Talens Oliag, P.; Mora, L.; Morsy, N.; Barbin, DF.; Elmasry, G.; Sun, D. (2013). Prediction of water and protein contents and quality classification of Spanish cooked ham using NIR hyperspectral imaging. Journal of Food Engineering. (117):272-280. doi:10.1016/j.jfoodeng.2013.03.014.

Prediction of water and protein contents and quality classification of Spanish cooked ham using NIR hyperspectral imaging

Pau Talens[a*]; Leticia Mora[b]; Noha Morsy[c]; Douglas F. Barbin[d]; Gamal ElMasry[c]; Da-Wen Sun[d*]

a) Departamento de Tecnología de Alimentos, Universitat Politècnica de València. Camino de Vera, s/n 46022, Valencia, Spain
b) Ashtown Food Research Centre, Teagasc, Ashtown, Dublin 15, Ireland
c) Suez Canal University, Faculty of Agriculture, Ismailia, Egypt
d) FRCFT Group, Biosystems Engineering, University College Dublin, National University of Ireland, Belfield, Dublin 4, Ireland

(*) Contact information for Corresponding Authors

Pau Talens. Departamento de Tecnología de Alimentos. Universitat Politècnica de València. Camino de Vera, s/n. 46022. Valencia. Spain. Phone: 34-963879836, Fax: 34-963877369, e-mail: pautalens@tal.upv.es

Da-Wen Sun. Tel: +353-1-7167342, Fax: +353-1-7167493, E-mail: dawen.sun@ucd.ie, Website: www.ucd.ie/refrig; www.ucd.ie/sun

ABSTRACT

This study was carried out to investigate the ability of hyperspectral imaging technique in the NIR spectral region of 900–1700 nm for the prediction of water and protein contents in Spanish cooked hams. Multivariate analyses using partial least-squares regression (PLSR) and partial least squares-discriminant analysis (PLS-DA) were applied to the spectral data extracted from the images to develop statistical models for predicting chemical attributes and classify the different qualities. Feature-related wavelengths were identified for protein (930, 971, 1051, 1137, 1165, 1212, 1295, 1400, 1645 and 1682 nm) and water (930, 971, 1084, 1212, 1645 and 1682 nm) and used for regression models with fewer predictors. The PLS-DA model using optimal wavelengths (966, 1061, 1148, 1256, 1373 and 1628 nm) successfully classified the examined hams in different quality categories. The results revealed the potentiality of NIR hyperspectral imaging technique as an objective and non-destructive method for the authentication and classification of cooked hams.

Keywords: Chemical image; chemical attributes; PLSR; PLS-DA; Spanish cooked ham; hyperspectral imaging.

1. INTRODUCTION

Cooked ham is a meat product with high levels of consumption in Spain and other countries. In 2009, the consumption of cooked meat products including different categories of cooked ham resulted in total expenses of 185.77 million Euros and a consumption of 4.05 kg per capita in Spain (MARM, 2010). The main stages in the processing of cooked hams are the brine injection, which is different according to the desired final quality; the tumbling and massaging step, which is useful to distribute the brine solution through the entire piece and to extract the proteins from the fibers; the cooking step, which requires a rigorous temperature and time control to assure the wholesomeness of the product; and finally, the cooling step, after which hams are taken out of the mold and packaged (Frentz, 1982; Toldrá et al., 2010). The salt added in the brine solution reaches a final content of around 2% in the ham. Nitrite is also added at levels of 120-150 mg/kg as a contributor to color formation, antioxidant, and to preserve against pathogens. In order to avoid risks of nitrosamine formation, the addition of ascorbates at levels of 200-400 mg/kg is recommended. Phosphates given as P_2O_5 at a maximum concentration of 7500 ppm are also allowed in Spain in all type of cooked ham as a contributor to ham's water retention. Sugar in dextrose form is also added for taste purposes (Toldrá et al., 2010)

The final quality of the product depends on the raw material used (pH, microbial content, or fat) and the processing conditions. There is a broad range of types of cooked ham, which generally are classified depending on different characteristics. Therefore, Spanish Government has established quality regulations for cooked meat products with the aim to define the conditions and characteristics that must be complied by all Spanish cooked hams (BOE, 1983). This quality regulation also

classifies and categorizes the cooked ham in three main categories: ‘extra cooked ham’ (or extra category), ‘first class cooked ham’ (or first category), and ‘second class cold cut ham’ (or second category). Main differences among these categories are based on the protein and water contents of the cooked hams. In fact, according to the legislation, the water/protein ratio in ‘Extra category’ should be less than 4.13, whereas 4.65 would be the allowed ratio for ‘First category’ cooked ham. In ‘Second category’, a minimum of 14% of meat protein in the total product is allowed, although it is possible to add up to 1% of external proteins.

In the industry, the quality of ham in terms of chemical composition is generally assessed by experienced personnel using analytical techniques that are time consuming and sample-destructive such as the gravimetric measurements of water or the most commonly used nitrogen determination methods that include Kjeldahl (AOAC, 1999a), Dumas (AOAC, 1999a), and combustion (AOAC, 1999b) methods. In this sense, the hyperspectral imaging system is an emerging technique that integrates both conventional imaging and spectroscopy technologies for attaining spatial and spectral information of the product (ElMasry et al., 2012). This technique has been proved to be useful in quality evaluation and classification of different types of products such as raw meat (Prevolnik et al, 2004; Qiao et al., 2007; Liu et al., 2010; Barbin et al., 2012; Kamruzzaman et al., 2011; ElMasry et al., 2011a), meat products (ElMasry et al., 2011b), or fruits and vegetables (Karimi et al., 2009; Cubero et al., 2011; Rajkumar et al., 2012). This technique is considered as low time-consuming, non-destructive, and requiring a minimum of human intervention. Hyperspectral imaging technique has the ability to capture internal constituent gradients within the product, which is a useful tool for non-homogeneous multi-

component products. The main objective of the present work was to investigate the potential application of the NIR hyperspectral imaging technique for evaluating the quality of Spanish cooked hams. The specific objectives were to (a) predict water and protein contents in Spanish cooked ham and (b) classify the ham according to its initial quality. These objectives were achieved by (1) establishing a NIR hyperspectral imaging system with a spectral region of 910-1700 nm, (2) building robust calibration models using partial least-squares regression (PLSR) to quantitatively predict protein and water contents and partial least-squares discriminant analysis (PLS-DA) to classify the different ham qualities, (3) identifying the most informative wavelengths for prediction and classification purposes, and (4) building chemical images by developing image processing algorithms for mapping the concentration of protein and water contents in the ham slices.

2. MATERIALS AND METHODS

2.1. Cooked ham samples

Four types (Ham 1, Ham 2, Ham 3 and Ham 4) of Spanish cooked ham of different categories of quality (extra category, first category, and second category), were studied. Ham 1 (H1) and ham 2 (H2) were both ‘extra category’, but H1 had a lower fat content than H2. Ham 3 (H3) and ham 4 (H4), represent the ‘first category’ and the ‘second category’, respectively. Samples from cooked hams labeled according to the Spanish Quality Regulation were purchased from a local retailed market. A total of sixty-three slices, 1cm thick, of Spanish cooked ham were analyzed. All cooked hams were accurately labeled according to their categories of quality and stored in a

fridge at 4ºC. Before image acquisition, cooked hams were removed from the fridge and kept for 30 min at room temperature (22 ºC) to be acclimatized with the surrounding environment. Each slice was then imaged individually in the hyperspectral imaging system. Three cylindrical subsamples from each cooked ham slice were cored using a 2.5 cm diameter cylindrical hollow drill with a sharp cutting edge. The subsamples were chosen from different locations in the slice, comprising a broad range of protein (n=126), water (n=126) and fat (n=63) concentrations.

