An integrated and multidisciplinary system of analysis, using innovative and emerging technologies, combined with chemometric data evaluation, is becoming a valuable tool for food analysis. The main lines of this approach and the exploitation of innovative possibilities applied to milk products are reviewed and discussed. Special attention is paid to spectroscopic techniques combined with multivariate statistical data analysis. Moreover, a case study of a milk-based dish is reported: the IR spectroscopic characterization and assignment of the main bands of Béchamel sauce is shown. This can serve as a basis for developing multivariate calibration models for useful applications in order to collect information having specific required characteristics, e.g. for discrimination of products and the investigation of fraud. The main perspective is the increased awareness of the importance of sharing data and the creation of an integrated system for calibration and standardization of milk product spectra in the Multicentre Research Network. It aims at making the analysis faster to ensure economic and health safety in a global context.

**Keywords:** Integrated Food Research; Chemometrics; New markers; Milk; Béchamel sauce.

**Key Points and Descriptors of a Modern, Integrated and Innovative Approach of Food Research**

An integrated and multidisciplinary system of analysis is becoming a valuable tool to analyze and study food, taking into account the various aspects of food quality and composition by employing innovative and emerging technologies. Such a multi-technique approach to food research enables utilizing all the resources of quality, safety and traceability in food systems.

A lot of advanced and emerging analytical techniques such as infrared (IR) spectroscopy, multi-elemental analysis, isotopic ratio mass spectrometry have been applied to verify different items, i.e. origin and provenance of foods and authentication concerning the food system [1,2]. The “Integrated Approach” is the key of modern food research and the innovative challenge for analyzing and modelling agro-food systems in their totality.

The experimental methodology must be complemented by suitable data analysis tools. Chemometrics is defined as the relevant discipline to extract the right information from chemical experimental data using mathematical and statistical methods. Nowadays, a large quantity of analytical data is produced from a simple experiment. To handle these data different statistical approaches have been developed and applied to achieve a comprehensive and meaningful analysis [3-5]. Generally, the application of statistical methods in food science allows to highlight effective trends, study and investigate relationships and make conclusions from experimental data [6].

*alessandra.durazzo@crea.gov.it*

[http://orcid.org/0000-0002-7747-9107](http://orcid.org/0000-0002-7747-9107)
The combined uses of experimental data lead to a comprehensive user database for multivariate statistical analysis, a valid complex tool for the development of models and applications of data study: starting from data analysis, through the extraction of qualitative and quantitative information, and possibly leading to enhanced food research in various directions and boost knowledge. Indeed, the multivariate statistical approach applied to the combined chemical data i.e. physicochemical, spectral, multi-elemental, isotopic parameters, allows the evaluation of qualitative, safety and nutraceutical aspects as well as the development and definition of new quality and safety controls, e.g. through the identification of suitable biomarkers.

The following preliminary plan helps to define the principal lines of this approach and the tasks involved:

- Planning and designing experiments in food science: due to the fact that food products are highly complex systems containing hundreds of chemical species, there is a need for analytical technology that can deal with this complexity.
- Conducting experiments and applying experimental methods to get data containing information about food composition and state.
- Processing analytical data and understanding the linkage and inter-relationship among various measurements.
- Using computational methods for data evaluation to extract meaningful information about food composition and state.
- Combining and associating individual parameters to cover different aspects and features of foods.
- Connecting User Databases.
- Using and taking advantage of Chemometrics.
- Producing and Making use of Model Development and Applications.
- Identifying new markers.
- Selecting management tools (farm, food and health).

There is a high demand for new rapid and green analytical methodologies that allow the direct analysis of a sample. The list of potential experimental techniques includes fluorescence, near infrared (NIR), mid infrared (MIR), and nuclear magnetic resonance (NMR) spectroscopies in combination with multivariate data analysis methods. Pooling and combining these methods could be considered as an emerging integrated tool in food research. The applications of spectroscopic techniques coupled with chemometric analysis [7] encompass qualitative and quantitative analysis, food classification, discrimination and authentication, as well as monitoring of contaminants and adulterants [8].

