



# An exploratory study for the technological classification of egg white powders based on infrared spectroscopy

Silvia Grassi<sup>a</sup>, Raffaele Vitale<sup>b, c</sup>, Cristina Alamprese<sup>a, \*</sup>

<sup>a</sup> Department of Food, Environmental, and Nutritional Sciences (DeFENS), University of Milan, Via G. Celoria 2, 20133 Milan, Italy

<sup>b</sup> Molecular Imaging and Photonics Unit, Department of Chemistry, Katholieke Universiteit Leuven, Celestijnenlaan 200F, B-3001 Leuven, Belgium

<sup>c</sup> Laboratoire de Spectrochimie Infrarouge et Raman – UMR 8516, Université de Lille – Sciences et Technologies, Bâtiment C5, 59655 Villeneuve d'Ascq, France

## ARTICLE INFO

### Keywords:

Gelling properties  
Foaming properties  
Duplex algorithm  
PLS-DA  
Variable selection

## ABSTRACT

This work aims at the evaluation of FT-NIR and FT-IR spectroscopy as rapid, easy, and cost-effective tools for the classification of egg white powder (EWP) based on its technological properties. Up to 100 commercial spray-dried EWP samples with known gelling and foaming properties were used to acquire FT-NIR and FT-IR spectra. An appropriate data-splitting algorithm (Duplex) was applied in order to create, for each dataset, a calibration set and a representative validation test set for prediction. Different spectral pre-treatments and their combinations were investigated for the calculation of Partial Least Squares–Discriminant Analysis models in order to classify samples according to gel strength, foam height, and foam instability. A variable selection strategy based on the so-called Variable Importance in Projection scores was also evaluated. Both FT-NIR and FT-IR spectroscopy showed good potential in discriminating EWP samples with different technological properties. Correct classification percentages in prediction ranging from 59% to 89% were obtained with the best models calculated with selected wavenumbers. These results show a promising industrial perspective, demonstrating the possibility of developing cheap and fast instruments spanning a limited spectral range, which can be implemented on the production lines for EWP sorting and quality control.

## 1. Introduction

The excellent foaming and gelling properties of egg white powder (EWP) make it an optimal ingredient for many processed foods, such as bakery products, meringues, pasta, and meat products. Different kinds of proteins, representing more than 80g/100g of the dry matter, are responsible for the technological performance of albumen (Li-Chan, Powrie, & Nakai, 1995). However, during the production of EWP, proteins are subjected to several processing steps with thermal, physical, interfacial, and chemical treatments that may jeopardize technological properties (Mine, 1995). Besides the raw material characteristics, each stage of EWP manufacturing can influence the protein quality (Lechevalier, Jeantet, Arhaliass, Legrand, & Nau, 2007; Talansier et al., 2009). In particular, raw materials greatly influence egg white protein conformation and gelling properties, while the most critical process step is spray-drying, which strongly damages the foaming properties.

Pumping and filtering have also considerable negative effects, whereas concentration and desugarization exert interesting improving effects on EWP foaming properties (Lechevalier et al., 2007). In order to offset functional property losses resulting from the spray-drying step, heating of EWP at 75–80 °C for 10–15 days is widely used. It improves both gelling and foaming properties thanks to protein mild conformational changes and polymerization caused by disulphide formation and/or sulfhydryl-disulphide interchange (Handa, Hayashi, Shidara, & Kuroda, 2001; Kato, Ibrahim, Watanabe, Honma, & Kobayashi, 1989; Mine, 1996).

EWPs are often marketed as “high gel” or “high foam” products, in order to help end-users in choosing the right ingredient based on the final application. However, due to the several factors influencing EWP technological properties, this labelling is not always correct, leading to a negative impact on the quality of the final food products. In order to correctly classify EWPs, the producers must analyze every production

*Abbreviation list:* EWP, egg white powder; FT, Fourier transform; IR, infrared; MIR, mid infrared; MSC, multiplicative scatter correction; NIR, near infrared; PLS, partial least squares regression; PLS-DA, partial least squares-discriminant analysis; VIP, variable importance in projection.

\* Corresponding author.

Email addresses: silvia.grassi@unimi.it (S. Grassi); rvitale86@gmail.com (R. Vitale); cristina.alamprese@unimi.it (C. Alamprese)

<https://doi.org/10.1016/j.lwt.2018.05.065>

Received 14 February 2018; Received in revised form 21 May 2018; Accepted 27 May 2018

Available online xxx

0023-6438/© 2018.

lot by means of traditional methods, with a large waste of time and resources. The use of infrared (IR) spectroscopy, both in the near (NIR) and mid (MIR) range, can totally change this scenario because simple, rapid, and cheap classification methods can be developed by analyzing the powder samples without re-dissolving them. The impact of the manual ability of the analyst becomes lower and, in the case of NIR spectroscopy, in-line applications can be implemented directly in the production chain, for instance by setting a NIR probe just before the EWP packaging step.

The application of IR spectroscopy to food classification has been studied for different matrices, mainly with the aim of authentication and fraud fighting (Danezis, Tsagkaris, Camin, Brusica, & Georgiou, 2016; Esslinger, Riedl, & Fauhl-Hassek, 2014; Lohumi, Lee, Lee, & Cho, 2015). In the field of egg products, some papers explore the possibility of evaluating egg quality and freshness, as well as egg product composition (Dalle Zotte, Berzaghi, Jansson, & Andrighetto, 2006; Lin, Zhao, Sun, Chen, & Zhou, 2011; Zhao et al., 2010). In other works, IR spectroscopy is used to study the denaturing effects of different treatments on albumen proteins (Németh et al., 2011; Seregély, Farkas, Tuboly, & Dalmadi, 2006; Uygun-Sarıbay, Ergun, Kalaycı, & Köseoğlu, 2017) or to evaluate the egg amount in fresh pasta (Bevilacqua, Bucci, Materazzi, & Marini, 2013). However, to the best of our knowledge, no studies deal with the use of IR spectroscopy for the prediction of EWP technological properties. Thus, the aim of this work is the assessment of NIR and MIR spectroscopy as rapid, easy, and cost-effective tools for the classification of EWPs based on their technological properties.