2.2. Chemical Analysis

2.2.1. Fat and water contents

The intramuscular fat and water contents for each sample were determined using the CEM analysis system described by Bostian et al., (1985). The subsamples were blended and subsequently analyzed on a Smart Trac System (CEM Corporation, Matthews, NC, USA) which consists of two modules: the Smart System 5 to determine water content, and the Smart Trac module for fat content determination. A sample between 2 to 3 grams of blended cooked ham was weighted in the Smart System 5, where its water content was gravimetrically analyzed by determining the weight loss. Then, the same dried samples were rolled in CEM's Trac Film™ and placed in the Smart Trac (Rapid Flow Analyzer,), which is a nuclear magnetic resonance (NMR) module. In this equipment, the sample is pulsed with radio frequency (RF) energy while within a static magnetic field. The resulting signal is recorded and analyzed for the total proton activity of fat present in the sample. Results of water and fat contents are given in percentage.

2.2.2. Protein content

Protein content was measured according to the method of Sweeney and Rexroad (1987) on the LECO FP-328 total nitrogen determinator (LECO R Corporation, St. Joseph, MI, USA) calibrated with ethylene diamine tetraacetic acid (EDTA) calibration solutions. This method is based on the measurement of nitrogen by combustion. For the analysis, a sample of 0.25±0.05 g was weighted in tin foil cups (LECO R Corporation, St. Joseph, MI, USA) in triplicate and loaded in the autoloader of the system. The factor used to convert the amount of detected nitrogen to a percent of crude protein was 6.25.

2.3. Image acquisition

A line-scan hyperspectral imaging system in the NIR range of 890-1750 nm with 256 spectral bands, described previously by ElMasry et al., 2011b, was used for acquiring hyperspectral images of ham slices (Figure 1). During image acquisition, each slice of each ham was placed on the translation stage and moved at a constant speed of 2.8 cm/s. The speed of the translation stage was synchronized with the image acquisition by the SpectralCube software to obtain a spectral image with a spatial resolution of 0.58 mm/pixel. The system scans the sample line by line and the reflected light is dispersed by the spectrograph and captured by the CCD array detector of the camera in spatial-spectral axes. The camera has 320×256 (spatial × spectral) pixels and the spectral increment between the contiguous bands was 3.36 nm in the spectral range of 890-1750 nm yielding 256 bands. Once the hyperspectral image has been acquired, it was sent to the computer for storage in a raw format before being processed. The main key steps for the whole procedure of image analysis are presented in Figure 2 and briefly described in the following section.

2.4. Image processing

160 The raw images were processed using the Environment for Visualizing Images
161 (ENVI) software (ENVI 4.6.1) (Research Systems Inc., Boulder Co., USA). Because
162 the response of the CCD detector in the ranges of 897–910 and 1700–1753 nm was
163 rather low and the resulting spectral images at these two particular ranges were rather
164 noisy, the hyperspectral images were resized to the spectral range of 910 nm to 1700
165 nm with a total of 237 bands.
166 To remove the effect of dark current of the camera sensor from the acquired images
167 (R_0), white and dark reference images were concurrently captured. The white
168 reference image (R_W) was acquired from a white Teflon calibration tile and the dark
169 reference image (R_D) was obtained by turning off the light source along with
170 completely closing the lens of the camera with its opaque cap. A relative reflectance
171 image (R) was then calculated using the following equation:

172 $$R = \frac{R_0 - R_D}{R_w - R_D} \qquad \text{(eq. 1)}$$

173 Final images with a dimension of 320 pixels × 500pixels × 237 bands were obtained
174 and subsequently used to extract the spectral information.

175 **2.5. Region of interest selection and spectral data extraction**

176 Two different regions of interest were selected in each slice to establish the spectral
177 data sets. Figure 3 shows the procedure for selecting the regions of interest for the
178 spectral datasets used for water, protein and fat prediction (Figure 3a) and for the
179 spectral dataset used for classification of cooked hams (Figure 3b-f). For predicting
180 protein, water and fat contents, the spectral data were extracted from the image pixels
181 corresponding to the circular areas shown in Figure 3a where the reference
182 subsamples had been collected. Pixel spectra within each circular region were
183 averaged and saved in three matrices (X_1, X_2 and X_3). The extraction of this spectral

184 information was carried out manually using ENVI software (Research Systems Inc.,
185 Boulder, CO, USA).
186 For classification purposes of the examined ham qualities, the spectral data were
187 extracted from the lean region of each ham slice. Therefore, a segmentation routine
188 was applied before extracting spectral data to separate the lean part of ham from the
189 background and the fat or gelatin covering layer. The process started by subtracting a
190 low-reflectance band from a high-reflectance band (Figure 3c) followed by a simple
191 thresholding. This step produces a segmented image for the whole ham sample
192 including the lean part, gelatin and fat portions of the sample (Figure 3d). Again,
193 segmentation was performed for detecting the gelatin and fat by simple thresholding
194 to produce a binary image of fat and gelatin pixels (Figure 3e). The lean portion was
195 isolated by subtracting the second segmented image (Figure 3e) from the first
196 segmented image (Figure 3d) to produce a mask containing only the lean part in a
197 black background (Figure 3f). The isolated lean portion was then used as the main
198 region of interest (ROI) to extract the average spectral data from only the lean part of
199 the ham samples and avoid fat and other undesired components that can affect the
200 prediction values. The extracted spectral data were then arranged in another matrix
201 (X). All extraction routines were performed using the software Matlab
202 7.11.0.584(R2010b) (The Mathworks Inc., Natick, MA, USA).
203 **2.6. Spectral data analysis and wavelength selection**
204 The extracted spectral data were then arranged in two matrices where the rows
205 represent the number of samples and the columns represent the number of variables
206 (237 wavelengths). Partial least-squares regression (PLSR) was applied to the first
207 matrix to develop separate models for each chemical constituent. PLSR technique is

208 particularly useful when it is necessary to predict a set of dependent variables from a
209 large set of independent variables (Abdi, 2010). In such case, the values of one
210 attribute (protein, water or fat content) of the dataset (Y_1, Y_2 and/or Y_3) were used to
211 represent the dependent variable and the reflectance values at 237 wavelengths (X_1,
212 X_2 and/or X_3) represented the independent variables. Performance of the prediction
213 models was evaluated using the root-mean-square error of calibration (RMSEC), the
214 root-mean-square error of cross-validation (RMSECV), the coefficients of
215 determination (R^2), and the numbers of the latent variables required (#LV). The
216 number of latent variables was determined using the minimum value of predicted
217 residual error sum of squares (PRESS) (ElMasry et al., 2007; Esquerre et al., 2009).
218 When the number of latent factors in the model increased, the value of PRESS
219 decreased until its lowest value corresponding to the ideal number of latent factors.
220 Leave-one-out cross-validation method was used to validate the prediction models.
221 Moreover, partial least-squares discriminant analysis (PLS-DA) (Prats-Montalban et
222 al., 2006; Berrueta et al., 2007; Gaston et al., 2010), developed with leave-one-out
223 cross-validation, was applied to the second matrix(X-Matrix) to build a qualitative
224 model for ham classification. For this purpose, a dummy *Y*-variable (Y-Matrix) was
225 assigned to each ham class, 1 for H1, 2 for H2, 3 for H3 and 4 for H4. Performance
226 of the classification model was evaluated using the same parameters used for the
227 prediction models (RMSEC, RMSECV, R^2, and the #LV). Samples for which the
228 difference between actual and predicted values exceeded three times of the standard
229 deviation were considered as outliers (Brimmer and Hall, 2001; Chen et al., 2005). In
230 order to reduce the high dimensionality of the extracted spectral data, to avoid the
231 presence of noise or information that is not related to the chemical features, and to

make the PLSR model more robust, the most important wavelengths for the prediction of the chemical attributes and classification were selected (ElMasry et al., 2007). Weighted β-coefficients resulting from the PLSR models established using the whole spectral range consisting of 237 wavelengths were used for identifying the optimal wavelengths.

2.7. Mapping of water and protein content

The PLS regression models were used to predict water and protein concentrations in each pixel of the spectral image. This was done by multiplying the model regression coefficients by the spectrum of each pixel in the image at selected wavelengths. The resulting prediction image (called 'chemical image') was displayed in colors, where different colors represent different concentrations of protein or water within the sample. Thus, the prediction was done by developing a calibration model and then an interpolation was applied to extend the model to all pixels in the image.

