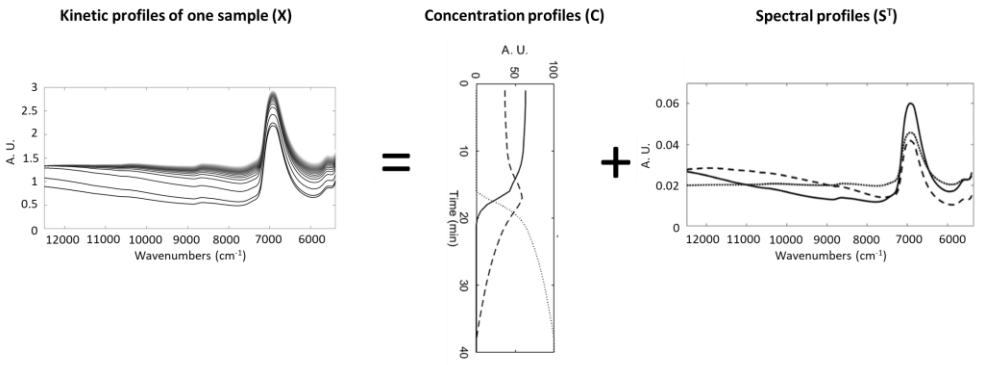


Figure 15: MCR decomposition of a NIR spectra dataset.

However, the first step of the algorithm looks for the optimal number of components to describe the data. Normally, this number is found by PCA algorithm, with the Single Value Decomposition (SVD) method. After this step, it is crucial to choose the proper initial estimates in the concentration or spectral profiles for the model computation. This could be done by choosing the most N different spectra present in the dataset, where N is the number of relevant components. At this point, Alternating Least Squares (ALS) optimization starts the calculation of **C** and $\mathbf{S^T}$ matrices by means of an iterative least square calculation, until convergence is achieved. The criterion imposed to stop the computation is the percentage of Lack of Fit (%LOF). In particular, when the relative difference between two consecutive iterations is equal or less than 0.1%, the algorithm stops the calculation. Equation 19 describes how to determine %LOF:

Equation 19

$$\%\boldsymbol{LOF} = \mathbf{100}\ \boldsymbol{x}\ \sqrt{\frac{\sum_i^I \sum_j^J e_{ij}^2}{\sum_i^I \sum_j^J d_{ij}^2}}$$

where e_{ij} is the ij^{th} component of the residual matrix **E**, whereas d_{ij} is the ij^{th} element of the **D** matrix. As the $\mathbf{CS^T}$ product is exposed to rotational and intensity ambiguities, the model computation can be performed under different constraints, both on concentration and spectral profiles, depending on the knowledge of the systems and on type of analytical data. For spectral

application, it is crucial the use of non-negativity on spectral profiles, as it is known that a negative absorbance value does not have any scientific meaning. It is also common to apply this constrain on concentration profiles, in case they express real concentrations that cannot assume negative values. Other commonly used constraints are unimodality, that can be applied to impose only one maximum in each profile, and closure, used for mass balance conditions. Finally, MCR-ALS method can also be used using different dataset, e.g. set of spectra related to different batches. This technique is called Multiblock, and it is performed by augmenting the D matrix column-wise. In this fashion, the concentration profile matrix C will be composed by C_i submatrices, where *i* is the number of different batches, but the spectral profile matrix S^T will still be a single matrix, describing each of the batches present in the D augmented matrix. This allows the concentration profiles of the different chemical species present in the system to change independently from one batch to the other.

2.3.5 – ANOVA-Simultaneous Component Analysis *(Jansen et al., 2005; Smilde et al., 2005)*

ANOVA-Simultaneous Component Analysis (ASCA) is a chemometric exploratory technique designed to handle multivariate data resulting from an experimental design. As a matter of fact, it can be considered as a multivariate extension of ANOVA, as it allows to evaluate the significance of one or more experimental factors, or their interactions, in a multivariate way, being suitable for spectroscopic data. To do so, ASCA splits the variability contained in the data matrix into the contributions of each controlled factor or interaction and analyze them with Simultaneous Component Analysis (SCA), in order to interpret the results more intuitively. For a given centered data matrix X_c, the partitioning into the individual effect matrices and interaction matrices occurs according to Equation 20:

Equation 20 $$X_c = X - 1m^T = X_1 + X_2 + \cdots + X_i + X_{1x2} + X_{2x3} + \cdots + X_{ixj} + X_{res}$$

where **X** is the original data matrix, $\mathbf{m}^T$ is the samples mean profile, $\mathbf{X}_{(1, 2,\ldots,i)}$ are the matrices linked to the main effects, $\mathbf{X}_{(1x2,\ 2x3,\ldots,ixj)}$ are the matrices corresponding to the interaction effects and $\mathbf{X}_{res}$ is the residual matrix. In particular, for each factor, the rows of the $\mathbf{X}_c$ matrix related to the different levels of the design are averaged, and the obtained mean values are utilized to calculate the effect associated matrix. As an example, if the i^{th} sample matches to the factor 1 first level, the i^{th} row of the $\mathbf{X}_1$ matrix will contain an average profile of all samples for which factor 1 is at first level, and so on. The interaction matrix is calculated similarly, after the subtraction of the main effect matrices. The magnitude of the effects is assessed by the sums of squares of the matrix elements:

Equation 21 $$SSQ_i = |X_i|^2$$

Then, in order to assess if an effect is statistically significant, the sum of square of the corresponding matrix is compared to its distribution under the null hypothesis, evaluated by a permutation test. Then, a SCA is done to perform the bilinear decomposition of each effect matrix, modelling the variability linked to each effect. In this context, i.e. under the constraints of ANOVA, SCA works exactly as a normal PCA. Therefore, the matrices obtained by the ANOVA partitioning are decomposed according to Equation 22:

Equation 22 $$X_i = T_i P_i^T + E_i$$

where $\mathbf{T}_i$ are the scores and P_i are the loading matrices of the i^{th} partition, and E_i the related residual matrix. In this way, it is possible to have a better visualization of the data considering each experimental factor or interaction separately.

References

Chen, Q., Kruger, U., Meronk, M., Leung, A. Y. T., 2004. Synthesis of T2 and Q statistics for process monitoring. Control Eng. Pract., 12(6), 745-755.

De Juan, A., Tauler, R., 2006. Multivariate curve resolution (MCR) from 2000: Progress in concepts & applications. Crit. Rev. Anal. Chem., 36, 163–176

Geladi, P., Kowalski, B. R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta, 185, 1-17.

Jackson, J.E., 1981. Principal component and factor analysis: Part 1 – Principal Component. J. Qual. Tech., 13(1), 46-58.

Jansen, J. J., Hoefsloot, H. C. J., van der Greef, J., Timmerman, M. E., Westerhuis, J. A., Smilde, A. K., 2005. ASCA: analysis of multivariate data obtained from an experimental design. J. Chemom., 19, 469–481.

Kourti, T., 2006. The process analytical technology initiative and multivariate process analysis, monitoring and control. Anal. Bioanal. Chem., 384, 1043-1048.

Smilde, A. K., Jansen, J. J., Hoefsloot, H. C. J., Lamers, R. J. A. N., van der Greef, J., Timmerman, M. E.. 2005. ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. Bioinformatics, 21, 3043–3048.

Tauler, R., Smilde, A., Kowalski, B.R., 1995. Selectivity, Local Rank, Three-Way Data Analysis and Ambiguity in Multivariate Curve Resolution. J. Chemom, 9, 31–58.

Wold, S., Esbensen, K.H., Geladi, P., 1987: Principal component analysis. Chemom. Intell. Lab. Syst., 2, 37-52.

Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method. In Matrix Pencils. Springer. Berlin. Heidelberg, pp. 286-293

2.4 - PAT applications for dairy industry

Dairy industry has a huge relevance in the world's food sector, especially in Italy, where is the first food area and its value of production worth 15 billion Euros. Every year, dairy companies produce 3 million tons of milk, a million tons of different cheeses (460000 are DOP products), more than a billion of yogurt cups and 160000 tons of butter (www.senato.it). Obviously, the production of these manufactures is characterized by several physical-chemical processes, such as lactic fermentation, milk coagulation, protein degradation, lipid oxidation and many more. All these critical steps of the production have to be monitored and controlled, in order to handle the pressure to provide products of high and constant quality into the market.

Even though PAT founds more applications in pharmaceutical or chemical sectors, there are many studies that confirm the reliability and the usefulness of this approach in the food sector, especially in dairy area. The main PA used in these studies are the spectroscopic sensors, especially NIR and MIR ones, thanks to the fiber optic probes they can be equipped with (Cullen et al., 2014). Several studies demonstrate how FT-NIR and FT-MIR spectrometers allow to assess milk composition both at-line and on-line in terms of fat, protein, lactose content and total solids determination (Cattaneo & Holroyd, 2013; Diaz-Olivares et al., 2020; Inglingstad et al., 2016; Iweka et al., 2020; Logan et al., 2015; Katz et al., 2015; Malacarne et al., 2014), successfully replacing wet chemistry. Furthermore, these techniques have been applied for the assessment of quality parameters during cheese ripening (Downey et al., 2004; Skeie et al., 2005), to study shelf-life (Cattaneo et al., 2005) and to monitor lactic fermentation (Navratil et al., 2004; Svedsen et al., 2016). FT-NIR spectroscopic probes have been also used to monitor and control coagulation, one of the most critic stages in the cheesemaking sector. These sensors are used to replace Formagraph, the golden standard in this field (Pretto et al., 2011), as well as other recognized methods for the monitoring of coagulation process, i.e. fundamental and empirical rheology (Nassar et

al., 2020; Salvador et al., 2019). FT-NIR probes have the advantage of being placed directly in the coagulation vat, providing real-time, fast and non-destructive measurements, and allow to skip passages that could lead to errors like sampling and sample pretreatment (Cimander et al., 2002; Laporte et al., 1998; Lyndgaard et al., 2012). Besides, other PA can be used to extract relevant information from dairy industry processes. Through imaging systems, it has been possible to estimate size, shape and problems of homogeneity or consistency of samples very rapidly (Kucheryavskiy, et al., 2014; Kuo, et al., 2003). E-nose and E-tongue techniques have been used to recognize both simple and complex classes of volatile compounds (Kalit et al., 2014) and to provide different responses to chemical compounds based on their taste (Falchero et al., 2009), respectively. According to these studies, a PAT approach, coupled with the proper chemometric tool, can be applied on every type of dairy process or product, providing promising results for industrial-scale applications.

More details are given in the introductions of the attached papers.

References

Cattaneo, T. M. P., Giardina, C., Sinelli, N., Riva, M., Giangiacomo, R., 2005. Application of FT-NIR and FT-IR spectroscopy to study the shelf-life of Crescenza cheese. Int. Dairy J., 15(6-9), 693-700.

Cattaneo, T. M. P., Holroyd, S. E., 2013. The use of near infrared spectroscopy for determination of adulteration and contamination in milk and milk powder: updating knowledge. J. Near Infrared Spectrosc., 21(5), 341-349.

Cimander, C., Carlsson, M., Mandenius, C.F., 2002. Sensor fusion for on-line monitoring of yoghurt fermentation. J. Biotechnol., 99, 237–248.

Cullen, P.J., O'Donnell, C.P., Fagan, C.C., 2014. Benefits & challenges of adopting PAT for the food industry. In: O'Donnell, C.P., Fagan, C., Cullen,

P.J. (Ed.), Process Analytical Technology for the Food Industry. Springer, New York, NY, USA, pp. 1–5.

Diaz-Olivares, J. A., Adriaens, I., Stevens, E., Saeys, W., Aernouts, B., 2020. Online milk composition analysis with an on-farm near-infrared sensor. bioRxiv.

Downey, G., Sheehan, E., Delahunty, C., O'Callaghan, D., Guinee, T., Howard, V., 2005. Prediction of maturity and sensory attributes of Cheddar cheese using near-infrared spectroscopy. Int. Dairy J., 15(6-9), 701-709.

Falchero, L., Sala, G., Gorlier, A., Lombardi, G., Lonati, M., Masoero, G., 2009. Electronic Nose analysis of milk from cows grazing on two different Alpine vegetation types. J. Dairy Res., 76(3), 365.

Inglingstad, R. A., Eknæs, M., Brunborg, L., Mestawet, T., Devold, T. G., Vegarud, G. E., Skeie, S. B., 2016. Norwegian goat milk composition and cheese quality: The influence of lipid supplemented concentrate and lactation stage. Int. Dairy J., 56, 13-21.

Iweka, P., Kawamura, S., Mitani, T., Kawaguchi, T., Koseki, S., 2020. Online Milk Quality Assessment during Milking Using Near-infrared Spectroscopic Sensing System. ECB, 58(1), 1-6.

Kalit, M., Marković, K., Kalit, S., Vahčić, N., Havranek, J., 2014. Application of electronic nose and electronic tongue in the dairy industry. Mljekarstvo: časopis za unaprjeđenje proizvodnje i prerade mlijeka, 64(4), 228-244.

Kucheryavskiy, S., Melenteva, A., Bogomolov, A., 2014. Determination of fat and total protein content in milk using conventional digital imaging. Talanta, 121, 144-152.

Kuo, M. I., Anderson, M. E., Gunasekaran, S., 2003. Determining effects of freezing on pasta filata and non-pasta filata Mozzarella cheeses by nuclear magnetic resonance imaging. J. Dairy Sci., 86(8), 2525-2536.

Laporte, M. F., Martel, R., Paquin, P., 1998. The near-infrared optic probe for monitoring rennet coagulation in cow's milk. Int. Dairy J., 8(7), 659-666.

Logan, A., Leis, A., Day, L., Øiseth, S. K., Puvanenthiran, A., Augustin, M. A., 2015. Rennet gelation properties of milk: Influence of natural variation in milk fat globule size and casein micelle size. Int. Dairy J., 46, 71-77.

Lyndgaard, C. B., Engelsen, S. B., van den Berg, F. W., 2012. Real-time modeling of milk coagulation using in-line near infrared spectroscopy. J. Food Eng., 108(2), 345-352.

Katz, G., Merin, U., Bezman, D., Lavie, S., Lemberskiy-Kuzin, L., Leitner, G., 2016. Real-time evaluation of individual cow milk for higher cheese-milk quality with increased cheese yield. J. Dairy Sci., 99(6), 4178-4187.

Malacarne, M., Franceschi, P., Formaggioni, P., Sandri, S., Mariani, P., Summer, A., 2014. Influence of micellar calcium and phosphorus on rennet coagulation properties of cows milk. J. Dairy Res, 81(2), 129-136.

Nassar, K. S., Lu, J., Pang, X., Ragab, E. S., Yue, Y., Zhang, S., Lv, J., 2020. Rheological and microstructural properties of rennet gel made from caprine milk treated by HP. J. Food Eng., 267, 109710.

Navrátil, M., Cimander, C., Mandenius, C. F., 2004. On-line multisensor monitoring of yogurt and filmjölk fermentations on production scale. J. Agric. Food Chem., 52(3), 415-420.

Pretto, D., Kaart, T., Vallas, M., Jõudu, I., Henno, M., Ancilotto, L., Cassandro, M., Pärna, E., 2011. Relationships between milk coagulation property traits analyzed with different methodologies. J. Dairy Sci., 94(9), 4336-4346.

Salvador, D., Arango, O., Castillo, M., 2019. In-line estimation of the elastic module of milk gels with variation of temperature protein concentration. Int. J. Food Sci. Tech., 54(2), 354-360.

Skeie, S., Feten, G., Almøy, T., Østlie, H.,Isaksson, T., 2006. The use of near infrared spectroscopy to predict selected free amino acids during cheese ripening. Int. Dairy J., 16(3), 236-242.

Svendsen, C., Cieplak, T., van den Berg, F. W., 2016. Exploring process dynamics by near infrared spectroscopy in lactic fermentations. J. Near Infrared Spectrosc., 24(5), 443-451.

3 – Aims and Objectives

In this section are reported the main objectives of this thesis work.

The first aim was directed to monitor rennet coagulation process in real time through FT-NIR coupled with MCR-ALS algorithm, developing Multivariate Statistical Process Control (MSPC) charts.

Secondly, the influence of different process factors and conditions on the rennet coagulation process using ASCA method was assessed.

The third objective aimed to investigate the use of various milk powder on the coagulation step, evaluating the effects of different sample composition and experimental conditions.

Finally, a PAT approach was developed in order to monitor the Galacto-oligosaccharides (GOS) production from dairy industry by-products.

4 – Results

4.1 PDF PAPER I

foods

MDPI

Article

Control and Monitoring of Milk Renneting Using FT-NIR Spectroscopy as a Process Analytical Technology Tool

Silvia Grassi *, Lorenzo Strani, Ernestina Casiraghi and Cristina Alamprese

Department of Food, Environmental, and Nutritional Sciences, Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy

* Correspondence: silvia.grassi@unimi.it

Received: 9 July 2019; Accepted: 9 September 2019; Published: 12 September 2019

check for updates

Abstract: Failures in milk coagulation during cheese manufacturing can lead to decreased yield, anomalous behaviour of cheese during storage, significant impact on cheese quality and process wastes. This study proposes a Process Analytical Technology approach based on FT-NIR spectroscopy for milk renneting control during cheese manufacturing. Multivariate Curve Resolution optimized by Alternating Least Squares (MCR-ALS) was used for data analysis and development of Multivariate Statistical Process Control (MSPC) charts. Fifteen renneting batches were set up varying temperature (30, 35, 40 °C), milk pH (6.3, 6.5, 6.7), and fat content (0.1, 2.55, 5 g/100 mL). Three failure batches were also considered. The MCR-ALS models well described the coagulation processes (explained variance ≥99.93%; lack of fit <0.63%; standard deviation of the residuals <0.0067). The three identified MCR-ALS profiles described the main renneting phases. Different shapes and timing of concentration profiles were related to changes in temperature, milk pH, and fat content. The innovative implementation of MSPC charts based on T^2 and Q statistics allowed the detection of coagulation failures from the initial phases of the process.

Keywords: dairy industry; milk renneting; in-line control; near infrared spectroscopy; MCR-ALS; multivariate control chart

Foods **2019**, *8*, 405; doi:10.3390/foods8090405 www.mdpi.com/journal/foods

PAPER I *(published version in open access)*

Control and monitoring of milk renneting stage using FT-NIR spectroscopy, based on a process analytical technology approach

Silvia Grassi[1]*, Lorenzo Strani[1], Ernestina Casiraghi[1] and Cristina Alamprese[1]

[1]Department of Food, Environmental, and Nutritional Sciences (DeFENS), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy

*Correspondence: silvia.grassi@unimi.it

Received: 9 July 2019; Accepted: 9 September 2019; Published: 12 September 2019

Abstract: Failures in milk coagulation during cheese manufacturing can lead to decreased yield, anomalous behaviour of cheese during storage, significant impact on cheese quality and process wastes. This study proposes a Process Analytical Technology approach based on FT-NIR spectroscopy for milk renneting control during cheese manufacturing. Multivariate Curve Resolution optimized by Alternating Least Squares (MCR-ALS) was used for data analysis and development of Multivariate Statistical Process Control (MSPC) charts. Fifteen renneting batches were set up varying temperature (30, 35, 40 °C), milk pH (6.3, 6.5, 6.7), and fat content (0.1, 2.55, 5 g/100 mL). Three failure batches were also considered. The MCR-ALS models well described the coagulation processes (explained variance ≥ 99.93%; lack of fit < 0.63%; standard deviation of the residuals < 0.0067). The three identified MCR-ALS profiles described the main renneting steps. Different shapes and timing of concentration profiles were related to changes in temperature, milk pH, and fat content. The innovative implementation of MSPC charts based on T^2 and Q statistics allowed the detection of coagulation failures from the initial phases of the process.

Keywords: dairy industry; milk renneting; in-line control; near infrared spectroscopy; MCR-ALS; multivariate control chart.

1. Introduction

Milk coagulation is one of the most critical steps during cheese manufacturing. Failures in this operation can lead to decreased yield, anomalous behaviour of cheese during storage, significant impact on cheese quality and process wastes. Until recently, curd formation progress and cutting time setting have been mainly managed by skilled personnel through process variable control (e.g. vat temperature, rennet activity and concentration, calcium salt concentration, pH) [1]. However, systems for designing, analysing, and controlling manufacturing processes through timely measurements are spreading rapidly among the dairy industries [2]. These systems fall under the Process Analytical Technology (PAT) principles, borrowed from the pharmaceutical industry, which aimed to answer the challenge of ensuring final product quality by real-time process monitoring [3].

The PAT approach requires process analysers to be implemented in dynamic conditions. For the dairy industry, in-line mechanical and/or optical devices have been proposed instead of the gold standard lactodynamograph based on the oscillation recording of a small stainless-steel pendulum immersed in milk [2]. For instance, acoustic wave sensors [4], small angle neutron scattering techniques [5], and ultrasonic analyses [6, 7] have been evaluated. Spectroscopic sensors are among the most promising process analysers in food industry [8] and different PAT solutions have been proposed for their implementation in cheese manufacturing [9]. For instance, fluorescence spectroscopic sensors have been widely used to study and monitor rennet coagulation of milk [10], also assessing coagulum strength and gelation time [11]. However, near-infrared (NIR) spectroscopy presented numerous advantages compared to other spectroscopic techniques. In fact, NIR

spectroscopy has been demonstrated to be suitable for fast, non-destructive and low-invasive real-time measurements of both quality parameter evolution [12] and process dynamics [13, 14], due to the possibility to extract both chemical and physical information from a NIR spectrum. Moreover, NIR measurements can be realized also with *ad hoc* optic probes placed directly into the coagulation vats [14, 15, 16, 17, 18].

After the proper process analyser implementation, PAT requires robust data management and analysis tools for providing platform solutions [19]. In particular, chemometrics can be applied to handle spectroscopic data for process monitoring, control, and endpoint determination, by replacing univariate and bivariate statistical techniques. For instance, Multivariate Curve Resolution (MCR) was applied to elucidate process-related physico-chemical changes and to extrapolate process kinetic information. Indeed, MCR infers the contribution of each single component involved in the studied system, allowing its quantification over the process development [20]. Its successful application in the spectroscopic field is due to the ability in decomposing spectroscopic data that are characterized by overlapped spectral bands, especially when recorded from complex systems such as milk [18, 21].

The existing PAT approaches for milk renneting monitoring are able to detect the occurrence of coagulation point or to evaluate the curd setting rate, but they assume kinetic models on a case-by-case basis [2]. Moreover, the existing models often study reconstituted milk from skim milk powder and do not consider process variables such as coagulation temperature, or milk pH and/or fat content, which can vary depending on the cheese to be produced.

In this context, the present study proposes a PAT approach for milk renneting monitoring based on Fourier Transform (FT)-NIR spectroscopy coupled with MCR-ALS. A wide range of normal operating conditions (NOC) adopted for cheese production have been considered in order to make the proposed

models more robust. Through this approach, Multivariate Statistical Process Control (MSPC) charts have been implemented for fault diagnosis and ongoing process management, no matter the applied operating conditions. The development of the MSPC charts represents an important innovation since, to the best of our knowledge, in the scientific literature there are not works dealing with multivariate control charts implemented for milk rennet coagulation.

2. Materials and Methods

2.1. Experimental plan

Twelve milk renneting batches (named from NOC_1 to NOC_{12}) were set up under different normal operating conditions commonly applied in cheese production, in order to describe process changes due to different industrial settings.

In particular, three levels of coagulation temperature (30 °C, 35 °C, 40 °C), milk pH (6.3, 6.5, 6.7) and fat content (0.1 g/100 mL, 2.55 g/100 mL, 5 g/100 mL) were combined as reported in Table 1, taking into account real operating ranges. In addition, three replicates of the NOC_{13} batch (named NOC_{13a}, NOC_{13b}, and NOC_{13c}) that combines the intermediate levels of the considered operating factors (i.e. temperature, 35 °C; pH, 6.5; fat content, 2.55 g/100 mL), were performed to assess the reproducibility of FT-NIR measurements and MCR-ALS models.

Table 1. Milk renneting batches set up under normal operating conditions (NOC) for the development of a process monitoring tool. * NOC_{13} conditions were performed in three replicates (NOC_{13a}, NOC_{13b}, and NOC_{13c}).

Batch.	***Sample ID***	***Temperature (°C)***	***Fat content (g/100 mL)***	***pH***
NOC_1	*T30 F0.10 pH6.5*	*30*	*0.10*	*6.5*
NOC_2	*T35 F0.10 pH6.3*	*35*	*0.10*	*6.3*
NOC_3	*T35 F0.10 pH6.7*	*35*	*0.10*	*6.7*
NOC_4	*T40 F0.10 pH6.5*	*40*	*0.10*	*6.5*
NOC_5	*T30 F2.55 pH6.3*	*30*	*2.55*	*6.3*
NOC_6	*T30 F2.55 pH6.7*	*30*	*2.55*	*6.7*
NOC_7	*T40 F2.55 pH6.3*	*40*	*2.55*	*6.3*
NOC_8	*T40 F2.55 pH6.7*	*40*	*2.55*	*6.7*
NOC_9	*T30 F5.00 pH6.5*	*30*	*5.00*	*6.5*
NOC_{10}	*T35 F5.00 pH6.3*	*35*	*5.00*	*6.3*
NOC_{11}	*T35 F5.00 pH6.7*	*35*	*5.00*	*6.7*
NOC_{12}	*T40 F5.00 pH6.5*	*40*	*5.00*	*6.5*
*NOC_{13}**	*T35 F2.55 pH6.5*	*35*	*2.55*	*6.5*

Three failure batches (named FB_1, FB_2, and FB_3) were also set up according to NOC_{13} operating conditions but forcing anomalies as follows: in FB_1 only half the amount of rennet was added; in FB_2 milk heating was turned off just after rennet addition; in FB_3 only half the amount of $CaCl_2$ was added.

2.2. Milk preparation and renneting

Pasteurized skimmed milk (Centrale del Latte di Milano, Milan, Italy; fat, 0.1 g/100 mL) was purposely combined with pasteurized milk cream (Centrale del Latte di Milano, Milan, Italy; fat, 35 g/100 mL) in order to reach the different fat concentrations imposed by the experimental plan (Table 1). Each mixture (150 mL) was kept under stirring on a magnetic plate for 12 h in a cold room

(4±1 °C). After reconditioning to 20 °C and $CaCl_2$ addition (final concentration, 6 mM), pH was monitored, by using a previously calibrated pH-meter (mod. 3510; JENWAY, Dunmow, England), and adjusted to the desired value by the addition of a concentrated solution of citric acid (100 g/100 mL). Each sample was then heated and maintained at the desired operating temperature (±0.1 °C) by means of a thermostatic bath (Heidolph, MR Hei-Standard, Schwabach, Germany). After the addition of 52.5 µL liquid bovine rennet (75% chymosin, 25% rennin; Linea Rossa, 175 IMCU/mL, Caglificio Clerici, Cadorago, Como, Italy), the sample was stirred for 60 s before starting monitoring by FT-NIR spectroscopy and rheology.

2.3. FT-NIR spectroscopy

Each milk sample (100 mL), prepared as previously reported, was maintained in a thermostatic bath (Heidolph, MR Hei-Standard, Schwabach, Germany) at the required temperature (30 °C, 35 °C, or 40 °C) and monitored for 30 min through a FT-NIR spectrometer (MPA, Bruker Optics, Milan, Italy) equipped with a transflectance optic probe (1 mm pathlength) inserted directly in the sample. Spectra were collected every minute over the 12500 - 4000 cm^{-1} range, with a resolution of 8 cm^{-1}, and 64 scans for both sample and background in order to obtain a good signal-to-noise ratio. The instrument control was managed by OPUS software (v. 6.0 Bruker Optics, Milan, Italy).

2.4. Rheological behaviour

Milk renneting batches were monitored in continuous also by rheology measurements. In particular, a time curing test in oscillation was performed by means of a Physica MCR 300 rheometer (Anton Paar GmbH, Graz, Austria), controlled by the software Rheoplus/32 (v. 3.00, Physica Messtechnik GmbH, Ostfildern, Germany). Each milk sample (19 mL), prepared as previously reported, was poured in the concentric cylinders (CC27) of the rheometer heated at the desired temperature (30 °C, 35 °C, or 40 °C). Elastic (G') and viscous (G'') modulus were measured each minute

over a 30 min period, applying constant strain (0.01%) and frequency (1 Hz) values. Strain and frequency settings were chosen based on preliminary strain and frequency sweep tests carried out on both liquid and coagulated milk samples.

G' values of each batch were modelled as a function of coagulation time, using the following sigmoid curve (Eq. 1) implemented in Table Curve software (v. 4.0, Jandel Scientific, San Rafael, CA, USA):

$$y = a/(1 + \exp(\frac{-(x - b)}{c})) \quad (1)$$

In order to identify kinetic critical points during renneting - i.e. times related to the maximum rate, acceleration, and deceleration of the step - the first and second derivatives of the sigmoid functions were calculated. Afterwards, the times corresponding to maximum and minimum values of derivatives were extrapolated [18].

2.5. Data analysis

FT-NIR spectra collected during milk coagulation trials were reduced in spectral range (12500-5824 cm^{-1}) and batch-wise pre-processed with Standard Normal Variate (SNV). Spectral region from 5823 to 4000 cm^{-1} was excluded due to high noise and signal saturation.