For these purposes, spectroscopic data should be collected and inserted as routine analysis in each food chain at different levels in order to construct a comprehensive library or database of spectra that can be linked and connected to give an Integrated User Database (IUD) of each food product.

An Integrated Food Research Approach Combining IR Spectroscopy and Advanced Chemometrics Applied to Milk

The implementation of standards and criteria of some aspects concerning food safety and quality as well as the ensuring of product authenticity represent the core issues in the dairy sector. The analysis of milk and other dairy products can be performed through advanced techniques including near- and mid-infrared spectroscopy, mass spectrometry as well as chromatography, immune enzymatic assays, Polymerase Chain Reaction (PCR), and electrophoresis [8, 9]. The present study focuses on an integrated food research approach combining IR spectroscopy and advanced chemometrics applied to a milk product.

The application of the multivariate data analysis for the quantification of dairy products has already been reported a couple of decades ago. This was coincident with the advent of new paradigms in analytical chemistry, environmental protection and workplace safety requirements, as well as the demand for food traceability and the added value of products thanks to the introduction of the protected designation of origin (PDO) [10]. Moreover, increasing labor cost and the availability of computational power promoted the rise of chemometrics in the entire food sector [11].
As for nearly all analytical fields, dairy analysis began with univariate methodologies, which proved to be successful for working on the old paradigms; however, multivariate data analysis has gained wider acceptance during the last three decades. The success of chemometrics in this area stems from its ability to extract only the essential and relevant portions of a data set after filtering out noise or unnecessary information [4]. Additionally, chemometrics allows better quality controls. It is fast and eco-friendly, and it offers more information as well as an improved process understanding [12].

This section aims at reviewing the relevant literature on this topic, by giving an introduction to various chemometrics methods and an overview of their applications to milk.

**Chemometrics tools: description and outputs - Exploitation of possibilities**

Chemometrics quantitative methods can be classified in two general groups, first order methods and higher order methods. First order methods use a data vector for each sample as an input signal. This data vector can be provided by any analytical technique that is capable of giving non-scalar results for each sample. For example, this vector could be obtained by varying a parameter such as wavelength and recording intensity or absorbance values as function of the varied parameter like it is done in NIR, MIR, and UV-Vis spectroscopy. Another example of such a vector is a thermogram from DSC analysis, in which the temperature is varied.

**Overview of first and higher order methods**

The general working principle of first order methods is illustrated in Figure 1. Usually, calibration data from well-known samples are recorded and used to train the chemometric algorithm. Thereafter, data from unknown samples can be processed to determine the parameters of interest. Partial least squares (PLS) regression has become the golden standard in chemometrics applied to dairy due to its wide applicability and large availability of the algorithm on many software packages [13, 14]: mathematical [Matlab minitab, R], statistical [SAS, Excel], equipment [OPUS, Shimadzu,], chemometrics dedicated software [Unscrambler, SIMCA]).

**Figure 1. General working principle of first order methods in milk derived products.**

Data flow and usual scopes.

PLS is a "full-spectrum" method, where compression of the data is made using both instrumental signals (X matrix) and analyte concentrations (y vector). The non-linear iterative partial least squares algorithm provides two types of loadings contained in the matrices W (weight loadings) and P (loadings). Both help to explain the maximum covariance among X (instrumental data) and Y (concentration data). The matrices W and P are used to find regression coefficients (r), and finally to predict the analyte concentration in an unknown sample. Like its ancestor, principal component regression (PCR) [15], PLS was used for the determination of chemical parameters [16]: total fat, protein, and carbohydrates were determined using PLS coupled to FTIR with dispersion lower than 0.1% in all cases. Additionally, an indirect parameter, i.e. as kcal/100 ml, was successfully accessed with the same pair. Another indirect measurement that deserves to be mentioned is the determination of the solid non-fat in raw milk using PLS-MIR [17]. The entire fatty acid profile of milk samples was determined using PLS-MIR regression powered by a genetic algorithm in the variable selection, carried out by the group of Ferrand [18].
Other first-order methods based on natural computing methods are backpropagation artificial neural networks (ANN) [19]. ANNs are indicated to operate a non-linear mapping between an input and a target space. ANNs operate using a large number of simple parallel connected arithmetic units (neurons, Figure 2). Mathematically, a neuron can be defined as a non-linear, parameterized and bounded function (i.e. \( f(x) = 1/e^{\sum w_i x_i} \)). The variables of the signal vector \( x (x_1, x_2, x_3, \ldots, x_n) \) are called "neuron inputs" and the calculated value of the function is called "neuron output" \( (y) \). ANNs have a flexibility that makes them adaptable to different kinds of data structure, and liable to being custom-designed.