## 2. Materials and methods

### 2.1. Egg white powders

A total of 100 EWP samples (200 g each) were kindly provided by Lactosan-Sanovo Ingredients Group (Zeven/Aspe, Germany) together with a measure of their gelling properties, whereas foam height and instability values were provided only for 75 and 74 samples out of the 100, respectively. All samples were commercial spray-dried products belonging to different production lots.

### 2.2. Foaming and gelling properties

Technological performances of EWPs were assessed by Lactosan-Sanovo Ingredients Group (Zeven/Aspe, Germany) following internal standards. Briefly, gelling properties were measured by dissolving EWP in distilled water (11 g/100 mL) and adjusting pH at 6.6–6.7 with sodium hydroxide or hydrochloric acid. The solution was then transferred into synthetic casings (30 mm diameter) and put in a water bath at 90 °C for 30 min. After cooling down at 25 °C, the obtained gels were cut into 30 mm slices and the gel strength (kPa) was measured in triplicate by penetration tests, using a Fudoh rheometer (Rheotech Co., Ltd., Tokyo, Japan) equipped with a 4 mm cylindrical probe.

For the determination of foaming properties, EWP was suspended in distilled water (10 g/100 mL) and, after 15 min resting, it was whipped for 3 min by using a mixer (C-100 Hobart, Troy, OH). The foam was then levelled, and its height (mm) was measured at three different positions. After whipping, 75 g of foam were weighed into a funnel with a perforated plate and the dripping albumen was collected in a graduated cylinder for 70 min. Foam instability was expressed as mL of dripped albumen after 70 min.

### 2.3. Spectra acquisition

For Fourier transform (FT)-NIR analysis, a MPA spectrometer (Bruker Optics, Milan, Italy) equipped with an integrating sphere was used. Two aliquots of 15 g for each sample were homogeneously distributed into Petri dishes and spectra were acquired in the range 12500–3800  $\text{cm}^{-1}$ , with 16  $\text{cm}^{-1}$  resolution, and 128 scans for both samples and background. The same aliquots were used for FT-IR spectra collection, using a Vertex 70 spectrometer (Bruker Optics, Milan, Italy) equipped with a single reflection ATR cell (4000–600  $\text{cm}^{-1}$  spectral range, 16  $\text{cm}^{-1}$  resolution, 64 scans).

### 2.4. Data analysis

Prior to data analysis, an appropriate data splitting by a Euclidean distance-sampling algorithm (Duplex) (Snee, 1977) was performed in order to obtain, for each dataset, a calibration set consisting of two third of the samples and a representative validation set for prediction, containing the remaining samples (Bevilacqua, Bucci, Magri, Magri, & Marini, 2012; Vitale et al., 2013).

Different spectral pre-treatments and their combinations were investigated: de-trending, linear and quadratic baseline correction (Barnes, Dhanoa, & Lister, 1989), multiplicative scatter correction (MSC) (Geladi, MacDougall, & Martens, 1985; Martens, Jensen, & Geladi, 1983), and Savitzky-Golay first and second derivatives (Savitzky & Golay, 1964).

The pre-treated datasets were subjected to Partial Least Squares–Discriminant Analysis (PLS-DA) (Barker & Rayens, 2003; Wold et al., 1983) to classify samples from spectral information according to gel strength, foam height, and foam instability. Two classes were created for each technological property, based on threshold values suggested by Lactosan-Sanovo Ingredients Group (Zeven/Aspe, Germany): 58.8 kPa for gel strength, 160 mm for foam height, and 35 mL for foam instability.

PLS-DA is the direct extension of Partial Least Squares regression (PLS) (Wold, Sjöström, & Eriksson, 2001), developed for classification problem solving: the data matrix,  $\mathbf{X}$ , here containing in each one of its rows the spectrum of one of the analyzed samples, is regressed via PLS on a dummy binary-coded response array, say  $\mathbf{Y}$ , made up by a set of piled  $Z$ -dimensional vectors constructed so that if their corresponding objects are members of the  $z$ -th class, they have a 1-value in their  $z$ -th entry and 0-values in all the other ones. For instance, in a 2-class problem as the one under study, samples belonging to the first category will be described by the dependent vector [1 0], while samples belonging to the second one by the vector [0 1]. Whenever new measurements become available, the *a posteriori* probabilities that the objects they refer to belong to the  $Z$  concerned classes are calculated. The assignation is finally carried out according to the higher-than-a-threshold probability rule, being such a threshold estimated as detailed in Pérez, Ferré, and Boqué (2009).

PLS-DA does not need to assume linearly independent regressors and thus, in theory, the entire  $\mathbf{X}$  matrix can be employed for calibration purposes. However, in the attempt of enhancing the predictive ability and the robustness of the final classification models, the most discriminant spectral wavenumbers of the considered experimental ranges can be preliminarily selected and PLS-DA applied on the resulting reduced data array. For addressing such a selection, a variant of the approach described in Folch-Fortuny, Prats-Montalbán, Cubero, Blasco, and Ferrer (2016) was applied. For each training set, a PLS-DA model was built taking into account all the spectral variables available and their importance for the classification assessed according to the so-called Variable Importance in Projection (VIP) scores (Wold, Johansson, &

Cocchi, 1993), expressing their influence in the model building phase. The VIP scores were compared to those obtained by permuting (shuffling) 300 times the order of the Z-dimensional vectors within the dummy response matrix and re-calibrating the PLS-DA classifier. The idea behind this strategy is to evaluate which VIP scores, among the ones returned by the original model, were systematically higher than those retrieved by artificially assigning the samples to random classes, thus destroying any existing relationship between the nature of their spectral fingerprint and their class belonging. Such VIP scores can be therefore considered associated to those wavenumbers at which the classes under study can be more clearly discriminated. In this case, the global set of spectral channels was sorted by the number of random PLS-DA models resulting in VIP values larger than their corresponding original ones: the lower this number, the higher their discriminant power. At this point, new PLS-DA models were calibrated utilizing a growing number of wavenumbers (from the most to the least discriminant, in intervals of 5) until encompassing a maximum of 50 light channels. The models exhibiting the best classification performance (in terms of classification sensitivity and specificity) were retained and compared. The parameters of all the PLS-DA classifiers were tuned by 11-fold cross-validation (i.e. each training set was split into 11 different random groups of observations, which were iteratively eliminated in the cross-validation loop).