MCR-ALS analysis was performed by using a toolbox [22] implemented in MatLab v. 7.4 (MathWorks, Natick, MA, USA). Spectral data for NOC batches 1-12 were arranged in twelve **D** (M × N) sub-matrices, where the *M* rows correspond to the 30 spectra obtained at different renneting times and the *N* columns refer to the 1730 considered wavenumbers. Spectral data obtained from the three replicates of NOC_{13} and the failure batches (FB_1, FB_2, and FB_3) were arranged in a **D** matrix composed of 6 independent sub-matrices, each referring to one of the 6 trials.

MCR-ALS allowed the decomposition of each **D** matrix into two sub-matrices, **C** (*M* × *F*) and $\mathbf{S}^T$ (*F* × *N*), named concentration and spectral profiles respectively (Eq. 2). C describes the *F* components affecting the modification of the *M* spectra over time, whereas $\mathbf{S}^T$ contains the *F* component variations with respect to the considered *N* wavenumbers. **E** (*M* × *N*) is the residual matrix.

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (2)$$

Before applying MCR, the number of components (*F*) was defined by Principal Component Analysis (PCA). Then, the ALS optimization was performed by using a previously stated initial estimation of the spectral profiles [23].

The general steps of MCR-ALS are the following [24]:

Definition of the component number (*F*) for **D**.

Development of non-random initial estimates of either **C** or $\mathbf{S}^T$.

Given **D** and $\mathbf{S}^T$, least-squares calculation of C under given constraints.

Given **D** and **C**, least-squares calculation of $\mathbf{S}^T$ under given constraints.

Reconstruction of D as the product $\mathbf{C}\mathbf{S}^T$.

The last three steps have been repeated until the quality in data reconstruction was satisfactory and convergence in the iterative optimization was achieved. The proper final concentration and spectral profiles were determined by using a stopping criterion based on the relative difference of the Lack of Fit percentage (LOF; Eq. 3), i.e. when the LOF difference in two consecutive iterative cycles was lower than 0.1% [25].

$$LOF\ (\%) = 100\ \times \sqrt{\frac{\sum_{ij} e_{ij}^2}{\sum_{ij} d_{ij}^2}} \quad (3)$$

In this equation, e_{ij} is each *ij*th element of the residual matrix **E,** i.e. the related residual obtained from the difference between the input element and the MCR-ALS reproduction, and d_{ij} is each *ij*th element of the **D** matrix. Non-negative concentration and unimodality constraints were imposed on concentration profiles to solve MCR-ALS ambiguities due to rotational and scale intensity.

The concentration profiles calculated by MCR-ALS were compared with the rheological results to assess the reliability of the approach. In particular, a Pearson correlation matrix was calculated for kinetic critical times extrapolated from the G' curves and the time corresponding to the maximum value of the second MCR-ALS profile.

PCA-based MSPC charts were built using MCR-ALS concentration profiles obtained for the three replicates of NOC_{13} batch. The concentration profiles of NOC_{13c} and failure batches (FB_1, FB_2, and FB_3) were then used to detect if each considered batch was in or out of control based on the MSPC charts previously built. The PCA-based MSPC charts [26] were developed calculating Hotelling's T^2 and Q statistics for the PCA models constructed with the MCR-ALS concentration profiles. The Hotelling's T^2 chart represents the estimated Mahalanobis distance from the centre of the PCA model. The Q-statistic chart analyses the residuals, i.e. the process variations not represented in the PCA model. Sensitivity (i.e. true positive rate) and specificity (i.e. true negative rate) were calculated to evaluate the reliability of each control chart.

All the chemometric analyses were performed by using MatLab v. 7.4 (MathWorks, Natick, MA, USA).

3. Results and discussion

3.1. FT-NIR spectra interpretation

The FT-NIR spectra collected during the renneting trials showed a clear increment in absorbance over time, highly affected by scattering effects due to physical changes of milk caused by coagulation [27, 28], as well as to instrument characteristics. As an example, spectra trend during coagulation of batch NOC_{13a} is shown in Figure 1a. The main absorption band, observable at 6900 cm^{-1}, is due to the combination of symmetric and asymmetric stretching of O-H bond of water, as already observed in previous works [14, 17]. Other relevant bands at 10800 and 8600 cm^{-1} are ascribable to the lipid C-H bonds [29].

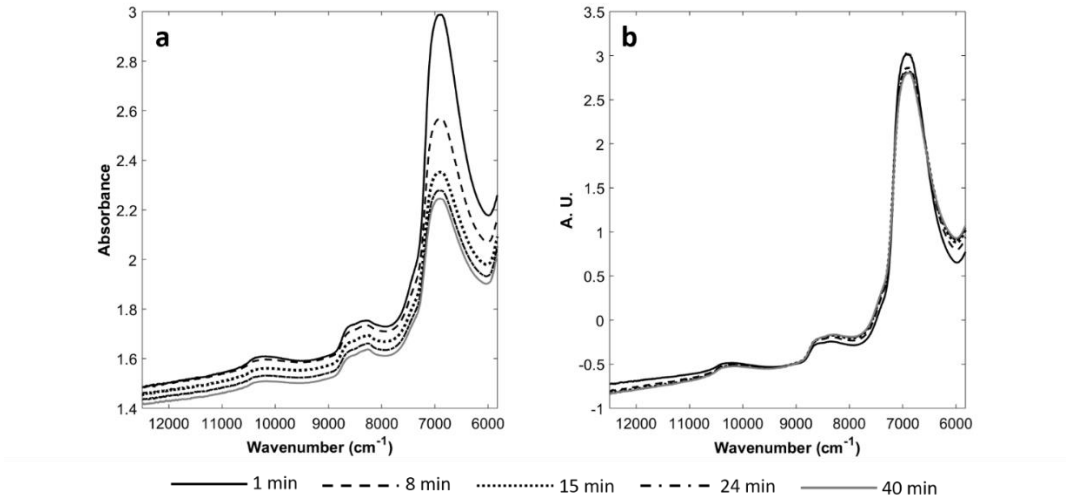


Figure 1. FT-NIR spectra collected for one milk renneting batch developed under normal operating conditions (NOC_{13a}): temperature, 35 °C; milk pH, 6.5; fat content, 2.55 g/100 mL. a) raw spectra; b) SNV pre-treated spectra.

After SNV pre-treatment (Figure 1b), it was possible to notice that the spectrum collected at the beginning of the process (just after 1 min) had a higher absorbance within the 12000 - 9000 cm^{-1} range if compared to the spectra acquired from 8 min on. The opposite occurs between 9000 cm^{-1} and

7000 cm^{-1}. In correspondence of O-H water symmetric and asymmetric stretching (6900 cm^{-1}) the absorbance of the liquid milk resulted again higher than that of the spectra collected on coagulating milk. Similar spectra and trends were observed for all the performed experimental trials.

3.2. Principal Component Analysis (PCA)

Each dataset collected from a single milk renneting batch was explored by PCA after spectral range reduction and pre-treatment with SNV and mean-centering. The number of significant components *F* to be used in the following MCR-ALS analysis was chosen by inspecting scores, loadings, and explained variance of PCA models. Figure 2 shows, as an example, the trend of the first and second principal component (PC1 and PC2) scores *vs* coagulation time and the corresponding loadings plot obtained for batch NOC_{13a}. Similar results were obtained for all the NOC batches evaluated. PC1 explained at least 98.54% of variance and described changes occurring in milk during renneting. Actually, the PC1 scores *vs* time presented a sigmoid-like distribution with a fast increase in score values up to 10 min of coagulation (●, Figure 2a). Then, the score increment rate decreased, reaching a steady state at the end of coagulation monitoring. At the beginning of the monitoring, samples were characterized by slightly negative values of PC2 (▲, Figure 2a); afterwards, they rapidly moved to a maximum in correspondence of the maximum increasing rate of PC1 (i.e. the maximum slope of PC1 curve), and then they decreased and stabilized. A very similar PC1 score trend was also observed for milk fermentation by Grassi et al. [17]. Moreover, Lyndgaard et al. [14] found a similar time-related distribution of PC1 scores for milk renneting processes, highlighting three coagulation stages: k-casein proteolysis, paracasein aggregation, and gel network formation. The intensity of PC1 (solid line) and PC2 (dashed line) loadings (Figure 2b) showed that the wavenumbers mainly responsible for score distribution are associated with the major bands already observed and commented for raw and SNV pre-treated spectra (Figure 1).

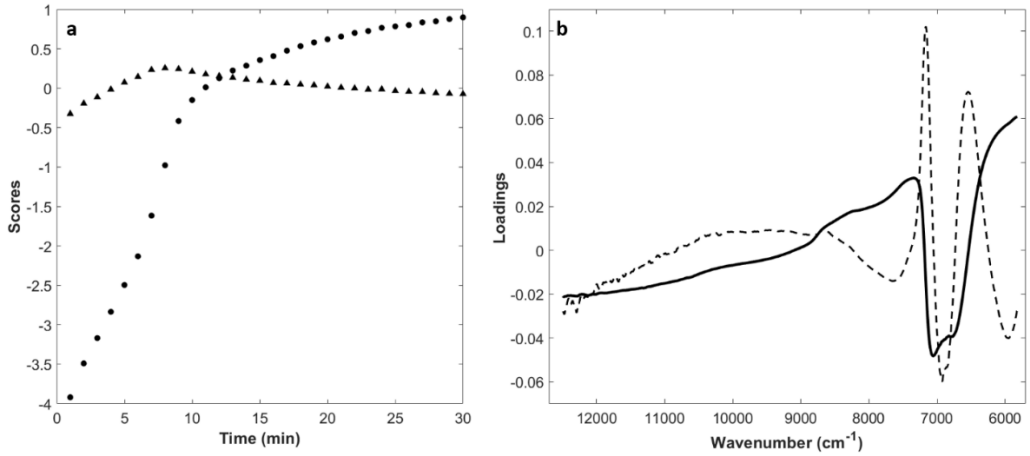


Figure 2. Principal component analysis results of the SNV transformed and centered FT-NIR spectra collected for one milk renneting batch developed under normal operating conditions (NOC_{13a}): temperature, 35 °C; milk pH, 6.5; fat content, 2.55 g/100 mL. a) plot of PC1 (●) and PC2 (▲) scores vs time; b) PC1 (solid line) and PC2 (dashed line) loading plot.

3.3. MCR-ALS results for NOC batches

Before performing MCR analysis, the selection of the proper number of components was performed through singular value decomposition (SVD algorithm on which PCA is based) by the ALS approach. Since two components were used to describe the process changes observed in PCA of mean-centered data, it is consistent that three components were found relevant by ALS applied to the not-centered data, because the rank decreases by one when mean-centering is performed [30].

After determining the number of significant components, it was necessary to choose the initial estimates for starting the MCR-ALS analysis. Three spectra were selected for each NOC batch by means of Pure Variable detection, which selects the purest column variables of a dataset based on the SIMPLISMA method [31]. In all cases, the pure spectra selected by the algorithm corresponded to signal collected (1) at the beginning of the process,

(2) at the time of significant reduction in absorbance at 6900 cm^{-1}, related to the end of the paracasein aggregation, and (3) at the end of the monitoring. The three selected spectra are thus linked to the different renneting phases, as already reported for yogurt coagulation [18].

The product $\mathbf{CS^T}$ of the obtained models explained at least 99.93% of the data variance; LOF was lower than 0.44%, and the standard deviation of the residuals was lower than 0.0067.

Figure 3 reports the MCR-ALS concentration profiles obtained for the 1-12 NOC batches of milk renneting. The trend of the three concentration profiles contains information about the well-known changes occurring during milk coagulation, already highlighted by the trend of PC1 and PC2 score values *vs* renneting time and also confirmed by the rheological measurement results. In particular, the first MCR-ALS concentration profile (solid lines) had a sigmoidal shape in all the performed renneting trials, with higher values at the beginning of the step as already observed by Grassi et al. [18] for milk fermentation. Actually, the first concentration profile describes the primary phase of coagulation, which involves the k-casein proteolysis [14, 32]. In this step, no changes in the liquid-like structure of milk can be observed. Later, when the first aggregation of paracasein occurs, a steep decrease in the first MCR-ALS profile was observed, because milk was no more liquid. In the second phase of renneting, described by the second MCR-ALS concentration profile (dashed lines in Figure 3), the massive aggregation of the rennet-altered casein micelles occurs due to the glycomacropeptide detachment and the consequent loss of colloidal stability, thus forming chains and clusters [14, 32]. Time corresponding to the maximum value of the second MCR-ALS profile can therefore be considered as the milk sol-gel transition point. Afterwards, coagulation enters the third phase, when the protein clusters grow until they form a continuous, three-dimensional gel network incorporating water and fat [14, 33]. This phase is represented by the third MCR-ALS concentration profile (dotted lines in Figure 3), which had a

sigmoidal shape opposite to the first profile. The three coagulation phases are highly interconnected and partially overlap over the coagulation time.

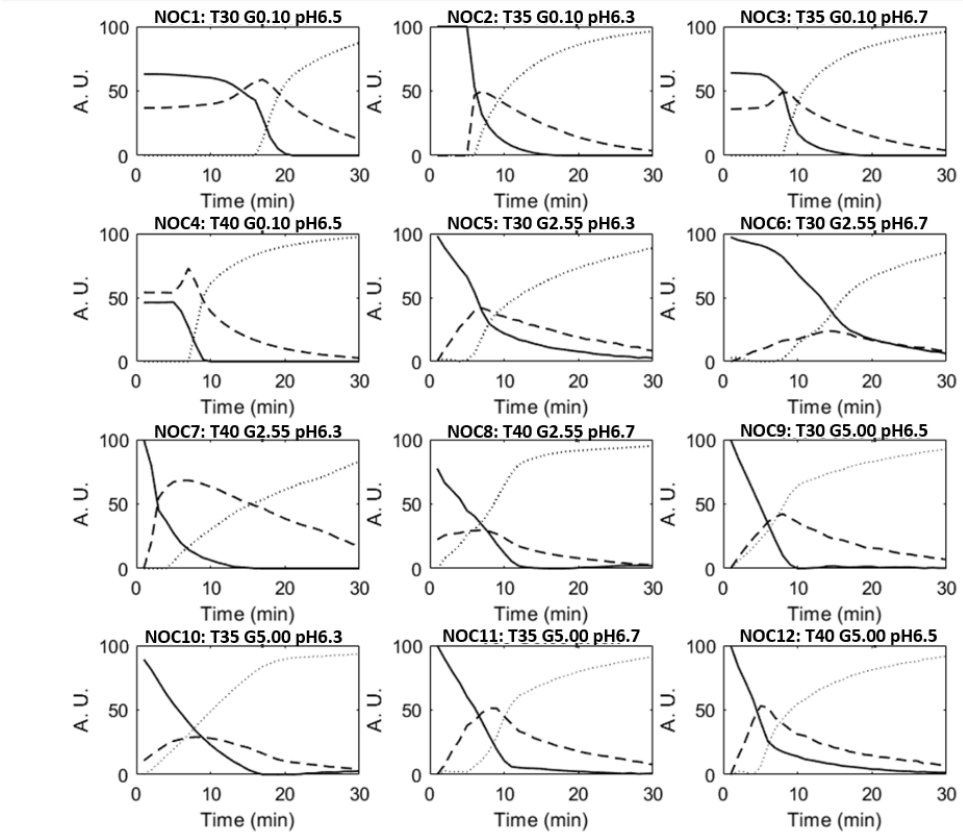


Figure 3. MCR-ALS concentration profiles of the milk renneting batches developed under normal operating conditions (NOC1-NOC12). Solid lines describe the liquid-like behavior of milk, dashed lines represent the transition phase of the renneting process, and dotted lines reflect the solid-like behavior of coagulated milk. See Table 1 for sample identification.

The different shapes and timing of the concentration profiles were strictly related to the combination of the investigated processing factors, i.e. temperature, milk pH and fat content. As expected, milk coagulation resulted faster in trials performed at 40 °C with a pH value of 6.3, as evidenced by the

advance transition time highlighted by the second concentration profile (dashed line) of NOC_7 batch (sample ID: T40 F2.55 pH6.3). NOC_1 (sample ID: T30 F0.10 pH6.5) and NOC_6 (sample ID: T30 F2.55 pH6.7) batches showed a slower coagulation, due to the combination of low temperature (30 °C) and high pH values (6.5 and 6.7, respectively). These observations were also confirmed by the time curing profiles reported as an example in Figure 4, where the three renneting phases can be clearly distinguished. At the beginning, no changes in the liquid-like behaviour of milk were observed, as the enzymatic modification of casein micelles did not affect the elastic (G') and viscous (G'') moduli trend. Then, an increase in both G' and G'' values was registered, corresponding to the formation of a three-dimensional protein network incorporating water and fat. Even though the time curing profiles of all the performed trials showed a similar behaviour, each operative condition led to characteristic trends and shapes of G' and G'' curves. In particular, for NOC_7 batch (sample ID: T40 F2.55 pH6.3) the process was so fast that also the coagulum break was evident, with a decrease in G' and G'' values after less than 10 min of renneting. Such a decrease corresponds to a strong whey syneresis and it was evident only for this sample. A similar behaviour of G' values during milk renneting at pH 6.1 was also reported by Ong et al. [34].

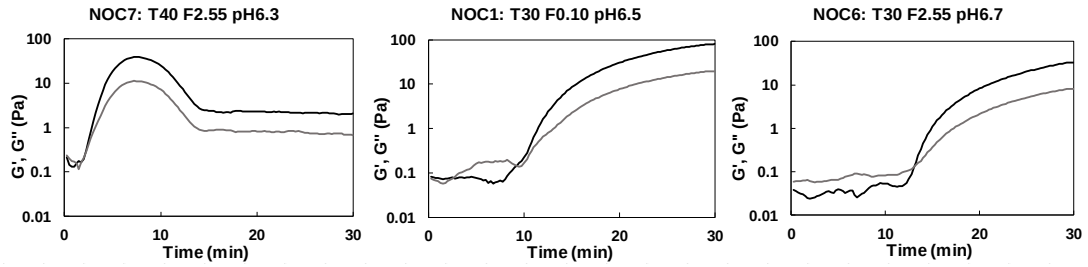


Figure 4. Time curing profiles of some of the evaluated milk renneting batches. Black and grey lines represent the elastic (G') and viscous (G'') modulus, respectively. See Table 1 for sample identification.

To better highlight the relationship between FT-NIR spectroscopy and rheology data, a Pearson correlation matrix was calculated for kinetic critical times extrapolated from G' curves and the time corresponding to the maximum value of the second MCR-ALS profile. The latter resulted to be highly correlated ($r=0.96$; $p < 0.001$) with the acceleration time of renneting calculated from the G' curves (Table 2), indicating that both the parameters describe the sol-gel transition of milk. However, the transition times measured by FT-NIR spectroscopy occurred always few minutes before those calculated from rheological curves, thus demonstrating a higher sensitivity of the spectroscopic technique to the coagulation phenomena. This higher sensitivity of FT-NIR spectroscopy with respect to rheological data was reported also by Grassi et al. [17] for milk fermentation and it can be relevant for a better renneting step control at industrial level.

Table 2. Milk renneting: comparison of time corresponding to the maximum value of the second MCR-ALS concentration profile calculated from FT-NIR data (CP2) and time corresponding to acceleration in elastic modulus (G') increase extrapolated from time curing curves (AT_G').

Batch.	***Sample ID***	***CP2 (min)***	***AT_G' (min)***
NOC_1	*T30 F0.10 pH6.5*	*14.5*	*18.9*
NOC_2	*T35 F0.10 pH6.3*	*6.0*	*7.9*
NOC_3	*T35 F0.10 pH6.7*	*9.0*	*13.3*
NOC_4	*T40 F0.10 pH6.5*	*8.0*	*11.2*
NOC_5	*T30 F2.55 pH6.3*	*7.3*	*10.0*
NOC_6	*T30 F2.55 pH6.7*	*17.0*	*19.4*
NOC_7	*T40 F2.55 pH6.3*	*6.5*	*8.8*
NOC_8	*T40 F2.55 pH6.7*	*7.0*	*9.5*
NOC_9	*T30 F5.00 pH6.5*	*8.3*	*13.1*
NOC_{10}	*T35 F5.00 pH6.3*	*8.0*	*11.0*
NOC_{11}	*T35 F5.00 pH6.7*	*8.5*	*13.7*
NOC_{12}	*T40 F5.00 pH6.5*	*5.5*	*6.8*
NOC_{13a}	*T35 F2.55 pH6.5*	*5.8*	*8.9*
NOC_{13b}	*T35 F2.55 pH6.5*	*7.5*	*9.8*
NOC_{13c}	*T35 F2.55 pH6.5*	*7.0*	*9.4*

These findings reveal that FT-NIR spectroscopy, combined with MCR-ALS, represents a robust approach for the description of milk renneting under different cheese-making conditions. Indeed, this approach was able to distinguish the three-main coagulation phases, no matter the operating temperature, and the milk pH and fat content.

3.4. MCR-ALS results for NOC_{13} and FB batches

The reliability of the proposed FT-NIR method based on MCR-ALS models to assess possible milk renneting failures was verified using the spectral data

matrix containing FT-NIR data collected from the three replicates of the NOC_{13} batch and the three failure trials FB_1, FB_2 and FB_3. Figure 5 shows the three MCR-ALS concentration profiles obtained for the six considered batches and their corresponding spectral profiles. The $\mathbf{CS^T}$ product explained 99.99% of the data variance; LOF was 0.63%, and the standard deviation of the residuals was lower than 0.0063. The concentration profiles obtained for the three replicates of NOC_{13} batch (reported in black in Figs. 5a-c) were almost overlapped and the small differences observed should be considered as the expected variability in a milk renneting process carried out with different lots of raw materials.

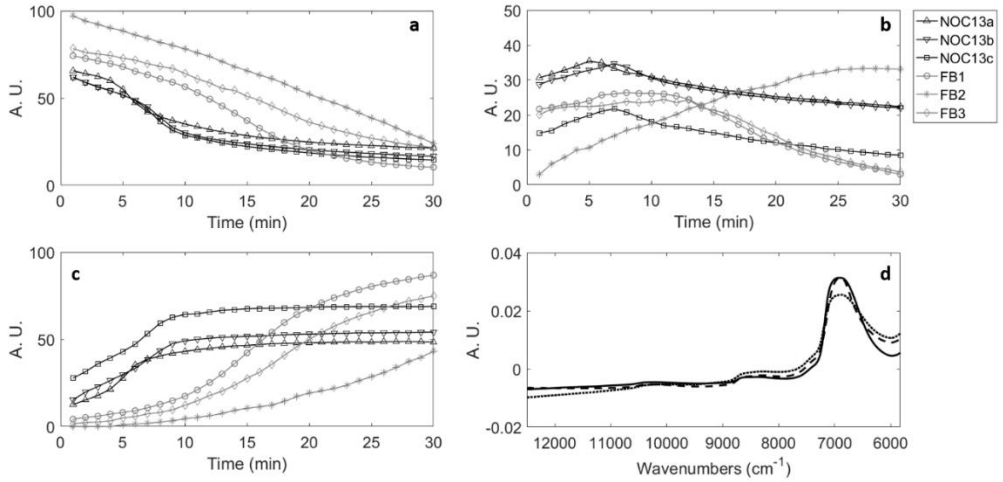


Figure 5. MCR-ALS results for the three replicates of NOC_{13} batch and for the three failure batches FB: a) concentration profiles of the liquid-like behavior of milk; b) concentration profiles of the transition phase of renneting; c) concentration profiles of the solid-like behavior of coagulated milk; d) spectral profiles of liquid-like behavior of milk (solid line), transition phase (dashed line) and solid-like behavior of milk (dotted line).

The concentration profiles of NOC_{13} replicates showed the same trends already observed for the NOC batches 1-12, thus, also in this case, the sol-gel transition time can be extrapolated as the time corresponding to the

maximum value of the second profile. MCR-ALS profiles for FB batches 1-3 (reported in grey in Figs. 5a-c) were clearly separated in timing and shape from the profiles of NOC_{13} replicates, according to the imposed failures in the process. The pure spectral profiles obtained (Figure 5d) are representative of the different coagulation phases, as already observed for $NOC_{1:12}$ (data not shown): the solid-line profile represents the liquid-like behaviour of milk; the dotted-line profile represents the solid-like behaviour of coagulated milk; the dashed-line profile stands for the transition phase.

3.5. MSPC charts on MCR-ALS results for NOC_{13} and FB batches

PCA-MSPC charts were built applying PCA to the three MCR-ALS concentration profiles calculated for the replicates “a” and “b” of NOC_{13} batch. The first principal component was considered, accounting for 80% of the variance. A 99% confidence interval was considered in order to calculate the chart limits (dashed lines in Figure 6). Then, the MCR-ALS concentration profiles of replicate “c” of NOC_{13} batch were also projected into the PCA model, as well as the profiles calculated for the failure batches (FB_1, FB_2, and FB_3). The relative Hotelling’s T^2 and Q statistics were calculated and implemented in the control charts. The NOC_{13c} batch resulted in-control over all the renneting time, i.e. within the calculated confidence interval. On the contrary, the FB batches showed off-control values of T^2 up to 12 min after the beginning of the trials and off-control values of Q residuals for most of the sampling points.

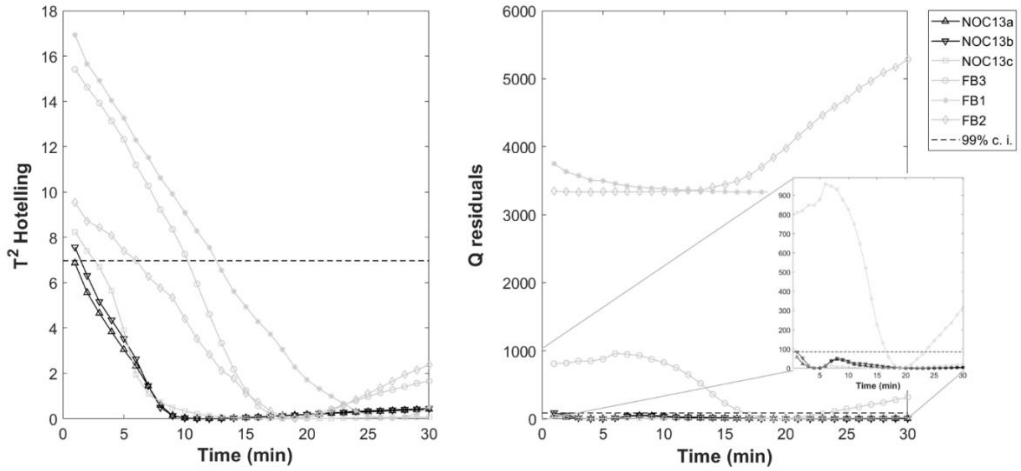


Figure 6 MSPC charts for the three replicates of the milk renneting batch NOC_{13} (in-control) and for the three failure batches FB (off-control): a) Hotelling's T^2; b) Q-statistic. Dashed line represents the 99% confidence interval (c.i.).

The goodness of the built charts was evaluated also by sensitivity and specificity values. T^2 chart showed quite poor specificity (19%) due to the in-control values reached by the tested batches with coagulation progress; on the contrary, sensitivity was high (95%). These results can be due to the fact that the variables used, i.e. the concentration profiles, are highly correlated, thus monitoring process compliance through T^2 based on the first PC could be not sufficient [26]. It is therefore advisable to refer also to the Q-statistic [35]. In this case study, Q residual chart gave both high specificity (94%) and sensitivity (100%), suggesting the reliability of this approach for the monitoring of milk renneting. By the combination of T^2 and Q control charts, it was possible to detect the in-control tested batch (NOC_{13c}) and to distinguish failure batches just from the first minutes of the process. Similarly, de Oliveira et al. [30] concluded that the combination of T^2 and Q charts gave specificity and sensitivity results more reliable than their single check when applying MSPC to synthetic commercial gasoline distillation.

4. Conclusions

The work demonstrated that coupling FT-NIR spectroscopy with MCR-ALS data elaboration allows the development of a useful tool for the in-line control of milk renneting during cheese manufacturing. The very innovative approach suggested for the implementation of MSPC charts is able to detect possible coagulation failures from the first minutes of the process. This is of fundamental importance for modern dairy industries, because of the urgent needs for automation in order to improve product quality and production yields. This kind of industrial control systems perfectly fit with the Industry 4.0 roadmap towards a fully digital enterprise.

Author Contributions: Conceptualization, C.A., E.C. and S.G.; methodology, C.A., E.C. and S.G.; formal analysis, S.G. and L.S.; data curation, C.A., S.G. and L.S.; writing—original draft preparation, C.A., S.G. and L.S.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

Fox, P. F.; McSweeney, P. L. Cheese: an overview. In *Cheese: Chemistry, Physics & Microbiology: General Aspects,* Fox, P. F., McSweeney, P. L., Cogan, T. M., Guinee, T. P., Eds.; Elsevier, London, UK, 2017; pp. 5-21.

O'Callaghan, D. J.; O'Donnell, C. P.; Payne, F. A. Review of systems for monitoring curd setting during cheesemaking. *Int J Dairy Technol* **2002***, 55,* 65-74.

Grassi, S.; Alamprese, C. Advances in NIR spectroscopy applied to Process Analytical Technology in food industries. *Curr Opin Food Sci* **2018***, 22,* 17-21.