2.8. Statistical Analysis

Analysis of variance (ANOVA) was conducted to determine significant differences in the measured protein, water, and fat contents, as well as the ratio water/protein among the analyzed cooked ham samples (see Table 1) and for the predicted values for protein and water using the PLS model (see Table 3), using the software Statgraphics Plus for Windows 5.1 (Manugistics Corp., Rockville, Md.). All multivariate analyses (PLS and PLS-DA) were conducted using The Unscrambler v9.7 (CAMO Software AS, OSLO, Norway).

3. RESULTS AND DISCUSSION

3.1. Cooked ham analysis

The protein, water, and fat contents (Y_1, Y_2 and Y_3) of the tested Spanish hams determined by instrumental methods are shown in Table 1. According to Spanish quality regulations for cooked meat products, the examined ham samples could be easily classified as extra category class (H1 and H2), first category (H3), and second category (H4). Water/protein ratio values of the analyzed H1 and H2 were under 4.13 whereas H3 showed a value of water/protein ratio above 4.13 but below 4.68. The protein concentration of H4 was 15.15 ± 1.87 (Table 1). This result agrees with the minimum protein amount of 14% required in second category cooked hams considering the addition of 1% extra protein. The content of fat is not considered in the quality regulations but a considerable difference between H1 (low fat content) and the rest of the hams was observed. Significant differences ($p<0.01$) between extra category (H1 and H2), first category (H3) and second category (H4) were detected in protein and water content.

3.2. Spectral characteristics of hams

Figure 4 shows the average reflectance spectra in the spectral range of 910–1700 nm of the four examined hams. In general, all recorded spectra had the same shape and showed characteristic bands of water at 970 nm and 1440 nm, and characteristic bands of fat at 1200 nm (Leroy et al., 2004; Barlocco et al., 2006; Cen and He, 2007; Andres et al., 2008; Prieto et al., 2008). Although the spectral curves show a similar pattern, small differences can be observed among the spectral profiles of the samples in terms of reflectance magnitude. It can be observed that H4 had the lowest reflectance values (higher absorbance) throughout the whole spectral range. Also, it was clear to notice that H2 had highest reflectance values compared with the other ham categories. These differences are clearly observed in the entire spectral range

especially in the range between 900 nm and 1400 nm and could be attributed to the differences observed on the water/protein ratio among the hams (Table 1). As can be observed in Figure 4, when the water content of ham samples increased (H4) the reflectance values decreased. In the same manner, when the water content decreased (H2) the reflectance values increased. In this sense it is interesting to try to correlate the spectral data and the chemical attributes measured in order to see if the NIR hyperspectral imaging technique can be used for classifying different ham qualities.

3.3. Prediction of protein, water and fat contents

Multivariate analyses, developed with leave-one-out cross-validation, were used to find the most accurate PLSR models for the prediction of protein, water and fat contents. This analysis involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data. Thus, using this validation procedure, one sample was left out and the multivariate model was constructed by the rest of the samples using exactly the same process as the modelling procedure for feature selection and model construction. The obtained model was used for the prediction of protein, water and fat contents in the sample. Table 2 shows the root-mean-square error of calibration (RMSEC), the root-mean-square error of cross-validation (RMSECV), the coefficients of determination (R^2), and the numbers of the latent variables required (#LV) for protein, water and fat contents in ham samples by using the full spectral and the important wavelengths for protein and water contents. The results indicated that PLSR models for protein and water exhibited low values of RMSEC and RMSECV and high values of coefficients of determination (R^2), indicating good performance of the models for predicting

304 protein and water. Regarding fat results, a bad performance of the predicting model
305 was obtained probably due to the narrow range of fat contents observed in the
306 cooked ham samples. Figure 5 shows the efficiency of PLRS models for predicting
307 protein, water and fat contents of all ham samples.
308 Wavelengths corresponding to the highest absolute values of weighted β-coefficients
309 were considered as optimal wavelengths. Ten individual wavelengths (930, 971,
310 1051, 1137, 1165, 1212, 1295, 1400, 1645 and 1682 nm) for protein and 6 individual
311 wavelength (930, 971, 1084, 1212, 1645 and 1682 nm) for water were identified as
312 important wavelengths. According to the literature, wavelengths at 971 and 1400 nm
313 are related to the absorbance of O–H bonds and could be associated to water content
314 (Cozzolino and Murray, 2004; Barlocco et al., 2006), meanwhile, wavelengths at 930
315 nm and in the range of 1100-1400 nm are related to the absorption of the C–H bonds
316 of fatty acids and could be associated with fat content (Alomar et al., 2003; Prieto et
317 al., 2009). Wavelengths at 1645 and 1682 nm could be related to C–H stretching first
318 overtones (Osborne et al., 1993).
319 After identifying the optimal wavelengths for each attribute, spectral data were then
320 reduced from 237 wavelengths to 10 wavelengths in the case of protein and to 6
321 wavelengths in the case of water (Table 2). For each attribute, the reduced spectral
322 data were used to build a new PLSR model using the reflectance at these particular
323 wavelengths as independent variables, and the measured values of protein or water as
324 dependent variables. As shown in Table 2, the optimized PLSR models had
325 comparable results to the original PLSR models. Both models had good performance
326 in predicting protein and water contents in ham samples, indicating that it is possible

to use fewer wavelengths than 237 wavelengths to predict the chemical compositions in Spanish cooked hams.

3.4. Distribution maps of protein and water

Although it was possible to predict the contents of protein and water of the examined ham samples, it was quite vital to see the difference in these chemical attributes within the same ham samples or even among ham samples in a visualized form called ‘chemical images’. This could be achieved by applying the resulting PLSR model in a pixel-wise manner by a simple interpolation to show the distribution of these chemical attributes and their gradients from point to point in the sample. The PLS regression coefficients calculated using the optimal wavelengths were used to create chemical images to show the distribution of protein and water contents in the ham samples. In the resulting chemical images, pixels having similar spectral features give the same predicted value of protein or water and will be visualized in a similar color; whereas pixels having different spectral patterns will exhibit different values of protein or water contents and will be visualized indifferent colors. Figure 6 shows an example of the resulting chemical images of water and protein contents. In these figures the changes in protein or water contents were assigned to a linear color scale and the numbers below each sample are the average protein or water content predicted (using the PLSR models) in the whole slice of ham. Although it was impossible to know the distribution of protein or water contents within the ham slices by the simple visual inspection (by either naked eyes or RGB camera), it was quite easy using the final chemical images to discern the change in these attributes from point to point in the same sample or even among ham samples as shown in Figure 6b and 6c. The contents of protein and water vary among different parts of the ham

351 slices and among hams. Differences in protein content and water distribution
352 observed among categories of ham are mainly due to differences in the quality and
353 integrity of the ham muscles, as well as the amount and composition of the brine
354 injected. In general, products of higher quality are made with a low level of injected
355 brine (Carisaghi et al., 2007) whereas an increase in the amount of retained water in
356 low quality cooked products is achieved using higher amounts of brine solution.
357 Thus, compared with those hams that have been injected with higher amounts of
358 brine solution, extra category cooked hams show lower levels of water and,
359 subsequently, higher protein content than first and second category cooked hams.
360 Regarding the differences in protein and water distributions observed in Figure 6
361 within each slice of cooked ham, it is important to consider that raw ham is a
362 multiphase system with a hierarchic arrangement of protein fibers from different ham
363 muscles in its structure. That structure is different in each ham, which makes it very
364 difficult for the equal distribution of the brine solution through the ham (Toldrá et al.,
365 2010). Samples might have been homogenized for the chemical analysis as these
366 results are in accordance with previously published analysis of homogenized samples
367 of cooked ham (Jiménez et al, 1980; Del Campo et al, 1998) although in this study
368 several samples from each slice were taken by separate with the aim to imitate the
369 on-line procedure used in the processing lines.
370 As declared in Figure 6, H4 had lower protein content (15.69%) and higher water
371 content (75.78%); meanwhile H1 and H2 exhibited the highest values of protein
372 (20.23% and 19.64%, respectively) and the lowest values of water (74.65% and
373 74.70%, respectively) than H3 that has intermediate protein and water values
374 (16.87% and 74.91%, respectively) between the extra category hams (H1 and H2)

and the low category ham (H4). Regarding the distribution of protein and water contents within the ham slice, H4 shows the highest homogeneity compared to the rest of the hams. This could be due to the use of smaller pieces of deboned ham legs for the processing as well as to the higher amounts of brine injection that are used in low quality cooked products and the important denaturation of the proteins during cooking (Cheng and Sun, 2006) that results in the welding of the muscles and the loss of their integrity (Toldrá et al., 2010).