Data analysis was carried out in Matlab environment (v. 2016a, Mathworks, Inc., Natick, MA).

### 3. Results and discussion

#### 3.1. Technological properties

Fig. 1 shows the Box-and-Whisker plots for the technological properties of the EWP samples. Gel strength ranged from 34.9 to 72.8 kPa, with average and standard deviation values of 57.5 and 8.0 kPa. Half of the samples (52 out of 100) had a gel strength included between 52.0 and 63.2 kPa. As for foaming properties, foam height and instability ranged from 92 to 175 mm and from 20 to 53 mL, with average and standard deviation values of  $158 \pm 14$  mm and  $35 \pm 7$  mL, respectively. About 55% of samples showed a 155–165 mm foam height range (43 samples out of 75) and a foam instability included from 29 to 40 mL (42 samples out of 74).

Since there are neither official methods nor standard materials for the evaluation of EWP technological properties, each producer adopts internal procedures and threshold values in order to distinguish the best performing products. Thus, according to Lactosan-Sanovo Ingredients Group's suggestions, EWP samples were classified as follows:

- high gel products: EWP samples with gel strength  $> 58.8$  kPa
- high foam products: EWP samples with foam height  $> 160$  mm
- high foam stability products: EWP samples with foam instability  $< 35$  mL

Based on these threshold values, 44 samples out of 100 were considered as “high gel” products, 34 samples out of 75 as “high foam” products, and 36 samples out of 74 as high foam stability products.

Summarizing, the evaluated technological properties showed limited variation ranges, highlighting a good standardization of the EWP production process.

#### 3.2. FT-NIR and FT-IR spectra

The average FT-NIR spectrum of the EWP samples is shown in Fig. 2a. EWP is mainly composed of proteins (82.4 g/100 g), followed by carbohydrates (4.5 g/100 g), fat (0.04 g/100 g), ash (4.6 g/100 g), and

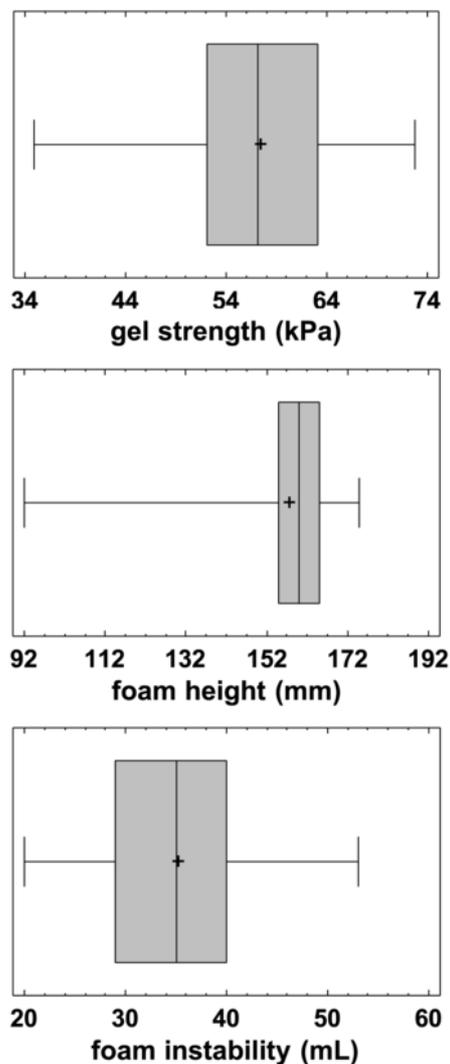
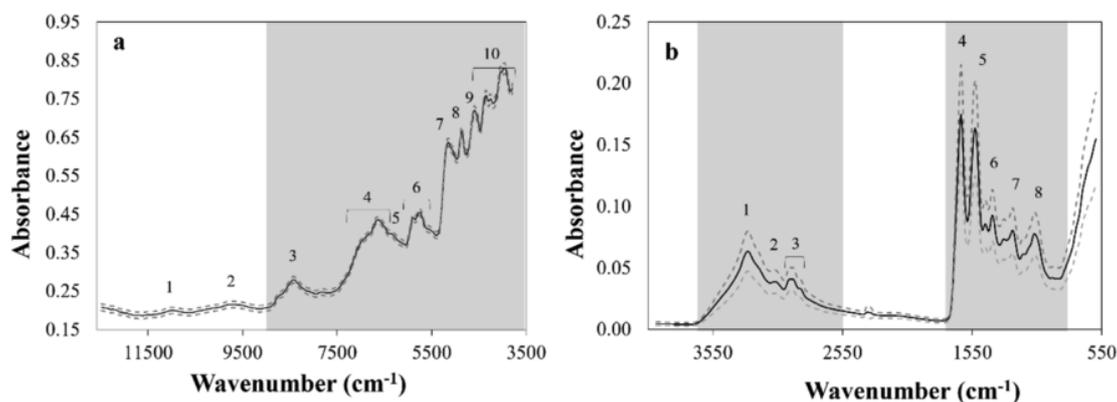


Fig. 1. Box-and-Whisker plots for the technological properties of the egg white powder samples.