In principle, ANNs have the same field of action as PLS. However, they are particularly suited to deal with highly non-linear problems, where the traditional statistical methods usually fail. On the other hand, they are prone to overfitting when the complexity of the relationship between input and output variables is low. As all first order methods, ANNs are capable of dealing with interferences, which are present in the calibration step. This feature is often referred to as the “first order advantage”. Support vector machines (SVM) are also supervised learning algorithms that analyze data used for regression analysis [20]. SVMs project a given data set into a higher-order mathematical space, where clear correlations between the observables and measurands are present [21].

It the dairy sector, ANNs were used to determine the total protein content in yogurt using MIR data with an error during calibration about 7% [22]. In the other extreme, ANNs were used to model an automated biosensor for the quantification of a binary organophosphate mixture in milk using a highly non-linear electrochemical signal [23].

Less widespread are the “high order methods”, where data for a single sample are contained in a multidimensional array (a matrix). High order methods (HOMs) can perform determinations in the presence of interferences not taking into account the calibration step. This property is known as “the second-order advantage” and makes HOMs particularly attractive for applications to complex samples. In this context, parallel factor analysis (PARAFAC) and multivariate curve resolution coupled with alternating least squares (MCR-ALS) are the most used techniques and present an inherent second-order advantage. For PARAFAC and MCR-ALS, the calibration samples and test (unknown) samples are put together and are decomposed by the model. Consequently, the number of factors necessary to perform the regression model is determined at once. In theory, they are equal to the number of independent chemical species in the calibration samples and the unknown interferences in the test samples. PARAFAC assumes that the data array for a group of samples follows a trilinear model [24] (Bro, 1997), while MCR-ALS places second-order data for a group of samples adjacent to each other along a qualitative dimension (usually the time dimension). Furthermore,
it assumes that the augmented matrix follows a bilinear model [25]. For quantitative analysis, areas or amplitude of the sample profiles (in the augmented dimension) are computed, and used to build a pseudo-univariate calibration graph. On the other hand, the PLS residual bi-linearization procedure (PLS-RBL) calibrates a PLS model, and assumes that the signals from the interference in the test sample follow a bilinear model. The analyte scores are produced by modeling the residuals of the fit of the test sample data array to the bilinear model, hence the name residual bi-linearization [26].

Morales et al. [27] reported an analytical procedure to determine sulfathiazole in milk by using molecular fluorescence spectroscopy and PARAFAC decomposition [27]. The use of the second order advantage avoided the need of fitting a new calibration model each time that the experimental conditions change. Tetracycline in whey milk was also determined by using PARAFAC and fluorescence [28]. Forchetti and Poppi [29] used MCR-ALSE on NIR hyperspectral imaging data to detect and quantify whey powder, starch, urea, and melamine as adulterants in milk powder. They achieved an LOD of the order of 0.05%. Finally, the quantification of melamine in milk samples was carried out applying PLS-RBL, PARAFAC and MCR-ALS to UV-Vis absorbance measurements at varied pH value [30]. In this comparison, PLS-RBL outperformed the other methods with superior mean recoveries and relative errors of prediction (100.1 ± 2.3%).

Unsupervised Qualitative methods

Unsupervised “classification”, often called clustering, occurs when the algorithm has no a priori knowledge of the groups present in the population and the samples have a set of features with unclear relationships with a class or category. The purpose of unsupervised methods is the discovery of groups of samples which have related features, to allow their separation into different classes.

