LWT - Food Science and Technology 174 (2023) 114429

Contents lists available at ScienceDirect

LWT

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

Application of UV-VIS-NIR spectroscopy in membrane separation processes for fast quantitative compositional analysis: A case study of egg products

Gema Puertas, Patricia Cazón, Manuel Vázquez *

Department of Analytical Chemistry, Faculty of Veterinary, University of Santiago de Compostela, 27002, Lugo, Spain

ARTICLE INFO

Keywords:
Membrane filtration
Spectroscopy
UV-VIS-NIR
Chemometrics
Egg products

ABSTRACT

Membrane separation technology is achieving broad applications in research, food and pharmaceutical industry. In this study, the application of UV-VIS-NIR spectroscopy to quantify the composition of the fractions obtained in a membrane separation process was assessed. An egg membrane filtration was employed as a case study to develop predictive models for a rapid determination of dry matter, protein and cholesterol composition in the feed and the fractions (retentate and filtrate). Whole egg plasma and egg yolk plasma obtained by centrifugation, and egg white were tangential filtrated using different pore size and membrane materials. Transmittance UV-VIS-NIR spectra showed significant differences between samples. Cholesterol, dry matter and proteins were predicted with UV-VIS-NIR spectroscopy with good statistical significance (RPD_{EV} over 3). UV-VIS-NIR spectroscopy combined with chemometric tools demonstrates to be an efficient green method for composition analysis in the membrane filtration process. It is reliable, non-destructive, quick and environmentally friendly. Moreover, it can be applied on-line for real-time quantitative compositional analysis.

1. Introduction

The use of membrane filtration technology as a processing and separation method in food industry is gaining wide applications. It is considered a green technology because it could avoid the use of additives and chemicals like shelf-life extenders or solvents for extraction, which is better for environment and human health. In addition, membrane separation processes are more efficient in economic and energetic terms compared to high temperatures treatments like pasteurization and sterilization. Furthermore, they preserve the natural taste and nutritional value of food products with heat-sensitive components (Dhineshkumar & Ramasamy, 2017). Something that society expects from food producers to accomplish a healthy lifestyle, while, at the same time, help them to adapt to increasingly strict environmental requirements. Like that membrane techniques are an interesting alternative to many conventional methods of food production (Staszak & Wieszczycka, 2022). Moreover, membrane separation is an attractive technique in other fields, such as research or pharmaceutical industry. Membrane processes can reduce the content of water (concentrate) or isolate compounds.

Eggs are a source of macro and micronutrients as well as the vessel of many bioactive compounds. Most of them are still unexplored and they may be of major interest for human health in preventing/curing diseases (Réhault-Godbert, Guyot, & Nys, 2019). Therefore, egg research is compelling and membrane separation could be essential, because it has the potential of isolating while maintaining protein activity (Li et al., 2022). For example, ultrafiltration has already been employed to separate egg white proteins (Li et al., 2022), to remove added water from egg yolk systems (Primacella, Wang, & Acevedo, 2018), or to isolate egg yolk antioxidants from protein enzymatic hydrolysates (Chay Pak Ting et al., 2011). Recently, the combination of membrane separation with chromatography (called membrane chromatography) has shown efficient results in egg white protein purification (Li et al., 2022).

The membrane separation processes in crossflow are also called tangential filtration and included ultrafiltration (<100 nm) and microfiltration (≥0.1 μm). Both of them are pressure-driven. At each process, two products are obtained from the feed introduced into the membrane system. Retentate is the fraction retained by the membrane. Filtrate or permeate is the fraction that passes through the membrane. It is crossflow because the feed is pumped parallelly with the membrane and it was possible to recirculate the retentate back to the feed flow. 'Pressure-driven' means that the main driving force for separation of these processes is the pressure discrepancy between retentate side and filtrate side, called transmembrane pressure (TMP) (Dhineshkumar &

* Corresponding author.
E-mail address: manuel.vazquez@usc.es (M. Vázquez).

https://doi.org/10.1016/j.lwt.2023.114429
Received 13 September 2022; Received in revised form 15 December 2022; Accepted 3 January 2023
Available online 8 January 2023

Ramasamy, 2017).

A filtration process seeks to produce two fractions that differ in composition from the starting sample. Then, the content of the new fractions must be known. If it were possible to determine the precise concentration of these fractions at any given time, it would be conceivable to halt the separation process as soon as the desired composition is found. Nevertheless, measurements could be tedious, and they can difficult the protocol implementation, as some authors described (Allègre, Moulin, Gleize, Pieroni, & Charbit, 2006). UV-VIS-NIR spectroscopy could be the solution. Spectroscopy has proved to be a useful technique combined with chemometrics to substitute traditional quantification methods. Several components can be determined simultaneously from a single spectrum with the help of the multivariate calibration process (Puertas, Cazón, & Vázquez, 2023). The UV spectroscopy is used to quantify proteins because peptide bond and aromatic amino acids are detected at 205 nm and 280 nm, respectively (Simonian & Smith, 2001). Visible spectroscopy targets color assessment and pigment analysis, while NIR spectroscopy allows the evaluation of macro constituents (Walsh, Blasco, Zude-Sasse, & Sun, 2020). For example, NIR transmission spectroscopy on other liquid egg products had already shown successfully results for the determination of composition parameters. However, those studies were sample destructive because employed some chemicals, since the spectrophotometer required a sample preparation (Osborne & Barret, 1984). Then egg product manufacturers were missing possibilities for fast raw material control (Karoui et al., 2009). Nowadays, industrial on-line sensors are highly feasible, with no sample preparation required. These advances in conjunction with the chemometric tools made possible on-line spectroscopic applications that provide quick answers (Porep, Kammerer, & Carle, 2015). Spectroscopic methods are currently energy-efficient, non-destructive, non-invasive, easy-to-use and inexpensive. They are a formidable green chemistry tool and environmentally sustainable analytical technique capable of handling a large sample size in short time and without solvents. For example, NIR spectroscopy together with appropriate chemometrics has become a routine analytical tool for the determination of intact olive drupes composition, in the field and at the mill (Grassi et al., 2021).

Fourier-transformed mid-infrared spectroscopy has been assessed to monitor the composition of the fractions produced at ultra- and nanofiltration of milk and whey (Franzoi, Manuelian, Rovigatti, Donati, & De Marchi, 2018; Solís-Oba et al., 2011). However, to our knowledge, UV-VIS-NIR spectroscopy has never been assessed for composition quantification in a membrane separation process. Therefore, the main objective of this research was to assess the application of UV-VIS-NIR spectroscopy to quantify the composition of the fractions in membrane separation processes. In this work, two membrane separation processes were studied: ultrafiltration (<100 nm) and microfiltration (≥0.1 μm). Egg products were employed as case study due to the potential application of these processes on them in research and industry. Whole egg, egg yolk and egg white were studied separately to explore the influence of their components on both techniques.