a remaining portion of water (about 8.5 g/100 g) (USDA, 2015). The information recoverable from the FT-NIR spectra was essentially linked to protein signals due to amides and various C-H functional groups. The key absorption bands were found at  $6670\text{--}6536\text{ cm}^{-1}$  (N-H stretching first overtone), at around  $4880\text{ cm}^{-1}$  (N-H stretching combinations), and at  $4613\text{--}4587\text{ cm}^{-1}$  (N-H bending second overtone and C=O stretching/N-H in-plane bending/C-N stretching combination bands) (Workman & Weyer, 2008). Other important bands linked to proteins were also visible: the N-H first overtone of primary amides around  $6990\text{ cm}^{-1}$ , the C=O/N-H combination of polyamides at  $6280\text{ cm}^{-1}$ ,  $2 \times$  C-H stretching (symmetric and antisymmetric) between 5700 and  $5850\text{ cm}^{-1}$ , the O-H stretching associated to the OH deformation of water at  $5135\text{ cm}^{-1}$ , and  $2 \times$  stretching C=O + peptide group mode around  $4590\text{ cm}^{-1}$  (Hecht & Wood, 1956). Carbohydrates signals were present with weak bands between  $11000$  and  $9680\text{ cm}^{-1}$ , where the second overtone of saccharide O-H was observable, and between  $4386$  and  $4283\text{ cm}^{-1}$  (combination of C-H stretching and  $\text{CH}_2$  deformation).

Liu, Cho, Sakuri, Miura, and Ozaki (1994) associated the FT-NIR peak at  $4855\text{ cm}^{-1}$  to the signal present at  $3300\text{ cm}^{-1}$  in the FT-IR



**Fig. 2.** Fourier transform-near infrared (FT-NIR; a) and Fourier transform-infrared (FT-IR; b) average (straight lines) and standard deviation (dashed lines) spectra of the egg white powders. In grey the selected spectral ranges used for data analysis. FT-NIR peak identification: 1)  $11000\text{ cm}^{-1}$ , second overtone OH; 2)  $9680\text{ cm}^{-1}$ , second overtone OH; 3)  $8310\text{ cm}^{-1}$ , combination band of water; 4)  $6670\text{--}6536\text{ cm}^{-1}$ , first stretching overtone N-H; 5)  $6280\text{ cm}^{-1}$ , C=O/N-H combination of polyamide; 6)  $5700$  and  $5850\text{ cm}^{-1}$ ,  $2 \times$  C-H stretching; 7)  $5135\text{ cm}^{-1}$ , O-H stretching; 8)  $4880\text{ cm}^{-1}$ , N-H stretching; 9)  $4613\text{--}4587\text{ cm}^{-1}$ , second overtone N-H bending and C=O stretching/N-H in-plane bending/C-N stretching; 10)  $4386$  and  $4283\text{ cm}^{-1}$ , combination of C-H stretching and  $\text{CH}_2$  deformation. FT-IR peak identification: 1)  $3300\text{ cm}^{-1}$ , N-H stretching mode; 2)  $3047\text{ cm}^{-1}$ , O-H stretching; 3)  $2926\text{--}2869\text{ cm}^{-1}$ ,  $\text{CH}_3$  stretching mode; 4)  $1650\text{ cm}^{-1}$ , amide I; 5)  $1550\text{ cm}^{-1}$ , amide II; 6–7)  $1400\text{--}1200\text{ cm}^{-1}$ , amide III; 8)  $1056\text{ cm}^{-1}$ , CO stretching.

spectrum assignable to the N-H stretching mode (Fig. 2b). Moreover, in the FT-IR spectrum, peaks in the  $3050\text{--}2800\text{ cm}^{-1}$  region were ascribable to the stretching modes of O-H ( $3047\text{ cm}^{-1}$ ) and  $\text{CH}_3$  ( $2926$  and  $2869\text{ cm}^{-1}$ ). Signals in the fingerprint region around  $1650\text{ cm}^{-1}$ ,  $1550\text{ cm}^{-1}$  and  $1400\text{--}1200\text{ cm}^{-1}$  were mainly linked to the amide I, II and III vibrations, respectively (Barth, 2007).

Based on such an identification, data analysis was performed considering the FT-NIR spectral range from  $8900$  to  $3800\text{ cm}^{-1}$  and the FT-IR regions between  $3700$  and  $2600\text{ cm}^{-1}$  and  $1700$  and  $860\text{ cm}^{-1}$  (Fig. 2).

### 3.3. Egg white powder classification

The ability of FT-NIR and FT-IR spectroscopy to discriminate EWP samples based on their technological properties was evaluated by PLS-DA. Sample classes were defined according to the producer's internal thresholds (see §3.1) and class 2 was always associated to the high-performance products (i.e. high gel, high foam, high foam stability). Thus, for gelling properties, the 56 samples with gel strength  $\leq 58.8\text{ kPa}$  were assigned to class 1, whereas the 44 “high gel” samples were included in class 2. The figures of merit for the models obtained with FT-NIR spectra after different pre-treatments are reported in Table 1. Most calculated PLS-DA models showed a good discriminant ability in cross-validation, expressed as classification sensitivity and specificity. Sensitivity, representing the true positive rate, ranged from 79 (23/29 samples) to 86% (25/29 samples) for class 2. A lower performance was observed for class 1, since the highest sensitivity in cross-validation was 78% (29/37 samples) for raw and second derivative data. Specificity, i.e. the true negative rate, was good for both classes; the lowest value obtained in cross-validation was 72% (27/37 samples for class 2). The calculated models did not perform so good in external validation, leading in prediction to a scarce sensitivity for class 2 and a low specificity for class 1 (40–53%, i.e. 6/15–8/15 samples). The best model, chosen considering both cross-validation and prediction results, was the one constructed with first derivative data yielding, for the test set, sensitivity and specificity values of 84 (i.e. 16/19 samples) and 46% (i.e. 7/15 samples) for class 1, and *vice versa* for class 2.