Principal Component Analysis (PCA) is the most widespread unsupervised tool in the field of food [31]. PCA is a technique used to reduce the dimensionality of a data set. It is primarily used as a display methodology in exploratory analysis. PCA involves calculating the eigenvalue decomposition of the covariance matrix, usually after data centering. Samples that are placed in the new space show closer positions when they have a strong relationship. On the other hand, Hierarchical Clustering Analysis (HCA) [32] allows to see grouping relations between the data. The successive hierarchical subdivisions provide an idea about the grouping criteria in a dendrogram. It is a tree-shape chart that organizes data into subcategories, which are divided to reach the desired level of detail.

Brescia et al. [33] typified the geographical origin of buffalo milk and mozzarella cheese using data of high performance ion chromatography and inductively coupled plasma emission spectroscopy analyzed through PCA and HCA. They compared their results with those obtained by nuclear magnetic resonance and isotope ratio mass spectrometry, which was also evaluated through both multivariate technics. PCA was also applied to data from laser induced breakdown spectroscopy (LIBS) in order to detect milk adulteration with melamine. ANN was subsequently used to access the level of contamination in positive samples [34].

Self-organizing maps (SOM), also named Kohonen neural networks [35], provide a way of representing multidimensional data in a two-dimensional map. The map consists of components called neurons, which have defined positions and are associated with a weight vector of the same dimension as the input data vectors. The procedure for placing a sample from onto the new map is to find the neuron with minimum distance. The weight vector of this neuron (the winner-one) is corrected towards the object position. The procedure is iteratively repeated a number of times (named ages) pre-established by the user. To give an example, SOMs were used to analyze the profile of twelve PCBs (obtained by headspace solid-phase microextraction/gas chromatography electron capture detection) present in breast milk. The resulting data matched to general habits, breastfeeding and environmental conditions of the living place [36].

Factor analysis (FA) [37], Gaussian mixture models (GMM) and K-means [38], completed the commonly used unsupervised methods.
Supervised methods (Classification methods)

Supervised methods are the true Classification methods, where the information about the belonging of each sample to a certain class is known. This knowledge is used for the algorithm to extract the model information, or assign a new sample to a class.

Linear discriminant analysis (LDA), like PCA, is a linear feature reduction method. LDA method that focuses on finding optimal boundaries (parametric) between classes. LDA, the widely used supervised classification method, selects the space directions that achieve a maximum separation among the different classes and uses Euclidean distance to classify unknown samples [39].

LDA successfully recognizes the milk samples according to their thermal processing, i.e. pasteurized milk, sterilized milk, UHT fresh milk and recombined milk (UHT milk having milk powder), with 100% classification accuracy in a cross validation, based on fluorescence response 400–600 nm when samples were excited at 375 nm [40]. PLS-discriminant analysis (PLS-DA) is another linear parametric method, which carries out a PLS of a set Y of binary variables, 0 or 1 (when the sample belongs to the class), on a set X of predictor variables [41]. Back propagation neural network (ANN) and support vector machines (SVM) are the most frequently used analysis [42]. As in the case of PLS-DA, the algorithm makes a correlation among a binary variable when data is non-linearly distributed. The work of Rodriguez and co-workers [43] showed how PLS-DA can discriminate among powder milk samples with normal and low lactose. In a second step, the same algorithm was used to detect sample adulterated with maltodextrine [43]. PCA was used in the early step to check if separation of the groups was possible.

Soft independent modelling of class analogy (SIMCA) puts more emphasis on similarity within a class than on discrimination among classes [44]. SIMCA considers each class separately and performs a PCA on each one, which leads to one model per class. An object is challenged to every class model and it is assigned to the class that produces the smallest residue during the prediction. SIMCA coupled to MIR spectroscopy was evaluated as a rapid method for the detection and quantification of anionic detergent (lissapol) in milk. The classification efficiency for test samples was recorded to be >93% [45]. SIMCA was also coupled to MIR spectroscopy for the detection of soymilk in cow-buffalo milk [46].

For completeness, Canonical Discriminant Analysis (CDA) [47] and extended canonical variate analysis (ECVA) [48] also belong to the linear supervised classification methods.