2. Materials and methods

2.1. Membrane separation process

Eggs were bought at local supermarkets. A total of 89 shell eggs were employed to prepare 11 starting egg products. These were 7 liquid whole eggs, 2 liquid egg yolks and 2 liquid egg whites. Each whole egg product was made with 8 or 10 eggs. For albumen and egg yolk products 12 and 13 eggs were broken and manually separated as white and yolk. All starting egg products were homogenised using a high-performance homogenizer (Ultra Turrax® T18 19G, IKA, Staufen, Germany) at 1064 g for 1 min. Before the membrane separation process, granules from egg yolk were removed by centrifugation to reduce membrane fouling. But liquid egg yolk was diluted 1:1 with tap water previously to centrifugation. Like that all the samples centrifuged had around 75% water. Following, liquid whole egg and diluted liquid egg yolk samples were centrifuged (Sorvall Contifuge Stratos, Thermo Fisher Scientific, Waltham, MA, USA). Centrifugation conditions were settled to 12,000 g, 40 min and 8 °C according to previous study (Puertas & Vázquez, 2021a; 2021b). After, plasmas were manually separated from granules.

Therefore, a total of 11 initial egg products were investigated for membrane separation: 7 whole egg plasmas, 2 egg yolk plasmas and 2 liquid egg whites. To simplify, liquid egg white is called egg white henceforth. Egg products were diluted 1:1 with water before filtration. KrosFlo® Research IIi Tangential Flow Filtration (TFF) System, (Repligen, California, USA) was used. This system employs hollow fiber modules, and it was settled in a batch filtration configuration. Fig. 1 describes the characteristics of the membranes employed with every initial egg product. To facilitate the interpretation of the results, each membrane separation process or run was given a number from 1 to 11 (Fig. 1). Microfiltration was achieved with three pore sizes: 0.1 μm, 0.2 μm and 0.65 μm. While ultrafiltration was performed with a molecular weight cut-off (MWCO) of 750 kDa. Three hydrophilic organic polymeric membranes were studied: Polyethersulfone (PES), modified polyethersulfone (mPES) and mixed cellulose ester (ME). According to the hollow fiber inner lumen and their effective length, five surface areas were analysed from 500 cm^2 to 2600 cm^2.

2.2. Reference methods for composition determination

The composition parameters studied on the initial feed (plasmas and egg white), retentate and filtrate were: dry matter (DM), proteins and cholesterol. DM content was determined by the gravimetric method. Between 2 and 5 g of sample were weighted, dried for 24 h in a vacuum oven at 105 °C and then weighted. DM content was expressed as g dry matter per 100 g fresh sample.

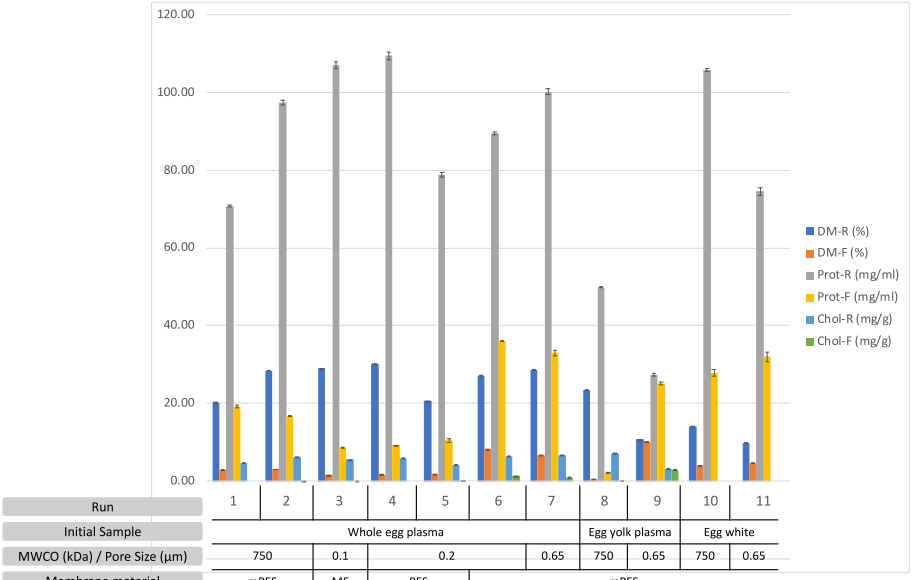
Protein content was measured with the microvolume spectrophotometer Nanodrop 2000 (Thermo Scientific™, Waltham, MA, USA). The method followed is described elsewhere (Puertas & Vázquez, 2021a). Protein content was expressed as mg per ml of sample. An enzymatic kit method (Enzytec™, R-Biopharm AG, Darmstadt, Germany) was employed for cholesterol quantification based on previous studies (Puertas & Vázquez, 2019). The sizes of the samples for analysis were 0.5 g for egg yolk, 1 g for liquid whole egg, 1.5–2 g for plasmas, 1–2.5 g for retentates and 5 g for filtrates. These quantities were decided based on preliminary studies to work within the detection limits stablished by manufacturer. Cholesterol content was expressed as mg per 1 g of egg product. All the composition determinations were done at least in duplicate, except for protein content that measurements were read a minimum of three times.

2.3. Spectra collection

The UV-VIS-NIR spectra of each egg product and egg white was measured with Spectrophotometer Jasco V670 (Jasco Inc., Hachioji, Tokyo, Japan). The cuvette employed was made of Quartz Suprasil ® 300 with a 1 mm light path (Hellma GmbH & Co. KG, Mulheim, Germany). UV-VIS-NIR spectra were acquired in transmittance mode (T) at 2 nm intervals in the range of 190–2500 nm. Samples were measured at 22–24 °C. The UV region was considered from 190 to 380 nm, the VIS region from 380 to 780 nm and the NIR region from 780 nm up to 2500 nm (Porep et al., 2015). Each sample was measured twice making a total of 66 UV-VIS-NIR spectra with 2311 variables each. The spectral data were collected with Spectra Manager II software (Jasco Inc., Hachioji, Tokyo, Japan).

2.4. Statistical analysis and chemometrics

Unscrambler® software Version 10.5 (Camo, Oslo, Norway) was employed for pre-processed treatments and chemometric analysis of UV-



Run	1	2	3	4	5	6	7	8	9	10	11
Initial Sample	Whole egg plasma							Egg yolk plasma		Egg white	
MWCO (kDa) / Pore Size (µm)	750	0.1	0.2				0.65	750	0.65	750	0.65
Membrane material	mPES	ME	PES	mPES							
Surface area (cm²)	1600	720	500	2600	1075	1600	1075	1600	1075		

Fig. 1. Composition results (DM = dry matter; Prot = protein; Chol = cholesterol) of the fractions obtained (R = retentate; F = filtrate) and the parameters of the membrane separation systems. MWCO: molecular weight cut off.

VIS-NIR spectra. Due to, for example, variations in temperature, density and spectral noise of the spectrometer, NIR spectra are frequently characterized by unfavorable spectral variations and baseline shifts. Pre-processing of spectra can remove noise caused by the system error and the random error (Puertas & Vázquez, 2019). Different pretreatments on spectral data were examined to overcome these drawbacks: baseline correction, standard normal variate (SNV), multiplicative scatter correction (MSC), normalize, first derivative using Gap Segment transformation with gap and segment size of 3 and 2, respectively and detrend.