Pais, V. F.; Veríssimo, M. I.; Oliveira, J. A.; Gomes, M. T. S. Using acoustic wave sensors to follow milk coagulation & to separate the cheeses according to the milk origin. *Sensors Actuat B- Chem* **2015***, 207,* 1121-1128.

van Heijkamp, L. F.; de Schepper, I. M.; Strobl, M.; Tromp, R. H.; Heringa, J. R.; Bouwman, W. G. Milk gelation studied with small angle neutron scattering techniques and Monte Carlo simulations. *J Phys Chem A* **2010***, 114,* 2412-2426.

Derra, M.; Bakkali, F.; Amghar, A.; Sahsah, H. Estimation of coagulation time in cheese manufacture using an ultrasonic pulse-echo technique. *J Food Eng* **2018***, 216,* 65-71.

Dwyer, C.; Donnelly, L.; Buckin, V. Ultrasonic analysis of rennet-induced pre-gelation & gelation processes in milk. *J Dairy Res* **2005***, 72,* 303-310.

Cullen, P. J.; O'Donnell, C. P.; Fagan, C. C. Benefits & challenges of adopting PAT for the food industry. In *Process Analytical Technology for the Food Industry;* O'Donnell, C. P., Fagan, C., Cullen, P. J., Eds.; Springer: New York, USA, 2014; pp. 1-5.

Panikuttira, B.; O'Shea, N.; Tobin, J. T.; Tiwari, B. K.; O'Donnell, C. P. Process analytical technology for cheese manufacture. *Int J Food Sci Technol* **2018**, *53*, 1803-1815.

Fagan, C.C.; Ferreira, T.G.; Payne, F.A.; O'Donnell, C.P.; O'Callaghan, D.J.; Castillo, M. Preliminary evaluation of endogenous milk fluorophores as tracer molecules for curd syneresis. *J Dairy Sci* **2011**, *94*, 5350-5358.

Blecker, C.; Habib-Jiwan, J.M.; Karoui, R. Effect of heat treatment of rennet skim milk induced coagulation on the rheological properties and molecular structure determined by synchronous fluorescence spectroscopy and turbiscan. *Food Chem* **2012**, *135*, 1809-1817.

Wang, Y.; Guo, W.; Zhu, X.; Liu, Q. Effect of homogenisation on detection of milk protein content based on NIR diffuse reflectance spectroscopy. *Int J Food Sci Technol* **2019**, *54*, 387-395.

Fagan, C.C.; Castillo, M.; Payne, F.A.; O'Donnell, C.P.; Leedy, M.; O'Callaghan, D.J. Novel online sensor technology for continuous monitoring of milk coagulation and whey separation in cheesemaking. *J Agric Food Chem* **2007**, *55*, 8836–8844.

Lyndgaard, C. B.; Engelsen, S. B.; van den Berg, F. W. Real-time modeling of milk coagulation using in-line near infrared spectroscopy. *J Food Eng* **2012***, 108,* 345-352.

Laporte, M. F.; Martel, R.; Paquin, P. The near-infrared optic probe for monitoring rennet coagulation in cow's milk. *International Dairy Journal* **1998***, 8,* 659-666.

Cimander, C.; Carlsson, M.; Mandenius, C. F. Sensor fusion for on-line monitoring of yoghurt fermentation. *J Biotechnol* **2002**, *99*, 237-248.

Grassi, S.; Alamprese, C.; Bono, V.; Casiraghi, E.; Amigo, J. M. Modelling milk lactic acid fermentation using multivariate curve resolution-alternating least squares (MCR-ALS). *Food Bioprocess Tech* **2014***, 7,* 1819-1829.

Grassi, S.; Alamprese, C.; Bono, V.; Picozzi, C.; Foschino, R.; Casiraghi, E. Monitoring of lactic acid fermentation process using Fourier transform near infrared spectroscopy. *J Near Infrared Spec* **2013***, 21,* 417–425.

Glassey, J. Data management systems. In *Process Analytical Technology for the Food Industry;* O'Donnell, C. P., Fagan, C., Cullen P. J., Eds.; Springer: New York, USA, 2014; pp. 61-71.

de Juan, A.; Tauler, R. Multivariate curve resolution (MCR) from 2000: progress in concepts & applications. *Crit Rev Anal Chem* **2006**, *36,* 163-176.

Tauler, R.; Kowalski, B.; Fleming, S. Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Anal Chem* **1993***, 65,* 2040-2047.

MATLAB program MCR-ALS. Tauler, R.; de Juan, A. Available on line: http://www.ub.es/gesq/mcr/mcr.htm (accessed 10 March 2018).

Rodríguez-Rodríguez, C.; Amigo, J. M.; Coello, J.; Maspoch, S. An introduction to multivariate curve resolution-alternating least squares: Spectrophotometric study of the acid–base equilibria of 8-hydroxyquinoline-5-sulfonic acid. *J Chem Educ* **2007***, 84,* 1190-1192.

Amigo, J. M.; de Juan, A.; Coello, J.; Maspoch, S. A mixed hard-& soft-modelling approach to study & monitor enzymatic systems in biological fluids. *Anal Chim Acta* **2006***, 567,* 245-254.

Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometr Intell Lab* **2005***, 76,* 101-110.

MacGregor, J. F.; Kourti, T. Statistical process control of multivariate processes. *Control Eng Pract* **1995***, 3,* 403-414.

Frake, P.; Luscombe, C. N.; Rudd, D. R.; Gill, I.; Waterhouse, J.; Jayasooriya, U. A. Near-infrared mass median particle size determination of lactose monohydrate, evaluating several chemometric approaches. *Anal* **1998***, 123*, 2043-2046.

Horne, D. S.; Lucey, J. A. Rennet-induced coagulation of milk. In *Cheese: Chemistry, Physics & Microbiology: General Aspects;* Fox, P. F., McSweeney, P. L., T. M. Cogan, & T. P. Guinee, Eds.; Elsevier: London, UK, 2017, pp. 115-143.

Tsenkova, R.; Atanassova, S.; Itoh, K.; Ozaki, Y.; Toyoda, K. Near infrared spectroscopy for biomonitoring: cow milk composition measurement in a

spectral region from 1,100 to 2,400 nanometers. *J Anim Sci* **2000***, 78,* 515-522.

de Oliveira, R. R.; Pedroza, R. H.; Sousa, A. O.; Lima, K. M.; de Juan, A. Process modeling & control applied to real-time monitoring of distillation processes by near-infrared spectroscopy. *Anal Chim Acta* **2017***, 985,* 41-53.

Windig, W.; Guilment, J. Interactive self-modeling mixture analysis. *Anal Chem* **1991***, 63,* 1425–1432.

Uniacke-Lowe, T.; Fox, P. F. Chymosin, pepsins and other aspartyl proteinases: Structures, functions, catalytic mechanism and milk-clotting properties. In *Cheese: Chemistry, Physics & Microbiology: General Aspects*; Fox, P. F., McSweeney, P. L., T. M. Cogan, & T. P. Guinee, Eds.; Elsevier: London, UK, 2017, pp. 69-113).

Horne, D. S.; Davidson, C. M. Direct observation of decrease in size of casein micelles during the initial stages of renneting of skim milk. *Int Dairy J* **1993**, *3,* 61-71.

Ong, L.; Dagastine, R. R.; Kentish, S. E.; Gras, S. L. The effect of pH at renneting on the microstructure, composition and texture of Cheddar cheese. *Food Res Int* **2012***, 48,* 119-130.

Jackson, J. E. *A User's Guide to Principal Components.* John Wiley & Sons, Inc: New York, USA, 1991; pp. 26-62.

4.2 PDF PAPER II

Food and Bioprocess Technology (2019) 12:954–963
https://doi.org/10.1007/s11947-019-02266-2

ORIGINAL PAPER

Milk Renneting: Study of Process Factor Influences by FT-NIR Spectroscopy and Chemometrics

Lorenzo Strani[1] · Silvia Grassi[1] · Ernestina Casiraghi[1] · Cristina Alamprese[1] · Federico Marini[2]

Received: 2 January 2019 / Accepted: 10 March 2019 / Published online: 30 March 2019

Abstract
The dairy industry is continuously developing new strategies to obtain healthier dairy products preserving expected properties. However, when modifying a food process, the reassessment of each parameters and their interaction should be considered as highly influencing the final quality. Among others, rennet process features are fundamental for both sensory properties and typical characteristics of a cheese. In this contest, the research addresses the development of a FT-NIR spectroscopic method, coupled with chemometrics, for the study of the effect of process variables on milk renneting. The effects of temperature (30 °C, 35 °C, 40 °C), milk fat concentration (0.1, 2.55, 5 g/100 mL), and pH (6.3, 6.5, 6.7) were investigated by means of a Box-Behnken experimental design. FT-NIR data collected along the 17 trials were explored by interval-PCA (i-PCA) and ANOVA-simultaneous component analysis (ASCA). i-PCA revealed differences in the occurrence and trends of coagulation phases, related to the three considered factors. ASCA allowed the characterization of renneting evolution and the assessment of the factor role, demonstrating that main and interaction effects are significant for the process progress. The proposed approach demonstrated that i-PCA and ASCA on FT-NIR data, highlighting the effects of the operating factors, allow a rapid and accurate analysis of process modifications in cheese manufacturing.

Keywords Milk renneting · Dairy industry · Near infrared spectroscopy · Interval-PCA · ANOVA-simultaneous component analysis · ASCA

✉ Silvia Grassi
silvia.grassi@unimi.it

Lorenzo Strani
lorenzo.strani@unimi.it

Ernestina Casiraghi
ernestina.casiraghi@unimi.it

Cristina Alamprese
cristina.alamprese@unimi.it

Federico Marini
federico.marini@uniroma1.it

[1] Department of Food, Environmental, and Nutritional Sciences (DeFENS), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy

[2] Department of Chemistry, Università di Roma "La Sapienza", P.le Aldo Moro 5, 00185 Rome, Italy

Springer

PAPER II *(submitted version)*

Milk renneting: study of process factor influences by FT-NIR spectroscopy and chemometrics

Lorenzo Strani[a], Silvia Grassi[a*], Ernestina Casiraghi[a], Cristina Alamprese[a], Federico Marini[b]

[a]Department of Food, Environmental, and Nutritional Sciences (DeFENS), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy

[b]Department of Chemistry, Università di Roma "La Sapienza", P.le Aldo Moro 5, 00185 Rome, Italy

* Corresponding author: Silvia Grassi, E-mail address: silvia.grassi@unimi.it

Address: via G. Celoria 2, 20133 Milan, Italy

E-mail addresses:

Lorenzo Strani, lorenzo.strani@unimi.it

Silvia Grassi, silvia.grassi@unimi.it

Ernestina Casiraghi, ernestina.casiraghi@unimi.it

Cristina Alamprese, cristina.alamprese@unimi.it

Federico Marini, federico.marini@uniroma1.it

Received: 2 January 2019; Accepted: 10 March 2019; Published: 30 March 2019

DOI: https://doi.org/10.1007/s11947-019-02266-2

Abstract

The dairy industry is continuously developing new strategies to obtain healthier dairy products preserving expected properties. However, when modifying a food process, the reassessment of each parameters and their interaction should be considered as highly influencing the final quality. Among others, rennet process features are fundamental for both sensory properties and typical characteristics of a cheese. In this contest, the research addresses the development of a FT-NIR spectroscopic method, coupled with chemometrics, for the study of the effect of process variables on milk renneting. The effects of temperature (30 °C, 35 °C, 40 °C), milk fat concentration (0.1, 2.55, 5 g/100 mL), and pH (6.3, 6.5, 6.7) were investigated by means of a Box-Behnken experimental design. FT-NIR data collected along the 17 trials were explored by interval-PCA (i-PCA) and ANOVA–Simultaneous Component Analysis (ASCA). i-PCA revealed differences in the occurrence and trends of coagulation phases, related to the three considered factors. ASCA allowed the characterization of renneting evolution and the assessment of the factor role, demonstrating that main and interaction effects are significant for the process progress. The proposed approach demonstrated that i-PCA and ASCA on FT-NIR data, highlighting the effects of the operating factors, allow a rapid and accurate analysis of process modifications in cheese manufacturing.

Keywords: milk renneting; dairy industry; near infrared spectroscopy; interval–PCA; ANOVA–Simultaneous Component Analysis; ASCA

1. Introduction

Recently, consumers' requests about healthier foodstuffs, such as reduced-fat dairy products, with similar properties (flavor, texture and firmness) to the traditional ones, have increased (Johnson et al. 2001). For this reason, dairy companies need efficient tools to control the fundamental steps of cheese making processes and to tune the process parameters to optimize the quality of the final product. Among dairy processes, coagulation is one of the most critical steps, making its monitoring a very important task for the dairy industry (Sbodio et al. 2002). Indeed, it is crucial to assess the optimal curd coagulation time and, in general, the behavior of milk during renneting, especially in case of recipe modification, as these parameters are fundamental for both sensory properties and typical characteristics of cheese (Martin et al. 1997). Normally, these properties are evaluated both visually, by expert cheesemakers breaking manually a little fraction of the curd, or by laboratory analyses. Formagraph is one of the most used equipment to assess milk coagulation properties, able to describe rheological changes during renneting, such as modifications of curd firmness (Visentin et al. 2015). Another largely used instrument is Optigraph, a single wavelength near infrared (NIR) benchtop instrument that can provide results comparable to those obtained by Formagraph (Cipolat-Gotet et al. 2012). However, these techniques are not applicable on-line, thus they cannot provide information useful for a real time control of the process. Real time monitoring will give the possibility to reduce the number of subjective and/or complex analyses and, further, it can ensure a persistent final product quality (Kondakci & Zhou 2017; Henihan et al. 2018). NIR is a technique that can satisfy these requirements because it is able to assess the principal compounds involved in the process (Woodcock et al. 2008; Shao & He 2009), and to assure an efficient control of every stage of the process through the description of its trend (Grassi et al. 2014). Actually, NIR spectroscopy is a type of vibrational spectroscopy and, being fast, non-destructive, and non-invasive, it can be used for analyses on

the production line. Nearly any molecule containing CH, NH, SH, or OH bonds can be detected, and several constituents can be measured simultaneously. However, because of the wide, overlapping peaks and weak absorbances, chemometric techniques are required to extract the useful information (Nelson 2018). It has been demonstrated that the use of this technique could be extremely convenient for dairy industry. Indeed, by NIR analyses it is possible to assess quickly and efficiently the composition and the desired characteristics of cheese products, such as dry matter (Wittrup & Nørgaard 1998), as well as crude protein and fat content (Čurda & Kukačková 2004). Even aging, sensory attributes (Downey et al. 2005), and shelf life (Cattaneo et al. 2005) can be assessed by NIR approaches. There are also several works in which NIR spectroscopy is used to evaluate features and composition of dairy raw materials, namely milk (Kasemsumran et al. 2007) and milk powder (Cama-Moncunill et al. 2016). Besides, the development of NIR fiber optic probes to be placed directly into the coagulation vats, eliminating the need of sample pretreatment, allows to obtain real time information (Laporte et al. 1998).

When modifying a food recipe, process variables should be reassessed considering their high influence on the final product quality. To this aim, experimental design techniques are excellent tools to determine how process and product respond under different conditions and to assess the best operating settings. Data collected from designed experiments are usually examined by multi-factor Analysis of Variance (ANOVA) in order to evaluate whether the effect of each factor (and of factor-factor interactions) on the observed experimental variability could be deemed significant or not (Kirk 1982). Nevertheless, since ANOVA is a univariate method, it is not effective when applied on spectral data due to their inherent multivariate nature; in fact, the joint variability among different descriptors (covariance) must be considered to obtain comprehensive results. On the other hand, the systematic correlated variation in a multivariate dataset can be effectively

captured and summarized by Principal Component Analysis (PCA), through the projection of the observations onto a reduced (parsimonious) subspace of latent variables (Jackson 1980). Furthermore, when the variance provided by small bands is covered by variance of larger bands, an efficient approach is the interval-PCA (i-PCA), which permits to analyze small spectral ranges and to highlight the variability due to bands of interest, independently on the variance of the whole spectrum. However, since in the analysis of data coming from designed measurements PCA does not consider the underlying experimental scheme in parameter estimation, its use in such problems would not be effective without the support of other methods. Accordingly, several approaches coupling ANOVA decomposition with a bilinear description of the partitioned variance, such as MANOVA (Ståhle and Wold 1990), PC-ANOVA (Bratchell 1989), ANOVA-Simultaneous Component Analysis (ASCA) (Smilde et al. 2005), ANOVA-PCA (Harrington et al. 2005), ANOVA-Target projection (Marini et al. 2015), Regularized MANOVA (rMANOVA) (Engel et al. 2015), have been proposed in the literature for the analysis of multivariate data coming from designed experiments. However, MANOVA has been criticized due to the incapacity of handle datasets with a number of variables larger than samples. Similarly, the addition of the residual matrix to the effect matrices before PCA may result in a not completely straightforward interpretation of ANOVA-PCA models. The other methods have been developed to overcome these problems, and they have been used in several works (Imram 1999; Ullah and Jones 2015). In particular, ASCA allows to study the variance of data coming from an experimental design by splitting the variation and performing a Simultaneous Component Analysis (SCA), making possible to identify the most significant factors. First an ANOVA is carried out to obtain effect matrices from the response matrix of the design, and secondly a SCA is performed on the effect matrices (Jansen et al. 2005).

Obviously, to properly apply this method to spectral data, it is fundamental to choose the appropriate preprocessing techniques in order to minimize the

undesired variability (Grassi et al. 2017). The aim of this work is to study the effect of process variables on milk rennet coagulation by FT-NIR spectroscopy (FT-NIRs) coupled with i-PCA and ASCA methods. The proposed approach overcomes the existing PAT tools for quality assurance applied to the dairy industry (Woodcock et al. 2008; Henihan et al. 2018), thus filling a relevant knowledge gap in this field.

The hypothesis is that FT-NIRs can be a useful approach to provide dairy industry with an efficient methodology for process in-line monitoring and study of the operating condition contribution.

2. Materials and Methods

2.1. Design of Experiments

The study of milk renneting was carried out based on a 3-factor and 3-level Box-Behnken experimental design, including 12 trials and 5 replicates of the central point, performed in a random order to minimize the risk of systematic errors. The experimental factors and the levels taken into account were temperature (30 °C, 35 °C, 40 °C), milk fat content (0.1, 2.55, 5 g/100 mL) and pH (6.3, 6.5, 6.7). A schematic representation of the design is shown in Fig. 1.

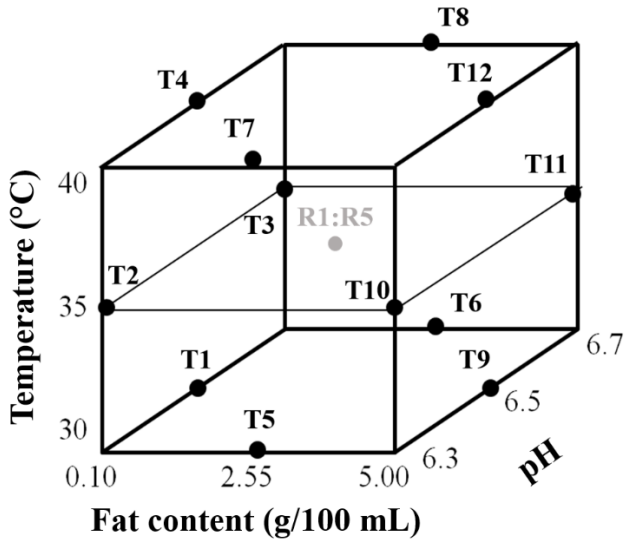


Fig. 1 Schematic representation of the Box-Behnken experimental design. The gray dot in the middle of the cube represents the five replicates (R) of the central point (temperature 40°C, milk fat content 2.55 g/100 mL, pH 6.5); experimental trials are indicated by the letter T followed by followed by an identification number

2.2. Milk preparation and coagulation

Fresh skimmed milk and fresh cream were suitably combined to obtain milk with different fat concentrations. Skimmed milk had a fat content of 0.1 g/100 mL, while cream, obtained by centrifugation, had a fat level of 35 g/100 mL. Skimmed milk-cream mixtures (100 mL) were poured in a Pyrex glass flacon and placed in a cold store room under stirring conditions on a magnetic plate for 12 h in order to obtain a homogeneous sample. Afterwards, samples were conditioned at 20 °C and added with $CaCl_2$ (final concentration, 3 µM). Citric acid (5 M) was used for pH correction to the desired value (monitored through a previously calibrated 3510 pH-meter, Jenway, Dunmow, England). To maintain the selected design temperatures (30 °C, 35 °C, 40 °C), samples were introduced in a thermostatic bath (MR Hei-Standard, Heidolph

Intruments GmbH, Schwabach, Germany). Then, 35 µL (175 IMCU/mL) of liquid rennet (Linea Rossa, Caseificio Clerici, Cadorago, Italy) composed of 75% chymosin and 25% bovine rennin, were added and coagulation was monitored for 40 min.

2.3. FT-NIR spectroscopy

Milk renneting was monitored by a FT-NIR spectrometer (MPA, Bruker Optics, Milan, Italy) through a fiber optic probe equipped with a transflectance adapter (1 mm pathlength) inserted directly in the sample. Spectra were collected every 60 s over the 12500 – 4000 cm^{-1} range, with 64 scans for both sample and background and a nominal resolution of 8 cm^{-1}. Instrument control was managed by using the OPUS software (v. 6.0 Bruker Optics, Milan, Italy).

2.4. Data analysis

Data preprocessing, PCA, i-PCA and ASCA models were performed with routines and toolboxes implemented in Matlab environment (the Mathworks Inc., Natick, MA, USA).

2.4.1. i-PCA

Data exploration was applied to extract useful information, linked to the tested experimental factors, about the behavior changes in the different coagulation phases. The spectra obtained from each trial were organized in as many datasets (40x2203); furthermore, a dataset containing the spectra of all trials (680x2203) was built. A PCA was performed to choose spectral ranges to be further considered. Lately, interval-PCA method was applied to extract relevant information from smaller and most significant parts of the spectrum and to exclude ranges that may contain noise and undesired signal. To this purpose, no matter the considered dataset, the FT-NIR spectra were divided in three different regions, each of them submitted to PCA. In particular, the regions between 7180 cm^{-1} and 6464 cm^{-1} and between 5823 cm^{-1} and 4000 cm^{-1} were discarded. The first range went from 12500 cm^{-1} to 9200 cm^{-1}, the

second from 9199 cm^{-1} to 7181 cm^{-1} and the third from 6463 cm^{-1} to 5824 cm^{-1}. Prior to i-PCA, spectral ranges were pretreated by standard normal variate (SNV), smoothing (Savitzky-Golay method, filter width: 9 points; polynomial order: 1) and mean centering.

2.4.2. ASCA

ASCA (Jansen et al. 2005) was used to detect possible significant effects of the experimental factors and of their interactions on the FT-NIR spectral profiles. In particular, to fully characterize the evolution of the coagulation and the role of the investigated factors (temperature, fat content and pH) across the process, time was included as the fourth design factor. Therefore, 10 different time levels, corresponding to a spectrum collected every 4 min, were considered, obtaining a final design matrix whose dimensions were 130x4.

In order to apply ASCA to a balanced set of measurements, for the five replicates of the central point five datasets were created, each containing one replicate along with the other 12 trials in the 12500 - 7181 cm^{-1} and 6463 - 5824 cm^{-1} ranges, i.e. merging the spectral ranges considered for i-PCA. The dataset for ASCA had thus the dimensions of 130x1545. ASCA was performed using the same pretreatments described in § 2.4.1 and the significance of the effects of each design term was tested by means of permutation tests with 1000 randomizations.

3. Results and Discussion

3.1. FT-NIR spectra

To better understand the influence of the three experimental factors (temperature, milk fat content, and pH) over the coagulation time, SNV pretreated spectra of the beginning (after 1 min), the transition phase (coagulum formation) and the end (40 min) of the coagulation averaged according to fat content, pH, or temperature for each considered time are

shown in Fig. 2. The different lines refer to different acquisition times: solid lines represent the averaged factor level spectra, acquired 1 minute after the beginning of the coagulation trial; dashed lines are averaged factor level spectra collected during the formation of the coagulum (transition phase of coagulation process); the dotted lines represent the averaged factor level spectra that have been acquired at the end of the process (40 min). Fig. 2 (a,b and c) show how temperature mostly affects the spectral ranges 12500-9000 cm^{-1} and 7000 - 5500 cm^{-1}, but mainly in a longitudinal fashion (i.e., across time). Indeed, the relative absorbance decreases with coagulation time in the range 12500-9000 cm^{-1} (Fig. 2b), whereas it increases in the region between 7000 and 5500 cm^{-1} (Fig. 2c). Furthermore, spectra collected after 40 min at 35 and 40 °C are almost identical, whereas the ones corresponding to 30 °C are slightly different in terms of absorbance. Differences among spectra acquired from skimmed milk (0.1 g/100 mL of fat) and milk samples with 2.55 g/100 mL and 5 g/100 mL of fat are highlighted in Fig. 2 (d, e and f). Large differences can be noticed between the coagulation trend of skimmed milk averaged spectra and samples with higher fat content in the region between 10800 and 9000 cm^{-1} (Fig 2e). The most relevant difference can be noticed examining the bands at 4332 and 4258 cm^{-1} (Fig. 2f) present in case of milk samples with fat concentration higher than 2.55 g/100 mL and directly linked to the fat absorbance (Brandao et al. 2010; Núñez-Sánchez et al. 2016). As far as pH is concerned, in Fig. 2 (g) it is possible to see that there are no visible differences between spectra collected at the beginning of the coagulation process at pH values of 6.5 and 6.7, whereas in the following times the differences are enhanced. In particular, relative absorbance decreases in the region 12500 – 15000 cm^{-1} along with coagulation progress for all the tested pH (Fig. 2h). Moreover, spectra collected at the beginning of the coagulation at pH 6.7 and at pH 6.5 show different spectral shape in correspondence of band with maxima at 5150 cm^{-1} (Fig.2 i).

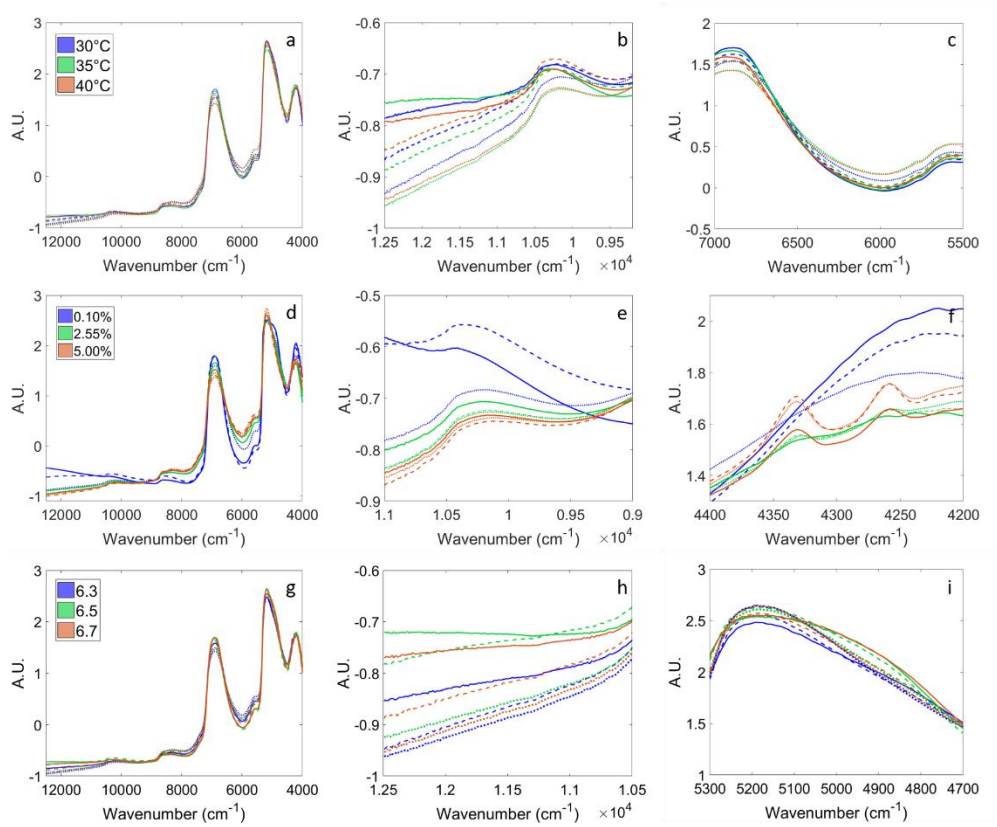


Fig. 2 SNV pretreated FT-NIR spectra of the beginning (solid lines), transition phase (dashed lines) and end (dotted lines) of milk renneting, averaged according to temperature (a, b and c), fat content (d, e and f), and pH (g, h and i). Different colors refer to a different level of the considered factor: blue lines, lowest level; green lines, medium level; orange lines, highest level

Even if some differences are visible in the spectra obtained with different experimental conditions and over coagulation time, it is difficult to get a perfect correlation between the absorption at a single wavenumber and the concentration of each milk component. Indeed, all the main milk components, such as fat, proteins, and water, absorb in the NIR region as they are constituted of C–H, N–H, O–H and C=O bonds, which arise bands between 12500 and 4000 cm^{-1} (Workman & Weyer 2007). Thus, molecule absorptions in the NIR region are overtones and combinations of fundamental vibrations,

resulting in broad and overlapped bands. Some attempts of chemical band assignment have been reported in literature. Bands at 10400, 6900, and 5150 cm^{-1} can be ascribed to the O-H first overtone of water and O-H combination bands. Signals at 5700 cm^{-1} are linked to the presence of lactic acid and lactose (Workman and Weyer 2007; Wang et al. 2015). Other relevant bands were found at 10800 and 8600 cm^{-1}, ascribable to the lipid C-H bonds (Tsenkova et al. 2000). A review by Holroyd (2013) deeply investigated band assignments in liquid milk. From a broad literature survey, the author assigned the protein N-H absorption to the regions at 1100-9700 cm^{-1}, 5690-5800 cm^{-1}, 4550-4350 cm^{-1}, and around 4300 cm^{-1}, and the lipid O-H and N-H absorptions to around 4800 cm^{-1} and 4200 cm^{-1} (Holroyd 2013).