Design of experiments

Design of experiments (DOE) is used both to determine how factors or parameters affect the production process and to optimize the process in a laboratory or a plant. During the screening phase of DOE, it is determined which factors may significantly affect the response of the experiment. Plackett-Burman or factorial fractional analysis are the most suitable approaches to find the factors which significantly influence the response [49] running a minimum number of experiments.

DOE was used to optimize the response [50], by finding factor values through a mathematical model. For this purpose, Factorial (Fractional Factorial), Central Composite, Box-Behnken and Mixture Models designs are the best choices to set off the objective and the planning of the experiment. The response surface (or space) for each variable to optimize could be calculated with multiple linear regression. The optimal value of the response in the surface generated by many responses could be calculated with Derringer's desirability function [51].

The work of Koc et al. [52] is a good example of DOE. The authors describe the optimization of whole milk powder processing variables [52]. DOE was used to design the experiments for collecting data that was required for modelling the response surfaces. An ANN model was developed to predict the responses of lactose crystallinity and free fat content from the processor screw speed, process temperature, milk powder feed rate and lecithin addition rate. Finally, a genetic algorithm was used to search for a combination of the process variables for maximum free fat content and maximum crystallinity.
Applications of Spectroscopic Techniques and Chemometrics to milk

Fourier-transform infrared (FTIR) spectrometry has been proposed for the determination of milk components, such as fatty acids [53-56], protein fractions [57-59], lactoferrin [60], enzymes [61] and minerals [62-66]. Concerning the quantification of minerals, Soyeurt et al. [62] used IR spectroscopy and prediction models for Ca, K, Mg, Na, and P on a multi-breed data set obtained from sampling over a period of one year. The results of this study showed the capability of quantifying the calcium and phosphorus content in bovine milk directly from the IR spectra.

The feasibility of using FTIR to detect heat induced conformational rearrangements of milk proteins, as well as protein-protein and protein-lipid interactions was studied by Grewal et al. [67, 68]. Moreover, they were able to identify changes in the structure and interactions of lipids, proteins, and carbohydrates and proposed them as markers of shelf life [67, 68]. The association of lipids with milk α- and β-caseins was studied by Bourassa et al. [69].

As generally true in the food sector, the recent advancements in instrumentation as well as chemometric pattern recognition techniques have amplified the range of IR spectroscopy applications to dairy products: monitoring contaminants and adulterants [70], classification [71, 72], discrimination, authentication [73], etc.

Kamal and Karoui [74] reviewed the applications of the spectroscopic-chemometric approach in the dairy sector. Their review impressively underlines that this approach represents a powerful tool for quality, safety, and authenticity analysis of milk. It is worth mentioning the work of Capuano et al. [72] that consisted of proving fresh grass feeding, pasture grazing and organic farming using FTIR spectroscopy of bovine milk. The authors concluded that organic and conventional milk can be discriminated with acceptable accuracy through FTIR.

Concerning applications on safety items, in a review on contaminants of liquid foods, Jha et al. 2016 [75] summarized the potential advantages and limitations of various techniques, including physicochemical methods, chromatography, immunoassays, molecular, electrical, spectroscopy with chemometrics, electronic nose, and biosensors. Their work highlighted how spectroscopy in combination with chemometrics represents a rapid, precise, and sensitive approach for these food products. The recent advances on determination of milk adulterants are also discussed in review articles of Poonia et al. [76] and Nascimento et al. [77]. Spectroscopic technique coupled with chemometrics have been successfully used to detect the adulterants in milk e.g. melamine, hydrogen peroxide, urea, glucose, whey and tetracycline hydrochloride [78-81]. Numerous applications were carried out on melamine in milk and in infant formula [82, 83].