The chemometric approach employed in this work included multivariate regression methods that analysed the spectral data for sample and instrumental calibration. Principal component analysis (PCA) was firstly used to elucidate the structural relationships in the data set for sample pattern recognition. PCA relies on the decomposition of x-variables (spectra data) and represents the data set in a new orthogonal coordinate system, eliminating the collinearity between the x-variables (Bwambok et al., 2020). The regression models studied were partial least squares regression (PLS) and principal component regression (PCR). All samples were mean centered before data analysis. The models obtained were validated with random cross validation (20 segments with 2 or 3 samples in each) and with an external validation (test set validation). Firstly, prediction equations were developed using the complete dataset (n = 33) and those were validated using cross validation. To evaluate the reliability and performance of the models for accurate determination of analyte concentrations different parameters were used. Specifically, calibration was evaluated with the values of root mean square error of calibration ($RMSE_C$) and the coefficient of determination of calibration (r^2_C). For cross validation, root mean square error of cross validation ($RMSE_{CV}$) and its coefficient of determination (r^2_{CV}) were estimated. In addition to those values, residual predictive deviation of cross validation (RPD_{CV}) was calculated. This parameter was calculated as the ratio between the standard deviation of the cross validation set and the $RMSE_{CV}$ values (Porep et al., 2015).

Secondly, the best performing calibration models, defined with the highest RPD_{CV} and the lowest $RMSE_{CV}$, were externally validated. Sample set (n = 33) was split for calibration and validation subsets, 75% and 25%, respectively. The Ms-Excel random function was employed to randomize the set. The 2 subsets were similar in mean and standard deviation. This calibration subset (n = 25) was employed to generate the prediction equations and the validation subset (n = 8) was used to test them. For cholesterol external validation the sample set (n = 27) was divided in 20 and 7, respectively. These external validation models were evaluated also in terms of coefficient of determination of external validation (r^2_{EV}), root mean square error of external validation ($RMSE_{EV}$) and RPD_{EV} (calculated dividing the standard deviation of the reference values at the validation subset between the $RMSE_{EV}$). In addition, slopes and mean differences of the external validation models were calculated to analyze the linearity and bias of the estimation (Ma, Babu, & Amamcharla, 2019; Manuelian, Currò, Penasa, Cassandro, & De Marchi, 2017).

3. Results

3.1. Composition results

Initial samples contained an average dry matter of 21.14 ± 0.77%, 21.26 ± 0.02% and 11.82 ± 0.56% for whole egg plasma, egg yolk plasma and egg white, respectively. In contrast, whole egg plasma and egg white showed similar protein concentration with 86.58 ± 2.95 mg/ml and 85.10 ± 2.36 mg/ml, respectively. While egg yolk plasma showed 46.82 ± 0.89 mg/ml of proteins. The amount of cholesterol at whole egg and egg yolk plasmas were 3.77 ± 0.21 mg/g and 6.16 ± 0.16 mg/g, respectively. The average composition values of the fractions for each membrane filtration run are shown in Fig. 1. Differences were mainly due to the initial feed and the membrane conditions employed.

The average dry matter and protein compositions at run 8 (egg yok plasma feed) were the lowest and merely null. Therefore, permeates from ultrafiltration (750 kDa) contained mainly egg white proteins. Cholesterol did not appear at filtrates till pore size was 0.2 µm and mPES was the membrane material. The other two membrane materials studied (ME and PES) barely filtrated egg compounds, hence their retentates were rich in dry matter, proteins and cholesterol. In contrast, filtrates from run 6, 7, 9 10 and 11 showed the highest protein and dry matter composition. Compositions of retentate and filtrate from run 9 were similar. It seems that the feed crossed the membrane freely to achieve an equilibrium concentration at both side of the membrane. This phenomenon was not detected when whole egg plasma or egg white (run 7

and 11, respectively) were employed with that membrane of 0.65 μm. In addition, filtrate from run 6 showed higher concentrations despite of the lower pore size (0.2 μm). The random formation of aggregates between egg yolk and egg white components could explain these results. Differences between run 1 and 2 were mainly in the composition of the retentates. Like if egg samples employed at run 2 contained more egg yolk; then more cholesterol, protein and dry matter were retained.

3.2. Spectra analysis

Firstly, PCA with raw data was built as data exploratory. It detected that the first component explained 92% of the variance of the spectra. Second and third components contained 6% and 1% of the variability, respectively. No outlier was detected. PCA score plot with first and second components is shown in Fig. 2A. Hotelling's T^2 ellipse with 0.5% significance level was included. Samples close to the ellipse are egg yolk retentates from microfiltration (run 9). In Fig. 2B, one of those samples was outside the limits for Q-residuals. A high value in Q and not in T^2 may be caused by a variability in the recorded signal that is not considered in the model (Porep et al., 2015). As these retentates included mostly particles with a size higher than 0.65 μm, it can be assumed that the variability in the signal was due to change in particle-size distribution that produced scattering. This phenomenon is minimized with adequate transformations (Pasquini, 2018). Then, the application of pretreatments was considered over sample removal. Sample elimination may apparently improve the model performance. However, as useful variance is removed, poorer predictions of new samples will be obtained. Real samples will also produce scattering and the aim is to obtain a model able to determine compositions, even under unexpected measurement conditions or with unusual samples (Porep et al., 2015). Hence all samples were maintained.

The 66 UV-VIS-NIR spectra from 33 samples plus water spectrum are shown in Fig. 3. Spectra are already baseline corrected in this Figure. Samples were classified according to the kind of initial sample (Fig. 3A: whole egg plasma, egg yolk plasma or egg white), the kind of process sample (Fig. 3B: initial, filtrate, retentate) and pore size or MWCO of the system (Fig. 3C). The peaks in the NIR are overtones and combinations of the sample functional groups and, for this reason, can be exhibited broadly and weakly (Krepper et al., 2018). Water is a strong absorber in the NIR spectrum region that suppress other constituent bands such as proteins (Reyhan Selin Uysal & Boyaci, 2020). Egg samples studied contained a minimum of 70% water, being around 97–99.5% moisture some filtrates (run 1 to 5 and 8). Therefore, water absorption in the NIR region was predominant as showed the band similarities between egg samples spectra and water spectrum in light blue. All NIR spectra studied comprised wide bands. Specially at long-wavelength NIR region (2500-1300 nm) where transmittance of water spectrum is almost the same to egg samples. All samples showed lack of transmission from 2500 to 2400 nm and 2000–1900 nm. In addition, transmission peaks detected around 2200 nm, 1850 nm, 1690 nm, 1450 nm, 1190 and 970 nm appeared in pure water spectrum and all the samples. In contrast, there were two peaks around 2309 nm and 1730 nm that differed from water spectrum.