Moreover, NIR response is affected by light scattering that is a physical effect due to radiation redistribution inside a medium characterised by specific microstructural properties. It has been previously investigated how light scattering affects the whole NIR range (Cattaneo et al. 2009). Scattering phenomena in milk are largely due to size and number of suspended fat globules (Cabassi et al. 2013). To a smaller extent, casein micelles are responsible for the increase of the bulk scattering coefficient in NIR spectra (Aernouts et al. 2015). In particular, Aernouts et al. (2015) found for both content of fat globules and casein micelles a negative correlation with the water absorption band at 6900 cm^{-1}, while a positive correlation characterised the NIR range between 12500 and 9000 cm^{-1}.

Due to the complex nature of NIR signal, multivariate approaches are required to better evaluate the differences among spectra and to assess the influence of the experimental factors and of their interactions on the coagulation process.

3.2. Occurrence and trends of milk renneting phases by iPCA

Prior to i-PCA, PCA was performed on the whole spectral range to discard spectral regions irrelevant to monitor the coagulation process. Fat bands at

4332 and 4258 cm^{-1}, even if relevant to evaluate fat influence in the milk coagulation performance, were discarded because they covered the greatest part of data variance. Then, for each of the 17 trials, three different PCA were carried out, one for every spectral range considered. Successively, the scores resulting from each i-PCA model were normalized from 0 to 1 in order to make a comparison among them possible. In parallel, a PCA was made on the block-wise augmented dataset containing all the trials; also in this case, the scores were extracted and normalized. Fig. 3 shows the score trends obtained from the single i-PCA models of all the trials: PC1 scores obtained from the first range models (solid lines) can be selected to describe the liquid behavior of milk, linked to the first phase of the coagulation process when the coagulum has not yet been formed (Grassi et al. 2014). In fact, in all the trials it is possible to see a decrease of PC1 scores with time evolution, indicating a progressive decrement of the liquid phase and the beginning of the curd formation. Trends highlighted by the dashed lines, related to the PC2 scores obtained from the second range models, can be ascribed to the second phase of the coagulation process, when the clot begins to form. In particular, the peak of these profiles corresponds to the coagulation time.

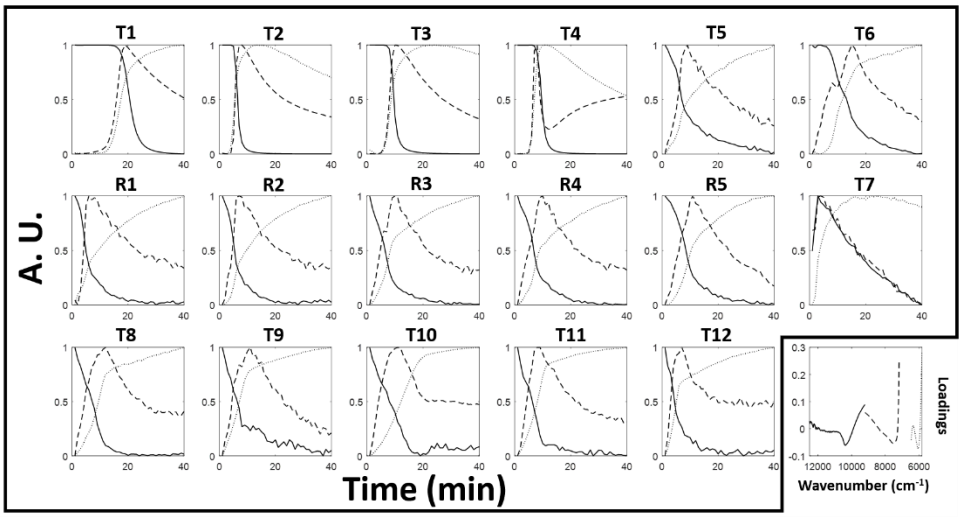


Fig. 3 Score profiles obtained from the i-PCA models calculated for each milk renneting trial. Solid lines, PC1 scores of the 12500 to 9200 cm^{-1}range; dashed lines, PC2 scores of the 9199 to 7181 cm^{-1} range; dotted lines, PC1 scores of the 6463 to 5824 cm^{-1}range. The lowest right block shows the loadings of the third central point replicate (R3) for each of the three i-PCA models: solid lines, PC1 loadings related to first range model; dashed lines, PC2 loadings related to second range models; dotted lines, PC1 loadings related to third range models

Trends and phase occurrence times are strongly related to temperature, fat, and pH levels. High temperatures and low pH values allow to reach this point faster (Zoon et al. 1988; Sbodio et al. 2002), as confirmed by T7 (Fig. 3), even if this trial presents overlapped trends for the first and second spectral range. Moreover, the comparison of renneting trials carried out with the same milk fat content confirms that, at lower temperature and higher pH values, the transition phase is delayed; this can be observed, for instance, for the first four trials (T1-T4, fat level = 0.1 g/100 mL). Furthermore, the trials with longer coagulation times are T1 and T6, and this is possibly caused by the lowest temperature level (30 °C) combined with the higher pH values (6.5 and 6.7).

The PC1 scores of the i-PCA models calculated with the third spectral range look promising in evaluating the last phase of the coagulation process, when the curd eventually reaches the maximum of its consistency (dotted lines in Fig. 3). These scores present a reverse trend in comparison with the ones related to the first range, confirming that they describe the solid-like behavior of milk during renneting. Three of the four trials carried out with a milk fat content of 0.1 g/100 mL (T1, T2, and T3) show a decreasing trend after 20 min, when the coagulum is completely formed, and this could be explained by the disruption of the curd resulting in a decrease of the solid component. However, further analysis must be carried out to have a more reliable explanation of this phenomenon. Also fat content affects the trends behavior

of these scores: indeed, the higher the fat content in milk samples, the higher the noise.

In the lower right part of Fig. 3, the PCA loadings related to one of the replicates of the central point (R3) are reported as an example. The band with the largest influence on the scores linked to the first spectral range is at 10400 cm^{-1}, assigned to the stretching of O-H bond in water. Concerning loadings of the second spectral range, the most important wavenumbers describing sample variability are between 7400 and 7181 cm^{-1}, connected to combination bands of C-H bonds of fatty acids and carbohydrates (Subramanian et al. 2011). Lastly, scores of the third spectral range models are mostly influenced by the region between 6100 – 5824 cm^{-1}, ascribed to the presence of lactic acid and lactose (Wang et al. 2015).

Scores of the models obtained from the i-PCA performed on the block-wise augmented dataset containing all the trials were studied in order to assess if a single model is able to give the same results of the models based on separated trials. Similar results were obtained for the first and the third coagulation phases (using the PC1 scores of the first and third range, respectively), but PC2 scores of the second spectral range model could not be used to describe the transition phase (results not shown).

These results reveal that i-PCA on FT-NIRs data was able to discriminate the three different phases of the rennet coagulation process already described by Grassi et al. (2014) during lactic acid fermentation of yoghurt. Thus, i-PCA can be used to efficiently describe and control milk renneting under different operating conditions.

3.3 Investigation of the effect of experimental factors by ASCA

ASCA was performed with the aim of verifying if factors considered in the experimental design and their interactions have a significant influence on milk spectral profiles during renneting. Since five datasets, one for each of the different replicates of the central point, were investigated with ASCA, only the

results of one of them are reported and commented in this work because the results obtained from the other datasets were extremely comparable.

The first step of ASCA is the decomposition of the total data variability into the individual contributions of as many effect matrices as the number of terms in the design. In particular, since time was also included as experimental factor (see § 2.4.2), variability in the spectral dataset was split into 16 arrays: 4 accounting for the main effect of the experimental factors, 6 corresponding to the effect of all possible two-way interactions, 4 describing the effect of three-way interactions, 1 for the effect of the only possible four-way interaction, and 1 for the residuals. However, based on the aim of the present work, the successive stages of the investigation were limited only to the contribution of the main effects and the two-way interactions. First of all, the significance of the effects of experimental factors and their interactions was assessed.

Permutation tests were performed to compare the experimental sum of squares for the effect matrices of the four main factors as well as “time x temperature” and “time x fat” interactions with their corresponding distributions under the null hypothesis.

Results concerning the other interactions are not reported, because they show the same pattern, meaning that all the effects are significant for the spectral profile trend description.

Interpretation of the effects of the significant terms on the multivariate spectral profiles was accomplished through a simultaneous component analysis performed on each factor (or interaction) matrix. The histograms accounting for the score distribution on SC1 (explaining more than 93% of the total variability) for the three Box-Behnken experimental factors (temperature, fat content, and pH) after back projection of the residuals are represented in Figs. 4a, 4c, and 4e, whereas Figs. 4b, 4d, and 4f show the corresponding loadings together with their 95% confidence interval. The clear separation of the score

distribution in the histograms is a further confirmation of the significant differences between the different levels of each experimental factor. Particularly, scores related to milk with 0.1 g/100 mL of fat are very far from the distributions of the other two levels, suggesting a relevant difference among those samples. The relevant effect of fat globules on the rheological properties of the rennet gels has been reported also by Logan et al. (2014). For each design term investigated, in order to evaluate which are the regions of the FT-NIR spectrum which are mostly affected by a particular factor or interaction, the corresponding loadings were inspected. To this purpose, for each SCA model, the 95% confidence interval around each loading vector was calculated by a nonparametric bootstrap procedure, as reported by De Luca et al. (2016). In Figs. 4b, 4d, and 4f, the statistically significant spectral regions are represented by a solid line, whereas a dotted line indicates the parts of the signal associated to loadings statistically indistinguishable from zero. As far as the models associated to the effects of milk fat content and pH, the whole spectral range is significant, and the loading profiles are almost equal (Figs. 4d and 4f). On the contrary, the important ranges for temperature are 6250 - 6450 cm^{-1}, 7200 - 8800 cm^{-1}, and 10400 - 11600 cm^{-1}; they can be mainly ascribed to the absorption of proteins, fat, and water. Obviously, their variation is influenced by the aggregation degree of casein micelles. Temperature loadings have a reverse trend if compared to the fat and pH ones (Fig. 4b, 4d and 4f), suggesting that this factor has an opposite influence on the process with respect to the other two.

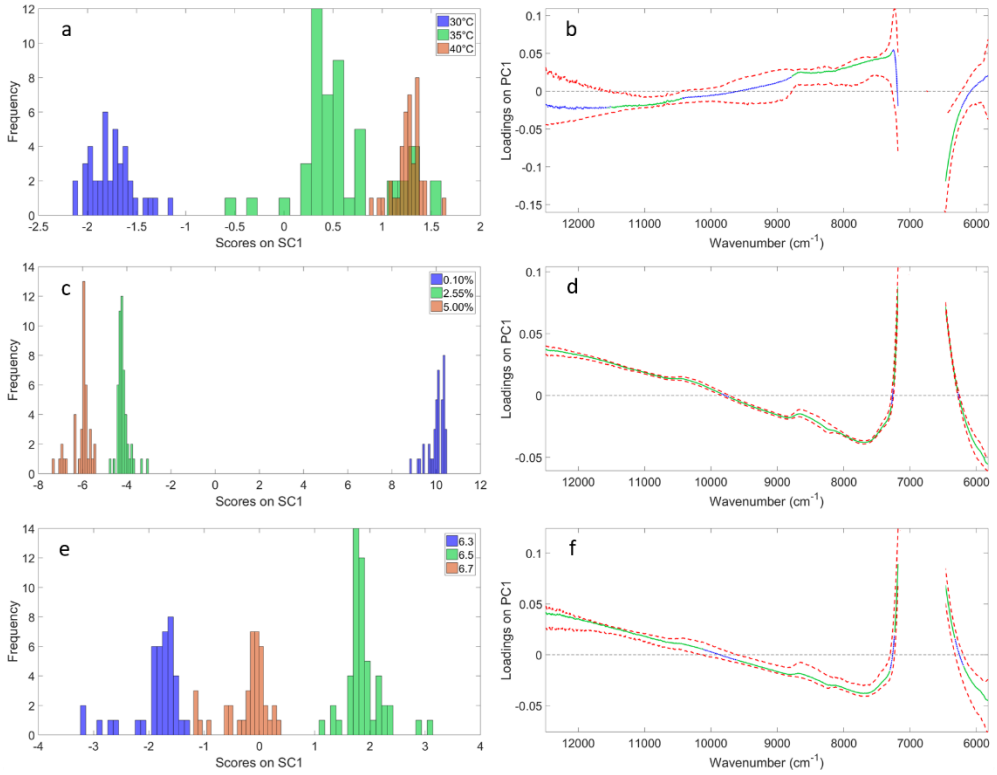


Fig. 4 Histograms of ASCA score frequency with projected residuals along SC1 for the different levels of experimental factors considered in the study of milk renneting: temperature (a), fat content (c), pH (e). The corresponding loadings are shown in panels b, d, and f (black lines) with 95% confidence interval (dashed gray lines). Statistically non-significant regions are represented with a black dotted line in the loading plots

Time was evaluated projecting scores on SC1 (96.69%) and SC2 (2.95%) with the corresponding residuals, as shown in Fig. 5a. Spectra of samples acquired at the beginning of coagulation highly differ from the ones collected at the end, in agreement with the results obtained by i-PCA. Moreover, scores on SC1 show a continuous decrease over time with a clear slowing down after 20 min, suggesting that the major changes occur at the beginning of the process, as confirmed by Fig. 5b. This trend is similar to the one obtained from scores on PC1 of first range i-PCA models (Fig. 3), which described the

liquid behavior of milk during the coagulation process. Besides, the SC2 score pattern resembles the trend of PC2 scores obtained from models related to the second range of the i-PCA. Indeed, they show a maximum value in correspondence to the transition time; then, they start to decrease again until reacquiring negative values. Fig. 5c and 5d show the loading plots for SC1 and SC2, respectively, with the corresponding 95% confidence intervals. Also in this case, for both the SCs, the whole spectral region results statistically significant in describing the effect of time on milk renneting, thus confirming that the selected ranges (12500 - 9200 cm^{-1}; 9199 - 7181 cm^{-1} and 6463 - 5824 cm^{-1}) are the ones to be considered for process monitoring.

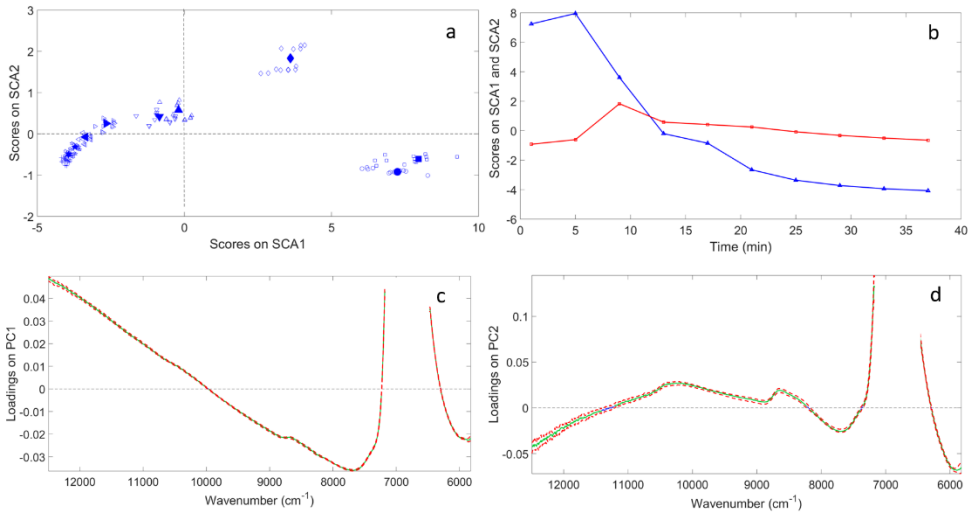


Fig. 5 a) Score plot for the effect of renneting time (filled symbols) with projected residuals (empty symbols); b) SC1 (blue line) and SC2 (red line) score profiles along renneting time; c) loadings for SC1 (green line) with 95% confidence interval (dashed red lines); d) loadings for SC2 (green line) with 95% confidence interval (dashed red lines). Statistically non-significant regions are indicated with a blue dotted line in the loading plots

4. Conclusions

In the present study the possibility of assessing the influence on milk renneting of different process conditions, i.e. temperature, milk fat content, and pH, was addressed by coupling chemometric techniques with FT-NIRs. Interval-PCA confirmed the ability of FT-NIRs in discriminating the three different phases of the renneting process. Indeed, it was possible to model the phase before the coagulum formation by PC1 scores obtained from the first spectral range (12500 - 9200 cm^{-1}); the trends of PC2 scores of the second range (9199 - 7181 cm^{-1}) well modelled the clotting beginning; the PC1 scores of the third spectral range (6463 - 5824 cm^{-1}) looked promising to describe the last phase of the coagulation process. Moreover, a strong effect of temperature, fat, and pH levels was highlighted by i-PCA trends and times of phase occurrence.

ASCA applied to the spectral data assessed that the effects of experimental factors and their interactions were statistically significant. In particular, the simultaneous component analysis clearly demonstrated that milk samples with the lowest fat content (0.1 g/100 mL) had a coagulation behavior significantly different from that of the other samples. Furthermore, the loadings evaluation, after a nonparametric bootstrap procedure, confirmed that the spectral ranges selected for i-PCAs are the strategic ones for milk renneting monitoring.

As assumed in the initial hypothesis, FT-NIRs, coupled with i-PCA and ASCA methods, demonstrated to be a valid approach to study the different phases of the renneting process and to assess the effect of temperature, fat content and pH. This study does the groundwork for the assessment of process parameter effects, thus giving to the dairy industry the opportunity of monitoring and studying the coagulation process when developing new strategies to obtain healthier dairy products.

Conflict of interest

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

Aernouts, B., Van Beers, R., Watté, R., Huybrechts, T., Lammertyn, J., & Saeys, W. (2015). Visible and near-infrared bulk optical properties of raw milk. *Journal of Dairy Science, 98,* 6727-6738.

Brandao, M. C. P., Carmo, A., Bell, M. J. V, & Anjos, V. C. (2010). Characterization of milk by infrared spectroscopy. *Revista do Instituto de Laticicinos "Cândido Tostes", 373*(65), 30–33.

Bratchell, N. (1989). Multivariate response surface modelling by principal component analysis. *Journal of Chemometrics, 3,* 579–588.

Cabassi, G., Profaizer, M., Marinoni, L., Rizzi, N., & Cattaneo, T. M. (2013). Estimation of fat globule size distribution in milk using an inverse light scattering model in the near infrared region. *Journal of Near Infrared Spectroscopy, 21,* 359-373.

Cama-Moncunill, R., Markiewicz-Keszycka, M., Dixit, Y., Cama-Moncunill, X., Casado-Gavalda, M. P., Cullen, P. J., & Sullivan, C. (2016). Multipoint NIR

spectroscopy for gross composition analysis of powdered infant formula under various motion conditions. *Talanta*, *154*, 423-430.

Cattaneo, T. M., Cabassi, G., Profaizer, M., & Giangiacomo, R. (2009). Contribution of light scattering to near infrared absorption in milk. *Journal of Near Infrared Spectroscopy, 17,* 337-343.

Cattaneo, T. M., Giardina, C., Sinelli, N., Riva, M., & Giangiacomo, R. (2005). Application of FT-NIR and FT-IR spectroscopy to study the shelf-life of Crescenza cheese. *International Dairy Journal*, *15*, 693-700.

Cipolat-Gotet, C., Cecchinato, A., De Marchi, M., Penasa, M., & Bittante, G. (2012). Comparison between mechanical and near-infrared methods for assessing coagulation properties of bovine milk. *Journal of Dairy Science*, *95*, 6806-6819.

Čurda, L., & Kukačková, O. (2004). NIR spectroscopy: a useful tool for rapid monitoring of processed cheeses manufacture. *Journal of Food Engineering*, *61*, 557-560.

De Luca, S., De Filippis, M., Bucci, R., Magrì, A. D., Magrì, A. L., & Marini, F. (2016). Characterization of the effects of different roasting conditions on coffee samples of different geographical origins by HPLC-DAD, NIR and chemometrics. *Microchemical Journal, 129,* 348–361.

Downey, G., Sheehan, E., Delahunty, C., O'Callaghan, D., Guinee, T., & Howard, V. (2005). Prediction of maturity and sensory attributes of Cheddar cheese using near-infrared spectroscopy. *International Dairy Journal*, *15*, 701-709.

Engel, J., Blanchet, L., Bloemen, B., Van den Heuvel, L. P., Engelke, U. H. F., Wevers, R. A., & Buydens, L. M. C. (2015). Regularized MANOVA (rMANOVA) in untargeted metabolomics. *Analytica Chimica Acta*, *899,* 1–12.

Grassi, S., Alamprese, C., Bono, V., Casiraghi, E., & Amigo, J. M. (2014). Modelling milk lactic acid fermentation using Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS). *Food and Bioprocess Technology, 7,* 1819–1829.

Grassi, S., Lyndgaard, C. B., Rasmussen, M. A., & Amigo, J. M. (2017). Interval ANOVA simultaneous component analysis (i-ASCA) applied to spectroscopic data to study the effect of fundamental fermentation variables in beer fermentation metabolites. *Chemometrics and Intelligent Laboratory Systems*, *163,* 86–93.

Harrington, P. de B., Vieira, N. E., Chen, P., Espinoza, J., Nien, J. K., Romero, R., & Yergey, A. L. (2006). Proteomic analysis of amniotic fluids using analysis of variance-principal component analysis and fuzzy rule-building expert systems applied to matrix-assisted laser desorption/ionization mass spectrometry. *Chemometrics and Intelligent Laboratory Systems*, *82*(S1–2), 283–293.

Henihan, L. E., O'Donnell, C. P., Esquerre, C., Murphy, E. G., & O'Callaghan, D. J. (2018). Quality assurance of model infant milk formula using a front-face fluorescence process analytical tool. *Food and Bioprocess Technology, 11*, 1402-1411.

Holroyd, S. E. (2013). The use of near infrared spectroscopy on milk and milk products. *Journal of Near Infrared Spectroscopy, 21,* 311-322.

Imram, N. (1999). Visual texture perception in formulated chilled dairy desserts. *British Food Journal*, *101*, 22–31.

Jackson, J. (1980). Principal components and factor analysis: part I-principal components. *Journal of Quality Technology, 12*(4), 201-213.

Jansen, J. J., Hoefsloot, H. C. J., van der Greef, J., Timmerman, M. E., Westerhuis, J. A., & Smilde, A. K. (2005). ASCA: analysis of multivariate data

obtained from an experimental design. *Journal of Chemometrics*, *19,* 469–481.

Johnson, M. E., Chen, C. M., & Jaeggi, J. J. (2001). Effect of rennet coagulation time on composition, yield, and quality of reduced-fat cheddar cheese. *Journal of Dairy Science, 84,* 1027–1033.

Kasemsumran, S., Thanapase, W., & Kiatsoonthon, A. (2007). Feasibility of near-infrared spectroscopy to detect and to quantify adulterants in cow milk. *Analytical Sciences*, *23*, 907-910.

Kirk, R. E. (1982). Experimental design. Hoboken: John Wiley & Sons.

Kondakci, T., & Zhou, W. (2017). Recent applications of advanced control techniques in food industry. *Food and Bioprocess Technology, 10,* 522-542.

Laporte, M. F., Martel, R., & Paquin, P. (1998). The near-infrared optic probe for monitoring rennet coagulation in cow's milk. *International Dairy Journal, 8,* 659–666.

Logan, A., Day, L., Pin, A., Auldist, M., Leis, A., Puvanenthiran, A., & Augustin, M. A. (2014). Interactive effects of milk fat globule and casein micelle size on the renneting properties of milk. *Food and Bioprocess Technology, 7,* 3175-3185.

Marini, F., de Beer, D., Joubert, E., & Walczak, B. (2015). Analysis of variance of designed chromatographic data sets: The analysis of variance-target projection approach. *Journal of Chromatography A, 1405,* 94–102.

Martin, B., Chamba, J. F., Coulon, J. B., & Perreard, E. (1997). Effect of milk chemical composition and clotting characteristics on chemical and sensory properties of Reblochon de Savoie cheese. *Journal of Dairy Research, 64,* 157–162.

Nelson, D. L. (2018). Introduction to spectroscopy. In A. S. Franca & L. Nollet (Eds.), Spectroscopic methods in food analysis (pp. 3-33). Boca Raton: CRC Press.

Núñez-Sánchez, N., Martínez-Marín, A. L., Polvillo, O., Fernández-Cabanás, V. M., Carrizosa, J., Urrutia, B., & Serradilla, J. M. (2016). Near Infrared Spectroscopy (NIRS) for the determination of the milk fat fatty acid profile of goats. *Food Chemistry, 190,* 244–252.

Sbodio, O. A., Tercero, E. J., Coutaz, R., & Martinez, E. (2002). Optimizing processing conditions for milk coagulation using the hot wire method and response surface methodology. *Journal of Food Science, 67,* 1097–1102.

Shao, Y., & He, Y. (2009). Measurement of soluble solids content and pH of yogurt using visible/near infrared spectroscopy and chemometrics. *Food and Bioprocess Technology, 2,* 229-233.

Smilde, A. K., Jansen, J. J., Hoefsloot, H. C. J., Lamers, R. J. A. N., van der Greef, J., & Timmerman, M. E. (2005). ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics, 21,* 3043–3048.

Ståhle, L., & Wold, S. (1990). Multivariate analysis of variance (MANOVA). *Chemometrics and Intelligent Laboratory Systems, 9,* 127–141.

Subramanian, A., Prabhakar, V., & Rodriguez-Saona, L. (2011). Analytical methods: Infrared spectroscopy in dairy analysis. In Encyclopedia of dairy sciences (2nd ed., pp. 115–124). Cambridge: Academic Press.

Tsenkova, R., Atanassova, S., Itoh, K., Ozaki, Y., & Toyoda, K. (2000). Near infrared spectroscopy for biomonitoring: cow milk composition measurement in a spectral region from 1,100 to 2,400 nanometers. *Journal of Animal Science, 78,* 515-522.

Ullah, I., & Jones, B. (2015). Regularised manova for high-dimensional data. *Australian and New Zealand Journal of Statistics, 57,* 377–389.

Visentin, G., McDermott, A., McParland, S., Berry, D. P., Kenny, O. A., Brodkorb, A., Fenelon, M. A. & De Marchi, M. (2015). Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. *Journal of Dairy Science*, *98*, 6620-6629.

Wang, Y., Ding, W., Kou, L., Li, L., Wang, C., & Jurick, W. M. (2015). A non-destructive method to assess freshness of raw bovine milk using FT-NIR spectroscopy. *Journal of Food Science and Technology, 52,* 5305–5310.

Wittrup, C., & Nørgaard, L. (1998). Rapid near infrared spectroscopic screening of chemical parameters in semi-hard cheese using chemometrics. *Journal of Dairy Science*, *81*, 1803-1809.

Woodcock, T., Fagan, C. C., O'Donnell, C. P., & Downey, G. (2008). Application of near and mid-infrared spectroscopy to determine cheese quality and authenticity. *Food and Bioprocess Technology, 1,* 117-129.

Workman, J., & Weyer, L. (2007). Practical guide to interpretive Near-Infrared Spectroscopy. Boca Raton: CRC Press.

Zoon, P., T. van Vliet, & P. Walstra. (1988). Rheological properties of rennet-induced skim milk gels. 2. The effect of temperature. *Netherlands Milk and Dairy Journal*, *42,* 271-294.