Also, a method for the detection of aflatoxin M1 in milk using spectroscopy and multivariate data analyses was developed [84]. Kümmel et al. [85] used chemometric-assisted FTIR spectroscopy for tracking Staphylococcus aureus bacteria from dairy cow to cheese. Furthermore, mid infrared spectroscopy was used for investigating the physiological status of cows [86, 87], their metabolic profile [88], their body energy status and energy intake [89, 90], and their pregnancy status [91]. Another example of a useful application was reported by Chessa et al. [92]: the study of milk coagulation properties by FTIR spectroscopy was proposed as a potential selection criterion for the Italian Holstein. Dagnachew et al. [93] studied the genetic and environmental components of goat milk with FTIR. They found substantial genetic differences/variations and pointed out the possibility of using FTIR spectra as a direct prediction approach and as a monitoring tool in herd management. In this regard, Bonfatti et al. [94] compared the common method of exploiting IR data in animal breeding (estimating the breeding values for the traits predicted by infrared spectroscopy) and an alternative approach. The latter is the direct use of the spectral information (direct prediction) in estimation of breeding values for fine composition and technological properties of milk. The authors concluded that the direct approach is more likely to be effective for traits more related to the main sources of spectral variation (i.e., protein and fat). However, they also concluded that further research is needed to study and understand spectral genetic variations.
Another example of the integrated food research approach applied to the milk chain is worth mentioning: Visentin et al. [95] identified factors associated with a range of milk processing characteristics that were predicted by IR spectroscopy analysis. The objectives of this study included the construction of a large database of seasonal calving, grass-based dairy cows, and the development of a model to be used as a support tool that can be employed by the dairy industry to predict and manage their product portfolio influenced by milk process ability.

Aside from individual studies, there are several multicenter projects being carried out in the European Network concerning the use of spectroscopy in the food and dairy sector. One aim is the creation of a universal spectroscopic calibration by the standardization of milk samples. All spectra will be merged in a common spectral database in order to create more robust calibrations that can be used throughout the Center Network. This will allow the retroactive application of calibration models for the prediction of milk traits. Eventually, the direct use of the spectral data without restrictions will promote the development of new indicators for dairy farm management, animal health and efficiency [96-98].

A Case Study of Milk-Based Dish: Qualitative Approach
Case Study Description and Methods

Milk is a basic ingredient of many industrial products and home-made food preparations. An example of a milk-based recipe is a popular sauce frequently used in international and also in Italian dishes: Béchamel sauce. Among other dishes, Béchamel sauce was studied within “QUALIFU” [99], concerning the composition analysis and the investigation of nutritional properties of some Italian dishes commonly consumed and representing traditional Italian cuisine. Samples were experimentally prepared in a dedicated lab-kitchen following a validated and standardized protocol developed within the EuroFIR Network and described in Durazzo et al. [100]. In detail, for every selected recipe, a document collection was carried out from the most popular and traditional cookbooks in Italy (Il cucchiaio d'argento; La cucina italiana, etc). A “standard recipe” was identified and one “preparation protocol” was elaborated in order to just establish ingredients, amounts, preparation and cooking techniques (time, temperature, utensils, etc.). The sampling plan considered the collection of single ingredients in different retail stores and supermarkets. Ingredients were purchased by collecting the main food brands and/or varieties of the same product. Béchamel sauce was made using: 6 brands of whole milk, 6 brands of butter and 2 brands of soft wheat flour (00 type). Each ingredient from all brands was properly weighed and then combined to make a composite sample (pool) before using it for the preparation of the ready-to-eat dish. The dish was assembled and cooked by trained persons according to the “preparation protocol” of the “standard recipe”, by using common household methods and utensils. Identical batches, about 2 kg of each dish, were produced twice in different periods. After cooking, the prepared dishes were weighed once more. Then, they were homogenized, frozen at -30 °C, and then lyophilized for subsequent analyses. Each analysis was carried out in triplicate. In the previous study [100], the focus was on traditional analytical methods to determine proximate composition, mineral content, total polyphenol content and antioxidant properties. In contrast, the present case study aims at applying and evaluating FTIR spectroscopy for food analysis taking Béchamel sauce as an example.
More details about the Béchamel sauce preparation (ingredients, quantity, methods and time of cooking) are reported in Table I.