Around 1450 nm, the first overtone of water belong that band (Xiaobo, Jiewen, Povey, Holmes, & Hanpin, 2010), while the bands described from 1900 to 2000 nm, around 2200 and 2400–2500 nm could be associated to water combinations bands as well as to lipid and proteins (Reyhan Selin Uysal & Boyaci, 2020; Xiaobo et al., 2010). At the peaks described around 1850 nm and 1690 nm, quantitative differences in transmission were detected between samples. Samples with lower transmissions at these peaks also showed the two peaks not detected at water spectrum at 2309 nm and 1730 nm. These samples were: whole egg plasma and retentates, and egg yolk samples, except for filtrate at 750 kDa. The composition of these samples produced a decrease in transmission around 2309 nm, 1850 nm, 1730 nm and 1690 nm. At these wavelengths besides the absorption of water, the absorption of CH, CH_2 and CH_3 is also described and it has been associated to fat content (Krepper et al., 2018; Reyhan Selin Uysal & Boyaci, 2020; Xiaobo et al., 2010). These absorptions were clearly detected in samples with a dry matter over 20% and a probably high lipid content (not egg white). However, for run 9 (egg yolk plasma - 0.65 μm), where dry matter was slightly over 10%, the peak around 1730 nm started to disappear and

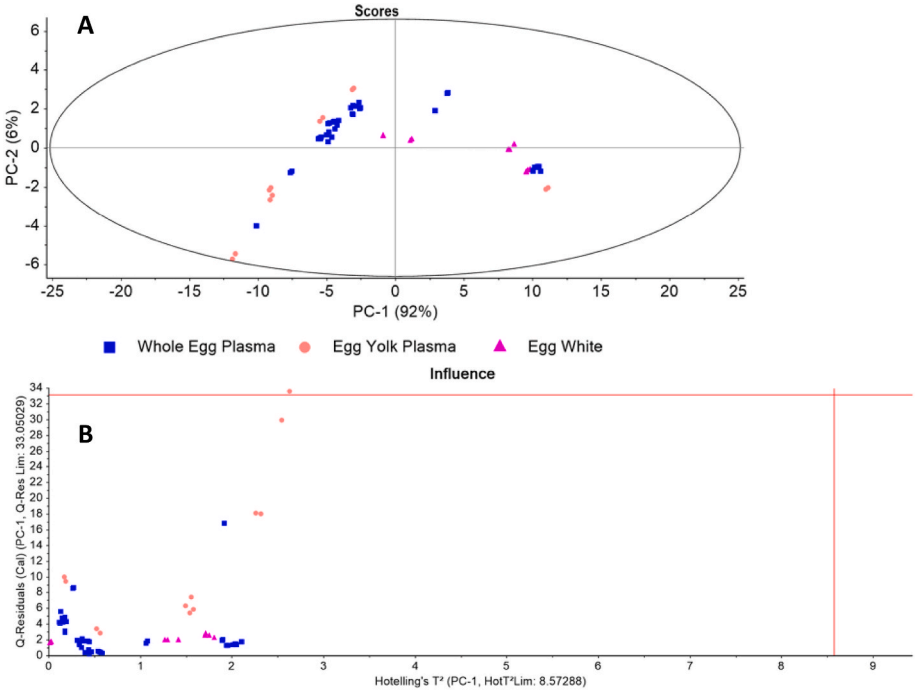


Fig. 2. A) PCA score plot with Hotelling's T^2 ellipse at 0.5% significance level. B). Plot of the Hotelling's T^2 and Q-residuals for principal component 1 (PC-1) with limits at 0.5%. Samples classified according to initial sample filtrated: blue square for whole egg plasma; light orange circle for egg yolk plasma and pink triangle for egg white. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

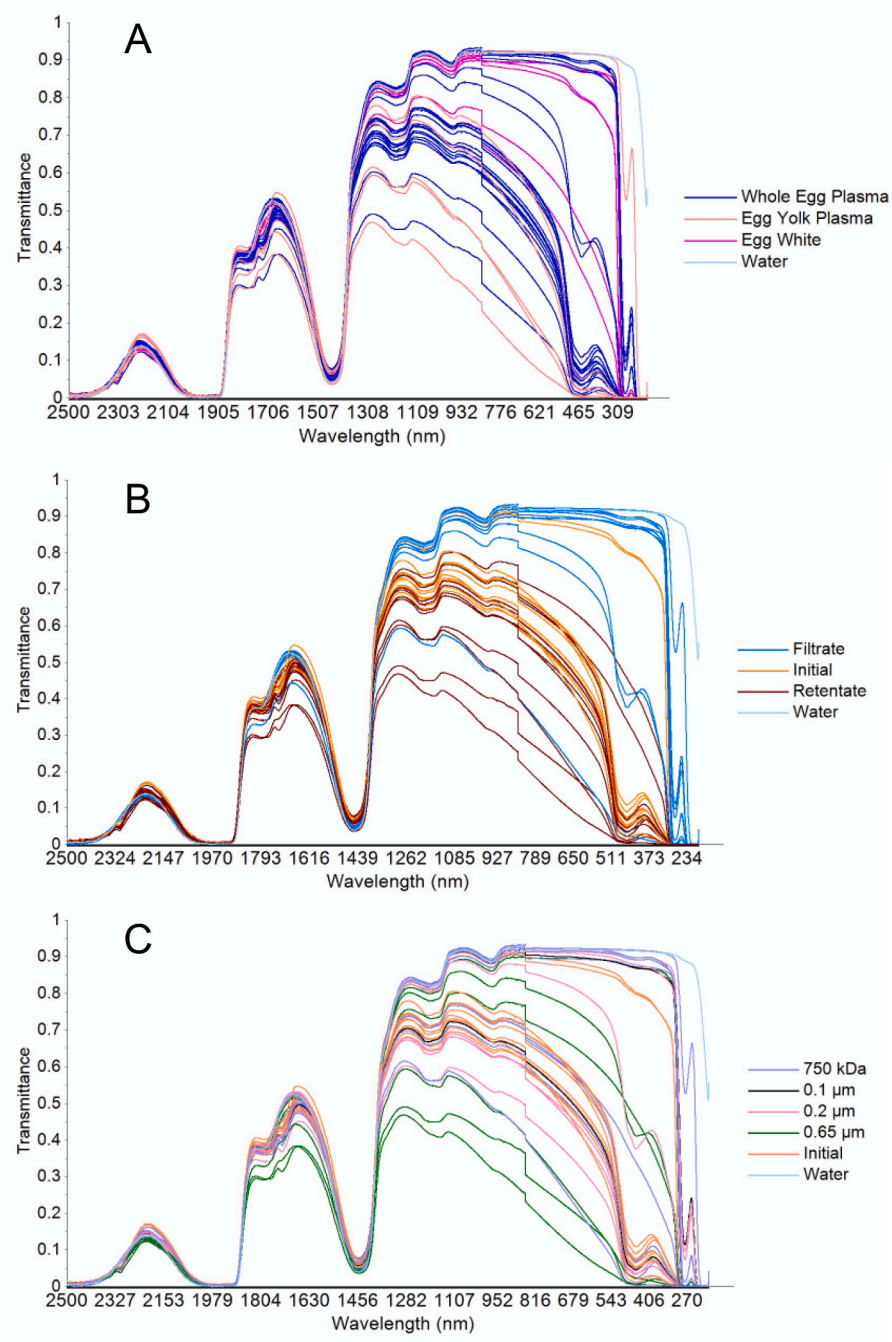


Fig. 3. Transmittance UV-VIS-NIR spectra with baseline correction. A) spectra classified by the kind of initial sample: whole egg plasma, egg yolk plasma or egg white. B) spectra according to filtration process: initial sample, retentate or filtrate. C) spectra classified according to different pore size. All figures include water spectrum in light blue. Initial samples in figures B and C are in orange. Separations conditions are defined in Fig. 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

was detected as a valley from 1737 to 1726 nm. Hence, the lack of this change in transmission at filtrates cannot be associated with the absence of fat but with the high moisture content (Fig. 1).