The good potential of FT-IR in discriminating EWPs depending on their gelling performances was proved above all by the figures of merit of the model obtained after second derivative spectral transformation (Table 1), giving in prediction 78 (15/19 samples) and 60% (9/15 sam-

ples) sensitivity, and 60 (9/15 samples) and 78% (15/19 samples) specificity for class 1 and 2, respectively. Compared to the previous models, those constructed with the FT-IR spectra generally guaranteed a slightly better performance in terms of sensitivity for class 2 and specificity for class 1 in prediction. This could be ascribed to the fact that gelling properties are mainly linked to EWP protein content and conformation, which are connected to the FT-IR amide I band ( $1600\text{--}1700\text{ cm}^{-1}$ ) (Li et al., 2018). In particular, FT-IR proved to be reliable for the determination of protein secondary structure (Meziani et al., 2011), which could be here linked to  $\beta$ -sheet ( $1612\text{--}1625\text{ cm}^{-1}$ ),  $\beta$ -turn ( $1667\text{--}1678\text{ cm}^{-1}$ ) and  $\alpha$ -helix ( $1650\text{--}1660\text{ cm}^{-1}$ ) of egg white proteins (Shen et al., 2014). Indeed, the three-dimensional network generated by the interactions among the different proteins contained in EWP (more than 40) is the main responsible for gelling performance. However, heat treatments, including spray drying used to obtain EWP, can lead to extensive SH–SS exchanges in proteins, resulting in a more rigid  $\beta$ -sheet structure and higher molecular weight complexes, which are indications of protein aggregation (Meziani et al., 2011) that can lower gelling performance.

PLS-DA models based on FT-NIR data were calibrated also to distinguish “low-foam” (class 1, foam height  $\leq 160\text{ mm}$ ) from “high-foam” (class 2, foam height  $> 160\text{ mm}$ ) EWP samples (Table 2). The models calculated after different spectral pre-treatments gave good results in terms of sensitivity and specificity in cross-validation. The best one was obtained after MSC and first derivative spectral transformation, being sensitivity and specificity equal to or higher than 54% (i.e. 12/22 samples) in cross-validation, and equal to or higher than 64% (i.e. 9/14 samples) in prediction. However, it is clear that such a model is not industrially applicable to discriminate EWPs based on whipping properties. Indeed, the low sensitivity observed for class 1 (“low foam” samples) in prediction (64%, i.e. 9/14 samples) might indicate that a scarce number of samples with low whipping performance would be recognized as such. Practically this could entail the mislabeling of commercial EWPs. On the other hand, the best PLS-DA model constructed with the FT-IR spectra was obtained after MSC transformation, but classification sensitivity in prediction for class 1 (42%, i.e. 6/14 samples) was still found too poor for industrial purposes.

Foam instability is another parameter useful to discriminate EWPs with different whipping properties, thus classification models for differentiating EWP samples as unstable (class 1, foaming instability  $\geq 35\text{ mL}$ ) or stable (class 2, foam instability  $< 35\text{ mL}$ ) foam products were also trained (Table 3). The PLS-DA models calculated with

**Table 1**

Results of PLS-DA models computed with FT-NIR and FT-IR data for egg white powder discrimination based on gel strength (class 1  $\leq$  58.8kPa; class 2  $>$  58.8kPa): sensitivity and specificity (given in counts of samples) after different mathematical pre-treatments. The best models are reported in bold.

Range	Pre-treatment	Cross-validation				Prediction			
		Class 1		Class 2		Class 1		Class 2	
		Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
FT-NIR	Raw data	29/37	23/29	23/29	29/37	14/19	6/15	6/15	14/19
	De-trending	28/37	24/29	24/29	28/37	14/19	8/15	8/15	14/19
	<b>1st derivative</b>	<b>28/37</b>	<b>24/29</b>	<b>24/29</b>	<b>28/37</b>	<b>16/19</b>	<b>7/15</b>	<b>7/15</b>	<b>16/19</b>
	2nd derivative	29/37	24/29	24/29	29/37	15/19	6/15	6/15	15/19
	MSC	28/37	24/29	24/29	28/37	14/19	8/15	8/15	14/19
	MSC + 1st derivative	28/37	25/29	25/29	28/37	14/19	7/15	7/15	14/19
	MSC + 2nd derivative	28/37	25/29	25/29	28/37	14/19	7/15	7/15	14/19
	MSC + de-trending	27/37	24/29	24/29	27/37	14/19	8/15	8/15	14/19
	Please, be sure that in the final print Table there is enough space between FT-NIR and FT-IR results (also for all the other tables)								
FT-IR	Raw data	28/37	21/29	21/29	28/37	15/19	8/15	8/15	15/19
	1st derivative	28/37	22/29	22/29	28/37	15/19	8/15	8/15	15/19
	<b>2nd derivative</b>	<b>27/37</b>	<b>24/29</b>	<b>24/29</b>	<b>27/37</b>	<b>15/19</b>	<b>9/15</b>	<b>9/15</b>	<b>15/19</b>
	Linear base line	28/37	23/29	23/29	28/37	13/19	8/15	8/15	13/19
	Quadratic base line	28/37	22/29	22/29	28/37	14/19	8/15	8/15	14/19
	MSC	31/37	23/29	23/29	31/37	14/19	9/15	9/15	14/19
	MSC + linear baseline	30/37	23/29	23/29	30/37	12/19	9/15	9/15	12/19
	MSC + quadratic baseline	30/37	22/29	22/29	30/37	13/19	10/15	10/15	13/19
	MSC + 1st derivative	29/37	22/29	22/29	29/37	14/19	8/15	8/15	14/19
MSC + 2nd derivative	28/37	21/29	22/29	28/37	13/19	10/15	10/15	13/19	

PLS-DA, partial least square-discriminant analysis; FT-NIR, Fourier transform-near infrared spectroscopy; FT-IR, Fourier transform-infrared spectroscopy; MSC, multiplicative scatter correction.