Table 3 shows the average and the standard deviation values of the protein and water predicted values calculated for all ham analyzed slices. The analysis of variance revealed no significant differences ($p>0.01$) in the predicted values of protein and/or water between H1 and H2 extra category hams using the PLSR. Significant differences ($p<0.01$) in the water/protein ratio were observed between extra category (H1 and H2), first category (H3) and second category (H4) hams. These values are in agreement with those obtained using traditional analysis (Table 1). In fact, the analysis of variance of protein and water main values obtained using both methodologies showed no significant differences ($p>0.01$) between the two sets of data.

3.5.Classification

The result of the PLS-DA model, using the whole spectral range consisting of 237 wavelengths is presented in Table 4. It can be seen that the validation test gave similar result as the calibration set, and present low values of RMSEC and RMSECV and high R^2 indicating good performance of the model for ham classification. The weighted β-coefficients resulting from the PLS-DA model (Figure 7) were used for identifying the optimal wavelengths responsible for discrimination among ham

399 quality classes. These six wavelengths (966, 1061, 1148, 1256, 1373 and 1628 nm)
400 are close to some of the optimal wavelengths obtained for the protein and water
401 prediction by the PLSR models and indicate that ham classification by PLS-DA
402 could be associated with the protein and water content of ham slices.
403 Once the optimal wavelengths were selected, the spectral dataset was reduced from
404 237 wavelengths to 6 wavelengths and a new PLS-DA was developed on the
405 reflectance spectral data using only the optimum wavelengths instead of the full
406 spectral range (237 wavelengths) (Table 4). Figure 8 shows the score plot of the first
407 and the second principal components of the optimized PLS-DA using the 6 optimal
408 wavelengths for all spectra of the tested hams. The first two principal components
409 (PCs) resulting from PLS-DA, which explained 98.96% (94.81% for the first PC and
410 4.15% for the second PC) of the total variance among the samples, can be used to
411 classify the hams. The first principal component evidently separates hams samples
412 into two main groups, H1 and H2 (extra category) in the negative part of PC1 and H3
413 and H4 (first and second category) in the positive part of PC1. In addition the second
414 principal component clearly separates the ham samples into two groups, H1 in the
415 negative part of PC2 and the other three hams (H2, H3 and H4) in the positive part of
416 PC2. In general, the tested ham samples can be obviously distinguished from each
417 other in the principal component space. These results suggested that PC1 could be
418 related with the water and protein content whereas PC2 could be related with the fat
419 content of the hams, indicating that it is possible to classify the Spanish cooked ham
420 on the basis of the water, protein and fat contents.
421

422 **4. CONCLUSIONS**

Chemical images obtained from the experiments explain the robustness of the PLSR models and the validity of this technique to visualize the water and protein distribution in the ham slices. Regarding fat composition, a bad performance of the predicting model was obtained probably due to the narrow range of fat contents observed in the cooked ham samples. The PLS-DA model developed using some optimal wavelengths (966, 1061, 1148, 1256, 1373 and 1628 nm) was successfully used to classify the examined hams to different quality categories. This study demonstrated the potential capability of NIR hyperspectral imaging as an objective, rapid, and non-destructive technique for evaluating the physicochemical properties of meat products as well as the authentication and classification of these products. With some simple modification and using the most informative wavelengths, this technique could be adapted in the ham industry for quality evaluation during processing or for quality control and quality assurance programs.

5. ACKNOWLEDGEMENT

Author Pau Talens acknowledges the Spanish Ministry of Education for the financial support of his fellowship to do a period abroad at University College Dublin, National University of Ireland (Orden EDU/3378/2010 de 21 de diciembre).

6. REFERENCES

AOAC, 1999a. Official Methods of Analysis of AOAC International, 16th ed., 4, 13-14, Association of Official Analytical Chemists, Gaithersburg, MD.

AOAC, 1999b. Official Method of Analysis of AOAC International, 16th ed., 4, 18-19, Association of Official Analytical Chemists, Gaithersburg, MD.

447 Abdi, H. (2010). Partial least squares regression and projection on latent structure
448 regression (PLS Regression). *Wiley Interdisciplinary Reviews: Computational*
449 *Statistics*, 2: 97–106.
450 Alomar, D., Gallo, C., Castañeda, M., & Fuchslocher, R. (2003). Chemical and
451 discriminant analysis of bovine meat by near infrared reflectance spectroscopy
452 (NIRS). *Meat Science*, 63 (4) 441-450.
453 Andrés, S., Silva, A., Soares-Pereira, A.L., Martins, C., Bruno-Soares, A.M., &
454 Murray, I. (2008). The use of visible and near infrared reflectance spectroscopy to
455 predict beef M. Longissimus thoracis et lumborum quality attributes. *Meat Science*,
456 78, 217–224.
457 Barbin, D., ElMasry, G.,Sun, D.-W., & Allen, P. (2012). Near-infrared hyperspectral
458 imaging for grading and classification of pork. *Meat Science*, 90 (1), 259-268
459 Barlocco, N., Vadell, A., Ballesteros, F., Galietta, G., & Cozzolino, D. (2006).
460 Predicting intramuscular fat, moisture and Warner-Bratzler shear force in pork
461 muscle using near infrared reflectance spectroscopy. *Animal Science*, 82 (1), 111-
462 116.
463 Berrueta, L. A., Alonso-Salces, R. M., & Heberger, K. (2007). Supervised pattern
464 recognition in food analysis. *Journal of Chromatography A*, 1158, 196–214.
465 BOE Num. 159 Orden de 29 de junio de 1983 por las que se aprueban las normas de
466 calidad para jamón cocido y fiambre de jamón, paleta cocida y fiambre de paleta y
467 magro de cerdo cocido y fiambre de magro de cerdo.
468 Bostian, M.L., Fish, D.L., Webb, N.B., & Arey, J.J. (1985). Automated methods for
469 determination of fat and moisture in meat and poultry products: Collaborative study.
470 *Journal of the Association of Official Analysis of Chemistry,* 68 (5), 876-881.

471 Brimmer, P.J., & Hall, J.W. (2001). Method development and implementation of
472 near-infrared spectroscopy in industrial manufacturing support laboratories. In:
473 Williams, P.K., Norris, K. (Eds.), Near-Infrared Technology in the Agricultural and
474 Food Industries, second ed. American Association of Cereal Chemists Inc., St. Paul,
475 Minnesota, USA, pp. 187–198.
476 Casiraghi, E., Alamprese C., & Pompei, C. (2007) Cooked ham classification on the
477 basis of brine injection level and pork breeding country. LWT Food Science and
478 Technology, 40, 164-169.
479 Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance
480 spectroscopy in determination of food quality. *Trends in Food Science & Technology*
481 18, 72–83.
482 Chen, D., Shao, X., Hu, B., & Su, Q. (2005). Simultaneous wavelength selection and
483 outlier detection in multivariate regression of near-infrared spectra. *Analytical*
484 *Sciences*, 21, 161-166.
485 Cheng, Q., & Sun, D-W. (2006). Improving the quality of pork ham by pulsed
486 vacuum cooling in water. *Journal of Food Process Engineering*, 29, 119-133.
487 Cozzolino,D., & Murray, I. (2004). Identification of animal meat muscles by visible
488 and near infrared reflectance spectroscopy. *Lebensmittel Wissenschaft und*
489 *Technologie*, 37, 447-452.
490 Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011).
491 Advances in Machine Vision Applications for Automatic Inspection and Quality
492 Evaluation of Fruits and Vegetables. *Food and Bioprocess Technology*, 4 (4), 487-
493 504.