Along the short-wavelength NIR region (1300 - 780 nm), transmissions varied from 0.9 to 0.25. Besides the bands already related to water absorption at 1190 nm and 970 nm, spectra varied significantly in intensity between samples. Absorption of lipids, fatty acids and carbohydrates can be associated, as reported by other researchers. Because at this region the absorption is also related to CH, CH_2 and CH_3 from the saturated fatty acids and carbohydrates. As well as the absorption of cis double bonds from unsaturated fatty acids (oleic acid) (Hoffman, Ni, Dayananda, Ghafar, & Cozzolino, 2022). Fig. 3 showed these associations with higher transmittances for filtrates (except from run 9) and egg white products. Transmission never achieved a value of 1 because the cuvette employed stopped slightly the light pathway. The gap detected at 850 nm was due to the grating change. The source change at 340 nm was not so marked. At VIS and UV regions samples were also easily differentiated. The color discrepancies visually detected were quantified at the VIS region. The presence of carotenoids in some whole egg and egg yolk samples justified their decrease in transmission detected from 500 to 380 nm. These compounds absorbed green-blue-purple light, producing samples visibly orangey-yellowy. In Fig. 3B, all filtrates except from those of runs 6, 7 and 9 showed maximum transmittances. Carotenoids and cholesterol appeared at same filtrates: membrane systems with mPES and a minimum pore size of 0.2 μm (Fig. 3C).

Around 300 nm there is an abrupt decrease in transmittance for all the samples except for water spectra. The presence of peptide bonds and conjugated double bonds can be measured at UV region. Most proteins exhibit a distinct ultraviolet light absorption maximum at 280 nm. That was the basis for the protein quantification method employed. Then

these falls in transmissions are proportional to the proteins quantified. From the abrupt decrease till the end all the samples obtained zero transmission, except for filtrates from 750 kDa, 0.1 μm, 0.2 μm but PES and 0.65 μm with egg white (Fig. 3C). At these filtrates some light transmission was detected with an increase around 255-251 nm. Then they did not encompass or included less amounts of those compounds present in their initial samples that absorb at that frequency. For example, nucleic acids, phenylalanine or cysteine have absorption maxima between 260 nm and 250 nm (Simonian & Smith, 2001).

Transformed spectra are shown in Fig. 4. Three scatter correction methods were employed: SNV, MSC and normalization (Rinnan, Berg, & Engelsen, 2009). Normalize spectra were previously baseline corrected, while the other transformations performed themselves this correction (Porep et al., 2015). MSC and normalization transformed the raw spectra keeping the original scale of the variables, while SNV, derivative and detrend altered the scale. All transformations reduced differences at short-wavelength NIR and VIS region compared to Fig. 3A. Whereas differentiation between samples increased at long-wavelength NIR region (2500 - 1300 nm), except for derivative transformation. This transformation remarked with sharp peaks spectra regions already commented at Fig. 3; where differences between spectra were detected. Fig. 4 demonstrated that the transformations employed resolved overlapped and hidden bands in NIR spectra (Porep et al., 2015).

3.3. Calibration and cross validation

PLS and PCR models were built to quantify dry matter, proteins and cholesterol of the initial samples (plasmas and egg white), retentates and filtrates. Cholesterol models were obtained only for whole egg and yolk samples. Then, the number of UV-VIS-NIR spectra used for cholesterol determination was 54. PLS and PCR best models based on determination coefficients and errors are reported for each parameter on Table 1. The number of latent factors (LF) for PLS and principal components (PC) for PCR was chosen as the minimum number allowing to minimize the cross validation error and maximize the correlation coefficient (Grassi et al., 2021; Krepper et al., 2018). To understand the results, it is necessary to consider the described differences detected at spectral analysis and the wide range of compositions analysed. Average for dry matter was 15.22 ± 9.44%, proteins 60.67 ± 34.21 mg/ml and cholesterol 3.46 ± 2.38 mg/g. Calibration and cross validation statistics were better for dry matter and cholesterol when pretreatments were performed on the spectral data. All models obtained cross validation correlation coefficients over 0.99. $RMSE_C$ and $RMSE_{CV}$ were below 0.4% for dry matter and below 0.2 mg/g for cholesterol content. Protein content obtained $RMSE_{CV}$ values below 3 mg/ml. Every composition parameter studied can be predicted better with PLS as showed the higher RPD_{CV} values.

3.4. External validation

Then for external validation analysis only PLS was studied. New models with the calibration subset were explored. To understand the optimum number of LF employed for this external validation, the evolution graph of $RMSE_C$ and $RMSE_{EV}$ over LF is included in Fig. 5. The aim was to reduce the number of LF to the minimum because the maximum number of samples included for external validation was 16 (8 x 2). Therefore 7 LF were studied in the new calibration models. However, this number was insufficient for protein PLS regression and a maximum of 10 LF were tested. $RMSE_C$ results are shown with circles. RMSE values for dry matter and proteins are ascribed to the left axis. While RMSE results from cholesterol models are ascribed to the right axis. Due to the lower values for cholesterol (between 2 and 0 mg/g) compared to those of proteins (36–2 mg/ml) and dry matter (7 to almost 0%). The new calibration models suggested 5 LF for dry matter and cholesterol. As Fig. 5 shows, the difference between LF 5 and 6 for $RMSE_C$ was 0.29% and 0.03 mg/g for dry matter and cholesterol, respectively. While protein calibration model suggested 8 LF. In this case, the difference between LF 8 and 9 was 0.76 mg/ml. $RMSE_{EV}$ results are shown with

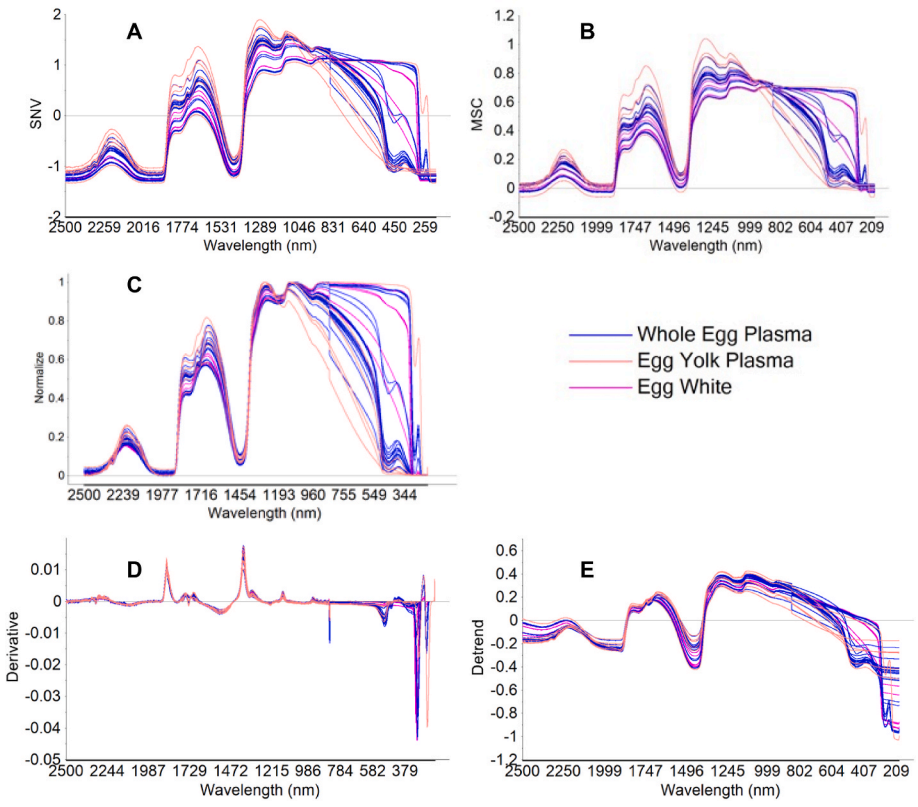


Fig. 4. Transformed spectra after pretreatments. A-standard normal variate (SNV); B- multiplicative scatter correction (MSC); C- normalize; D-first derivative gap segment transformation with gap and segment size of 3 and 2, respectively; E-detrend. Samples classified according to initial sample filtrated: blue whole egg plasma; light orange for egg yolk plasma and pink for egg white. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Calibration and cross-validation statistics for the determination of dry matter (n = 33), proteins (n = 33) and cholesterol (n = 27) by transmittance UV-VIS-NIR spectroscopy.