4.3 PDF PAPER III

Food Control 119 (2021) 107494

Contents lists available at ScienceDirect

Food Control

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

Effect of physicochemical factors and use of milk powder on milk rennet-coagulation: Process understanding by near infrared spectroscopy and chemometrics

Lorenzo Strani [a], Silvia Grassi [a,*], Cristina Alamprese [a], Ernestina Casiraghi [a], Roberta Ghiglietti [b], Francesco Locci [b], Nicolò Pricca [b], Anna De Juan [c]

[a] Department of Food, Environmental, and Nutritional Sciences (DeFENS), Università degli Studi di Milano, Via G. Celoria 2, 20133, Milan, Italy
[b] CREA-ZA, Research Centre for Animal Production and Aquaculture, Via A. Lombardo 11, 26900, Lodi, Italy
[c] Chemometrics Group. Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona, Diagonal 645, 08028, Barcelona, Spain

ARTICLE INFO

Keywords:
Milk powder
Coagulation
Near infrared spectroscopy
Multivariate curve resolution
Alternating least squares
Rheological properties

ABSTRACT

The effect of physicochemical factors and use of skim milk powder on milk rennet-coagulation was investigated combining near infrared (NIR) spectroscopic monitoring and Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS). Coagulum formation has been studied by reference approaches (Formagraph and fundamental rheology) and with NIR spectroscopy on unaltered reconstituted milk samples, pasteurized samples, samples with calcium chloride addition and samples of reconstituted milk mixed with fresh milk. The MCR-ALS models successfully described the process evolution, explaining more than 99.9% of variance. The MCR-ALS profiles revealed to be significantly directly correlated with Formagraph and rheological data ($p < 0.001$) and allowed assessing the significant effect ($p < 0.05$) of the milk powder type on the coagulation occurrence and the non-significance ($p > 0.05$) of the $CaCl_2$ concentration level added and the heat treatment applied. The MCR-ALS models calculated for the coagulation trials of pasteurized skimmed milk mixed with reconstituted milk samples highlighted shorter coagulation times with the increasing of reconstituted milk amount (from 4.3-6.6 min to 2–5 min). Profiles extracted from MCR-ALS models developed for a wide range of coagulation conditions proved to be suitable non-destructive, non-invasive and on-line tools to evaluate the rennet-induced coagulation of reconstituted milks.

* Corresponding author. Address: Via Celoria 2, 20133, Milan, Italy.
E-mail addresses: lorenzo.strani@unimi.it (L. Strani), silvia.grassi@unimi.it (S. Grassi), cristina.alamprese@unimi.it (C. Alamprese), ernestina.casiraghi@unimi.it (E. Casiraghi), roberta.ghiglietti@crea.gov.it (R. Ghiglietti), francesco.locci@crea.gov.it (F. Locci), nicolo.pricca@gmail.com (N. Pricca), anna.dejuan@ub.edu (A. De Juan).

https://doi.org/10.1016/j.foodcont.2020.107494
Received 21 April 2020; Received in revised form 16 July 2020; Accepted 17 July 2020
Available online 25 July 2020

PAPER III *(submitted version)*

Effect of physicochemical factors and use of milk powder on milk rennet-coagulation: process understanding by near infrared spectroscopy and chemometrics

Lorenzo Strani[1], Silvia Grassi*[1], Cristina Alamprese[1], Ernestina Casiraghi[1], Roberta Ghiglietti[2], Francesco Locci[2], Nicolò Pricca[2], Anna De Juan[3]

[1]Department of Food, Environmental, and Nutritional Sciences (DeFENS), Università degli Studi di Milano, via G. Celoria 2, 20133 Milan, Italy

[2]CREA-ZA, Research Centre for Animal Production and Aquaculture, via A. Lombardo 11, 26900 Lodi, Italy

[3]Chemometrics group. Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain

* Corresponding author: Silvia Grassi, e-mail address: silvia.grassi@unimi.it, Address: Via Celoria 2, 20133 Milan, Italy

Authors' e-mail:

Lorenzo Strani, lorenzo.strani@unimi.it

Cristina Alamprese, cristina.alamprese@unimi,it

Ernestina Casiraghi, ernestina.casiraghi@unimi.it

Roberta Ghiglietti, roberta.ghiglietti@crea.gov.it

Francesco Locci, francesco.locci@crea.gov.it

Nicolò Pricca, nicolo.pricca@gmail.com

Anna De Juan, anna.dejuan@ub.edu

Received: 21 April 2020; Accepted: 17 July 2020; Available online: 25 July 2020

Abstract

The effect of physicochemical factors and use of skim milk powder on milk rennet-coagulation was investigated combining near infrared (NIR) spectroscopic monitoring and Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS). Coagulum formation has been studied by reference approaches (Formagraph and fundamental rheology) and with NIR spectroscopy on unaltered reconstituted milk samples, pasteurized samples, samples with calcium chloride addition and samples of reconstituted milk mixed with fresh milk. The MCR-ALS models successfully described the process evolution, explaining more than 99.9% of variance. The MCR-ALS profiles revealed to be significantly directly correlated with Formagraph and rheological data ($p < 0.001$) and allowed assessing the significant effect ($p < 0.05$) of the milk powder type on the coagulation occurrence and the non-significance ($p > 0.05$) of the $CaCl_2$ concentration level added and the heat treatment applied. The MCR-ALS models calculated for the coagulation trials of pasteurized skimmed milk mixed with reconstituted milk samples highlighted shorter coagulation times with the increasing of reconstituted milk amount (from 4.3-6.6 min to 2-5 min). Profiles extracted from MCR-ALS models developed for a wide range of coagulation conditions proved to be suitable non-destructive, non-invasive and on-line tools to evaluate the rennet-induced coagulation of reconstituted milks.

Keywords: Milk powder; Coagulation; Near Infrared Spectroscopy; Multivariate Curve Resolution - Alternating Least Squares; Rheological properties

1. Introduction

From an economic and technological point of view, skimmed milk powder (SMP) is pivotal for countries in which milk production is scarce or absent due to climatic and high cost issues (Kjaergaard-Jansen, 1990; Lelievret, Shaker, & Taylor, 1991; Omar & Buchheim, 1983; Písecký, 2005), but also for top milk-producing countries. Indeed, SMP addition is relevant for protein content standardization in milk intended for cheesemaking (Lin, Kelly, O'Mahony, & Guinee, 2017; Písecký, 2005) as well as for the production cost saving if fresh milk price increases (Papadatos, Berger, Pratt, & Barbano, 2002).

SMPs differ mainly due to the technology used for their production. In particular, depending on the severity of the heat treatment and on the milk pH, the dehydration process affects the extent of whey protein denaturation and the binding of denatured whey protein to the casein micelles (Singh & Waungana, 2001). These changes affect cheese processing and the use of reconstituted skim milk may cause problems, such as longer coagulation time, slower syneresis, and formation of weaker or finer curd (Singh & Waungana, 2001). These effects may modify typical texture, ripening and functionality of the resulting cheese (Gulati et al., 2019; Rynne, Beresford, Kelly, & Guinee, 2004; Singh & Waungana, 2001). Several studies (Gastaldi, Pellegrini, Lagaude, & de la Fuente, 1994, Lucey, Gorry, & Fox, 1994; Sandra, Ho, Alexander, & Corredig, 2012; Singh & Waungana, 2001; Udabage, McKinnon, & Augustin, 2000) have revealed that the adverse effects of heat treatment on rennet coagulation can be overcome, to some extent, by adding calcium chloride ($CaCl_2$). The higher concentration of Ca^{2+} ions probably reduces electrostatic resistance between casein micelles, thus increasing aggregation (Singh & Waungana, 2001).

In cheesemaking, clotting time and coagulum firmness are commonly measured by Formagraph (Pretto et al., 2011). Several works have recently demonstrated also the usefulness of fundamental and empirical rheology in

monitoring properties of rennet-induced milk gels (Grassi, Strani, Casiraghi, & Alamprese, 2019; Han, Mei, Li, Xu, & Wang 2018; Kern, Bähler, Hinrichs, & Nöbel, 2019; Nassar et al., 2020; Salvador, Arango, & Castillo, 2019). However, these analytical systems are quite intrusive and unsuitable for the in-line/on-line monitoring of commercial cheesemaking (O'Callaghan, O'Donnell, & Payne, 2002). On the contrary, near infrared (NIR) spectroscopy has been recognized as a reliable analytical technology for the assessment of several quality parameters of a wide range of dairy products (IDF, 2019). Moreover, NIR spectroscopy, coupled with the use of specific chemometric tools, proved to be successful also for the evaluation of physical changes (Cabassi, Profaizer, Marinoni, Rizzi, & Cattaneo, 2013; Marinoni, Monti, Barzaghi, & de la Roza-Delgado, 2013; Strani, Grassi, Casiraghi, Alamprese, & Marini, 2019), thus leading to milk coagulation monitoring applications (Grassi, Alamprese, Bono, Casiraghi, & Amigo, 2014; Grassi et al., 2019; Lyndgaard, Engelsen, & van den Berg, 2012; Panikuttira, O'Shea, O'Donnell, & Tobin, 2017). In particular, among the chemometric tools, Multivariate Curve Resolution combined with Alternating Least Squares (MCR-ALS) is able to decompose the spectroscopic signal collected during NIR-based process monitoring into the contribution of several components with distinct spectral signatures, which are related to the different physicochemical forms of the product studied during the process development (De Juan & Tauler, 2006). MCR-ALS provides a bilinear model formed by chemically meaningful spectral signatures of the components, which can help in product characterization, and the related process concentration profiles, useful to interpret the time-dependent variation of the product forms. In difference with other chemometric tools, such as Principal Component Analysis (PCA) or Independent Component analysis (ICA), that provide bilinear models of few abstract profiles, MCR provides chemically meaningful profiles that can be more clearly interpreted and connected to the chemistry of the studied process; hence, the choice of this algorithm for this work (Parastar, Jalali-Heravi, & Tauler, 2012). The NIR/MCR-ALS combination meets the global

dairy market need for sensor-based process monitoring and interpretation, which will necessarily increase process efficiency and improve product quality and yield (Pu, O'Donnell, Tobin, & O'Shea, 2019).

In this framework, the present work investigated the monitoring of the renneting phase of reconstituted milk obtained from different low and medium heat SMPs produced in different European countries. Coagulum formation was studied on unaltered reconstituted milk samples, pasteurized samples, samples with added calcium chloride and samples mixed with fresh liquid milk with the aim of evaluating the effect of the modifications in the raw reconstituted milk. Moreover, the potential ability of the combination NIR/MCR-ALS for the monitoring and understanding of cheesemaking process has been investigated foreseeing a suitable non-destructive, non-invasive and on-line tool to evaluate the rennet-induced coagulation of milks.

2. Materials and methods

2.1. Sample preparation

Six samples of SMP were recovered from wholesalers present on the Italian market and were compared to assess coagulation properties and to study the effect of added $CaCl_2$ on the progress of the coagulation process. The country of production and the characteristics of the products are shown in the first block of Table 1. SMPs were dissolved in 300 mL distilled water (24 °C) at nearly 10% (w/w) to yield reconstituted skim milk samples with the same protein content (3.4±0.1 g/100g). Complete powder dissolution was achieved by mixing with a magnetic stirrer bar at room temperature for 20 min. Some of these reconstituted milk samples were enriched with 1 (0.0035 g/L) or 0.5 (0.0018 g/L) mmol/L of $CaCl_2$, from a stock solution (50 g/L). After measuring the pH with a pH-meter 3627 (Mettler-Toledo, Columbus, OH, USA), the samples were divided in three subsamples of 100 mL each to be used for NIR monitoring, rheology and Formagraph tests. For each of these techniques, two technical replicates were analyzed.

Fresh milk samples and SMP samples from EPI (EPI ingredients, Nantes, France) and SCA (Società Coadiuvanti Alimentari, Piacenza, Italy) milk were used to test the effect of pasteurization on milk coagulation, as shown in the second block of Table 1. Pasteurization was carried out at 73 °C for 16 s, by means of a self-assembled lab scale tubular heat exchanger.

A third study was oriented to test the effect of adding reconstituted milk (EPI and SCA powders) to fresh pasteurized milk (Granarolo, Bologna, Italy) on the coagulation process. Each reconstituted sample was mixed with skimmed milk in 40:60 and 60:40 ratios, as shown in the third block of Table 1. For these experiments, pasteurized skimmed milk and non-pasteurized skimmed milk were provided by CREA-ZA (Lodi, Italy) farm.

For all the coagulation trials, a concentration of 1.5 mL/L of liquid rennet (Naturen® 220 CHR Hansen, Hoersholm, Denmark) was used.

Table 1. Description of the samples used for the coagulation trials.

Sample ID	Number of replicates	Country	Manufacturing company	Powder type	$CaCl_2$ (g/L)	Sample pasteurization
Effect of milk powder type and added $CaCl_2$ on coagulation time						
EPI 18	2	France	EPI ingredients	Low heat	0.018	no
EPI 35	2	France	EPI ingredients	Low heat	0.035	no
Lactalis 18	2	France	Lactalis	Medium-heat	0.018	no
Lactalis 35	2	France	Lactalis	Medium-heat	0.035	no
Rucker 18	2	Germany	Rucker	Instant	0.018	no
Rucker 35	2	Germany	Rucker	Instant	0.035	no
Safivo 18	2	France	Safivo	Medium-heat	0.018	no
Safivo 35	2	France	Safivo	Medium-heat	0.035	no
SCA 18	2	Spain	Lafuente	Medium-heat	0.018	no
SCA 35	2	Spain	Lafuente	Medium-heat	0.035	no
SIA 18	2	Ireland	Glanbia	Instant	0.018	no
SIA 35	2	Ireland	Glanbia	Instant	0.035	no
Effect of pasteurization on coagulation						
Skimmed P	2	CREA-ZA	-	Skimmed fresh milk	0.035	yes
Skimmed NP	2	CREA-ZA	-	Skimmed fresh milk	0.035	no
EPI P	2	France	EPI ingredients	Low heat	0.035	yes
EPI NP	2	France	EPI ingredients	Low heat	0.035	no
SCA P	2	Spain	Lafuente	Medium-heat	0.035	yes
SCA NP	2	Spain	Lafuente	Medium-heat	0.035	no
Effect of addition of reconstituted milk to fresh milk on coagulation						
EPI 40	3	France	EPI ingredients	Low heat	0.035	no
EPI 60	3	France	EPI ingredients	Low heat	0.035	no
SCA 40	3	Spain	Lafuente	Medium-heat	0.035	no
SCA 60	3	Spain	Lafuente	Medium-heat	0.035	no

1

2.2. Formagraph analysis

The renneting properties were studied by using the Formagraph instrument (Foss Electric, Hillerød, Denmark). Milk samples were conditioned to 37 °C for 20 min before analysis, then rennet (Naturen® 220 CHR Hansen, Hoersholm, Denmark) was added at a final concentration of 0.088 IMCU/g of

milk. Measurements of milk coagulation properties ended 60 min after the addition of the clotting enzyme. From the Formagraph curve, the rennet coagulation time (r, min) and curd firmness after 30 min (A_{30}, mm) were calculated.

2.3. Rheological measurements

Samples rheological behavior during renneting was studied by using a Physica MCR 102 rheometer (Anton Paar GmbH, Graz, Austria) supported with the software RheoCompass (v. 1.21, Anton Paar GmbH, Graz, Austria). Samples were heated in a thermostatic bath to 37 °C and the proper amount of rennet was added. Afterwards, 19 mL of sample were inserted in the preheated concentric cylinders (CC27) of the rheometer in order to start the analysis exactly 1 min after the rennet addition. A time curing test was performed at 37 °C applying a constant 0.1% strain at a fixed 1 Hz frequency, collecting data every 15 s for a total of 30 min.

2.4. FT-NIR spectroscopy

Rennet coagulation process of each sample, placed in a thermostatic bath at 37 °C, was continuously monitored for 30 min with a Fourier-Transform (FT)-NIR spectrometer (MPA, Bruker Optics, Milan, Italy). A fiber optic probe, equipped with a transflectance adapter with a 0.1 cm optical path (0.2 cm total effective pathlength), was inserted in the sample and spectra were collected every minute in the whole NIR spectral range, namely 12500 – 4000 cm^{-1}, starting exactly 1 min after adding the liquid rennet. For each sample, 30 spectra were acquired, with a resolution of 8 cm^{-1} and 64 scans for both background and sample. OPUS software (v. 6.0 Bruker Optics, Milan, Italy) was used to manage the instrument.

2.5. Data analysis

Spectral, rheological and Formagraph data were processed and analyzed with toolboxes and routines present in Matlab environment (the Mathworks Inc., Natick, MA, USA).

The range of FT-NIR spectra was reduced to 12500 – 5824 cm^{-1}, in order to remove spectral ranges with signal saturation and high noise. A toolbox (Jaumot, de Juan, & Tauler, 2015) implemented in Matlab was used to perform the MCR-ALS analysis on the spectra collected in each coagulation trial. The two technical replicates were not combined in a single multiset structure because the scattering was different in each run. Unlike other processes monitored with FT-NIR, scattering should not be removed by preprocessing in the coagulation process because it is an indication of the coagulation progress. Spectral data of each trial were arranged in a matrix $\boldsymbol{D}$ (MxN), where M are the number of spectra (30) and N the number of wavenumbers (1840). MCR-ALS decomposed each $\boldsymbol{D}$ matrix into two matrices: the concentration profiles $\boldsymbol{C}$ (MxF) and the spectral profiles $\boldsymbol{S^T}$ (FxN) matrices (Eq. 1). The $\boldsymbol{S^T}$ matrix contains the spectral signatures of the F components linked to the different milk coagulation forms found during the monitored process, whereas the $\boldsymbol{C}$ matrix contains information about the evolution in time of the concentration of the F components during the process. Finally, $\boldsymbol{E}$ (MxN) is the matrix that contains the residuals.

$$\boldsymbol{D} = \boldsymbol{CS^T} + \boldsymbol{E} \quad \text{(Eq. 1)}$$

Principal Component Analysis (PCA) was used prior to MCR-ALS in order to define the proper number of components (F) to describe the spectral variation recorded during the coagulation process. Next, the ALS optimization was started by using initial estimates of spectral profiles provided by a pure variable selection method based on SIMPLISMA (Windig & Stephenson, 1992). Once the initial estimates are obtained, $\boldsymbol{C}$ and $\boldsymbol{S^T}$ are alternatingly optimized in each iteration under the action of constraints. The constraints

selected were non-negativity for the spectral and concentration profiles and unimodality, i.e. the presence of a single maximum per profile, in the concentration profiles because the different milk forms during the coagulation process can be considered to follow a kinetic sequential emergence-decay pattern. This procedure was similar to that followed in many process analysis examples (de Juan & Tauler, 2016; de Oliveira, Pedroza, Sousa, Lima, & de Juan, 2017; Grassi et al., 2014) and shows analogies with the study by Amigo, de Juan, Coello, and Maspoch (2006).

The calculation of $\boldsymbol{C}$ and $\boldsymbol{S}^T$ was repeated until a satisfactory reproduction of the original data $\boldsymbol{D}$ through the MCR model $\boldsymbol{CS}^T$ was achieved. In order to obtain satisfactory concentration and spectral profiles, a stopping convergence criterion was used based on the relative difference between percentages of lack of fit (LOF) between consecutive iterations:

$$LOF\ (\%) = 100 * \sqrt{\frac{\sum_{ij} e_{ij}^2}{\sum_{ij} d_{ij}^2}} \quad \text{(Eq. 2)}$$

where e_{ij} is each ijth component of the residual matrix ($\boldsymbol{E}$) and d_{ij} is each ijth component of $\boldsymbol{D}$ matrix. When the difference of two consecutive LOF percentages was lower than 0.1%, convergence was considered achieved and the iterative cycle stopped (Jaumot et al., 2015).

The modeling of the studied processes by MCR-ALS needed several components linked to the progress of the milk coagulation. The last concentration profile in time, linked to the coagulated form (called solid-like form in Grassi, Alamprese, Bono, Casiraghi, & Amigo, 2014) was compared with rheological, after logarithmic transformation, and Formagraph curves, meant to describe the coagulation process as well. To do so, correlation coefficients among the three profiles (MCR, rheological and Formagraph curves) were calculated. Furthermore, the concentration profile linked to the coagulated form obtained from each MCR-ALS analysis was modelled as a function of coagulation time through the following sigmoid equation (Eq. 3)

implemented in Table Curve software (v. 4.0, Jandel Scientific, San Rafael, CA, USA) (Grassi et al., 2013):

$$Y = a * exp\left\{-exp\left[-\left(\frac{x-cln(ln(2))-b}{c}\right)\right]\right\} \quad \text{(Eq. 3)}$$

First and second derivatives of the MCR profile of the coagulated form were used to detect kinetic parameters characterizing the process, i.e. the maximum acceleration, velocity and deceleration of each coagulation process. The maximum of first derivative of these curves was used to individuate the maximum speed of the process, whereas maximum and minimum of second derivatives were used to detect maximum acceleration and maximum deceleration of the process, respectively.

One-way analysis of variance (ANOVA) was performed to compare powder coagulation performance in term of r and A_{30}, calculated from the Formagraph curve, gelation point extrapolated from fundamental rheology analysis and maximum velocity of the solid-like concentration profile resulting from the MCR-ALS analysis. In case of statistically significant differences, a Least Significant Difference (LSD) post hoc test was performed.

3. Results and discussion

A clear increment of absorbance over time was observed during coagulation as shown in Fig. 1, due to scattering effects caused by suspended fat globules and micelles aggregation (Aernouts, Van Beers,Watté, Huybrechts, Lammertyn, & Saeys, 2015; Cabassi et al., 2013). Most of the changes occurs during the early moments of coagulation, especially in the 12500-9000 cm^{-1} region, where it is possible to observe a clear change in the spectra baseline overtime related to scattering effects linked to changes in particle size then on particle composition (Cabassi et al., 2013). The absorption band at 6900 cm^{-1} is ascribable to symmetric and asymmetric stretching of O-H water bond, whereas bands at 10800 and 8600 are linked to C-H bonds of lipids

(Tsenkova, Atanassova, Itoh, Ozaki, & Toyoda, 2000; Workman, & Weyer, 2007).

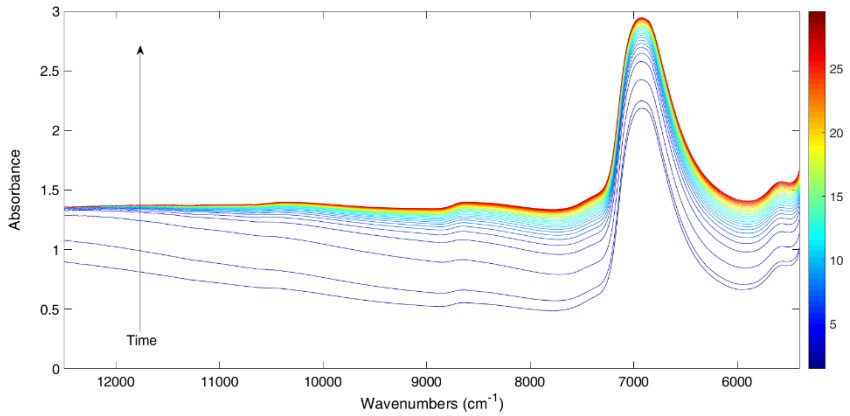


Fig 1: Spectra collected during coagulation of the sample EPI 35. Legend refers to coagulation time (min).

According to the PCA results, all coagulation processes could be described with three spectral contributions. The initial estimates of MCR-ALS were selected using a pure variable detection method, which helped to select the three most different spectra of the dataset ***D***, in order to describe the process. The algorithm selected always the first and the last spectrum of each coagulation trial, whereas the other component was chosen at different processing times, depending on the considered conditions.

The application of the ALS procedure to the milk renneting trials permitted the resolution of both concentration profiles (Fig. 2a) and pure spectra (Fig. 2b) of the milk forms involved. The obtained MCR-ALS models successfully described the process evolution. For all trials, the product $\boldsymbol{CS}^T$ explained more than 99.9% of variance of the data and the LOF was lower than 0.72%. Fig. 2a reports an example of concentration profiles for the coagulation performed with the sample Rucker 35. The obtained profiles contained information about the main changes occurring in milk. The first MCR-ALS concentration profile (represented by the solid line in Fig. 2a) had an inverse sigmoid shape in all

the performed runs: it described the first stage of the rennet coagulation of milk, where rheological data showed low and constant G' values (Fig. 2c) and curd firmness was equal to zero (Fig. 2d). This profile showed a clear decrease when the first aggregation of casein micelles caused a steep increase in the elasticity of the system, corresponding to the fast rise of G' in the time curing curve (Fig. 2c). It is also interesting to note that, when the first concentration profile reached a null value, the Formagraph curve started to increase, meaning that liquid non-coagulated milk was almost absent. Thus, the first profile was associated to the liquid-like milk behavior. The third concentration profile (Fig. 2a, dotted line) had a sigmoid shape inversely correlated to the liquid-like behavior: the decreasing evolution of the first profile corresponded to the increasing evolution of the latter profile that confirmed the transition of milk from liquid to a viscoelastic structure, thus describing the milk solid-like behavior as already reported by Grassi et al. (2014) and Grassi et al. (2019). The transition profile (second concentration profile represented by the dashed line in Fig. 2a) reached its maximum at the time in which G' crossed G'' (gelation point), due to the formation of a continuous protein network in the coagulated reconstituted milk. The observed differences in the time required to reach the maximum of the second profile were strictly related to the SMP used. The three pure spectra obtained in the MCR-ALS models (Fig. 2b) explained well the changes occurring in the FT-NIR data during the coagulation phases and confirmed what was previously found by Grassi et al. (2014) and Grassi et al. (2019). The solid line represented the typical spectrum recorded at the beginning of the renneting process, i.e. when the milk still had a liquid-like behavior. It showed a visible slope in the baseline at low wavenumbers and a high peak at 6900 cm^{-1} due to O-H combination band of symmetric and asymmetric stretching of water. The characteristic spectrum of coagulated milk, influenced by the characteristics of the continuous protein network, is represented by the dotted line and is characterized by an absence of slope between 12500 and 7500 cm^{-1} and a reduction in the absorption of the peak at 6900 cm^{-1} if compared

with the liquid-like behavior profile (solid line). The other pure spectrum (dashed line) stood for the transition phase, during which the first changes in the casein micelle structure took place, due to the solubilization of colloidal calcium phosphate. The relationship between casein coagulation and MCR-ALS profiles obtained by FT-NIR spectra analysis can be mainly ascribed to the spectra differences in baseline slope, caused by physical effects (Frake et al., 1998) such as the changes in number and size of casein micelles during rennet milk coagulation (Horne & Davidson, 1993), which are closely related to changes in light scattering. Furthermore, the significant reduction in the absorbance at 6900 cm^{-1} observed during renneting revealed the ability of the NIR probe of measuring the water retained by the curd, thus not poured out in the syneresis.

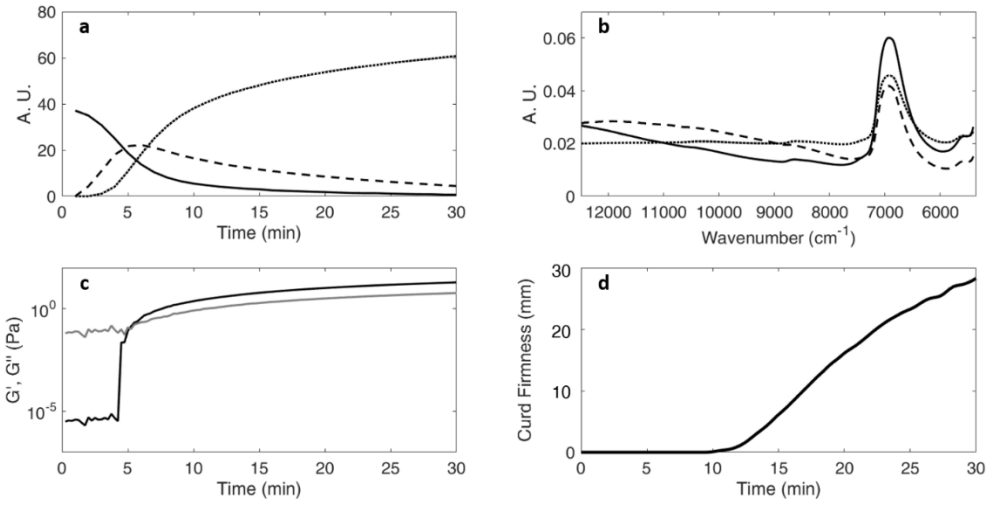


Fig 2: Coagulation trial of the sample Rucker 35: MCR-ALS concentration (a) and spectral profiles (b) of liquid-like behaviour of milk (solid lines), transition phase (dashed lines) and solid-like behaviour of milk (dotted lines); (c) rheological data (black: G', gray: G''); (d) Formagraph curve.

Similar behavior was obtained with all other coagulation experiments performed. The following subsections will address the study of different

factors on the coagulation process based on different subsets of experiments as described in Table 1.

3.1. Effect of milk powder type and $CaCl_2$ on coagulation time

This study was carried out with the 12 samples of the first block in Table 1, considering different SMP types and $CaCl_2$ concentrations.

To better investigate the relationship between the MCR-ALS profiles and the reference tests (rheology and Formagraph), a correlation analysis was performed between MCR-ALS data and both Formagraph and rheological data (Table 2). In detail, the MCR-ALS solid-like behavior profile, the last one appearing in time, was correlated with G', G'' and the Formagraph profile obtained for each coagulation trial.

Table 2. Results of the correlation analysis between MCR-ALS milk solid-like concentration profile and rheological data (logG' and logG'' curves) or Formagraph curve: correlation coefficients and statistical significance.