494 Del Campo, Gallego B., Berregi I. & Casado J.A. (1998) Creatinine, creatine and
495 protein in cooked meat products *Food Chemistry,* Vol. 63, No. 2, pp. 187-190.
496 ElMasry, G.,Wang, N., ElSayed, A., & Ngadi, M. (2007). Hyperspectral imaging for
497 nondestructive determination of some quality attributes for strawberry. *Journal of*
498 *Food Engineering,* 81, 98–107.
499 ElMasry, G., Sun, D.W., & Allen, P. (2011a). Non-destructive determination of
500 water-holding capacity in fresh beef by using NIR hyperspectral imaging. *Food*
501 *Research International*, 44 (9), 2624-2633
502 ElMasry, G., Iqbal, A., Sun, D.-W., Allen, P., & Ward, P. (2011b). Quality
503 classification of cooked, sliced turkey hams using NIR hyperspectral imaging
504 system. *Journal of Food Engineering*, 103 (3), 333-344.
505 ElMasry G., Barbin D., Sun D.W., & Allen, P. (2012). Meat Quality Evaluation by
506 Hyperspectral Imaging Technique: an overview. *Critical Reviews in Food Science*
507 *and Nutrition,* 52, 8, 689-711.
508 Esquerre, C., Gowen, A.A., O'Donnell , C.P., & Downey, G. (2009). Initial Studies
509 on the Quantitation of Bruise Damage and Freshness in Mushrooms Using Visible-
510 Near-Infrared Spectroscopy. *Journal of Agricultural and Food Chemistry*, 57, 1903–
511 1907.
512 Frentz, J.-C. (1982). Jambon cuit. In *L'Enciclopedie de la charcuterie*.Paris:
513 Sousanna.
514 Gaston E., Frias J.M., Cullen P.J., O'Donnell C.P., & Gowen A.A. (2010) Visible-
515 near infrared hyperspectral imaging for the identification and discrimination of
516 brown blotch disease on mushroom (Agaricus bisporus) caps. *Journal of Near*
517 *Infrared Spectroscopy,* 18 (5), 341–353.

518 Jiménez, R.; Huescar, J.; Yuste, T.; Barrio, J. del; León, F.; Luzón, F.; Reuvers, T.;
519 Ruiz, E. & Serrano, B. (1980) Composition of cooked ham prepared in Spain.
520 *Boletin del Centro Nacional de Alimentacion y Nutricion* (3/4): 12-17
521 Kamruzzaman, M., ElMasry, G., Sun, D.W., & Allen, P. (2011). Application of NIR
522 hyperspectral imaging for discrimination of lamb muscles. *Journal of Food*
523 *Engineering*, 104, (3), 332-340.
524 Karimi, Y., Maftoonazad, N., Ramaswamy, H.S., Prasher S.O., & Marcotte, M.
525 (2009). Application of hyperspectral Technique for Color Classification Avocados
526 Subjected to Different Treatments. *Food and Bioprocess Technology*, 1-13.
527 Leroy, B., Lambotte, S., Dotreppe, O., Lecocq, H., Istasse L., & Clinquart, A.
528 (2004). Prediction of technological and organoleptic properties of beef *Longissimus*
529 *thoracis* from near-infrared reflectance and transmission spectra. *Meat Science*, 66,
530 45-54.
531 Liu, L., Ngadi, M.O., Prasher, S.O., & Gariépy, C. (2010). Categorization of pork
532 quality using Gabor filter-based hyperspectral imaging technology. *Journal of Food*
533 *Engineering*, 99 (3), 284-293.
534 MARM (2010). Consumo Alimentario en España, Ministerio de Medio Ambiente, y
535 Medio Rural y Marino e Instituto Cerda, Madrid, www.marm.es
536 Osborne, B. G., Fearn, T., & Hindle, P. H. (1993). Near Infrared Spectroscopy in
537 Food Analysis. Harlow, Essex, UK: Longman Scientific and Technical.
538 Prats-Montalban, J. M., Ferrer, A., Malo, J.L., & Gorbena, J. (2006). A comparison
539 of different discriminant analysis techniques in a steel industry welding process,
540 *Chemometrics and Intelligent Laboratory Systems*, 80 (1), 109-119.

Prevolnik, M., Čandek-Potokar, M., & Škorjanc, D. (2004). Ability of NIR spectroscopy to predict meat chemical composition and quality–a review. Czech Journal Animal Science, 49, (11) 500-510.

Prieto, N., Andrés, S., Giraldez, F.J., Mantecon, A.R., & Lavin, P. (2008). Ability of near infrared reflectance spectroscopy (NIRS) to estimate physical parameters of adult steers (oxen) and young cattle meat samples. *Meat Science*, 79, 692–699.

Prieto, N., Roehe, R., Lavin, P., Batten, G., & Andres, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Science,* 83, 175–186.

Qiao, J., Wang, N., Ngadi, M.O., Gunenc, A., Monroy, M., Gariépy, C., & Prasher, S.O. (2007). Prediction of drip-loss, pH, and color for pork using a hyperspectral imaging technique. *Meat Science*, 76 (1), 1-8.

Rajkumar, P., Wang, N., ElMasry, G., Raghavan, G.S.V., & Gariepy, Y. (2012). Studies on Banana Fruit Quality and Maturity Stages using Hyperspectral Imaging. *Journal of Food Engineering*, 108 (1), 194-200.

Sweeney, R.A., & Rexroad, P.R. (1987). Comparison of Leco FP-228 'Nitrogen determinator' with AOAC copper catalyst Kjeldahl method for crude protein. *Journal of the Association of Official Analysis of Chemistry,* 70 (6), 1028-1030.

Toldrá, F., Mora, L., & Flores, M. (2010). Cooked Ham, Handbook of Meat Processing. F. Toldrá (Ed) Wiley-Blackwell Ch. 16, 301-311.

Figure 1

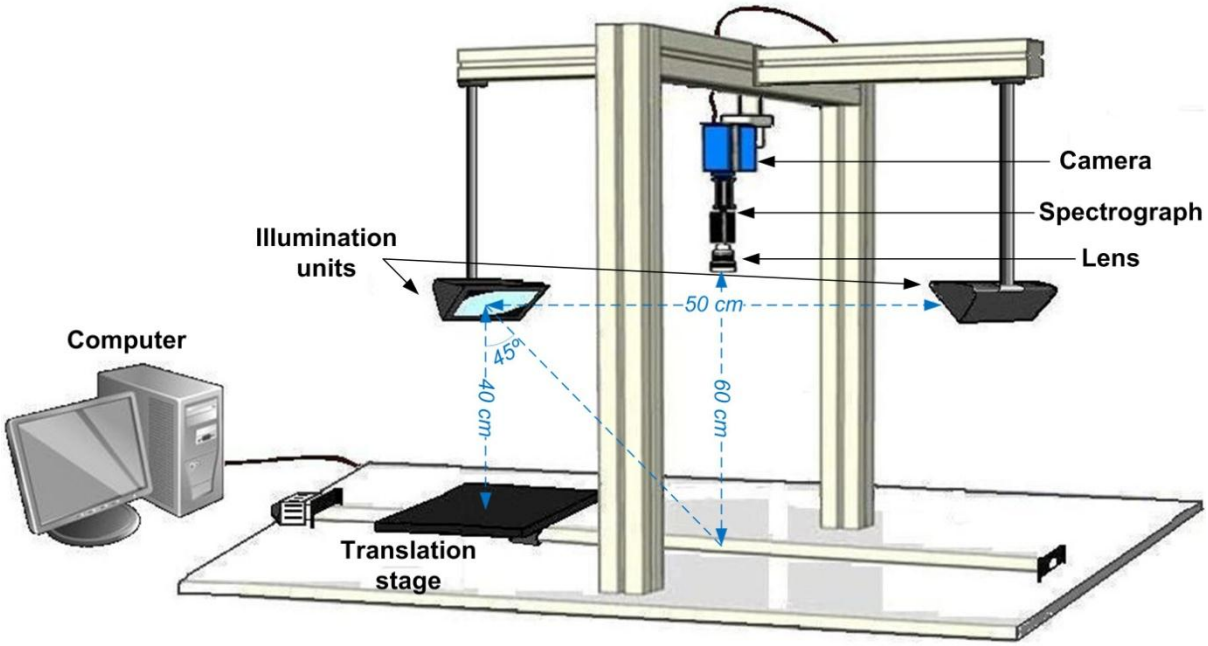

Camera
Spectrograph
Lens
Illumination units
50 cm
45°
40 cm
60 cm
Computer
Translation stage