Parameter	Regression model	Pretreatment	Number of LF/PC	Calibration		Cross validation		
				$RMSE_C$	r^2_C	$RMSE_{CV}$	r^2_{CV}	RPD_{CV}
Dry matter (%)	PLS	SNV	11	0.212	1.000	0.290	0.999	32.50
	PCR	Normalize	15	0.281	0.999	0.355	0.999	26.58
Proteins (mg/ml)	PLS	none	11	1.946	0.997	2.501	0.995	13.68
	PCR	MSC	16	2.253	0.996	2.988	0.992	11.45
Cholesterol (mg/g)	PLS	MSC	10	0.07	0.999	0.09	0.999	27.12
	PCR	SNV	10	0.13	0.997	0.16	0.996	14.90

LF: latent factors; PC: principal components; $RMSE_C$: root mean square error of calibration; r^2_C: coefficient of determination in calibration; $RMSE_{CV}$: root mean square error of cross validation; r^2_{CV}: coefficient of determination in cross validation; RPD_{CV}: residual predictive deviation of cross validation (SD/$RMSE_{CV}$); SNV: standard normal variate; MSC: multiplicative scatter correction; PLS: parcial least squares regression; PCR: principal component regression.

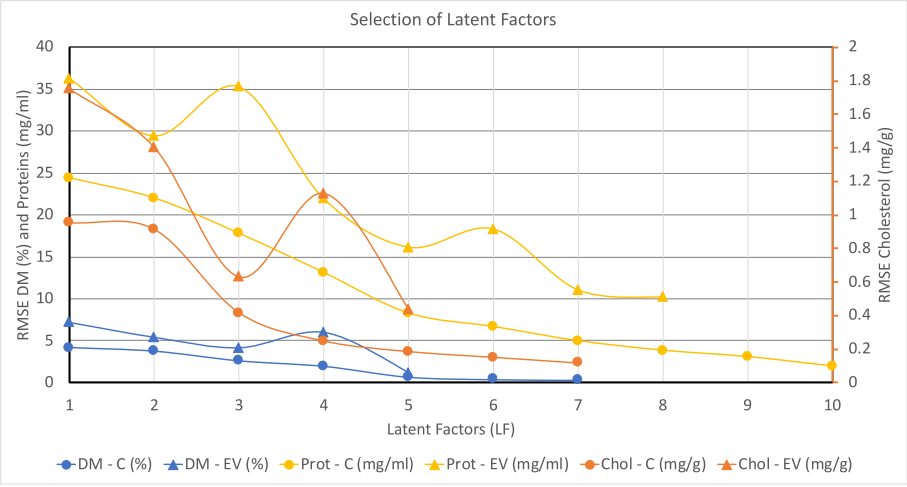


Fig. 5. Selection of Latent Factors (LF). The plot represents LF and RMSE (root mean square error) values for calibration (C) with circles and external validation (EV) with triangles. RMSE values from dry matter (DM) and protein models are ascribed to left axis, while RMSE values from cholesterol models are ascribed to right axis. DM is shown in blue, proteins in yellow and cholesterol in orange. RMSE units are in brackets. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

triangles. $RMSE_{EV}$ for dry matter and cholesterol revealed peaks at LF 4. These peaks remarked the need of 5 LF. For protein EV model, $RMSE_{EV}$ showed two increases: at LF 3 and 6. The difference between LF 7 and 8 was only 0.86 mg/ml. Hence, even 7 LF could have been employed, 8 LF were chosen to obtain a slightly lower $RMSE_{EV}$.

Table 2 shows calibration and external validation results with the number of LF suggested by the calibration model. For dry matter content, the $RMSE_{EV}$ was 1.14% and RPD_{EV} 8.81. The low value for the mean difference or bias (0.02%) agreed with the slope of 0.998. Same agreement was detected at cholesterol content where the bias was −0.002 mg/g and the slope was also around 1. The $RMSE_{EV}$ at cholesterol determination was 0.44 mg/g. And this value related to the standard deviation of the validation subset supposed a RPD_{EV} of 5.77. The model for protein content yielded a good prediction with RPD_{EV} of 3.55 with a difference of 1.21 mg/ml, although the $RMSE_{EV}$ was 10.19 mg/g.

4. Discussion

Spectroscopy measurements are hardly affected by temperature. Temperature varies electromagnetic radiation (Plank's law), solubility (Van't Hoff equation) and H bonds measured at NIR (Walsh et al., 2020), among others. Therefore, spectra must be obtained at a given temperature. Nevertheless, in a manufacturing process, this should not be a limitation as a narrow range of temperature is normally enforced to avoid pathogen growth. However, the effect of temperature on particle aggregation should be considered because the particle size is strongly connected to scattering (Solheim et al., 2022). Some authors detected higher absorbance for egg samples about 4 °C than at greater temperatures. The formation of larger lipoprotein aggregates was suggested as one of the causes of light scattering (Yao, Zhou, Wang, Liu, & Yu, 2014).

Herein, scattering was noticed at PCA analysis. Because samples with higher concentration of particles sized over 0.65 μm (retentate from run 9) were detected around Q-residuals limits. The scatter-correction methods applied (SNV, MSC and normalization) obtained the best results (Table 1). For dry matter and cholesterol determination the scatter-correction was mandatory, while protein quantification obtained a suitable model without this transformation. Dry matter measurements are related to NIR interaction with water from samples, therefore,

Table 2
Results from external validation of partial least squares regression models for the determination of dry matter, protein and cholesterol by transmittance UV-VIS-NIR spectroscopy.

Parameter	Pretreatment	Number of LF	$RMSE_C$	r^2_C	$RMSE_{EV}$	r^2_{EV}	RPD_{EV}	Slope	Bias
Dry matter (%)	SNV	5	0.69	0.99	1.14	0.986	8.81	0.998	0.02
Protein (mg/ml)	None	8	3.83	0.99	10.19	0.916	3.55	0.86	1.21
Cholesterol (mg/g)	MSC	5	0.19	0.99	0.44	0.968	5.77	1.10	−0.002

LF: latent factors; $RMSE_C$: root mean square error of calibration; r^2_C: coefficient of determination in calibration; $RMSE_{EV}$: root mean square error of external validation; r^2_{EV}: coefficient of determination in external validation; RPD_{EV}: residual predictive deviation of external validation (SD/$RMSE_{EV}$); SNV: standard normal variate; MSC: multiplicative scatter correction.

uncovering overlapped and hidden bands in this region with transformations was necessary. Cholesterol quantification depends on the main particles that produced scattering: low density proteins (LDL). As cholesterol was only detected at filtrates from membranes with a minimum pore size of 0.2 μm, it was always constituting these particles; together with proteins, phospholipids and triglycerides (Anton, 2013). In contrast, the development of protein quantification models with PLS and only mean-centered spectra transformation (default settings) achieved better results than those with scatter-corrected spectra. The reason is that the reference method employed the absorbance measurement at 280 and 260 nm (Puertas & Vázquez, 2021a). Then, sample interaction with UV-VIS-NIR radiation was compared to two wavelengths. Therefore, raw spectra resembled reference technique more. If another reference method were employed, the transformation would be required.