Sample ID	G'	G''	Formagraph
EPI 18 R1	0.99***	0.99***	0.99***
EPI 18 R2	0.99***	0.99***	0.99***
EPI 35 R1	0.99***	0.99***	0.99***
EPI 35 R2	0.99***	0.99***	0.99***
Lactalis 18 R1	0.75***	0.67***	0.99***
Lactalis 18 R2	0.73***	0.80***	0.99***
Lactalis 35 R1	0.93***	0.92***	0.97***
Lactalis 35 R2	0.92***	0.79***	0.97***
Rucker 18 R1	0.99***	0.98***	0.98***
Rucker 18 R2	0.99***	0.99***	0.99***
Rucker 35 R1	0.99***	0.98***	0.98***
Rucker 35 R2	0.99***	0.99***	0.98***
Safivo 18 R1	0.92***	0.97***	0.97***
Safivo 18 R2	0.96***	0.99***	0.97***
Safivo 35 R1	0.98***	0.98***	0.99***
Safivo 35 R2	0.99***	0.99***	0.99***
SCA 18 R1	0.99***	0.99***	0.99***
SCA 18 R2	0.99***	0.99***	0.99***
SCA 35 R1	0.99***	0.99***	0.98***
SCA 35 R2	0.99***	0.99***	0.98***
SIA 18 R1	0.95***	0.97***	0.99***
SIA 18 R2	0.97***	0.94***	0.99***
SIA 35 R1	0.99***	0.98***	0.88***
SIA 35 R2	0.99***	0.95***	0.88***

***p<0.001; R1 = first replicate; R2 = second replicate.

The correlations between the concentration profiles related to the solid-like behavior and G' and G'' curves were highly significant ($r > 0.92$; $p < 0.001$), with the exception of the trials performed with the two replicates of Lactalis 18 ($r < 0.80$), which were characterized by irregular coagulation profiles. However, correlations were significant in all cases, even for samples with

lower correlation coefficient values. These results are in agreement with those by Klandar, Lagaude, and Chevalier-Lucia (2007), who also found highly significant correlations ($r > 0.83$, $p < 0.001$) between parameters obtained from G' curves and kinetic parameters derived from the time-dependent evolution of some absorbances at individual wavelengths in raw NIR spectra acquired on milk reconstituted from low-heat or medium-heat skim milk powders. Moreover, good correlations were found also between the MCR-ALS solid-like profiles and the Formagraph curves ($r > 0.88$, $p < 0.001$), especially for EPI ($r > 0.99$), Rucker ($r > 0.98$) and SCA ($r > 0.98$), the powders that presented the best coagulation properties. These results confirmed the significant correlation between MCR-ALS data and both the reference analyses taken into account, which in their turn, were well correlated among them (correlation coefficients between rheological and Formagraph curves were always higher than 0.92, $p < 0.001$, with the exception of trials performed with Lactalis 18).

Results of this block of experiments were used to select the best SMPs and $CaCl_2$ concentration in terms of speed of coagulum formation and strength. The characteristic indexes of the Formagraph analysis (A_{30} and r), the rheological gelation point (G' and G'' cross-over) and the time corresponding to the maximum velocity (steepness) of the MCR-ALS solid-like behavior concentration profile, which was calculated from the maximum of the first derivative (Grassi et al., 2013), were calculated for all the performed trials and reported as the average of the two performed replicates (Table 3).

Table 3. Characteristic indexes from Formagraph (A_{30} and r), fundamental rheology (gelation point) and MCR-ALS analysis (time corresponding to the maximum velocity of the concentration profile related to the solid-like behavior). In a column, results with the same letter are not statistically different between each other ($p > 0.05$).

	0.018 g/L of $CaCl_2$				**0.035 g/L of $CaCl_2$**			
Sample	A_{30} (mm)	r (min)	Gelation point (min)	Solid-like behavior concentration profile (max velocity, min)	A_{30} (mm)	r (min)	Gelation point (min)	Solid beha concer profile velocit
EPI	42.1^{d}	8.0^{a}	2.0^{a}	3.9^{a}	39.6^{c}	10.1^{a}	2.3^{a}	4.
Lactalis	10.8^{a}	17.5^{d}	-	10.4^{c}	11.9^{a}	18.2^{c}	-	8.
Rucker	24.9^{b}	12.4^{b}	5.8^{b}	5.4^{b}	28.2^{b}	11.1^{a}	5.8^{b}	7.
Safivo	11.8^{a}	16.8^{cd}	16.5^{c}	10.4^{c}	14.1^{a}	17.9^{c}	16.3^{c}	8.
SCA	30.3^{c}	10.4^{ab}	5.0^{b}	6.5^{b}	31.2^{b}	13.1^{b}	4.8^{ab}	6.
SIA	16.1^{a}	15.4^{c}	7.5^{b}	-	16.1^{a}	16.1^{c}	6.3^{b}	5.

EPI and SCA resulted the best powders for milk enrichment as the A_{30} values were significantly higher ($p < 0.05$) than those obtained for the other SMPs; furthermore, the coagulation of reconstituted EPI and SCA milk samples occurred faster ($r < 10.5$ min, gelation point < 5 min). The times corresponding to the maximum velocity of the milk solid-like concentration profile agreed with Formagraph and rheological results, confirming the optimal behavior of the reconstituted powders in terms of coagulation occurrence.

The MCR-ALS analysis, in agreement with Formagraph and rheological data, showed that the concentration of calcium chloride did not significantly affect the coagulation process. Indeed, in Figure 3 it is possible to visually confirm that there was no difference between EPI 18 (Fig. 3a) and EPI 35 (Fig. 3c) samples nor between SCA 18 (Fig. 3b) and SCA 35 (Fig. 3d) samples, since the maxima of the transition concentration profiles (dark grey) occurred at the same time.

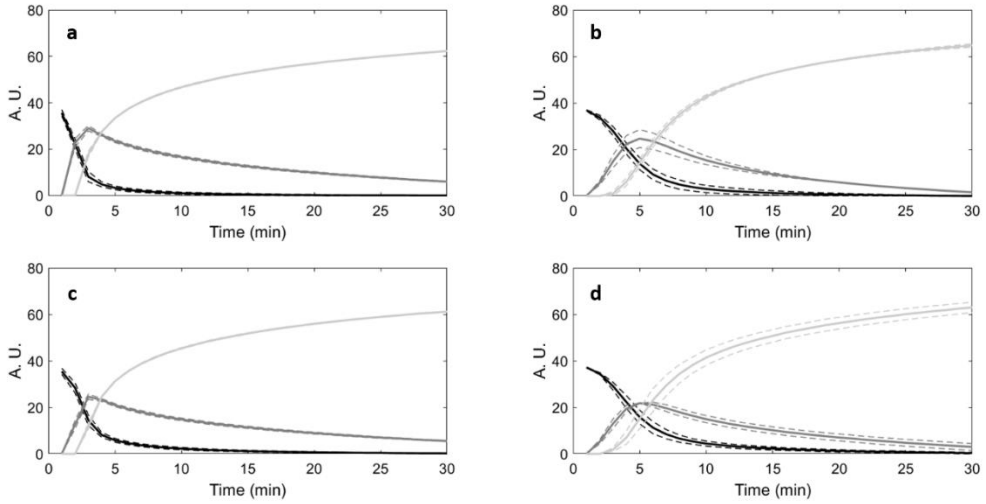


Fig 3: MCR-ALS concentration profiles of EPI 18 (a), SCA 18 (b), Epi 35 (c) and SCA 35 (d) samples. Dashed lines represent standard deviation interval.

Based on these results, EPI and SCA were selected as the best powders in terms of coagulation properties and used for further analyses. Considering the good agreement of the information obtained from rheological and Formagraph curves and MCR-ALS profiles issued from FT-NIR data, the second block of experiments were carried out using only FT-NIR monitoring (much faster) and MCR-ALS.

3.2. Effect of pasteurization in coagulation

To evaluate the possible effect of pasteurization on coagulation properties, six different coagulation trials were performed in duplicate using skimmed milk and EPI and SCA reconstituted samples, both non-pasteurized and pasteurized, maintaining the $CaCl_2$ concentration at 0.035 g/L (see second block in Table 1). As shown in Figure 4, in which the three MCR-ALS concentration profiles of all the trials are reported, there were no substantial differences between non-pasteurized and pasteurized samples, except for SCA P (Fig. 4f), whose coagulation appeared slightly delayed (1 min) with respect to SCA NP sample (Fig. 4e). The delay could be linked to thermally

induced changes in milk proteins, indeed SCA is a medium-heat SMP and a pasteurization procedure could influence the already modified interactions among caseins and denatured whey proteins (Kethireddipalli & Hill, 2015) causing a light delay in the coagulation.

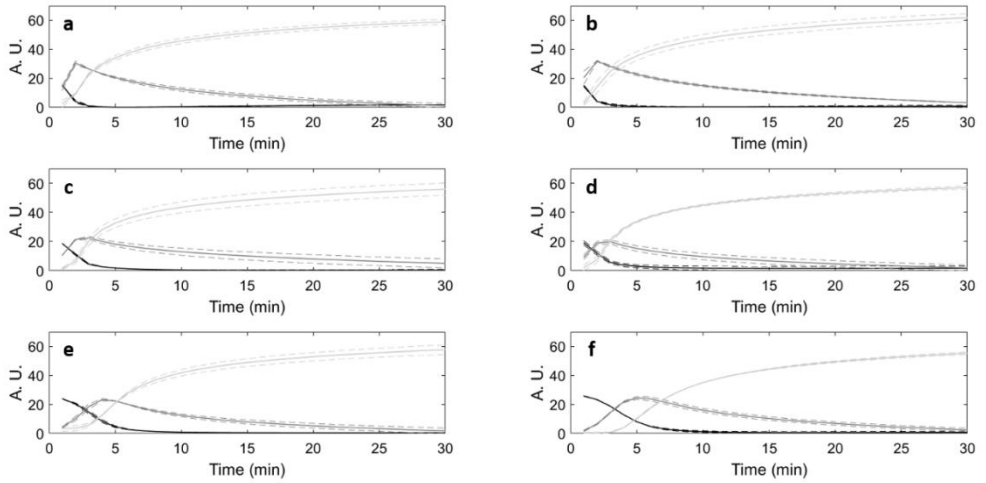


Fig 4: MCR-ALS concentration profiles of renneting trials carried out with non-pasteurized skimmed milk (S NP) (a), pasteurized skimmed milk (S P) (b), non-pasteurized reconstituted EPI sample (EPI NP) (c), pasteurized reconstituted EPI sample (EPI P) (d), non-pasteurized reconstituted SCA sample (SCA NP) (e) and pasteurized reconstituted SCA sample (SCA P) (f). Dashed lines represent standard deviation interval.

3.3. Effect of addition of reconstituted milk to fresh milk samples on coagulation

MCR-ALS was also applied to spectral data collected during the coagulation trials of pasteurized skimmed milk mixed with reconstituted milk samples. Both EPI and SCA reconstituted samples were mixed with 40% and 60% of pasteurized skimmed milk, and each experiment was replicated three times, for a total of twelve trials (third block in Table 1). The results (Fig. 5) highlighted that a lower amount of reconstituted milk in the mixture

corresponded to a delayed occurrence of the coagulum formation (indicated in the figure by the black vertical lines). In fact, in samples containing 60% of reconstituted milk (Fig. 5b and 5d) the peak of the transition concentration profiles occurred 1-2 min earlier than the ones of the samples with 40% reconstituted milk (Fig. 5a and 5c). In the experiments performed, the higher the amount of powder used, the faster the coagulum occurrence, approaching the performance of pasteurized skimmed milk (Fig. 4b) when 100% of reconstituted milk was used (Fig. 4d and 4f). Indeed, the maximum of the transition profiles recorded for the pasteurized skimmed milk occurred around 2 min, for reconstituted EPI and SCA samples (EPI P, SCA P) occurred at around 2 and 5 min (Fig. 4d and 4f), for EPI 60 and SCA 60 at 4.3 and 6.6 min, and for EPI 40 and SCA 40 at 5.3 and 8.1 min (Fig. 5), respectively.

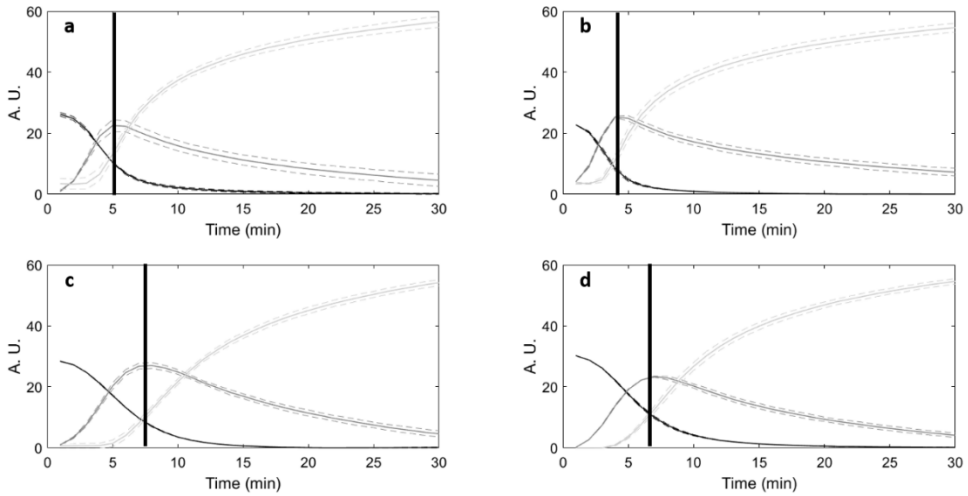


Fig 5: MCR-ALS concentration profiles of coagulation trials performed with the mixtures of pasteurized skimmed milk and reconstituted milk samples: EPI 40 (a), EPI 60 (b), SCA 40 (c) and SCA 60 (d). Dashed lines represent standard deviation interval.

4. Conclusions

A wide range of experimental conditions were studied in order to describe the coagulation process and the possible changes undergone due to different low and medium heat skim milk powders, amount of $CaCl_2$ added, use of pasteurization treatment and fraction of reconstituted milk powders.

MCR-ALS combined with FT-NIR monitoring during milk coagulation process was proposed as an alternative to standard methods, such as Formagraph and rheological methods. The process profile of the coagulated milk form obtained with the developed approach, highly correlated to the standard methods (always $r > 0.67$, and for the powders with the best coagulation properties $r > 0.98$, $p < 0.001$), allowed proving the significant effect of the milk powder type on coagulation occurrence, as opposed to the non-significance of the added $CaCl_2$ concentration and the heat treatment. Moreover, monitoring of coagulation trials of pasteurized skimmed milk mixed with reconstituted milk samples permitted the identification of shorter coagulation times when higher reconstituted milk percentage was used.

Profiles extracted from MCR-ALS models appear to be suitable as a fast, non-destructive, non-invasive on-line method to evaluate the rennet-induced coagulation of reconstituted milks and to assess changes in coagulation performance for a wide range of coagulation conditions.

Declaration of interest

The Authors declare no conflicts of interest.

Acknowledgements

The authors would like to thank Dr. Giovanni Cabassi of CREA-ZA, Research Centre for Animal Production and Aquaculture (via A. Lombardo 11, 26900 Lodi, Italy).

References

Aernouts, B., Van Beers, R., Watté, R., Huybrechts, T., Lammertyn, J., & Saeys, W. (2015). Visible and near-infrared bulk optical properties of raw milk. Journal of Dairy Science, *98*(10), 6727–6738.

Amigo, J. M., de Juan, A., Coello, J., & Maspoch, S. (2006). A mixed hard- and soft-modelling approach to study and monitor enzymatic systems in biological fluids. *Analytica Chimica Acta*, *567*(2), 245-254.Cabassi, G., Profaizer, M., Marinoni, L., Rizzi, N., & Cattaneo, T. M. (2013). Estimation of fat globule size distribution in milk using an inverse light scattering model in the near infrared region. *Journal of Near Infrared Spectroscopy*, *21*(5), 359-373.

De Juan, A.; Tauler, R. (2006). Multivariate curve resolution (MCR) from 2000: Progress in concepts & applications. *Critical Reviews in Analytical Chemistry*, *36*, 163–176

De Juan, A., & Tauler, R. (2016). Multivariate curve resolution-alternating least squares for spectroscopic data. In *Data Handling in Science and Technology* (pp. 5-51). Amsterdam, Netherlands: Elsevier.

De Oliveira, R. R., Pedroza, R. H., Sousa, A. O., Lima, K. M., & de Juan, A. (2017). Process modeling and control applied to real-time monitoring of distillation processes by near-infrared spectroscopy. *Analytica chimica acta*, *985*, 41-53.

Frake, P., Luscombe, C. N., Rudd, D. R., Gill, I., Waterhouse, J., & Jayasooriya, U. A. (1998). Near-infrared mass median particle size determination of lactose monohydrate, evaluating several chemometric approaches. *Analyst*, *123*(10), 2043-2046.

Gastaldi, E., Pellegrini, O., Lagaude, A., & de la Fuente, B. T. (1994). Functions of added calcium in acid milk coagulation. *Journal of Food Science*, *59*(2), 310-312.

Grassi, S., Alamprese, C., Bono, V., Picozzi, C., Foschino, R., & Casiraghi, E. (2013). Monitoring of lactic acid fermentation process using Fourier transform near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, *21*(5), 417-425.

Grassi, S., Alamprese, C., Bono, V., Casiraghi, E., Amigo, J.M. (2014). Modelling milk lactic acid fermentation using multivariate curve resolution-alternating least squares (MCR-ALS). *Food and Bioprocess Technology*, *7*, 1819–1829

Grassi, S., Strani, L., Casiraghi, E., & Alamprese, C. (2019). Control and monitoring of milk renneting using FT-NIR spectroscopy as a Process Analytical Technology tool. *Foods*, *8*(9), 405.

Gulati, A., Hennessy, D., O'Donovan, M., McManus, J. J., Fenelon, M. A., & Guinee, T. P. (2019). Dairy cow feeding system alters the characteristics of low-heat skim milk powder and processability of reconstituted skim milk. *Journal of Dairy Science*, *102*(10), 8630-8647.

Han, Y., Mei, Y., Li, K., Xu, Y., & Wang, F. (2019). Effect of transglutaminase on rennet-induced gelation of skim milk and soymilk mixtures. *Journal of the Science of Food and Agriculture*, *99*(4), 1820-1827.

Horne, D. S., & Davidson, C. M. (1993). Direct observation of decrease in size of casein micelles during the initial stages of renneting of skim milk. *International Dairy Journal*, *3*(1), 61-71.

IDF (2019). *Applications of Near Infrared Spectrometry for the Analysis of Milk and Milk products. IDF 497:2019.* Brussel, Belgium: International Dairy Federation.

Jaumot, J., de Juan, A., & Tauler, R. (2015). MCR-ALS GUI 2.0: new features and applications. *Chemometrics and Intelligent Laboratory Systems*, *140*, 1-12.

Kelly, A. L., & Fox, P. F. (2016). Manufacture and properties of dairy powders. In *Advanced dairy chemistry* (pp. 1-33). New York, USA: Springer.

Kern, C., Bähler, B., Hinrichs, J., & Nöbel, S. (2019). Waterless single screw extrusion of pasta-filata cheese: Process design based on thermo-rheological material properties. *Journal of Food Engineering*, *260*, 58-69.

Kethireddipalli, P., & Hill, A. R. (2015). Rennet coagulation and cheesemaking properties of thermally processed milk: overview and recent developments. *Journal of Agricultural and Food Chemistry*, 63(43), 9389-9403.

Klandar, A. H., Lagaude, A., & Chevalier-Lucia, D. (2007). Assessment of the rennet coagulation of skim milk: A comparison of methods. *International Dairy Journal*, *17*(10), 1151-1160.

Kjaergaard-Jansen, G. (1990). Milk powders: specification in relation to the products to be manufactures. In: *Recombination of milk and milk products*, Special Issue No. 9001. (pp. 104-125). Brussel, Belgium: International Dairy Federation.

Lelievret, J., Shaker, R. R., & Taylor, M. W. (1991). The influence of milk powder characteristics on the properties of halloumi cheese made from recombined milk. *International Journal of Dairy Technology*, *44*(2), 41-45.

Lin, Y., Kelly, A. L., O'Mahony, J. A., & Guinee, T. P. (2017). Addition of sodium caseinate to skim milk increases nonsedimentable casein and causes significant changes in rennet-induced gelation, heat stability, and ethanol stability. *Journal of Dairy Science*, *100*(2), 908-918.

Lucey, J. A., Gorry, C., & Fox, P. F. (1994). Methods for improving the rennet coagulation properties of heated milk. In *IDF seminar. Cheese yield and factors affecting its control, Cork (Ireland), Apr 1993*. Brussel, Belgium: International Dairy Federation.

Lyndgaard, C. B., Engelsen, S. B., & van den Berg, F. W. (2012). Real-time modeling of milk coagulation using in-line near infrared spectroscopy. *Journal of Food Engineering*, *108*(2), 345-352.

Marinoni, L., Monti, L., Barzaghi, S., & de la Roza-Delgado, B. (2013). Quantification of casein fractions and of some of their genetic variants in phosphate buffer by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, *21*(5), 385-394.

Nassar, K. S., Lu, J., Pang, X., Ragab, E. S., Yue, Y., Zhang, S., et al. (2020). Rheological and microstructural properties of rennet gel made from caprine milk treated by HP. *Journal of Food Engineering*, *267*, 109710.

O'Callaghan, D. J., O'Donnell, C. P., & Payne, F. A. (2002). Review of systems for monitoring curd setting during cheesemaking. *International Journal of Dairy Technology*, *55*(2), 65-74.

Omar, M. M., & Buchheim, W. (1983). Composition and microstructure of soft brine cheese made from instant whole milk powder. *Food Structure*, *2*(1), 6.

Panikuttira, B., O'Shea, N., O'Donnell, C. P., & Tobin, J. T. (2017). PAT approach for cheese manufacture. *TResearch*, *12*, 24-25.

Papadatos, A., Berger, A. M., Pratt, J. E., & Barbano, D. M. (2002). A nonlinear programming optimization model to maximize net revenue in cheese manufacture. *Journal of Dairy Science*, *85*(11), 2768-2785.

Parastar, H., Jalali-Heravi, M., & Tauler, R. (2012). Is independent component analysis appropriate for multivariate resolution in analytical chemistry? *TrAC Trends in Analytical Chemistry*, *31*, 134-143

Písecký, J. (2005). Spray drying in the cheese industry. *International Dairy Journal*, *15*(6-9), 531-536.

Pretto, D., Kaart, T., Vallas, M., Jõudu, I., Henno, M., Ancilotto, L., et al. (2011). Relationships between milk coagulation property traits analyzed with different methodologies. *Journal of Dairy Science*, *94*(9), 4336-4346.

Pu, Y. Y., O'Donnell, C., Tobin, J., & O'Shea, N. (2019). Review of near-infrared spectroscopy as a process analytical technology for real-time product monitoring in dairy processing. *International Dairy Journal*, 104623.

Rynne, N. M., Beresford, T. P., Kelly, A. L., & Guinee, T. P. (2004). Effect of milk pasteurization temperature and in situ whey protein denaturation on the composition, texture and heat-induced functionality of half-fat Cheddar cheese. *International Dairy Journal*, *14*(11), 989-1001.

Salvador, D., Arango, O., & Castillo, M. (2019). In-line estimation of the elastic module of milk gels with variation of temperature protein concentration. *International Journal of Food Science & Technology*, *54*(2), 354-360.

Sandra, S., Ho, M., Alexander, M., & Corredig, M. (2012). Effect of soluble calcium on the renneting properties of casein micelles as measured by rheology and diffusing wave spectroscopy. *Journal of Dairy Science*, *95*(1), 75-82.

Singh, H., & Waungana, A. (2001). Influence of heat treatment of milk on cheesemaking properties. *International Dairy Journal*, *11*(4-7), 543-551.

Strani, L., Grassi, S., Casiraghi, E., Alamprese, C., & Marini, F. (2019). Milk Renneting: Study of process factor influences by FT-NIR spectroscopy and chemometrics. *Food and Bioprocess Technology*, *12*(6), 954-963.

Tsenkova, R., Atanassova, S., Itoh, K., Ozaki, Y., & Toyoda, K. (2000). Near infrared spectroscopy for biomonitoring: cow milk composition measurement in a spectral region from 1,100 to 2,400 nanometers. Journal of Animal Science, *78*(3), 515–522.

Udabage, P., McKinnon, I. R., & Augustin, M. A. (2000). Mineral and casein equilibria in milk: effects of added salts and calcium-chelating agents. *Journal of Dairy Research*, *67*(3), 361-370.

Windig, W., & Stephenson, D. A. (1992). Self-modeling mixture analysis of second-derivative near-infrared spectral data using the SIMPLISMA approach. *Analytical Chemistry*, *64*(22), 2735-2742.

Workman, J., & Weyer, L. (2007). Practical guide to interpretive near-infrared spectroscopy. Boca Raton: CRC Press.

4.4 PDF PAPER IV

FOOD AND BIOPRODUCTS PROCESSING 126 (2021) 113–120

Contents lists available at ScienceDirect

Food and Bioproducts Processing

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

Study of Galactooligosaccharides production from dairy waste by FTIR and chemometrics as Process Analytical Technology

Fabián Rico-Rodriguez[a], *Lorenzo Strani*[b], *Silvia Grassi*[b,*], *Ruth Lancheros*[c], *Juan Carlos Serrato*[c], *Ernestina Casiraghi*[b]

[a] *Program of Food Engineering, Faculty of Engineering, Universidad de Cartagena – Cartagena de Indias D.T. y C., Colombia*

[b] *Department of Food, Environmental and Nutritional Sciences (DeFENS), Università degli Studi di Milano, Milano, Italy*

[c] *Department of Chemical and Environmental, Faculty of Engineering, Universidad Nacional de Colombia, Bogotá D.C, Colombia*

ARTICLE INFO

Article history:
Received 16 October 2020
Received in revised form 12 December 2020
Accepted 20 December 2020
Available online 30 December 2020

Keywords:
Prebiotic synthesis
Cheese by-products
Infrared spectroscopy
Process Analytical Technology
Chemometrics
Multivariate Curve Resolution
Partial Least Square regression

ABSTRACT

Galactooligosaccharides (GOS) production from whey, a relevant by-product of dairy industry, answers to the Circular Economy principle of extending the life cycle of products. Indeed, it allows the reuse of dairy waste to produce prebiotics to be used in functional food preparations. For this purpose, the effective monitoring of GOS production should be performed in real time and by environmentally friendly techniques. Thus, FTIR spectroscopy, combined with different chemometric approaches, has been tested to assess a Process Analytical Technology to follow GOS production from cheese whey. Partial Least Square regression models were reliable for lactose, glucose and galactose determination (Root Mean Square Error of Prediction of 21.9, 11.1 and 12.4 mg mL^{-1}, respectively). Furthermore, Multivariate Curve Resolution – Alternating Least Square models were proposed to describe trends of the reaction components along the process being an interesting alternative to chromatographic determinations. The real time implementation of the proposed approach will provide the dairy industry with a reliable and green Process Analytical Technology for dairy waste reallocation, avoiding sample pre-processing, large use of organic solvents and long times of analysis.

4.4 PAPER IV *(submitted version)*

Study of Galactooligosaccharides production from dairy waste by FTIR and Chemometrics as Process Analytical Technology

Fabián RICO-RODRIGUEZ[1], Lorenzo STRANI[2], Silvia GRASSI[2*], Ruth LANCHEROS[3], Juan Carlos SERRATO[3], Ernestina CASIRAGHI[2]

[1] Program of Food Engineering, Faculty of Engineering, Universidad de Cartagena – Cartagena de Indias D.T. y C., Colombia

[2] Department of Food, Environmental and Nutritional Sciences (DeFENS), Università degli Studi di Milano – Milano, Italy

[3] Department of Chemical and Environmental, Faculty of Engineering, Universidad Nacional de Colombia – Bogotá D.C, Colombia

*Author to whom correspondence should be addressed:

Università degli Studi di Milano

Department of Food, Environmental and Nutritional Sciences (DeFENS)

via Celoria, 2 - 20133, Milano, Italy

E-mail: silvia.grassi@unimi.it

Abstract

Galactooligosaccharides (GOS) production from whey, a relevant by-product of dairy industry, answers to the Circular Economy principle of extending the life cycle of products. Indeed, it allows the reuse of dairy waste to produce prebiotics to be used in functional food preparations. For this purpose, the effective monitoring of GOS production should be performed in real time and by environmentally friendly techniques. Thus, FTIR spectroscopy, combined with different chemometric approaches, has been tested to assess a Process Analytical Technology to follow GOS production from cheese whey. Partial Least Square regression models were reliable for lactose, glucose and galactose determination (Root Mean Square Error of Prediction of 21.9, 11.1 and 12.4 mg mL^{-1}, respectively). Furthermore, Multivariate Curve Resolution – Alternating Least Square models were proposed to describe trends of the reaction components along the process being an interesting alternative to chromatographic determinations. The real time implementation of the proposed approach will provide the dairy industry with a reliable and green Process Analytical Technology for dairy waste reallocation, avoiding sample pre-processing, large use of organic solvents and long times of analysis.

Keywords: Prebiotic synthesis, cheese by-products, Infrared spectroscopy, Process Analytical Technology, Chemometrics

1. Introduction

Galactooligosaccharides (GOS) comprise a group of oligomers, derived from lactose, possessing functional properties since they act as prebiotics and promote benefits in microbial gut and human health (Rodrigues Mano et al., 2019). Indeed, GOS are one of the main prebiotics used for functional food preparations. Among their functional properties, it is possible to mention their ability to increase beneficial gut microbiota, immune system modulation and/or antipathogenic effect (Byfield et al., 2010; Sangwan et al., 2011). Besides their functional characteristics, GOS possess desired technological properties which allows them to be used on different food matrixes like dairy, bakery or beverages, among others (Sangwan et al., 2011).