Figure 2

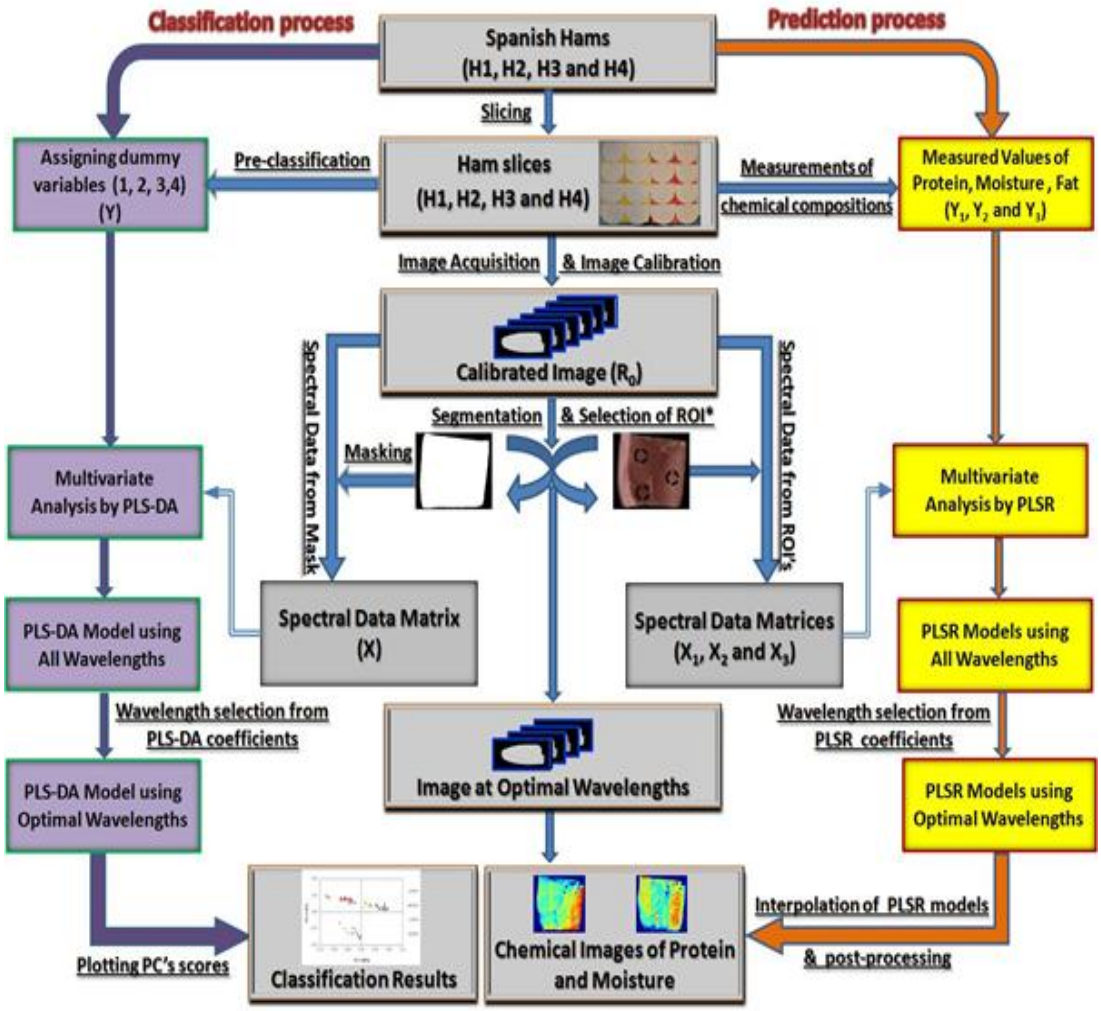
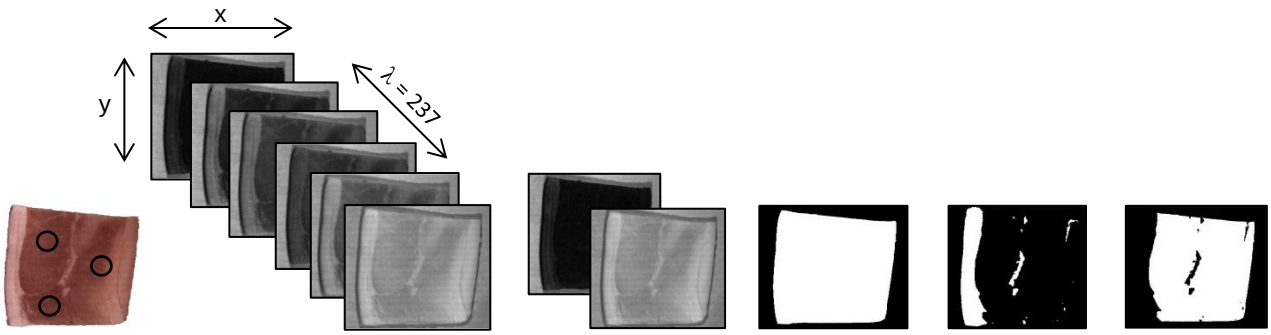

Classification process
Spanish Hams
(H1, H2, H3 and H4)
Prediction process
Slicing
Assigning dummy variables (1, 2, 3,4) (Y)
Pre-classification
Ham slices
(H1, H2, H3 and H4)
Measurements of chemical compositions
Measured Values of Protein, Moisture , Fat (Y_1, Y_2 and Y_3)
Image Acquisition
& Image Calibration
Calibrated Image (R_0)
Spectral Data from Mask
Segmentation
& Selection of ROI*
Masking
Spectral Data from ROI's
Multivariate Analysis by PLS-DA
Multivariate Analysis by PLSR
PLS-DA Model using All Wavelengths
Spectral Data Matrix (X)
Spectral Data Matrices (X_1, X_2 and X_3)
PLSR Models using All Wavelengths
Wavelength selection from PLS-DA coefficients
Image at Optimal Wavelengths
Wavelength selection from PLSR coefficients
PLS-DA Model using Optimal Wavelengths
PLSR Models using Optimal Wavelengths
Plotting PC's scores
Classification Results
Chemical Images of Protein and Moisture
Interpolation of PLSR models
& post-processing

Figure 3.



(a) (b) (c) (d) (e) (f)

Figure 4.