The wide range of compositions under study was reflected at spectra. Spectral analysis evidenced that samples were clearly differentiated with UV-VIS-NIR spectroscopy, specially from 1300 nm to 200 nm. Due to the high moisture of egg products, water dominated NIR region and prevented from associating changes in transmissions to certain compounds. However, thanks to multivariate regression models, all spectrum was used to predict the studied composition parameters. The models achieved satisfactory fits. At cross validation (Table 1), RPD_{CV} values above 8 for all the composition parameters indicated that these models could be used for any application. Below this value the model is suitable for screening (3.1–4.9), quality control (5–6.4) or process control (6.5–8) (Chitra, Ghosh, & Mishra, 2017). When sample set was divided for external validation, only RPD_{EV} value of dry matter was above 8. However, a minimum RPD value of 3 is recommended for a predictive model (Chitra et al., 2017). In addition, validation results for cholesterol and protein could be considered excellent with a r^2_{EV} greater than 0.91 (Manuelian et al., 2017).

It is known that an insufficient number of samples in the calibration set is detected by a $RMSE_{CV}$ or $RMSE_{EV}$ value substantially higher than the $RMSE_C$ value (Walsh et al., 2020). Herein, $RMSE_{CV}$ values were between 25.6 and 36.7% higher than $RMSE_C$ values (Table 1). These values are like those found in literature (Puertas & Vázquez, 2019; R.S. Uysal, Mentes Yilmaz, & Boyaci, 2019). Nevertheless, as this cross-validation was done using within-population sets, it only provides an indication of the model performance. A validation with independent test sets is needed for practical implementation (Walsh et al., 2020). Therefore, an external validation was performed. $RMSE_{EV}$ resulted 64%, 166% and 134% higher than $RMSE_C$ for dry matter, proteins and cholesterol, respectively. Although other studies described up to above 200% (R.S. Uysal et al., 2019), these ratios revealed an insufficient number of samples (20 and 25) for the calibration performed before the external validation. Consequently, before a practical implementation of these models, the calibration set must be increased.

Although only 33 samples were tested, they encompassed the variability of 89 eggs. Moreover, the 2311 variables from UV-VIS-NIR spectra can justify the development of the models. Thenceforth PLS models could be useful for dry matter, protein and cholesterol determinations in the egg products studied. The reason is the wide range of compositions under study. Because there was high variability between the kind of samples (whole egg, egg yolk or egg white) and the product of the membrane process (initial feed, filtrates and retentates). Although this variability is more frequent in a research project, in a membrane separation technique could be expectable a wide range of compositions. In addition, if the membrane is employed to remove water, NIR spectroscopy has demonstrated to be sensitive to small differences in water content.

Therefore, UV-VIS-NIR spectroscopy is a promising technique for composition determinations in a membrane separation process. There are other studies that analysed lipids and total soluble solids in liquid egg products (R.S. Uysal et al., 2019) or minerals in cheeses (Manuelian et al., 2017) using spectroscopy and chemometrics. Subsequently, further research could be guided to determine other egg components simultaneously with UV-VIS-NIR spectroscopy. The real-time composition measurement at the membrane process is a potential application of this technique. Moreover, it will save time, chemicals and egg products. For example, in an egg cholesterol measurement, as saponification step will be avoided, 3 h for saponification and chemicals will be saved. Subsequently, it will be more economical and environmentally sustainable than conventional quantification methods. This technique will ease filtrations in batch configuration that need to be stopped at a specific concentration. On the other side, the implementation of UV-VIS-NIR spectroscopy for on-line composition determinations in a continuous membrane filtration could increase process efficiency. However, before its practical implementation, an exhaustive calibration must be performed to create a robust model; with reference techniques and an independent test set. Other factors like seasons, storage conditions, locations, or production conditions should be taken into consideration if it is necessary given the nature of the samples being investigated (Walsh et al., 2020).

The application of UV-VIS-NIR spectroscopy to membrane separation process demonstrated a powerful symbiosis. Both techniques are low energy consumption and therefore green technologies. The employment of green technologies like filtration and UV-VIS-NIR spectroscopy suppose that the method describes herein ties in with the European Green Deal.

5. Conclusion

In this study, an egg membrane separation process was analysed through UV-VIS-NIR spectroscopy. The composition of the fractions was measured by reference methods and compared to UV-VIS-NIR spectroscopy combined with chemometric tools. The spectroscopy method had demonstrated to be an efficient and promising green technology for monitoring a membrane separation process. It could be applied with no sample preparation and without destruction of the sample. Moreover, it could be a low cost, environmentally friendly and rapid technique for on-line applications. It can measure real-time composition of the membrane fractions, giving the possibility of stopping the separation process at a desirable concentration factor.

Funding

This work was supported by the Spanish National Plan for Scientific and Technical Research and Innovation. A University Professor Education grant (FPU 16/05128) by the Spanish Ministry of Education, Culture and Sport to author Gema Puertas is gratefully acknowledged.

CRediT authorship contribution statement

Gema Puertas: Investigation, Writing – original draft, Visualization. **Patricia Cazón:** Validation, Writing – review & editing. **Manuel Vázquez:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Allègre, C., Moulin, P., Gleize, B., Pieroni, G., & Charbit, F. (2006). Cholesterol removal by nanofiltration: Applications in nutraceutics and nutritional supplements. *Journal of Membrane Science, 269*(1–2), 109–117. https://doi.org/10.1016/j.memsci.2005.06.025

Anton, M. (2013). Egg yolk: Structures, functionalities and processes. *Journal of the Science of Food and Agriculture, 93*(12), 2871–2880. https://doi.org/10.1002/jsfa.6247

Bwambok, D. K., Siraj, N., Macchi, S., Larm, N. E., Baker, G. A., Pérez, R. L., et al. (2020). QCM sensor arrays, electroanalytical techniques and NIR spectroscopy coupled to multivariate analysis for quality assessment of food products, raw materials, ingredients and foodborne pathogen detection: Challenges and breakthroughs. *Sensors, 20*(23), 1–42. https://doi.org/10.3390/s20236982

Chay Pak Ting, B. P., Mine, Y., Juneja, L. R., Okubo, T., Gauthier, S. F., & Pouliot, Y. (2011). Comparative composition and antioxidant activity of peptide fractions obtained by ultrafiltration of egg yolk protein enzymatic hydrolysates. *Membranes, 1*(3), 149–161. https://doi.org/10.3390/membranes1030149

Chitra, J., Ghosh, M., & Mishra, H. N. (2017). Rapid quantification of cholesterol in dairy powders using Fourier transform near infrared spectroscopy and chemometrics. *Food Control, 78*, 342–349. https://doi.org/10.1016/j.foodcont.2016.10.008

Dhineshkumar, V., & Ramasamy, D. (2017). Review on membrane technology applications in food and dairy processing. *Journal of Applied Biotechnology & Bioengineering, 3*(5), 399–407. https://doi.org/10.15406/jabb.2017.03.00077

Franzoi, M., Manuelian, C. L., Rovigatti, L., Donati, E., & De Marchi, M. (2018). Development of Fourier-transformed mid-infrared spectroscopy prediction models for major constituents of fractions of delactosated, defatted milk obtained through ultra- and nanofiltration. *Journal of Dairy Science, 101*(8), 6835–6841. https://doi.org/10.3168/jds.2017-14343

Grassi, S., Jolayemi, O. S., Giovenzana, V., Tugnolo, A., Squeo, G., Conte, P., et al. (2021). Near infrared spectroscopy as a green technology for the quality prediction of intact olives. *Foods, 10*(5), 1–12. https://doi.org/10.3390/foods10051042

Hoffman, L. C., Ni, D., Dayananda, B., Ghafar, N. A., & Cozzolino, D. (2022). Unscrambling the provenance of eggs by combining chemometrics and near-infrared reflectance spectroscopy. *Sensors, 22*(4988), 1–8.