<table>
<thead>
<tr>
<th>Name</th>
<th>Ingredients (g/100g)</th>
<th>Cooking</th>
<th>Timing (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Béchamel sauce</td>
<td>Milk (83), Butter (8), Flour (8), Salt (0.5)</td>
<td>Melt the butter, add flour, then the milk and cook gently, until mixture thickens</td>
<td>33.50</td>
</tr>
</tbody>
</table>

The FTIR spectra were recorded on a Nicolet iS10 FT-IR spectrometer equipped with a diamond crystal cell for ATR. Spectra were acquired (32 scans/sample or background) in the range of 4000-650 cm\(^{-1}\) at a nominal resolution of 4 cm\(^{-1}\). The spectra were corrected using the background spectrum of air. The analysis was carried out at room temperature. For each measurement, each lyophilized sample was placed onto the surface of the ATR crystal. Before acquiring a spectrum, the ATR crystal was carefully cleaned with wet cellulose tissue and dried using a flow of nitrogen gas. The cleaned crystal was checked spectrally in order to ensure that no residue was retained from the previous sample. The spectrum of every sample was collected 5 times.

**RESULTS AND DISCUSSION**

Figure 3a shows the FTIR spectra of two experimental Béchamel sauce preparations. It is possible to discern numerous peaks, which correspond to functional groups and modes of vibration of the individual components. The spectra are affected by the time and the type of cooking.

The spectra in Figure 3a contain a multitude of bands that are more or less characteristic of food samples. The broad band peaking at around 3290 cm\(^{-1}\) corresponds to the OH stretching modes. Water molecules dominate the overall band, but other components containing hydroxyl groups contribute as well. The broad appearance of the band is a result of the complex hydrogen-bonding network. Consequently, the broad band can be deconvolved into sub-bands corresponding to different hydrogen-bonding states. Usually, six individual sub-bands can be observed [101-103]. The lower the wavenumber of an individual sub-band, the stronger the hydrogen-bonding interactions [104]. A detailed analysis of this network, however, is beyond the scope of the present work.
In the remainder of the spectrum, more specific bands characteristic of Béchamel sauce were highlighted and identified and characteristics bands were assigned as follows.

- **CH stretching region:** asymmetric stretching vibrations of CH\(_2\) and CH\(_3\) groups are found at 2954 cm\(^{-1}\) and 2917 cm\(^{-1}\), respectively. The corresponding symmetric ones are located at 2871 cm\(^{-1}\) and at 2850 cm\(^{-1}\). These bands are mainly associated with the alkyl chains of the fatty acids [16, 105, 106]. - **Spectral region >1500 cm\(^{-1}\):** The spectral band at 1741 cm\(^{-1}\) and the shoulder band at 1728 cm\(^{-1}\) are attributed to the absorption of the C=O bonds of the ester groups. They are present in the fat fraction (triglycerides, in which the glycerol is connected with the fatty acids by ester groups) related to the milk and butter ingredients. Further C=O bonds are present in the carbonyl groups of carboxylic acids and proteins [16, 106-108]. The shoulder at 1728 cm\(^{-1}\), that overlaps the stretching vibration at 1741 cm\(^{-1}\) and leads to a broadening of this peak at the low wavenumber wing, indicates a different molecular oxidation and/or protonation state. For example, the presence of free fatty acids, i.e. COO- is a possible explanation here [108]. The peak at 1236 cm\(^{-1}\) is likely the corresponding C-O stretching mode of the carboxylic acid groups. This is in good agreement with previous work on protonated and de-protonated carboxylic acid groups [109]. The appearance and shape of these bands is probably related to the time and type of cooking and hence may serve as a suitable spectroscopic marker in future work. The region 1700-1500 cm\(^{-1}\) includes two main parts characteristic of proteins with amide groups: an amide I band located around 1700–1600 cm\(^{-1}\) that is related to the stretching of C=O stretching vibrations of peptide bonds (coupled with the deformation of the N–H amide bond I), and an amide II band at 1600–1500 cm\(^{-1}\) that is characteristic of C-N stretching vibrations combined to N–H bending modes [107; 110-112]. Detailed studies on specific signals of the protein secondary structure (i.e.α-helix, β-sheet, turn structure β-antiparallel, random coil) for the free α- and β-caseins, and also their interactions with lipids were reported elsewhere [69]. Bassbasi et al. [71] found bands in the spectra of milk that could be assigned to amino acid side chain vibrations due to tyrosine at about 1515 cm\(^{-1}\), phenylalanine at about 1498 cm\(^{-1}\), and proline at about 1454 and 1438 cm\(^{-1}\). As reported by Murphy et al. [113] the peak at absorbance at 1645 cm\(^{-1}\) could be related to Schiff base intermediates formed during the course of Maillard browning. The latter is a reaction of amino acids and glycosides.