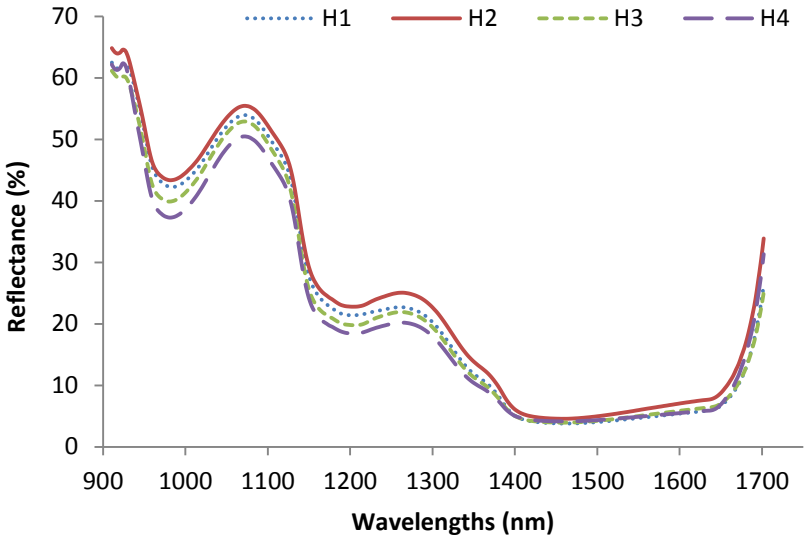

H1
H2
H3
H4
Reflectance (%)
70
60
50
40
30
20
10
0
900
1000
1100
1200
1300
1400
1500
1600
1700
Wavelengths (nm)

Figure 5.

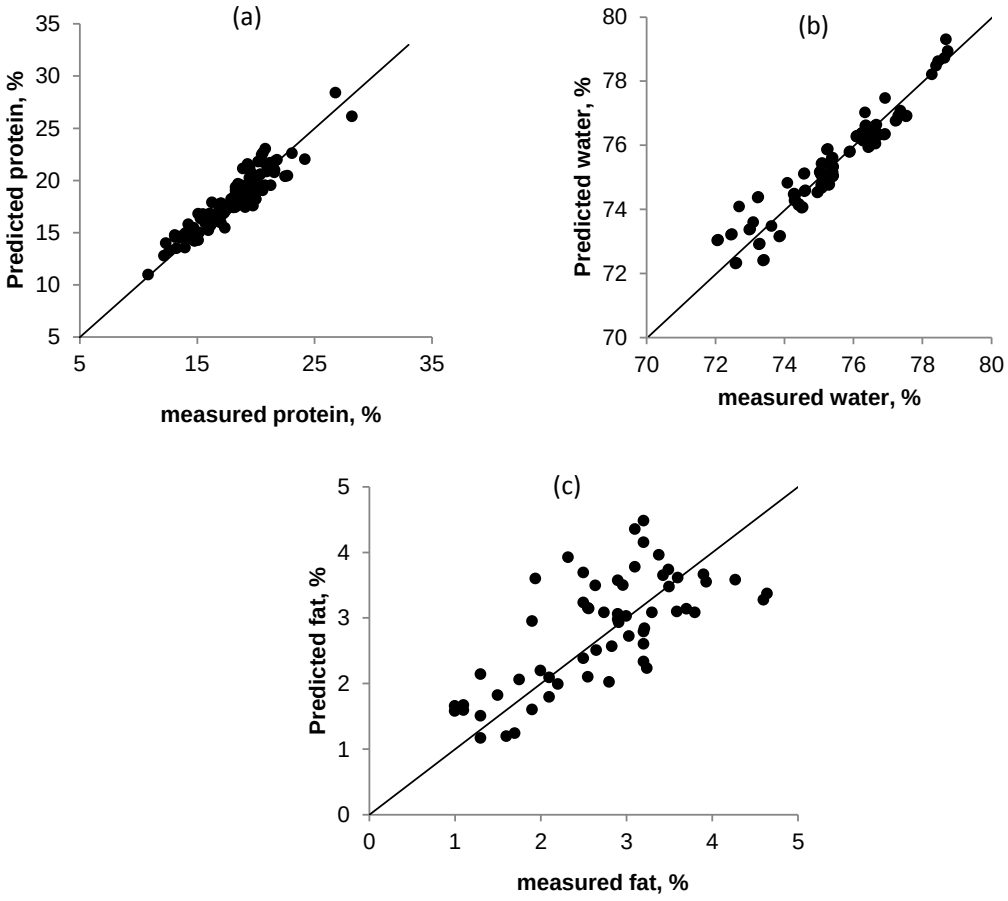

(a)
Predicted protein, %
measured protein, %
5
15
25
35
(b)
Predicted water, %
measured water, %
70
72
74
76
78
80
(c)
Predicted fat, %
measured fat, %
0
1
2
3
4
5

Figure 6

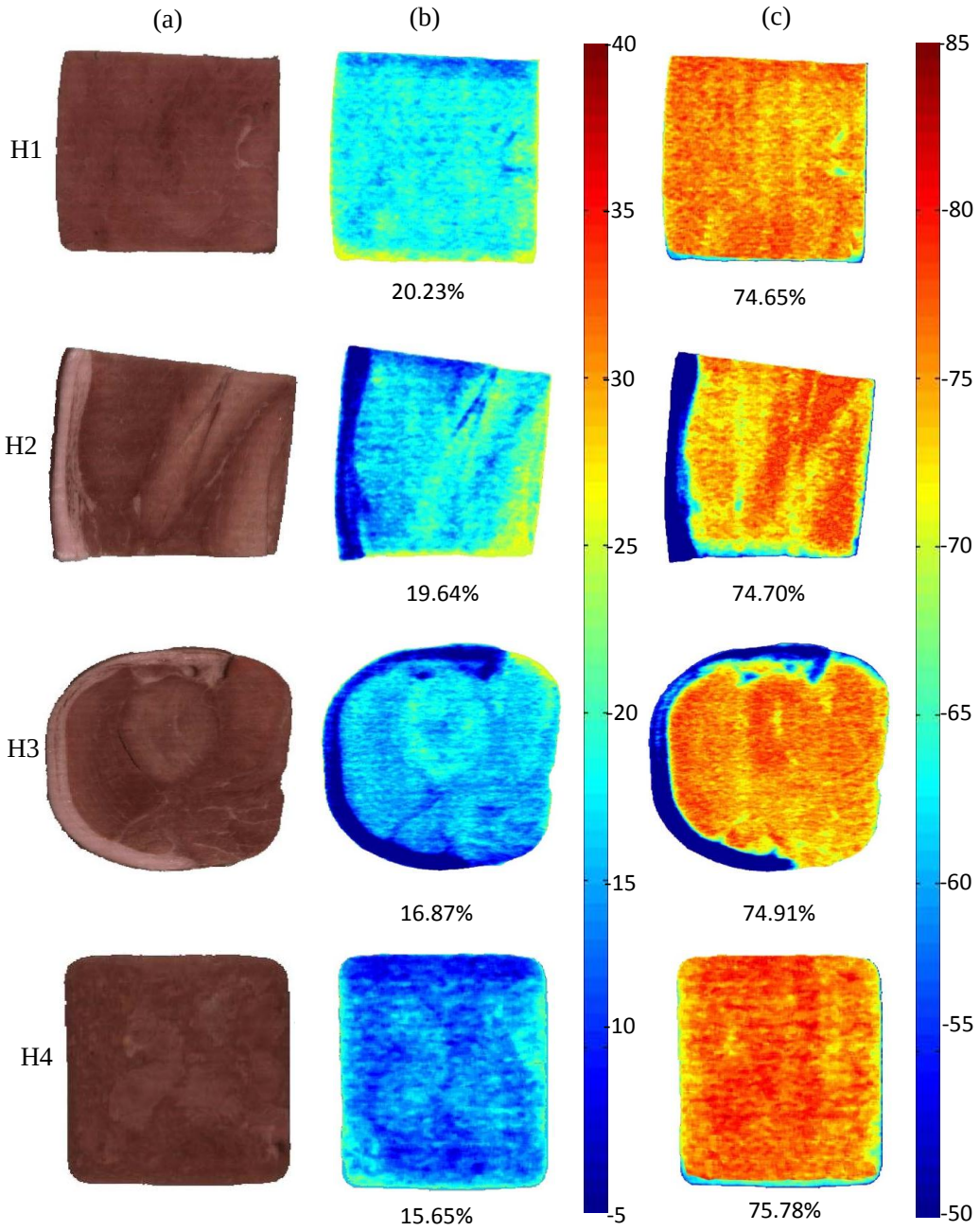

(a)
(b)
(c)
H1
H2
H3
H4
20.23%
74.65%
19.64%
74.70%
16.87%
74.91%
15.65%
75.78%
40
35
30
25
20
15
10
5
85
80
75
70
65
60
55
50