Karoui, R., De Ketelaere, B., Kemps, B., Bamelis, F., Mertens, K., & De Baerdemaeker, J. (2009). Eggs and egg products. Infrared spectroscopy for food quality analysis and control. https://doi.org/10.1016/B978-0-12-374136-3.00015-8.

Krepper, G., Romeo, F., Fernandes, D. D. de S., Diniz, P. H. G. D., de Araújo, M. C. U., Di Nezio, M. S., et al. (2018). Determination of fat content in chicken hamburgers using NIR spectroscopy and the Successive Projections Algorithm for interval selection in PLS regression (iSPA-PLS). *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 189*, 300–306. https://doi.org/10.1016/j.saa.2017.08.046

Li, Z., Huang, X., Tang, Q., Ma, M., Jin, Y., & Sheng, L. (2022). Functional properties and extraction techniques of chicken egg white proteins. *Foods, 11*(2434). https://doi.org/10.3390/foods11162434

Ma, Y. B., Babu, K. S., & Amamcharla, J. K. (2019). Prediction of total protein and intact casein in cheddar cheese using a low-cost handheld short-wave near-infrared spectrometer. *LWT - Food Science and Technology, 109*(December 2018), 319–326. https://doi.org/10.1016/j.lwt.2019.04.039

Manuelian, C. L., Currò, S., Penasa, M., Cassandro, M., & De Marchi, M. (2017). Prediction of minerals, fatty acid composition and cholesterol content of commercial cheeses by near infrared transmittance spectroscopy. *International Dairy Journal, 71*, 107–113. https://doi.org/10.1016/j.idairyj.2017.03.011

Osborne, B. G., & Barret, G. M. (1984). Compositional analysis of liquid egg products using infrared transmission spectroscopy. *International Journal of Food Science and Technology, 19*(3), 349–353. https://doi.org/10.1111/j.1365-2621.1984.tb00358.x

Pasquini, C. (2018). Analytica chimica acta near infrared spectroscopy : A mature analytical technique with new perspectives e A review. *Analytica Chimica Acta, 1026*, 8–36. https://doi.org/10.1016/j.aca.2018.04.004

Porep, J. U., Kammerer, D. R., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science & Technology, 46*(2), 211–230. https://doi.org/10.1016/j.tifs.2015.10.002

Primacella, M., Wang, T., & Acevedo, N. C. (2018). Use of reconstituted yolk systems to study the gelation mechanism of frozen-thawed hen egg yolk. *Journal of Agricultural and Food Chemistry, 66*(2), 512–520. https://doi.org/10.1021/acs.jafc.7b04370

Puertas, G., Cazón, P., & Vázquez, M. (2023). A quick method for fraud detection in egg labels based on egg centrifugation plasma. *Food Chemistry, 402*(October 2022), 1–7. https://doi.org/10.1016/j.foodchem.2022.134507

Puertas, G., & Vázquez, M. (2019). Cholesterol determination in egg yolk by UV-VIS-NIR spectroscopy. *Food Control, 100*, 262–268. https://doi.org/10.1016/j.foodcont.2019.01.031

Puertas, G., & Vázquez, M. (2021a). Evaluation of the composition and functional properties of whole egg plasma obtained by centrifugation. *International Journal of Food Science and Technology, 56*(10), 5268–5276. https://doi.org/10.1111/ijfs.15124

Puertas, G., & Vázquez, M. (2021b). Liquid whole egg fractionation: Effect of centrifugation on physicochemical attributes of quality. *Journal of Food Processing and Preservation, 45*(e15334). https://doi.org/10.1111/jfpp.15334

Réhault-Godbert, S., Guyot, N., & Nys, Y. (2019). The golden egg: Nutritional value, bioactivities, and emerging benefits for human health. *Nutrients, 11*(3), 1–26. https://doi.org/10.3390/nu11030684

Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC, Trends in Analytical Chemistry, 28*(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Simonian, M. H., & Smith, J. A. (2001). Spectrophotometric and colorimetric determination of protein concentration. *Current Protocols in Toxicology*, (Supplement 10)https://doi.org/10.1002/0471142727.mb1001as76. A.3G.1-A.3G.9.

Solheim, J. H., Zimmermann, B., Tafintseva, V., Dzurendová, S., Shapaval, V., & Kohler, A. (2022). The use of constituent spectra and weighting in extended multiplicative signal correction in infrared spectroscopy. *Molecules, 27*(6). https://doi.org/10.3390/molecules27061900

Solís-Oba, M., Teniza-García, O., Rojas-López, M., Delgado-Macuil, R., Díaz-Reyes, J., & Ruiz, R. (2011). Application of infrared spectroscopy to the monitoring of lactose and protein from whey after ultra and nano filtration process. *Journal of the Mexican Chemical Society, 55*(3), 190–193.

Staszak, K., & Wieszczycka, K. (2022). Membrane applications in the food industry. *Physical Sciences Reviews, 1–31*. https://doi.org/10.1515/psr-2021-0050

Uysal, R. S., Mentes Yilmaz, O., & Boyaci, I. H. (2019). Determination of liquid egg composition using attenuated total reflectance Fourier transform infrared spectroscopy and chemometrics. *Journal of the Science of Food and Agriculture, 99*(7), 3572–3577. https://doi.org/10.1002/jsfa.9578

Uysal, Selin, R., & Boyaci, I. H. (2020). Authentication of liquid egg composition using ATR-FTIR and NIR spectroscopy in combination with PCA. *Journal of the Science of Food and Agriculture, 100*(2), 855–862. https://doi.org/10.1002/jsfa.10097

Walsh, K. B., Blasco, J., Zude-Sasse, M., & Sun, X. (2020). Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. *Postharvest Biology and Technology, 168*(May), Article 111246. https://doi.org/10.1016/j.postharvbio.2020.111246

Xiaobo, Z., Jiewen, Z., Povey, M. J. W., Holmes, M., & Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta, 667*(1–2), 14–32. https://doi.org/10.1016/j.aca.2010.03.048

Yao, L., Zhou, W., Wang, T., Liu, M., & Yu, C. (2014). Quantification of egg yolk contamination in egg white using UV/Vis spectroscopy: Prediction model development and analysis. *Food Control, 43*, 88–97. https://doi.org/10.1016/j.foodcont.2014.02.037