GOS are produced by lactose transgalactosylation. This reaction is catalysed by β-galactosidase (βgal) enzymes, obtained from a wide variety of microorganisms: bacteria, yeast and/or fungi (Byfield et al., 2010). Two of the most used βgal in industry are obtained from *Kluyveromyces lactis* and *Aspergillus oryzae* (Fischer and Kleinschmidt, 2015). As βgal possesses double activity (hydrolytic and transgalactosylation), it is necessary to control the thermodynamics of the reaction by increasing lactose concentration in the media, which favours GOS formation over monosaccharide release (González-Delgado et al., 2016). GOS production depends on many factors like substrate and initial sugar concentration, pH and temperature of reaction, presence of enzyme inhibitors and source of enzyme (Chockchaisawasdee et al., 2004).

It has been suggested that a variation in the GOS degree of polymerisation and linkages plays an important role on the effect of gut microbiota (Akiyama et al., 2015). Different assays have shown that GOS produced by βgal from *K. lactis* and *A. oryzae* differs in yields and composition. Whilst *K. lactis* GOS are mainly composed of di- and tri-saccharides, *A. oryzae* βgal produces GOS from di- to hexa-saccharides (Gosling et al., 2010). However, yields for

lactose transgalactosylation with *K. lactis* βgal are around 45 – 50% of total products (Rico-Rodríguez et al., 2018). Nonetheless, with *A. oryzae* βgal this value is lower than 35% (Otieno, 2010).

On the other hand, GOS quantification is generally performed by high performance liquid chromatography (HPLC) (Hernández-Hernández et al., 2012) or gas chromatography (GC) (Ruiz-Matute et al., 2012), two of the most reliable analytical techniques for the purpose. However, these methods require sample pre-processing (GC), large volumes of organic solvents (HPLC) and long analysis time per sample.

Nowadays, the research for quick-environmentally friendly analytical techniques has focused on vibrational spectroscopic technologies, such as mid-infrared, near-infrared and Raman spectroscopy (Moros et al., 2010).

Fourier transformed mid-infrared (FTIR) has been used in enzymatic studies for structural characterization from β-galactosidase immobilization, encapsulation and conjugation (Gennari et al., 2019; Li et al., 2019; Eskandarloo & Abbaspourrad, 2018; Misson et al., 2016) to derivatives examination (Kumar et al., 2020). However, little investigation has been proposed for process monitoring in this field, mainly represented by the in-line anomer concentration measurements in solution proposed by Schiele et al. (2020). Furthermore, Romano et al. (2016) evaluated, by FTIR and Partial Least Squares (PLS) regression, the effect of sucrose concentration on the composition of enzymatically synthesized short-chain fructo-oligosaccharides; similarly, the lactose hydrolysis in milk has been studied by Cocciardi et al. (2004). On the other hand, FTIR proved to be reliable to control processes in a fast, real time and non-destructive way (Grassi et al., 2014), or as a tool for evaluating quality parameters in dairy products (Mohamed et al., 2020) among others.

Despite this, FTIR technique has few drawbacks. Among them, the difficult interpretation of the signals, which are often composed of overlapping

spectral bands resulting from absorption of multiple compounds in the sample. In this context, chemometric techniques can help to overcome interpretation problems and they allow to extract relevant information. Spectral data can be modelled by both hard- and soft-modelling methods. Among hard modelling methods, Partial Least Square (PLS) regression is one of the most used for predicting quality parameters from a single spectrum after a proper model calibration, thanks to its ability to handle highly overlapping and colinear data (Martens and Næs, 1989). Instead, Multivariate Curve Resolution – Alternating Least Square (MCR-ALS) (De Juan and Tauler, 2006), a soft-modelling method, is useful to describe trends of the reaction components along the process.

This work aims to develop an analytical technology process based on FTIR spectroscopy, coupled with multivariate analysis, for at-line monitoring of GOS production from whey.

2. Material and methods

2.1. Reagents

Cheese whey was purchased from CIMPA SAS (Bogotá, Colombia) with lactose content of 71.12 % and <1% of monosaccharides. Commercial β-galactosidase Enzeco fungal lactase from *A. oryzae* (Ao) was provided by Enzyme Development Corporation (New York, USA); HA-Lactase5200 from *K. lactis* (Kl) was provided by CHR Hansen (Bogotá, Colombia). Enzymes had a total protein content, determined by Bradford method (Bradford, 1976), of 38.9 ± 0.1% (Ao) and 4.8 ± 0.1% (Kl). Enzyme activities were measured as reported by Rico-Rodríguez et al. (2018) giving enzyme activity of 15045 U/g (Ao) and 5172 U/g (Kl). Bovine serum albumin (BSA), sodium dihydrogen phosphate monohydrate, disodium hydrogen phosphate dihydrate, and sodium hydroxide were purchased from Sigma–Aldrich (St. Louis, MO). Glucose (Glu), galactose (Gal), lactose (Lac) and raffinose (DP3), were purchased from Panreac® (Barcelona, Spain).

2.2. Enzyme reactions

All assays were carried out in 50 mL falcon tubes with an effective volume of 20 mL (370 mg mL^{-1} of lactose equivalent) in sodium phosphate buffer solution 0.01 M at pH 6.0 and temperature of 42 ± 1 °C. Enzyme doses for each assay are reported in Table 1. Sampling was done by duplicate after 0, 20, 40, 60, 120, 180, 240, 300 min of reaction under continuous stirring at 800 rpm. Samples were immersed in boiling water (5 min) for enzyme inactivation; afterwards, samples were cooled and stored at -20 °C until carbohydrate quantification.

Table 1 - Reaction arrays and dose for each assay

Assay	Alias	Hal Dose (U)	Enz Dose (U)
Enz:Hal (100:0)	WO	37.8	0
Enz:Hal (75:25)	W1/4	28.4	0.9
Enz:Hal (50:50)	W1/2	18.9	1.9
Enz:Hal (25:75)	W3/4	9.9	2.8
Enz:Hal (0:100)	WK	0	3.8

2.3. Carbohydrate quantification

Substrates and products of enzyme reactions were analysed in an Ultimate 3000 HPLC instrument (Themo Fisher Scientific, USA). For GOS3, GOS4, GOS5, lactose and monosaccharides a BP-100 Ca^{++} (300 mm x 7.8 mm – Benson polymerics, USA) was set at 80°C and 0.5 mL/min of deionised water flowing rate. GOS2 were separated from lactose with a Spherisorb S5-NH_2 (250 mm x 4.0 mm – Waters, USA) with acetonitrile:water (75:25) as eluent at 30°C and 0.75 mL/min flowing rate. Both measurements were done in a

Transgenomic® RI detector. Calibration curves were performed with appropriate sugar standards in the linear range 0.05 mg mL^{-1} to 5 mg mL^{-1}.

2.4. *FTIR measurements*

IR spectra, from different enzyme mixture reaction samples, were collected in transmission mode, with a FT-IR spectrometer (IR-Prestige21, Shimatsu Corporation, Japan) equipped with an ATR cell. The spectra information was collected in the range 4000 – 400 cm^{-1}, with a resolution of 4 cm^{-1} and 25 scans for both background and samples. For each duplicate of the enzymatic rection, the same samples were analysed twice; the total number of collected spectra was 160. Besides, samples of carbohydrate standards (400 mg mL^{-1}) were analysed following the same procedure.

2.5. Data processing

The whole wavenumber range (4000 – 400 cm^{-1}) was reduced to the fingerprint region (1200 – 900 cm^{-1}) and spectra were averaged on sample base. Before applying multivariate analysis methods, several pre-treatments were performed in order to assess which one could efficiently correct scattering effects and reduce noise: Standard Normal Variate (SNV), first and second derivatives, baseline correction (automatic weighted least squares) and smoothing (Savitzky–Golay, filter width 7 points; polynomial order 1).

To predict the amounts of lactose, GOS, glucose and galactose in whey samples, PLS regression was performed by using the PLS Toolbox software (Eigenvector Research, Inc., Wenatchee, Washington, USA) implemented in Matlab (MathWorks, Natick, MA, USA). For each evaluated parameter, data were partitioned in a calibration (including 64 averaged spectra) and a validation set (including 16 averaged spectra). The validation sets were constructed on reaction bases. This means that one out of five enzymatic reactions (16 averaged spectra) was kept out from the calibration set in an iterative way, thus developing five different models, each of them tested by one different reaction test set. Models were also internally validated by

Venetian Blind Cross-Validation, with 10 splits and 2 samples per split, being sure to keep duplicates together. To evaluate model reliability, Root Mean Square Error of Prediction (RMSEP) was compared with the Root Mean Square Error of Laboratory (RMSEL). Furthermore, residual prediction deviation (RPD), i.e. ratio of standard error of performance to standard deviation, was calculated to compare the precision of the prediction with the average composition of all the samples, as stated by Camacho-Tamayo et al. (2014).

MCR-ALS method was applied on spectra using MCR-ALS 2.0 Toolbox (Jaumot et al., 2015) implemented in Matlab (MathWorks, Natick, MA, USA). Spectral data were arranged in five matrices, one for each enzymatic reaction considering together the duplicates, and analysed separately. For each model the number of components was selected by Pure Variable detection method. Model optimization was achieved by iterative implementation of ALS algorithm, i.e. until the Lack of Fit (LOF) difference in two consecutive iterative cycles was lower than 0.1%. A systematic investigation on any constraint in spectral modes of data matrix has be performed via calculation of feasible solutions and interpretation of obtained results. Then, non-negativity constraint was imposed on both concentration and spectral profiles to solve MCR-ALS ambiguities. For further details about the MCR-ALS procedure and optimization the reader can refer to Grassi et al. (2019).

3. Results and discussion

3.1. Kinetics of GOS production by HPLC

Transgalactosylation reactions by βgal from two sources (*K. lactis* and *A. oryzae*) and their mixtures followed the characteristic lactose consumption (Fig. 1). Lactose conversion for βgal from the yeast *Kluyveromyces lactis* results in values over 90% with a GOS production above 40% - mainly disaccharides and trisaccharides - and few tetrasaccharides (<20% of the total GOS). However, combination with βgal from *A. oryzae* could lead to

better profiles of oligomers present in the final mixture. In Fig. 1A it is observed that lactose consumption is highly affected by βgal mixture, i.e. by the enzyme source. When *K. lactis* enzyme is predominant in the mixture, Kl and 1A/3K trials, lactose concentration dropped faster than in other cases to values below 50 mg mL^{-1}. Nonetheless, kinetic behaviour of sample 2A/2K, which had equal dose of both enzymes, showed an unusual trend. Lactose consumed (50% of initial lactose) by 2A/2K reaction was lower than Ao (63% of initial lactose), which corresponded to *A. oryzae* pure enzyme.

In this set of reactions, monosaccharides (Fig. 1C and 1D) had one of the most significant variations. Though in all reactions there was a typical increase in the concentration of monosaccharides (glucose and galactose), there was also, in each treatment, a significant variation in their amount.. *K. lactis* enzyme is known by its poor ability to include glucose in GOS molecule, hence, the concentration of this monosaccharide is above 40% of total carbohydrates present in media (Jenab et al., 2017; Rico-Rodríguez et al., 2018). Moreover, galactose release depends on the ability of the enzyme to transform lactose into GOS at defined conditions of reaction as the later are mainly composed of this monosaccharide. In this case, it is possible to observe that a gradually decreasing in final concentration of those monosaccharides occurs with the reduction in the concentration of initial *K. lactis* enzyme or with the increase of *A. oryzae* enzyme source.

Fig. 1B shows GOS kinetic concentration for each reaction. Samples Ao, 2A/2K and 3A/1K show a similar kinetics with GOS concentration increasing over the 5 hours of reaction. A typical trend is shown by Kl; in this reaction, GOS hydrolysis started after 1 hour. For 1A/3K the reaction exhibits similar trend, nonetheless, maximum concentration was found at 120 min and GOS hydrolysis is not as fast as Kl. Highest concentration of GOS was 36% of total carbohydrates and it was reached after 60 min for 1A/3K and after 300 min for 2A/2K. In Fig. 1B it is notable a clear differentiation in the kinetics of the pure enzyme *K. lactis*, if compared to all enzyme combinations.

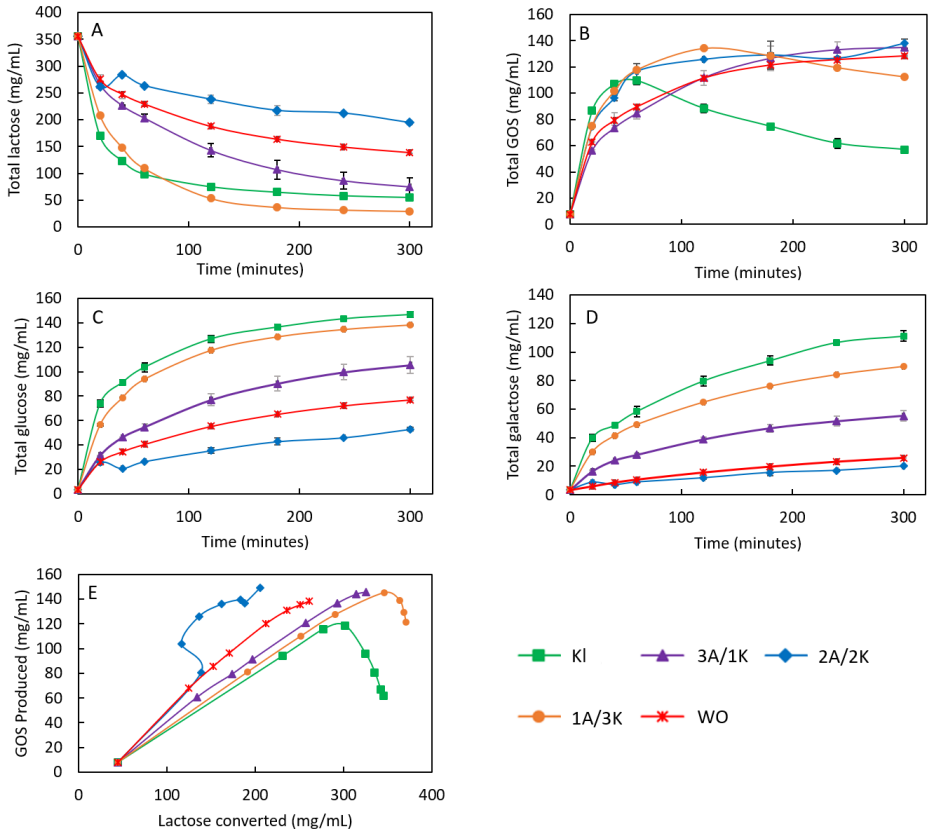


Fig. 1 - Lactose transgalactosylation in presence of β-galactosidase mixture: (A) Lactose consumption; (B) GOS production; (C) Glucose production; (D) Galactose production; (E) Lactose conversion to GOS. Error bars represent the standard deviations between

On the other hand, a graph of lactose conversion to GOS is presented in Fig. 1E. This graph shows the ability of βgal to transform lactose to GOS in reaction media. Although in Fig. 1E the conversion of lactose is well differentiated for pure enzymes and their mixtures, 2A/2K has a bulging trend, an unexpected behaviour possibly linked to the interaction between the enzymes.

Results suggest that βgal from *K. lactis* consumes high amounts of lactose (73%) to reach 30% of GOS in media with initial lactose of 370 mg mL^{-1}.

However, the amount of required lactose changes gradually when enzyme doses change in different reactions. Reaction Ao, in which only *A. oryzae* enzyme is present, required the lowest amount of lactose to produce the maximum concentration of GOS when initial lactose was 370 mg mL^{-1}. Fischer and Fischer and Kleinschmidt (2018) performed a similar assay, however, they evaluated time of enzyme addition instead of enzyme concentration. The yield they found was significantly lower than those obtained in the present work, even for pure enzymes. Another difference, very important for this kind of reactions, is the source of lactose employed in both works. While Fischer and Fischer and Kleinschmidt (2018) evaluated pure lactose as reaction media, our work assayed cheese whey, which as reported by same authors, might interfere significantly with enzyme transgalactosylation ability (Fischer and Kleinschmidt, 2015). Same results were found in a previous work from our group (Rico-Rodríguez et al., 2020).

3.2. FTIR spectroscopy

FTIR data obtained from enzymatic reactions followed a similar trend. As the composition of total carbohydrates in the media changed with time, so FTIR spectra changed. In Fig. 2 it is possible to observe data collected for treatments Kl (Fig. 2A), 1A/3K (Fig. 2B) and 3A/1K (Fig. 2C) as well as spectra of standards of lactose, glucose, galactose and GOS (370 mg mL^{-1}) (Fig. 2D).

Changes in spectra profiles of GOS reactions are more evident in the first minutes of the reactions, especially for Kl and 3A/2K reactions. These changes are linked to the fast drop of lactose concentration after 40 minutes of reaction and to the rise in glucose, galactose and GOS signals.

Indeed, the FTIR spectra of cheese whey (beginning of the reactions) are characterised by a broad absorption band from 1200 to 964 cm^{-1}, with two maxima at 1074 and 1034 cm^{-1} and a shoulder at 995 cm^{-1}. Those signals are mainly related to lactose presence (the main carbohydrate present in cheese whey - > 98% of total carbohydrates). Pure lactose (Fig. 2D) showed a profile

characterised by a band with a maximum at 1157 cm^{-1}, followed by a broad band between 1110 and 970 cm^{-1} with two maxima with similar absorbance at 1074 cm^{-1} and 1034 cm^{-1} and a shoulder at 995 cm^{-1}.

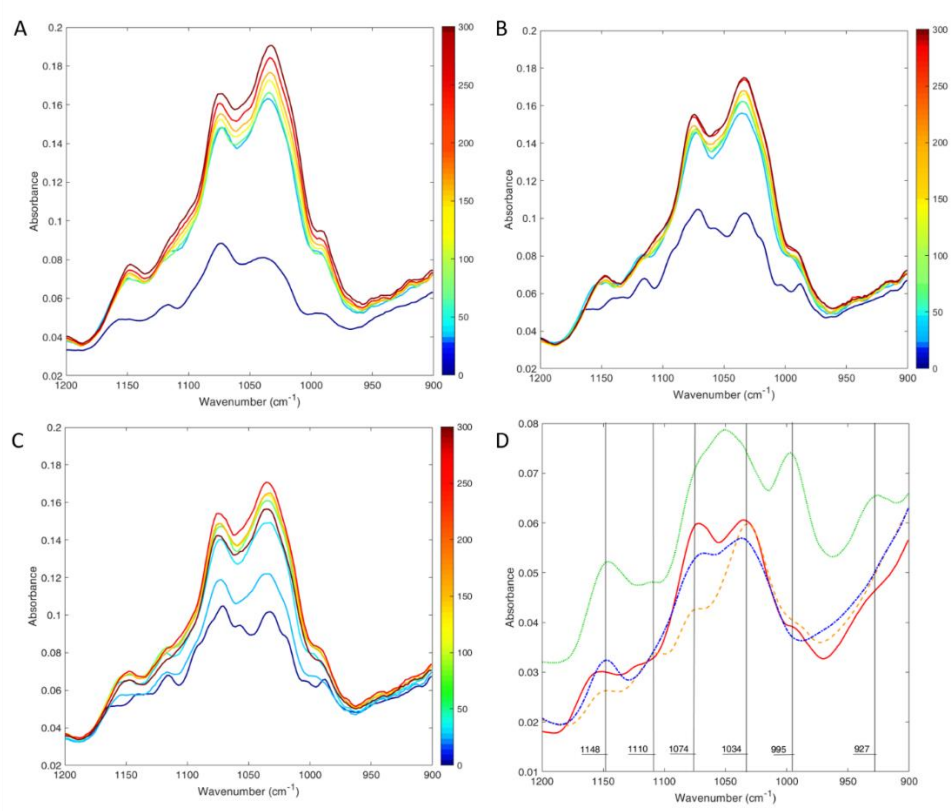


Fig. 2 - FTIR spectra in the region 1200–900 cm^{-1} collected from different enzymatic reactions (A, KI; B, 3A/1K; C, 1A/3K) and pure components spectra (D). In A, B and C spectra are coloured according to reaction time (min). In D spectra correspond to lactose (red, -), glucose (orange, ---), galactose (blue, -.-.) and GOS (green, ···).

Spectra collected along reactions showed slightly different profiles from those collected at the beginning. Particularly, with the progress of time, the peak at

1100 cm^{-1} disappears, the difference between the absorbance at 1074 and 1034 cm^{-1} increases and a small signal becomes visible at 1050 cm^{-1}. Those changes can be attributed to the consumption of lactose and the increase of glucose, galactose and GOS. Indeed, pure glucose and galactose FTIR spectra (Fig. 2D) showed a spectral profile slightly different from the one of lactose. Even if the maxima are similar, glucose FTIR profile differs from the other pure components due to the presence of a shoulder at 1110 cm^{-1} and a high difference between the maxima at 1074 and 1034 cm^{-1}; whereas the spectra collected for pure galactose is characterised by a higher absorbance at 1148 cm^{-1} and the absence of the shoulder at 995 cm^{-1}. For pure GOS spectra (Fig. 2D) the band between 1110 and 970 cm^{-1} does not show double maxima but a higher absorbance at 1050 cm^{-1}. Moreover, the shoulder observed at 995 cm^{-1} for lactose and glucose is a distinct peak followed by a peak at 927 cm^{-1}.

Even if the contribution of the pure compounds can be inferred by visual inspection, the complexity of FTIR data could be better managed by multivariate approaches such as PLS and MCR-ALS, with the aim of monitoring GOS formation.

3.3. PLS regression

The PLS models were calibrated using 64 averaged spectra and internally validated by Venetian Blind Cross-Validation. Furthermore, they were tested for prediction in an iterative way, i.e. by testing they prediction capability by a validation set consisting of 16 averaged spectra of one of the enzymatic reactions performed. In Table 2 are reported the best results obtained applying PLS regression on pre-treated spectral data, they were obtained when the validation set was constructed by KI, 3A/1K and 1A/3K data. Less promising results (data not shown) were obtained when Ao and 2A/2K enzymes were used for validation. These enzymes showed an unexpected behaviour, more specifically the reactions presented a lactose consumption

and its conversion to GOS significantly different from the other three enzymes (Fig. 1A and 1C).

Considering reactions with KI, 3A/1K and 1A/3K enzymes PLS results were good both in terms of RMSEP and R^2_P. In particular, lactose models (range 31.50 to 384.44 mg mL^{-1}) presented very high R^2 values in prediction (R^2_P > 0.96) and a RMSEP value approximately lower than 3 times the RMSEL, an acceptable error for a NIR application (Shenk and Westerhaus, 1996). Similarly, Cocciardi et al. (2004) predicted lactose during hydrolysis reaction in milk. They obtained a reliable PLS models (SEP of 0.20% w/v) considering a lactose variability between 0 and 5 % w/v. Our results, converted into percentage, are comparable being our lowest RMSEP for lactose prediction 6.2% (i.e. 21.9 mg/mL), that is comparable in magnitude with Cocciardi et al. (2004) SEP expressed in percentage (i.e. 4%).

Furthermore, in Table 2 are reported the models to predict in a non-destructive, fast and at-line approach the amount of GOS (range 8.06 -151.62 mg mL^{-1}). The model validated with 3A/1K spectra, pre-processed with the baseline correction, using 7 LV gave a RMSEP of 16.0 mg mL^{-1} and a R^2_P of 0.88. With the same validation set, good prediction ability was obtained for both glucose (range 3.90 to 160.25 mg mL^{-1}) and galactose (range 3.60 to 122.97 mg mL^{-1}). In these cases, RMSEP values are extremely close to RMSEL values, indicating an excellent estimation of these parameters (Shenk and Westerhaus, 1996).

Table 2 - PLS regression models for lactose, GOS, glucose and galactose content prediction from FTIR spectra.

	Test set	LV	Pre-treatment	RMSEL (mg mL^{-1})	RMSEC (mg mL^{-1})	RMSECV (mg mL^{-1})	RMSEP (mg mL^{-1})	R^2_{Cal}	R^2_{CV}	R^2_P
	Kl	**6**	**Baseline correction**		**14.9**	**19.2**	**21.9**	**0.98**	**0.97**	**0.98**
Lactose	3A/1K	3	1st derivative	8.2	24.7	28.3	28.4	0.94	0.92	0.98
	1A/3K	5	Baseline correction		16.6	20.3	22.3	0.98	0.97	0.96
	Kl	7	SNV		9.4	17.9	18.7	0.95	0.84	0.76
GOS	**3A/1K**	**7**	**Baseline correction**	3.2	**10.2**	**14.3**	**16.0**	**0.94**	**0.89**	**0.88**
	1A/3K	7	1st Derivative		11.7	17.3	19.5	0.92	0.84	0.92
	Kl	2	1st derivative		12.6	13.7	14.5	0.91	0.89	0.96
Glucose	**3A/1K**	**6**	**Raw**	8.0	**8.1**	**10.6**	**11.1**	**0.97**	**0.94**	**0.99**
	1A/3K	5	Baseline correction		8.7	10.3	12.1	0.97	0.96	0.93
	Kl	2	1st derivative		11.9	12.9	23.6	0.8	0.76	0.95
Galactose	**3A/1K**	**6**	**SNV**	8.1	**9.6**	**11.5**	**12.4**	**0.91**	**0.88**	**0.97**
	1A/3K	5	Baseline correction		10.1	12.5	12.6	0.92	0.88	0.75

LV, latent variables; SNV, Standard Normal Variate; RMSEL, Root Mean Square Error of Laboratory; RMSEC, Root Mean Square Error of Calibration; RMSECV, Root Mean Square Error of Cross-Validation; RMSEP, Root Mean Square Error of Prediction; R^2, coefficient of determination in calibration (Cal), cross-validation (CV) and prediction (P).

Referring to the work by Cocciardi et al. (2004), minimum SEP values of 0.13% w/v and 0.075% w/v for glucose and galactose prediction were reached, respectively. However, the authors performed a leave-one-out cross-validation within each enzymatic reaction, thus considering a small inter-enzyme variability and a small range of sugar content variability (0 - 2.5% w/v). Another relevant work in the field, by Romano et al. (2016), focused on the development of PLS models to predict the effect of sucrose concentration on fructo-oligosaccharides (FOS). They reached optimal results when predicting glucose production by a PLS model with a R^2_P of 0.98 and an RMSEP of 2.028 g of glucose/ 100 g of sucrose. Being their variability in a range of 0 – 56 g of glucose/ 100 g of sucrose, the %RMSEP will correspond to 3.6, which is lower than our (i.e. %RMSEP = 7.1).

Furthermore, our model performances were evaluated in terms of precision of the prediction with the average composition of all the samples, by means of residual prediction deviation (RPD). GOS and galactose models have an RPD of 2.8 and 2.7 respectively, suggesting a good predictive performance of the models, whereas lactose and glucose presented an RPD > 3 (5.9 and 4.9, respectively), indicating an excellent predictive performance (Saeys et al., 2005).

3.4. MCR-ALS models

The promising results obtained by PLS regression, suggested the possibility of building MCR-ALS models to follow the enzymatic reaction in vision of a Process Analytical Technology able to predict the process kinetics. Thus, an MCR-ALS model for the duplicates of each enzymatic combination has been developed.

The application of ALS procedure to the FTIR spectra collected allowed the resolution of both spectral and concentration profiles by four-component models, retaining at least 99.9% of the total variance, and a LOF lower than 0.44 %.

Three out of four MCR-ALS profiles were assigned to lactose, GOS and a combination of Glu and Gal; whereas the fourth profile was assumed to be an interference used to isolate possible noise in the process as reported by Ahmadi, Tauler, & Abdollahi (2015). This is in agreement with previous works, reporting that the quantification of monosaccharides, Glu and Gal, as one group (Palai et al., 2012) or as a ratio (González-Delgado et al., 2016) can ease the analysis of their effect on the kinetic model for GOS prediction. According to these works, it is reasonable that one of the MCR resulting profiles is associated to the sum of the monosaccharides.

The profile identity was confirmed by comparing the spectral features of the pure compounds (Fig. 2D) and their kinetics measured by HPLC (Fig. 1). Fig. 3 reports the comparison between MCR-ALS components and the measured trends for Kl. Lactose MCR-ALS spectral profile (Fig. 3D) showed the characteristics observed for pure lactose FTIR spectra (Fig. 2D) with a broad band between 1110 and 970 cm^{-1} with two maxima at similar absorbances (1074 cm^{-1} and 1034 cm^{-1}) and a shoulder at 995 cm^{-1}. The related MCR-ALS concentration profile described a fast decrease in lactose within the 40 minutes followed by a plateau until the end of the process monitoring. This behaviour is similar to the lactose content quantified by HPLC analysis, even

though by the chromatography determination lactose reached the plateau less fast, being the concentration close to 0% after 120 min (Fig. 3A). The MCR-ALS profile attributed to GOS showed a peak at 1148 cm^{-1}, a broad band between 1110 and 970 cm^{-1}, a distinct peak at 950 cm^{-1} and a small feature around 927 cm^{-1} (Fig. 3D). All these characteristics were observed for GOS pure spectra reported in Fig. 2D. Even if there was a spectral correspondence, the GOS concentration profile failed in describing the GOS behaviour as observed by HPLC analysis. Indeed, even if the GOS concentration profile described a fast increase in relative concentration within the first 20 minutes, its drop has been much higher than the one recorded by HPLC (Fig. 3B).