**Table 2**

Results of PLS-DA models computed with FT-NIR and FT-IR data for egg white powder discrimination based on foam height (class 1  $\leq$  160 mm; class 2  $>$  160 mm): sensitivity and specificity (given in counts of samples) after different mathematical pre-treatments. The best models are reported in bold.

Range	Pre-treatment	Cross - validation				Prediction			
		Class 1		Class 2		Class 1		Class 2	
		Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
FT-NIR	Raw data	17/27	12/22	12/22	17/27	8/14	7/12	7/12	8/14
	Detrending	19/27	14/22	14/22	19/27	4/14	9/12	9/12	4/14
	1st derivative	18/27	13/22	13/22	18/27	6/14	9/12	9/12	6/14
	2nd derivative	20/27	13/22	13/22	20/27	4/14	9/12	9/12	3/14
	MSC	17/27	14/22	14/22	17/27	7/14	8/12	8/12	7/14
	<b>MSC + 1st derivative</b>	<b>17/27</b>	<b>12/22</b>	<b>12/22</b>	<b>17/27</b>	<b>9/14</b>	<b>8/12</b>	<b>8/12</b>	<b>9/14</b>
	MSC + 2nd derivative	18/27	12/22	12/22	18/27	6/14	9/12	9/12	6/14
	MSC + detrending	19/27	15/22	15/22	19/27	8/14	8/12	8/12	8/14
	FT-IR	Raw data	18/27	12/22	12/22	18/27	6/14	7/12	7/12
1st derivative		16/27	13/22	13/22	16/27	7/14	8/12	8/12	7/14
2nd derivative		17/27	13/22	13/22	17/27	6/14	7/12	7/12	6/14
Linear baseline		16/27	14/22	14/22	16/27	7/14	9/12	9/12	7/14
Quadratic baseline		19/27	16/22	16/22	19/27	7/14	8/12	8/12	7/14
<b>MSC</b>		<b>18/27</b>	<b>13/22</b>	<b>13/22</b>	<b>18/27</b>	<b>6/14</b>	<b>8/12</b>	<b>8/12</b>	<b>6/14</b>
MSC + linear baseline		17/27	14/22	14/22	17/27	7/14	8/12	8/12	7/14
MSC + quadratic baseline		17/27	15/22	15/22	17/27	6/14	8/12	8/12	6/14
MSC + 1st derivative		17/27	13/22	13/22	17/27	7/14	6/12	6/12	7/14
MSC + 2nd derivative		17/27	14/22	14/22	17/27	7/14	6/12	6/12	7/14

PLS-DA, partial least square-discriminant analysis; FT-NIR, Fourier transform-near infrared spectroscopy; FT-IR, Fourier transform-infrared spectroscopy; MSC, multiplicative scatter correction.

**Table 3**

Results of PLS-DA models computed with FT-NIR and FT-IR data for egg white powder discrimination based on foam instability (class 1  $\geq$  35 mL; class 2 < 35 mL): sensitivity and specificity (given in counts of samples) after different mathematical pre-treatments. The best models are reported in bold.

Range	Pre-treatment	Cross - validation				Prediction				
		Class 1		Class 2		Class 1		Class 2		
		Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	
FT-NIR	Raw data	16/25	16/24	16/24	16/25	7/13	9/12	9/12	7/13	
	Detrending	17/25	19/24	19/24	17/25	8/13	11/12	11/12	8/13	
	<b>1st derivative</b>	<b>18/25</b>	<b>20/24</b>	<b>20/24</b>	<b>18/25</b>	<b>8/13</b>	<b>11/12</b>	<b>11/12</b>	<b>8/13</b>	
	2nd derivative	17/25	19/24	19/24	17/25	6/13	10/12	10/12	6/13	
	MSC	17/25	16/24	16/24	17/25	8/13	10/12	10/12	8/13	
	MSC + 1st derivative	18/25	18/24	18/24	18/25	7/13	11/12	11/12	7/13	
	MSC + 2nd derivative	18/25	19/24	19/24	18/25	7/13	10/12	10/12	7/13	
	MSC + detrending	18/25	18/24	18/24	18/25	8/13	9/12	9/12	8/13	
	FT-IR	Raw data	19/25	20/24	20/24	19/25	6/13	6/12	6/12	6/13
		<b>1st derivative</b>	<b>20/25</b>	<b>20/24</b>	<b>20/24</b>	<b>20/25</b>	<b>9/13</b>	<b>7/12</b>	<b>7/12</b>	<b>9/13</b>
2nd derivative		18/25	20/24	20/24	18/25	9/13	7/12	7/12	9/13	
Linear baseline		19/25	20/24	20/24	19/25	8/13	7/12	7/12	8/13	
Quadratic baseline		17/25	20/24	20/24	17/25	8/13	7/12	7/12	8/13	
MSC		19/25	20/24	20/24	19/25	8/13	6/12	6/12	8/13	
MSC + linear baseline		19/25	20/24	20/24	19/25	8/13	6/12	6/12	8/13	
MSC + quadratic baseline		19/25	20/24	20/24	19/25	8/13	6/12	6/12	8/13	
MSC + 1st derivative		19/25	21/24	21/24	19/25	9/13	6/12	6/12	9/13	
MSC + 2nd derivative		19/25	20/24	20/24	19/25	9/13	5/12	5/12	9/13	

PLS-DA, partial least square-discriminant analysis; FT-NIR, Fourier transform-near infrared spectroscopy; FT-IR, Fourier transform-infrared spectroscopy; MSC, multiplicative scatter correction.

FT-NIR and FT-IR data showed similar classification ability in terms of sensitivity and specificity, both in cross-validation and in prediction. In the case of FT-NIR spectroscopy, the best PLS-DA model was obtained with first derivative spectra, which showed sensitivity and specificity in prediction of 61 (8/13 samples) and 92% (11/12 samples) for class 1, and *vice versa* for class 2.