Figure 7

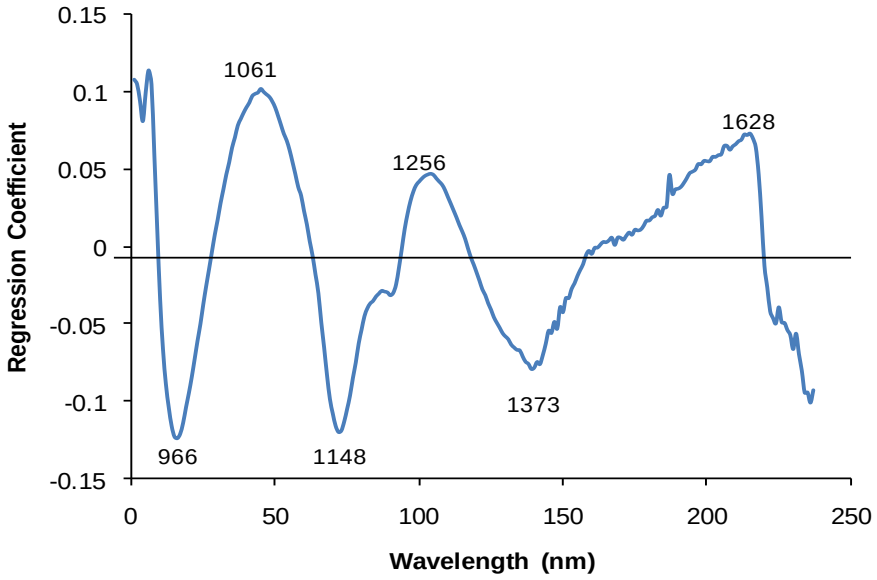

0.15
0.1
0.05
0
-0.05
-0.1
-0.15
Regression Coefficient
1061
1256
1628
966
1148
1373
0
50
100
150
200
250
Wavelength (nm)

Figure 8

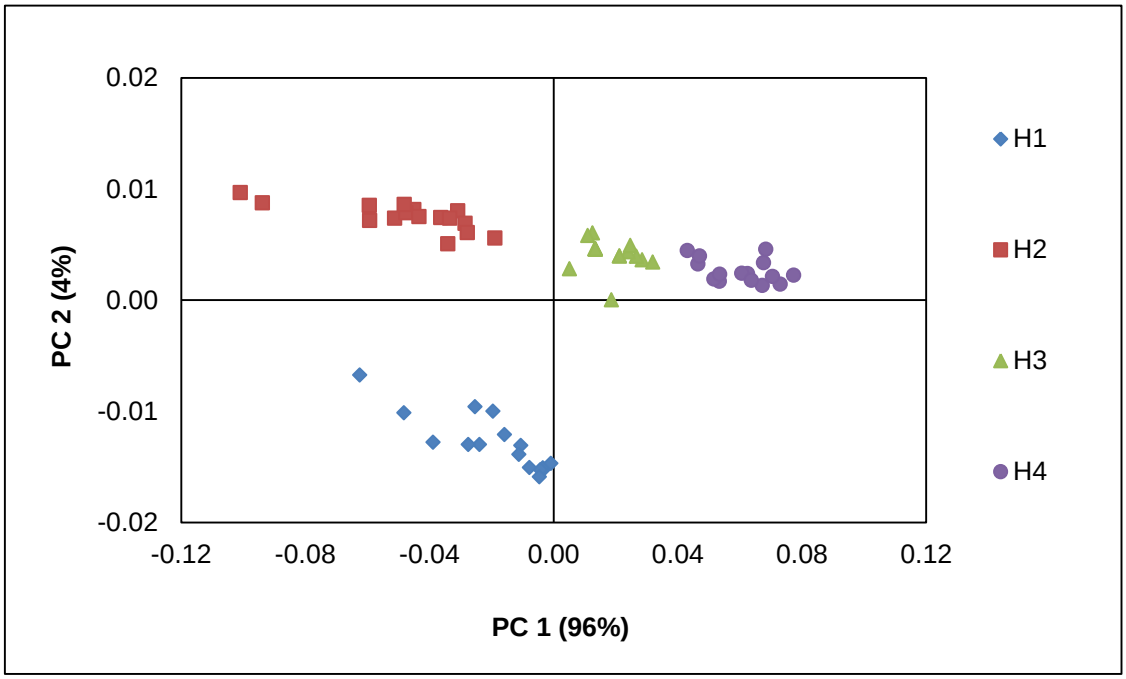

0.02
0.01
0.00
-0.01
-0.02
PC 2 (4%)
-0.12
-0.08
-0.04
0.00
0.04
0.08
0.12
PC 1 (96%)
H1
H2
H3
H4

Table 1. Measured water, protein and fat contents (mean ± standard deviation) of the examined commercial hams.

Ham	Protein (P, %) (n=126)	Water (W, %) (n=126)	W/P ratio	% Fat (n=63)
H1	$19.58^{a} \pm 0.74$	$75.33^{a} \pm 0.56$	$3.85^{a} \pm 0.14$	$1.62^{a} \pm 0.50$
H2	$19.99^{a} \pm 1.70$	$75.00^{a} \pm 0.66$	$3.75^{a} \pm 0.28$	$3.68^{b} \pm 0.93$
H3	$17.38^{b} \pm 0.96$	$75.47^{a} \pm 0.58$	$4.34^{b} \pm 0.26$	$3.05^{b} \pm 0.88$
H4	$15.15^{c} \pm 1.87$	$76.79^{b} \pm 0.53$	$5.07^{c} \pm 0.64$	$3.01^{b} \pm 0.58$

^{a,b,c} *Different letters in the same column mean significant differences ($p<0.01$).*

Table 2. Results of the PLSR models for predicting protein, water and fat contents in ham samples by using the full spectral range and the important wavelengths.

Attribute	#W	#LV	Calibration		Validation	
			R^2	RMSEC	R^2	RMSECV
Protein (n=126)	237	8	0.903	0.885	0.875	1.013
	10	8	0.877	0.994	0.855	1.09
Water (n=126)	237	6	0.947	0.376	0.925	0.456
	6	5	0.891	0.540	0.868	0.602
Fat (n=126)	237	7	0.607	0.662	0.396	0.834

#w: Number of Wavelengths, #LV: Number of Latent Variables, RMSEC: root-mean square error of calibration, RMSECV: root-mean square error estimated by cross-validation, R^2: coefficient of determination between the predicted and measured values.

Table 3. Predicted values for protein and water (mean ± standard deviation) using the PLS model.

Ham	Protein (P, %) (n=126)	Water (W, %) (n=126)	W/P ratio
H1	$20.30^{a} \pm 0.63$	$74.57^{a} \pm 0.42$	$3.67^{a} \pm 0.10$
H2	$20.16^{a} \pm 0.74$	$74.54^{a} \pm 0.36$	$3.70^{a} \pm 0.14$
H3	$16.95^{b} \pm 0.64$	$75.06^{a} \pm 0.34$	$4.51^{b} \pm 0.16$
H4	$15.66^{c} \pm 1.06$	$76.01^{b} \pm 0.28$	$4.85^{c} \pm 0.32$

[a,b,c] Different letters in the same column mean significant differences ($p<0.01$).

Table 4. Results of the partial least square discriminant analysis PLS-DA model for classification of ham samples by using the full spectral range and the important wavelengths.

			Calibration		Validation	
	#W	#LV	R^2	RMSEC	R^2	RMSECV
Classification (n=63)	237	5	0.973	0.176	0.966	0.198
	6	3	0.956	0.216	0.952	0.237

#W: Number of Wavelengths, #LV: Number of latent variables, RMSEC: root-mean square error of calibration, RMSECV: root-mean square error estimated by cross-validation, R^2: coefficient of determination between the predicted and assigned dummy values.