- **Spectral region <1500 cm\(^{-1}\):** This spectral is commonly referred to as the fingerprint region. This region is, on the one hand, very rich in information but, on the other hand, difficult to analyze due to its complexity. The fingerprint region includes the amide III spectral region that corresponds to the secondary structure of proteins and is related to N-H bending and C-N stretching vibrations as well as CH\(_2\) scissoring of the acyl chains of lipids [114]. Furthermore, it includes the region 1200-900 cm\(^{-1}\), which is associated to polysaccharide peaks related to C–C and C-O stretching modes. In detail, the bands at 1150 cm\(^{-1}\) and 1018 cm\(^{-1}\) are dominated by stretching vibrations of the C–O bond [115]. The peak at 1076 cm\(^{-1}\) is related to the bending of hydroxyl groups (OH) of lactose [16, 116]. Moreover, as reported by Etzon et al. [110], the signatures in the 1060 to 1100 cm\(^{-1}\) range are associated with phosphate groups covalently bound to casein proteins. Maamouri et al. [53] and Karoui et al. [117] used only the region 1500-900 cm\(^{-1}\) as fingerprint region for milk. Because of some interesting signatures, we extend this range for Béchamel sauce. The band at 966 cm\(^{-1}\) is characteristic of out-of-plane bending of –HC=CH– groups of disubstituted olefins in trans conformation [118]. The mode of the corresponding cis conformation is the band at 916 cm\(^{-1}\). The observation of both features in the spectra of Béchamel sauce is presumably due to the presence of trans and cis CLA conjugated fatty acids in butter and other dairy products [119]. The band at 702 cm\(^{-1}\) is likely due to overlapping peaks of the CH\(_2\) rocking vibration and the out-of-plane vibration of cis-disubstituted olefins, related to the presence of long chain hydrocarbons [108].

This qualitative approach aiming at the characterization and assignment of the main IR bands of Béchamel sauce represents the basis for the future development of multivariate calibration model. This will be the key for discrimination applications and the investigations of fraud. In this context, the study of Nedeljkovic et al. [120] examined the feasibility of the discrimination of dairy creams and cream-like
analogues (analogues with sunflower oil, coconut oil and palm oil in different milk fat/vegetable fat ratios) using Raman Spectroscopy and chemometric analysis. However, ATR-IR spectroscopy seems to be a more user-friendly and straightforward approach.

We have tested the application of PCA to the 10 individual spectra recorded from the two Béchamel sauce batches (5 spectra each). The results of the PCA are plotted in Figure 3b. The averaged spectra in Figure 3a shows only minor differences, which can be attributed to small differences in the preparation of the initial samples and the ingredients. The PCA, however, is capable of utilizing the subtle changes in the spectra and accomplishes a clear distinction between the two batches. Furthermore, this indicates that the homogenization procedure used after preparation of the dishes works well. Therefore, we can conclude that the combination of FTIR and PCA is a powerful method in the field of food research.

CONCLUSIONS

It is clear how the joined approach of advanced technologies with the multivariate statistical approach enlarges the possibilities to exploit a wide range of food traits and aspects: this direction is leading towards new quality and safety biomarkers. The main perspective is the increased awareness of the importance of sharing data and the creation of an integrated system for calibration and standardization of milk spectra in Multicentre Research Network.

The case study of applying FTIR and PCA to Béchamel sauce has shown that combined spectroscopy and chemometrics is a powerful tool for food research and beyond.

REFERENCES

An Innovative and Integrated Food Research Approach: spectroscopy applications to milk and a case study of a milk-based dish


An Innovative and Integrated Food Research Approach: spectroscopy applications to milk and a case study of a milk-based dish