Also for FT-IR spectroscopy, the highest prediction ability was achieved by the model built after first derivative transformation, being sensitivity and specificity in prediction for class 1, 69 (9/13 samples) and 58% (7/12 samples), respectively, and *vice versa* for class 2.

Finally, the best PLS-DA models were re-calculated after the selection of the most discriminant spectral wavenumbers carried out as described in §2.3. The outcomes can be considered rather similar to those yielded by the classifiers built employing the full spectra of the different samples (see Table 4 for an overview of the best classifications achieved). Slight improvements can be observed in the PLS-DA model constructed with the FT-NIR data for the discrimination based on gel strength (see class 2 sensitivity/class 1 specificity in prediction – Tables 1 and 4), in the PLS-DA model constructed with the FT-IR data for the discrimination based on foam height (see class 1 sensitivity/class 2 specificity in prediction – Tables 2 and 4) and in the PLS-DA model constructed with the FT-IR data for the discrimination based on foam instability (see class 1 sensitivity/class 2 specificity in prediction – Tables 3 and 4).

Nevertheless, even if promising, for a real industrial application the classification models should be improved by including in the datasets other EWP samples with higher variability in the technological properties.

#### 4. Conclusions

Overall, the obtained results show that both FT-IR and FT-NIR spectroscopy coupled with multivariate data analysis, and specifically with

PLS-DA, could represent feasible solutions for the classification of EWPs according to their technological properties. In addition, the possibility of performing this classification by utilizing only few channels of the original spectral range and obtaining still satisfactory results in most of the explored situations may constitute a promising industrial perspective especially for future studies that might involve hyperspectral imaging. In fact, cheaper and faster instruments spanning only a limited interval of wavelengths but guaranteeing almost the same performance for EWP sorting and/or quality control could be designed and exploited directly on the production lines. Such a possibility could greatly benefit producers and end-users by allowing the correct labeling and destination of each single EWP production lot.

#### Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

#### Declarations of interest

None.

#### Acknowledgements

The Authors wish to thank Lactosan-Sanovo Ingredients Group (Zeven/Aspe, Germany) for providing EWP samples and the values of their technological properties. A special acknowledgement to dr Giosuè Casagrande, R&D Food Technologist of Sanovo Group, for his technical support to the work.

**Table 4**

Results of PLS-DA models computed with the selected wavenumbers of FT-NIR and FT-IR spectra for egg white powder discrimination based on technological properties: pre-treatment, number of selected variables and sensitivity and specificity (given in counts of samples) in cross-validation and prediction.

Range	Technological property	Pre-treatment	Selected variables	Cross-validation				Prediction			
				Class 1		Class 2		Class 1		Class 2	
				Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
FT-NIR	Gel strength	1st derivative	20	32/37	23/29	23/29	32/37	16/19	10/15	10/15	16/19
	Foam height	MSC + 1st derivative	25	18/27	16/22	16/22	18/27	9/14	8/12	8/12	9/14
	Foam instability	1st derivative	20	21/25	17/24	17/24	21/25	8/13	11/12	11/12	8/13
FT-IR	Gel strength	2nd derivative	20	27/37	26/29	26/29	27/37	16/19	10/15	10/15	16/19
	Foam height	MSC	15	18/27	13/22	13/22	18/27	9/14	10/12	10/12	9/14
	Foam instability	1st derivative	50	21/25	22/24	21/25	22/24	10/13	7/12	7/12	10/13

PLS-DA, partial least square-discriminant analysis; FT-NIR, Fourier transform-near infrared spectroscopy; FT-IR, Fourier transform-infrared spectroscopy; MSC, multiplicative scatter correction.