Different reasons could explain the failure in kinetic prediction. First of all, it should be considered that in our approach GOS were evaluated as a unique product from the transgalactosylation reaction. However, the biochemical reactions for GOS production are rather a complex mechanism in which lactose is converted to one of a wide group of oligosaccharides derived from lactose (Gosling et al., 2010). Indeed, GOS are a carbohydrate-based molecule group, thus, they can be considered as a class, but the different proportion of di-, tri-, tetra-, penta- and hexa- saccharides might affect the kinetic modelling results. For this reason, trying to explain the reaction through a mathematical or a statistical model is not as simple as one can expect if the total amount of GOS is considered. Several attempts to model the kinetics have been assayed (Jenab et al., 2017; Palai et al., 2012; Warmerdam et al., 2014; Yin et al., 2017), nonetheless, uncertainness is always present.

Furthermore, the considered reaction media, i.e. cheese whey, is very complex. Indeed, it contains salts and whey proteins, as well as a mixture of sugars of different size and linkages; all of them leading to bands overlapping in the FTIR fingerprint region (1200 – 950 cm^{-1}). Even more, the total content of saccharides does not change over time, but saccharides are transformed by the transgalactosylation reaction; so, FTIR measurements could be less

reliable in detecting changes in GOS production. Moreover, to improve the efficacy of this technique, it would be necessary to test standards of every possible carbohydrate formed during the reaction.

Even if the MCR-ALS models failed in assessing GOS production, model reliability was confirmed for monosaccharides released by the reaction, i.e. glucose and galactose. In fact, the spectral profile (Fig. 3D) exhibits features resulting from the combination of pure glucose and pure galactose FTIR spectra (Fig. 2D). In this case, high correspondence was found between the MCR-ALS concentration profile and the monosaccharides concentrations measured by the chromatographic method (Fig. 3C).

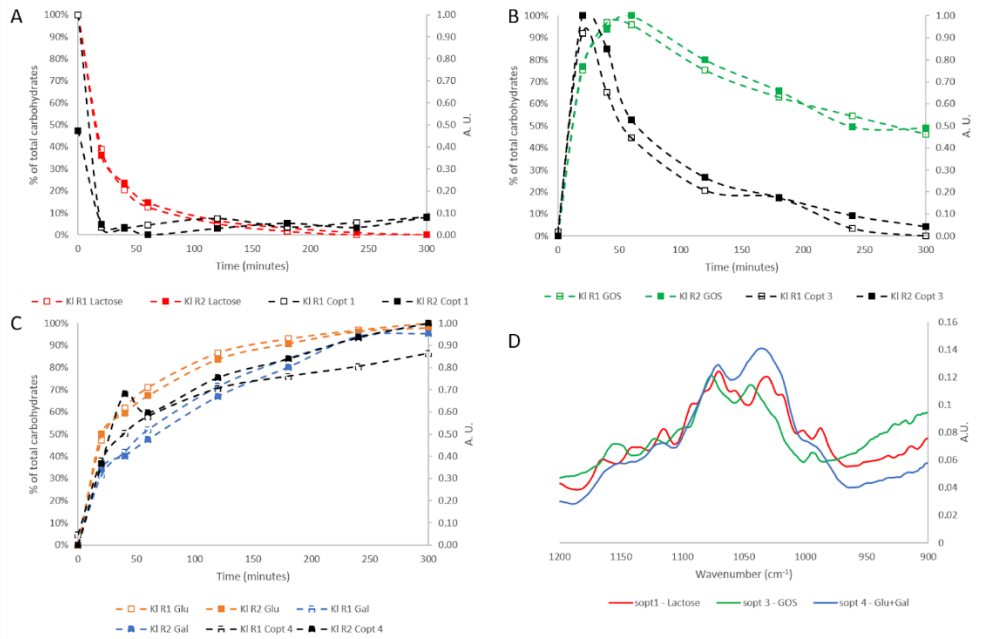


Fig. 3 - MCR-ALS results for the Kl enzymatic reaction: (A) concentration profiles of lactose and lactose concentration as determined by HPLC; (B) concentration profiles of GOS and GOS concentration as determined by HPLC; (C) concentration profiles of glucose-galactose and glucose and

galactose concentrations as determined by HPLC; (D) spectral profiles of lactose, GOS and glucose-galactose.

Similar results were obtained by MCR-ALS models developed for 1A/3K (Fig. 1S) and 3A/1K (Fig. 2S). Indeed, lactose decrease, as well as the glucose and galactose production, has been detected with a faster drop by the MCR-ALS models, especially for 1A/3K duplicates. However, the GOS production was not systematically modelled by MCR algorithm, especially for 1A/3K trials. Regardless GOS typical behaviour after reaching their maximum value is to follow a reduction in their concentration due to hydrolytic enzyme property, this depletion was depicted faster than the observed HPLC data.

4. Conclusions

Different chemometric approaches have been tested to assess a Process Analytical Technology to follow GOS production from lactose naturally present in dairy industry waste by FTIR spectroscopy. PLS regression models demonstrated to be reliable to assess lactose, glucose and galactose content in a non-destructive, fast and at line approach, which could be used for real time applications in the future. Indeed, their prediction by KI or 3A/1K enzymatic reaction led to good fit ($R^2_P > 0.97$) and acceptable errors (RMSEP of 21.9 mg mL^{-1} for lactose, 11.1 mg mL^{-1} for glucose and 12.4 mg mL^{-1} for galactose). Unfortunately, GOS prediction by regression models did not reach the same performance. This could be linked to the development of a model considering GOS as a unique class, whereas it is quite heterogeneous accounting for di-, tri-, tetra-, penta- and hexa- saccharides.

Furthermore, an MCR-ALS model for each enzymatic combination has been developed focusing more on the process kinetics rather than the prediction of each component. Three MCR-ALS profiles were assigned to lactose, GOS and a combination of glucose and galactose. The corresponding MCR-ALS concentration profiles gave trends similar to the sugar concentrations

measured by the chromatographic method, however they demonstrated not to be enough accurate to be implemented. Thus, this work should be considered as a preliminary analysis to demonstrate the potential of FTIR spectroscopy, combined with Chemometrics, to follow the kinetics of GOS production from lactose naturally present in dairy industry waste.

The future implementation of this approach to real time systems will answer the need of Circular Economy by providing a reliable monitoring of dairy waste reallocation, avoiding sample pre-processing, large volumes of organic solvents and long-time of analysis. In any case, further investigation is suggested to better assess specific GOS prediction by a single FTIR analysis.

5. ACKNOWLEDGEMENTS

F. Rico-Rodriguez thanked to Ministry of science, technology and innovation of Colombia – MINCIENCIAS by the grant for Colombian posdoctoral contract No.811/2018.

6. REFERENCES

Akiyama, T., Kimura, K., & Hatano, H., 2015. Diverse galactooligosaccharides consumption by bifidobacteria: implications of β-galactosidase—LacS operon. Biosci. Biotechnol. Biochem. 79, 664-672.

https://doi.org/10.1080/09168451.2014.987204

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254. https://doi.org/10.1016/0003-2697(76)90527-3

Byfield, J., Cardenas, S., Alméciga-Diaz, C.J., Sánchez, O.F., 2010. β-galactosidase and galactooligosaccharides production and applications. Recent Patents Chem. Eng. 3, 17–29.

Chockchaisawasdee, S., Athanasopoulos, V.I., Niranjan, K., Rastall, R.A., 2004. Synthesis of Galacto-oligosaccharide From Lactose Using Beta-

Galactosidase from Kluyveromyces lactis: Studies on Batch and Continuous UF Membrane-Fitted Bioreactors. Biotechnol. Bioeng. 89, 9. https://doi.org/10.1002/bit.20357

Cocciardi, R. A., Ismail, A. A., Van De Voort, F. R., & Sedman, J., 2004. Monitoring of lactose hydrolysis in milk by single-bounce attenuated total reflectance Fourier transform infrared spectroscopy. Milchwissenschaft 59(7-8), 403-407.

Eskandarloo, H., & Abbaspourrad, A., 2018. Production of galacto-oligosaccharides from whey permeate using β-galactosidase immobilized on functionalized glass beads. Food Chem. 251, 115-124. https://doi.org/10.1016/j.foodchem.2018.01.068

Fischer, C., & Kleinschmidt, T., 2015. Synthesis of galactooligosaccharides using sweet and acid whey as a substrate. Int. Dairy J. 48, 15–22. https://doi.org/10.1016/j.idairyj.2015.01.003

Fischer, C., & Kleinschmidt, T., 2018. Combination of two β-galactosidases during the synthesis of galactooligosaccharides may enhance yield and structural diversity. Biochem. Bioph. Res. Co. 506(1), 211-215. https://doi.org/10.1111/1541-4337.12344

Gennari, A., Mobayed, F. H., Catto, A. L., Benvenutti, E. V., Volpato, G., & de Souza, C. F. V., 2019. Kluyveromyces lactis β-galactosidase immobilized on collagen: catalytic stability on batch and packed-bed reactor hydrolysis. React. Kinet. Mech. Cat. 127(2), 583-599. https://doi.org/10.1007/s11144-019-01598-6

González-Delgado, I., López-Muñoz, M.-J., Morales, G., Segura, Y., 2016. Optimisation of the synthesis of high galacto-oligosaccharides (GOS) from lactose with β-galactosidase from Kluyveromyces lactis. Int. Dairy J. 61, 211–219. https://doi.org/http://dx.doi.org/10.1016/j.idairyj.2016.06.007

Gosling, A., Stevens, G.W., Barber, A.R., Kentish, S.E., Gras, S.L., 2010. Recent advances refining galactooligosaccharide production from lactose. Food Chem. 121, 307–318. https://doi.org/http://dx.doi.org/10.1016/j.foodchem.2009.12.063

Grassi, S., Amigo, J.M., Lyndgaard, C.B., Foschino, R., Casiraghi, E., 2014. Assessment of the sugars and ethanol development in beer fermentation with FT-IR and multivariate curve resolution models. Food Res. Int. 62, 602–608. https://doi.org/10.1016/j.foodres.2014.03.058

Grassi, S., Strani, L., Casiraghi, E., & Alamprese, C., 2019. Control and Monitoring of Milk Renneting Using FT-NIR Spectroscopy as a Process Analytical Technology Tool. Foods 8, 405. https://doi:10.3390/foods8090405

Hernández-Hernández, O., Calvillo, I., Lebrón-Aguilar, R., Moreno, F.J., Sanz, M.L., 2012. Hydrophilic interaction liquid chromatography coupled to mass spectrometry for the characterization of prebiotic galactooligosaccharides. J. Chromatogr. A 1220, 57–67. https://doi.org/10.1016/j.chroma.2011.11.047

Jaumot, J., de Juan, A., & Tauler, R., 2015. MCR-ALS GUI 2.0: new features and applications. Chemom. Intell. Lab. Syst. 140, 1-12. https://doi.org/10.1016/j.chemolab.2014.10.003

Jenab, E., Omidghane, M., Mussone, P., Hernandez Armada, D., Cartmell, J., Montemagno, C., 2017. Enzymatic conversion of lactose into galacto-oligosaccharides: The effect of process parameters, kinetics, foam architecture, and product characterization. J. Food Eng. 222, 63-72. https://doi.org/10.1016/j.jfoodeng.2017.11.015

Kumar, V., Sharma, D. K., Sandhu, P. P., Jadaun, J., Sangwan, R. S., & Yadav, S. K., 2020. Sustainable process for the production of cellulose by an *Acetobacter pasteurianus* RSV-4 (MTCC 25117) on whey medium. Cellulose 1-14. https://doi.org/10.1007/s10570-020-03519-6

Li, H., Cao, Y., Li, S., Jiang, Y., Chen, J., & Wu, Z., 2019. Optimization of a dual-functional biocatalytic system for continuous hydrolysis of lactose in milk. J. Biosci. Bioeng. 127, 38-44. https://doi.org/10.1016/j.jbiosc.2018.07.009

Misson, M., Du, X., Jin, B., & Zhang, H., 2016. Dendrimer-like nanoparticles based β-galactosidase assembly for enhancing its selectivity toward transgalactosylation. Enzyme Microb. Technol. 84, 68-77. https://doi.org/10.1016/j.enzmictec.2015.12.008

Mohamed, H., Nagy, P., Agbaba, J., & Kamal-Eldin, A., 2020. Use of near and mid infra-red spectroscopy for analysis of protein, fat, lactose and total solids in raw cow and camel milk. Food Chem. 334, 127436. https://doi.org/10.1016/j.foodchem.2020.127436

Moros, J., Garrigues, S., & de la Guardia, M., 2010. Vibrational spectroscopy provides a green tool for multi-component analysis. TrAC Trend. Anal. Chem. 29(7), 578-591. https://doi.org/10.1016/j.trac.2009.12.012

Otieno, D.O., 2010. Synthesis of β-Galactooligosaccharides from Lactose Using Microbial β-Galactosidases. Compr. Rev. Food Sci. Food Saf. 9, 471–482. https://doi.org/10.1111/j.1541-4337.2010.00121.x

Palai, T., Mitra, S., Bhattacharya, P.K., 2012. Kinetics and design relation for enzymatic conversion of lactose into galacto-oligosaccharides using commercial grade β-galactosidase. J. Biosci. Bioeng. 114, 418–23. https://doi.org/10.1016/j.jbiosc.2012.05.012

Rico-Rodríguez, F., Serrato, J.C.J.C., Montilla, A., Villamiel, M., 2018. Impact of ultrasound on galactooligosaccharides and gluconic acid production throughout a multienzymatic system. Ultrason. Sonochem. 44, 177–183. https://doi.org/10.1016/j.ultsonch.2018.02.022

Rico-Rodríguez, F., Villamiel, M., Ruiz-Aceituno, L., Serrato, J. C., & Montilla, A., 2020. Effect of the lactose source on the ultrasound-assisted enzymatic

production of galactooligosaccharides and gluconic acid. Ultrason. Sonochem., 67, 104945. https://doi.org/10.1016/j.ultsonch.2019.104945

Rodrigues Mano, M.C., Paulino, B.N., Pastore, G.M., 2019. Whey permeate as the raw material in galacto-oligosaccharide synthesis using commercial enzymes. Food Res. Int. 124, 78–85. https://doi.org/https://doi.org/10.1016/j.foodres.2018.09.019

Romano, N., Santos, M., Mobili, P., Vega, R., & Gómez-Zavaglia, A., 2016. Effect of sucrose concentration on the composition of enzymatically synthesized short-chain fructo-oligosaccharides as determined by FTIR and multivariate analysis. Food Chem. 202, 467-475. http://dx.doi.org/10.1016/j.foodchem.2016.02.002

Ruiz-Matute, A.I., Corzo-Martínez, M., Montilla, A., Olano, A., Copovi, P., Corzo, N., 2012. Presence of mono-, di- and galactooligosaccharides in commercial lactose-free UHT dairy products. J. Food Compos. Anal. 28, 164–169. https://doi.org/http://dx.doi.org/10.1016/j.jfca.2012.06.003

Sangwan, V., Tomar, S.K., Singh, R.R.B., Singh, A.K., Ali, B., 2011. Galactooligosaccharides: Novel Components of Designer Foods. J. Food Sci. 76, R103–R111. https://doi.org/10.1111/j.1750-3841.2011.02131.x

Schiele, S. A., Meinhardt, R., Eder, C., & Briesen, H., 2020. ATR-FTIR spectroscopy for in-line anomer concentration measurements in solution: A case study of lactose. Food Control 110, 107024. https://doi.org/10.1016/j.foodcont.2019.107024

Warmerdam, A., Zisopoulos, F.K., Boom, R.M., Janssen, A.E.M., 2014. Kinetic characterization of galacto-oligosaccharide (GOS) synthesis by three commercially important β-galactosidases. Biotechnol. Prog. 30, 38–47. https://doi.org/10.1002/btpr.1828

Yin, H., Bultema, J.B., Dijkhuizen, L., van Leeuwen, S.S., 2017. Reaction kinetics and galactooligosaccharide product profiles of the β-galactosidases

from Bacillus circulans, Kluyveromyces lactis and Aspergillus oryzae. Food Chem. 225, 230–238. https://doi.org/https://doi.org/10.1016/j.foodchem.2017.01.030

5 – General Conclusion

In this thesis project, innovative PAT approaches have been proposed in order to provide dairy industry with efficient tools to improve production processes in a green and sustainable way. Specifically, the work focused on the study of the rennet coagulation process, one of the most critical and important steps in the cheesemaking.

FT-NIR spectroscopy coupled with MCR-ALS algorithm proved to be a reliable technique to control and understand the rennet coagulation process. In fact, the development of MSPC charts permitted a fast detection of possible abnormalities and failures from the very beginning of the process, avoiding wastes in terms of raw materials, product reallocation, energy, time and money. Furthermore, ASCA method allowed to assess, in a multivariate way, the statistical significance of the studied process variables and their interactions. A loading evaluation permitted to evaluate which are the spectral region mainly influenced by each factor, with the aim of developing a simplified sensor to be placed directly along the process chain.

The second part of the work regarded the assessment of FT-NIR spectroscopy reliability in replacing golden standard methods, such as Formagraph, for the monitoring and the control of coagulation process, even using reconstituted milk powder samples. Through this technique, coupled with MCR-ALS algorithm once again, it was possible to assess, among the tested products, the best milk powder in terms of coagulation attitude and occurrence. It was also proved that neither $CaCl_2$ concentration or heat treatment had a significant effect on coagulation attitude and occurrence in the studied conditions. Moreover, the monitoring of coagulation trials of milk reconstituted with skimmed milk powders showed a longer coagulation time when a lower reconstituted milk percentage was used.

In the last part of the work a non-destructive approach based on MIR spectroscopy was developed to monitoring GOS production from cheese

whey, a product commonly discarded from dairy industry processes. In this context, PLS regression method was used to create models that guaranteed the prediction of the specific compounds resultant from the enzymatic reaction, providing high R^2 values and low errors in both cross-validation and prediction. The proposed method could be useful for the quantification of the studied compounds in a faster and cheaper way, avoiding wastes and optimizing the GOS production

6 – Implications and Future Directions

In general, the application of the proposed methods will implicate, with a modest environmental impact, an efficient control of the process, satisfying at the same time law requirements and consumers' needs.

This thesis work provides the scientific groundwork for the future implementation of a PAT approach to an actual industrial-scale dairy process, monitoring and improving the understanding of its critical steps. Unfortunately, it was not possible to finalize a real application of the presented methods to an industrial process. First of all, it is difficult to collaborate with food companies, especially in the dairy sector, where the transition between the traditionally manual work of the "casari", the experts of cheese making, and automation is not easily accepted, mainly due to the difficulty in comprehending PAT techniques. Moreover, to bridge the gap between laboratory experiments and industrial implementation, an engineering step is still missing. Indeed, in order to validate the industrial applicability of the proposed methods, it is crucial to develop simplified systems, easy to use on-line, that can guarantee a reliable and efficient process monitoring and understanding.

Since this thesis work has, with no doubt, fully investigated the reliability of these methods, the lack of industrial applications is an incentive for future works and collaborations. Actually my research work is addressed, in collaboration with Dr. Giovanni Cabassi of CREA-ZA, Research Centre for Animal Production and Aquaculture, to the development of a spectroscopic sensors to be installed directly in a cheese production vat in order to provide in-line measurements, for a real PAT approach. Data, so far collected, are promising but not yet adequate to be part of this thesis work.

The ultimate aim, i.e. a real PAT approach to cheese-making, including by-product valorization, will be crucial for cheese quality management, money

saving and wastage reduction, in a circular economy approach. In this direction future work will be addressed.

7 – Appendices

Scientific dissemination

Oral presentations

- Strani L., Casiraghi E. Near Infrared spectroscopy for process monitoring and quality assessment in dairy sector. Advanced School of Food Proteins. Bergamo, Italy; May 2nd – 4th 2018

- Strani L., Grassi S., Casiraghi E., Alamprese C. FT-NIR spectroscopy to monitor rennet coagulation in milk with different fat levels. VIII Italian Symposium of Near Infrared Spectroscopy. Genoa, Italy; 30th – May 31st 2018

- Grassi S., Strani L., Alamprese C., Marini F., Casiraghi E. Milk renneting: a process study by Near Infrared Spectroscopy. 9th biennial meeting of the International Council for NIR Spectroscopy (ICNIRS). Gold Coast, Australia; September 14th – 20th 2019

- Strani L. Process Analytical Technology Approaches for Dairy Industry. Workshop on the PhD research in Food Systems. Milan, Italy; September 14th – 18th 2020

Poster presentation

- Strani L., Casiraghi E. PAT application in dairy industry using sensing techniques. XXIII Workshop on the developments in the Italian PhD research on Food Sciences, Technology and Biotechnology. Oristano, Italy; September 19th – 21st 2018

UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE AGRARIE E ALIMENTARI

Pat application in dairy industry using sensing techniques

Lorenzo Strani (lorenzo.strani@unimi.it)
Department of Food, Environmental and Nutritional Sciences, University of Milan, Via Celoria 2, Milan, Italy
Tutor: Prof. Ernestina Casiraghi

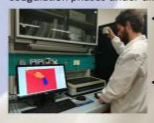
1. State-of-the-Art

Dairy industry needs efficient but, at the same time, inexpensive and fast methods to improve the quality of the production processes, with the aim to satisfy the increase of demand and consumers requests for assurance on origin, nutritional values, environmental impact and quality of the products.

Sensing techniques, used as a tool for a Process Analytical Technology (PAT) approach, are the most promising methods for process control, as it is possible to place instrument probes directly in the production line, allowing to acquire data in real time. In doing so, all steps of food processing and process dynamic can be followed, reducing the possibility of malfunctions, skipping passages that could lead to errors (such as sample pretreatment), and ensuring an high quality final product. For what concerns data elaboration, it is fundamental the use of chemometrics, a set of multivariate statistical methods for process design, data acquisition and data analysis (Kourti, 2006).

- Infrared spectroscopic sensors are the most used in food industry, due to their high-speed analysis and relatively low costs compared to traditional techniques, allowing to obtain a physical and chemical characterization of products.
  NIR spectroscopy coupled with PLS algorithm have been successfully used to determine protein and carbohydrate content in powdered infant formula samples under different motion conditions, with the aim to simulate an industrial application (Cama-Moncunil et al., 2016).

This technique has also been used to control the lactic acid coagulation in milk. In this case, MCR-ALS chemometric algorithm was capable to give an efficient description of the curd formation, providing specific trends for all the coagulation phases under different operative conditions and assessing the fermentation end point (Grassi et al. 2013).



- Imaging can be used as Process Analyzer for the evaluation of shape, size and problems of homogeneity or consistency of samples in a very short time (Kucheryavskiy et al., 2014).
- Electronic nose has been successfully used for the monitoring of cheese ripening process (Trihaas et al., 2005).

This project aims at PAT development in dairy industry processes, using sensing technologies combined with multivariate data analysis, to develop strong and efficient methods for process optimization and food quality characterization and authentication.

2. PhD Thesis Objectives and Milestones

Below are reported the activities in which this PhD project can be divided, according to the Gantt chart in Table 1.

A1) **Process optimization of cheeses and light dairy products.** Study of coagulation process, curd syneresis, texture, and levels of fat, protein and moisture of fresh cheese predicted by FT-NIR, FT-IR, e-nose and Imaging, and compared to reference analysis (A1.1). Lead the process in a larger scale, from laboratory to industrial plant passing through a pilot plant, optimizing it at every stage. (A1.2).

A2) **Application of PAT in a dairy industry process.** The large amount of data collected in A1 will require the use of advanced multivariate data analysis strategies. PCA will be used for exploratory data analysis and, together with MCR, for the real-time monitoring of the industrial process. Based on exploratory results, multivariate control charts, either LV's or MCR resolved components based, will be developed. The results will be compared to conventional statistical process control (SPC) charts (A2.1). PLS algorithm will be implemented to predict quality variables of the selected product in real time (A2.2).

A3) **Implementation of models to monitor cheeses ripening.** Evaluation of the maturation of ripened cheese via FT-NIR, FT-IR, Imaging and e-nose and comparison with traditional methods for protein degradation and lipid oxidation analysis (A3.1). Development of classification models; best chemometrics approach will be assessed during research, based on different ripening stages (A3.2).

A4) **Regression methods for the evaluation of cheese quality variables.** The datasets generated in A3 will be elaborated to extract relevant information on the quality of ripened cheese. PLS regression will be applied to evaluate and predict quality variables (A4.1). Authentication of ripened cheese with Chemometrics methods of class-modeling (e.g. SIMCA) based on fingerprinting approach. Classification of the different quality categories (e.g. based on ageing, or other marketing strategies) trough Chemometrics discriminant tools (e.g. PLS-DA) (A4.2).

A5) **Preparation of scientific papers, oral/poster communications and PhD thesis.**

Table 1: *Gantt chart for the current PhD project*

Activity \ Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A1) ***Process optimization of light dairy products***																								
1) Coagulation, texture, moisture (etc.) assessment																								
2) Upgrading the research scale																								
A2) ***Application of PAT***																								
1) Multivariate Control Charts development																								
2) Real-time monitoring and prediction determination																								
A3) ***Monitoring of cheese ripening***																								
1) Protein degradation and lipid oxidation																								
2) Classification based on ripening stages																								
A4) ***Evaluation of ripened cheese quality variables***																								
1) PLS for quality evaluation																								
2) Authentication and classification																								
A5) ***Papers and thesis preparation***																								

3. Selected References

Kourti T (2006) The Process Analytical Technology initiative and multivariate process analysis, monitoring and control. Anal. Bioanal. Chem. 384: 1043-1048.
Cama-Moncunill R, Markiewicz-Keszycka M, Dixit Y, Cama-Moncunill X, Casado-Gavalda MP, Cullen PJ, Sullivan C (2016) Multipoint NIR spectroscopy for gross composition analysis of powdered infant formula under various motion conditions. Talanta 154: 423-430.
Grassi S, Alamprese C, Bono V, Casiraghi E, Amigo JM (2014) Modelling milk lactic acid fermentation using Multivariate Curve Resolution-Alternating Least Square. Food Bioprocess Technol 7(6): 1819-1829.
Kucheryavskiy S, Melenteva A, Bogomolov A (2014) Determination of fat and total protein content in milk using conventionall digital imaging. Talanta 121: 144-152.
Trihaas J, Vognsen L, Nielsen PV (2004) Electronic nose: New tool in modelling the ripening of Danish blue cheese. International dairy journal 15: 679-691.

XXIII WORKSHOP ON THE DEVELOPMENTS IN THE ITALIAN PhD RESEARCH ON FOOD SCIENCE, TECHNOLOGY AND BIOTECHNOLOGY
ORISTANO 19th · 20th · 21st SEPT 2018

Work in progress

Strani, L., Riccioli, C., Calero, A., Garrido-Varo, A., Perez-Marin, D. Frying oil quality assessment by Fourier Transform Near-Infrared (FT-NIR) Spectroscopy. In preparation

Strani, L., Grassi, S., Alamprese, C., Casiraghi, E., Pricca, M., De Juan, A., Cabassi, G. Milk renneting control by a Process Analytical Technology approach based on NIR spectroscopy. In preparation

PhD periods abroad

January – March 2019: Department of Chemistry (Chemometrics group) - University of Barcelona, Barcelona, Spain. Tutor: Prof. Anna De Juan. Research topic: using MCR-ALS method on milk coagulating NIR spectra

April – July 2019: Faculty of Agriculture & Forestry Engineering (ETSIAM) - Cordoba University, Cordoba, Spain. Tutor: Prof. Ana Garrido-Varo. Research topic: study of the evolution of frying oil quality using FT-NIR spectroscopy and chemometrics.

Awards

- Winner of the international “John Shenk travel grant” offered by International Council of Near Infrared Spectroscopy (ICNIRS) to attend the 19th biennial meeting of the International Council for NIR Spectroscopy (NIR2019)

- Winner of the SISNIR grant offered by Società Italiana di Spettroscopia NIR to attend the 19th biennial meeting of the International Council for NIR Spectroscopy (NIR2019)

- Winner of an Erasmus+ Traineeship Grant for spending four months (April-July 2019) at Faculty of Agriculture & Forestry Engineering (ETSIAM), Cordoba University, Supervisor Prof. Ana Garrido-Varo

Scientific training courses

- Online course: Fundamentals and Applications of Near Infrared Spectroscopy, University of Cordoba, November 2017 – January 2018

- School of Experimental Design (Research group of analytical Chemistry and Chemometrics), Department of Pharmacy, University of Genoa, November 2018

- Workshop: Process Analytical Technologies, Crowne Plaza, Gold Coast, Australia, September 2019

- Online course: Hyperspectral Image Analysis, Eigenvector – May 2020

- Online course: Multivariate Analysis of Spectroscopic Data, Camo Analytics – June 2020

- Online course: Multivariate Analysis in Regulated Environment, Camo Analytics – June 2020

Transferable skills courses (Università degli Studi di Milano):

- Open access – open data e il mondo delle pubblicazioni - November 2017

- La valutazione della ricerca: indicatori bibliometrici e peer review – January 2018

- Research integrity – June 2018

- Tutelare e valorizzare sul mercato i risultati della ricerca – November 2019

- Communication on new media – March 2020

- Valorizzazione creando impresa: fare spin-off all'Università degli Studi di Milano – May 2020

- Sostenibilità e innovazione – June 2020

- Data protection e attività di ricerca scientifica – September 2020