## References

- Barker, M., Rayens, W., 2003. Partial least squared for discrimination. *Journal of Chemometrics* 17, 166–173.
- Barnes, R.J., Dhanoa, M.S., Lister, S., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy* 43, 772–777.
- Barth, A., 2007. Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1767, 1073–1101.
- Bevilacqua, M., Bucci, R., Magrì, A.D., Magrì, A.L., Marini, F., 2012. Tracing the origin of extra virgin olive oils by infrared spectroscopy and chemometrics: A case study. *Analytica Chimica Acta* 717, 39–51.
- Bevilacqua, M., Bucci, R., Materazzi, S., Marini, F., 2013. Application of near infrared (NIR) spectroscopy coupled to chemometrics for dried egg-pasta characterization and egg content quantification. *Food Chemistry* 140, 726–734.
- Dalle Zotte, A., Berzaghi, P., Jansson, L.M., Andrighetto, I., 2006. The use of near-infrared reflectance spectroscopy (NIRS) in the prediction of chemical composition of freeze-dried egg yolk and discrimination between different n–3 PUFA feeding sources. *Animal Feed Science and Technology* 128, 108–121.
- Danezis, G.P., Tsagkaris, A.S., Camin, F., Brusic, V., Georgiou, C.A., 2016. Food authentication: Techniques, trends & emerging approaches. *Trends in Analytical Chemistry* 85, 123–132.
- Esslinger, S., Riedl, J., Fahl-Hassek, C., 2014. Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Research International* 60, 189–204.
- Folch-Fortuny, A., Prats-Montalbán, J.M., Cubero, S., Blasco, J., Ferrer, A., 2016. VIS/NIR hyperspectral imaging and N-way PLS-DA models for detection of decay lesions in citrus fruits. *Chemometrics and Intelligent Laboratory Systems* 156, 241–248.
- Geladi, P., MacDougall, D., Martens, H., 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy* 39, 491–500.
- Handa, A., Hayashi, K., Shidara, H., Kuroda, N., 2001. Correlation of the protein structure and gelling properties in dried egg white products. *Journal of Agricultural and Food Chemistry* 49, 3957–3964.
- Hecht, K.T., Wood, D.L., 1956. The near infrared spectrum of the peptide group. *Proceedings of the Royal Society of London: Mathematical, Physical and Engineering Sciences* 235, 174–188.
- Kato, A., Ibrahim, H.R., Watanabe, H., Honma, K., Kobayashi, K., 1989. New approach to improve the gelling and surface properties of dried egg white proteins in the dry state. *Journal of Agricultural and Food Chemistry* 37, 433–437.
- Lechevalier, V., Jeantet, R., Arhaliass, A., Legrand, J., Nau, F., 2007. Egg white drying: Influence of industrial processing steps on protein structure and functionalities. *Journal of Food Engineering* 83, 404–413.
- Li-Chan, E.C.Y., Powrie, W.D., Nakai, S., 1995. The chemistry of eggs and egg products. In: Stadelman, W.J., Cotterill, O.J. (Eds.), *Egg science and technology*. The Haworth Press Inc, Binghamton, pp. 105–176.
- Li, J., Li, X., Wang, C., Zhang, M., Xu, Y., Zhou, B., et al., 2018. Characteristics of gelling and water holding properties of hen egg white/yolk gel with NaCl addition. *Food Hydrocolloids*, 77 887–893.
- Lin, H., Zhao, J., Sun, L., Chen, Q., Zhou, F., 2011. Freshness measurement of eggs using near infrared (NIR) spectroscopy and multivariate data analysis. *Innovative Food Science & Emerging Technologies* 12, 182–186.
- Liu, Y., Cho, R.K., Sakuri, K., Miura, T., Ozaki, Y., 1994. Studies on spectra/structure correlations in near-infrared spectra of proteins and polypeptides. Part I: A marker band for hydrogen bonds. *Applied Spectroscopy* 48, 1249–1254.
- Lohumi, S., Lee, S., Lee, H., Cho, B.-K., 2015. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science & Technology* 46, 85–98.
- Martens, H., Jensen, S.A., Geladi, P., 1983. Multivariate linearity transformations for near infrared reflectance spectroscopy. In: Christie, O.H.J. (Ed.), *Proceedings of the Nordic symposium on applied statistics*. pp. 205–234. (Stavanger: Stokkland Forlag).
- Meziani, S., Jasniowski, J., Gaiani, C., Ioannou, I., Muller, J.-M., Ghoul, M., et al., 2011. Effects of freezing treatments on viscoelastic and structural behavior of frozen sweet dough. *Journal of Food Engineering* 107, 358–365.
- Mine, Y., 1995. Recent advances in the understanding of egg white proteins functionality. *Trends in Food Science & Technology* 6, 225–231.
- Mine, Y., 1996. Effect of pH during the dry heating on the gelling properties of egg white proteins. *Food Research International* 29, 155–161.
- Németh, C., Pataki, J., Jónás, G., Surányi, J., Friedrich, L., Pásztor-Huszár, K., et al., 2011. Near infrared spectroscopic measurements in liquid egg white products kept at 50, 55 and 60°C. *Journal of Food Agriculture and Environment* 9 (3&4), 49–52.
- Pérez, N., Ferré, J., Boqué, R., 2009. Calculation of the reliability of classification in Discriminant Partial Least-Squares binary classification. *Chemometrics and Intelligent Laboratory Systems* 95, 122–128.
- Savitzky, A., Golay, M., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36, 1627–1639.
- Seregély, Z., Farkas, J., Tuboly, E., Dalmadi, I., 2006. Investigating the properties of egg white pasteurised by ultra-high hydrostatic pressure and gamma irradiation by evaluating their NIR spectra and chemosensor array sensor signal responses using different methods of qualitative analysis. *Chemometrics and Intelligent Laboratory Systems* 82, 115–121.
- Shen, F., Niu, F., Li, J., Su, Y., Liu, Y., Yang, Y., 2014. Interactions between tea polyphenol and two kinds of typical egg white proteins-ovalbumin and lysozyme: Effect on the gastrointestinal digestion of both proteins in vitro. *Food Research International* 59, 100–107.
- Snee, R., 1977. Validation of regression models: Methods and examples. *Technometrics* 19, 415–428.
- Talansier, E., Loisel, C., Dellavalle, D., Desrumaux, A., Lechevalier, V., Legrand, J., 2009. Optimization of dry heat treatment of egg white in relation to foam and interfacial properties. *LWT-Food Science and Technology* 42, 496–503.
- USDA, 2015. National nutrient database for standard reference, release 28 (slightly revised). In: <https://www.ars.usda.gov/northeast-area/beltsville-md/beltsville-human-nutrition-research-center/nutrient-data-laboratory/docs/sr28-download-files/>, Accessed 2 February 2018.
- Uygun-Sarıbay, M., Ergun, E., Kalaycı, Y., Köseoğlu, T., 2017. The secondary structure of proteins in liquid, frozen, and dried egg-white samples: Effect of gamma irradiation treatment. *International Journal of Food Properties* 20, S1195–S1203.
- Vitale, R., Bevilacqua, M., Bucci, R., Magrì, A.D., Magrì, A.L., Marini, F., 2013. A rapid and non-invasive method for authenticating the origin of pistachio samples by NIR spectroscopy and chemometrics. *Chemometrics and Intelligent Laboratory Systems* 121, 90–99.
- Wold, S., Albano, C., Dunn III, W.J., Esbensen, K., Hellberg, S., Johansson, E., et al., 1983. Pattern recognition: Finding and using regularities in multivariate data. In: Martens, H., Russwurm Jr., H. (Eds.), *Food research and data analysis*. Applied Science Publishers, London, pp. 147–188.
- Wold, S., Johansson, E., Cocchi, M., 1993. 3D QSAR. In: Kubinyi, H. (Ed.), *Drug design: Theory, methods and applications*. Kluwer Academic Publishers, Dordrecht.
- Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58, 109–130.
- Workman Jr., J., Weyer, L., 2008. *Practical guide to interpretative near-infrared spectroscopy*. CRC Press, Boca Raton.
- Zhao, J., Lin, H., Chen, Q., Huang, X., Sun, Z., Zhou, F., 2010. Identification of egg's freshness using NIR and support vector data description. *Journal of Food Engineering* 98, 408–414.