Selection of Representative Hyperspectral Data and Image Pretreatment for Model Development in Heterogeneous Samples: A Case Study in Sliced Dry-Cured Ham

Gamal ElMasry*[1,2,3] Elena Fulladosa[1] Josep Comaposada[1] Salim Al-Rejaie[2] Pere Gou[1]

[1] *Institute of Agriculture and Food Research and Technology (IRTA), Finca Camps i Armet s/n, 17121, Monells, Spain*

[2] *Suez Canal University, Faculty of Agriculture, Agricultural Engineering Department, Ismailia, PO Box 41522, Egypt*

[3] *Department of Pharmacology & Toxicology, College of Pharmacy, King Saud University, Riyadh, 11564, Saudi Arabia*

Abstract

Sliced dry-cured ham arranged in ready-to-eat packages is a convenient and widely consumed commodity characterised by heterogeneity in composition not only among different industrial batches but also through their horizontal and vertical profiles, making precise nutrition labelling of the packages a difficult task. Hyperspectral imaging techniques can serve as a steadfast solution not only to predict the overall composition of the major constituents of dry-cured ham but also to visualise their distributions. The main aim of this study was to define the optimal protocol for pretreating hyperspectral images and selecting representative hyperspectral data for developing accurate predictive models in excessively heterogeneous samples, using sliced dry-cured ham as a case study. Hyperspectral images (400-1000 nm) were acquired for heterogeneous sliced dry-cured ham and homogeneous unsliced dry-cured muscles. Partial least squares (PLS) regression models to predict fat, water, salt and protein contents were developed and tested in an independent dataset. The PLS predictive models developed from the whole surface of sliced dry-cured ham were the most accurate ones for predicting fat, water, salt and protein contents with a determination coefficient in prediction (R_p^2) of 0.89, 0.85, 83 and 0.63 and standard error in prediction (SEP) of 1.43, 1.21, 0.51 and 1.57%, respectively. The chemical images resulting from the models gave advantages of hyperspectral imaging technique over traditional chemical methods to visualise the spatial distribution of different constituents within the packaged ham slices.

Keywords: Chemical imaging, dry-cured ham, hyperspectral imaging, multivariate analysis, PLS, ROI

* Corresponding author. Tel: +34 666293308, Fax: +34 972630980, E-mail: gamal.masry@irta.es

1. INTRODUCTION

Information provided by food labels is sometimes not precise because it only specifies batch nutritional composition instead of the composition of each individual package. Where the composition differs due to raw material variations (*e.g.* water, fat and protein contents) or processing conditions (*e.g.* salting, drying, thermal treatment and ageing), the nutritional facts specified in the label of individual packages may exceed the error limits established by the European Union (Kaur, Rayner & Heike, 2016). The traditional methods used for the characterisation of food products depend basically on ordinary wet-chemistry assays in well-equipped laboratories. Besides being destructive, costly and slow, these methods entail complex multi-phased procedures, require certain dangerous substances, and employ experienced personnel. Besides, the ordinary techniques give only one average value over the whole tested product without providing a measurement for every single portion in the product. Therefore, it becomes a great challenge to develop reliable, economic, fast, and environment friendly tools to overcome the limitations and disadvantages of these methods.

Optical methods have gained popularity and become good candidates and viable options to be implemented for on-line applications in the food industry (Kamruzzaman, ElMasry & Nakauchi, 2015). Considerable amount of research endeavours have been directed in the past two decades towards using optical techniques and spectral imaging methods for different quality assessment scenarios in food science and technology, such as reagent-less determination of chemical composition of raw and processed meat (Craigie et al., 2017; ElMasry, Sun & Allen, 2013; Reis et al., 2018; Velásquez, Cruz-Tirado, Siche & Quevedo, 2017), seed quality evaluation (Dumont et al., 2015; ElMasry, Mandour, Al-Rejaie, Belin & Rousseau, 2019a; ElMasry et al., 2019b; Wakholi et al., 2018), quality estimation of fruits and vegetables (ElMasry, Wang, ElSayed & Ngadi, 2007; ElMasry, Wang, Vigneault, Qiao & ElSayed, 2008; Pathmanaban, Gnanavel & Anandan, 2019), determination of food safety, authentication and microbiological evaluation (Barbin, ElMasry, Sun, Allen & Morsy, 2013; Crichton et al., 2017; ElMasry & Sun, 2010; Foca et al., 2016; Siripatrawan, 2018).

Dry-cured ham as a convenient and widely consumed meat product in many countries with special organoleptic characteristics has received considerable interest in research by using different spectral imaging modalities for predicting the concentration and distribution of essential constituents (Garrido-Novell, Garrido-Varo, Pérez-Marín, Guerrero-Ginel & Kim, 2015; Gou, Santos-Garcés, Høy, Wold, Liland & Fulladosa, 2013; Liu, Qu, Sun, Pu & Zeng, 2013; Pérez-Santaescolástica et al., 2019). Generally, the characteristics of the raw meat, salt

content, amount and distribution of intramuscular fat (marbling) and dryness of the final dry-cured hams have tremendous effect on the consumer preferences (Hersleth, Lengard, Verbeke, Guerrero & Næs, 2011; Resano, Sanjuán, Cilla, Roncalés & Albisu, 2010). In this sense, the recent research in examining dry-cured hams using hyperspectral imaging focused on monitoring the fundamental salting stages during its production as well as analysing the chemical composition during processing to correct any possible defects in the final product. Thus, developing a technology that explores the overall composition of the commercially packaged sliced ham to provide consumers with verified nutritional facts of the product is of great interest for the ham industry.

The analysis of the spectral information residing in the hyperspectral images is often not an easy task and usually requires specific mathematical and statistical modelling for accurate estimation of the attributes in interest. Thus, for handling the complex spectral data with collinearity phenomenon among their variables (wavelengths), it is quite important to develop chemometric calibration models that best fit such spectral data of the samples being analysed with their reference chemical measurements (Garrido-Novell et al., 2015). However, the developed models might not be general and must be adjusted to new samples because the model probably will not work if there is a difference between the samples used in developing the models and those used in testing or validating the models (Alander, Bochko, Martinkauppi, Saranwong & Mantere, 2013). The same problem may be encountered if the regions from which the spectral data were extracted were biased or not representative of the sample under investigation. In general, to employ spectral techniques for routine quality evaluation scenarios, it is not enough to develop the calibration models because the techniques should be also optimised by considering some other factors that affect the quality of the calibration models and their reliability such as atmospheric conditions, sample geometry and the way by which the spectral data were extracted from the images.

One of the challenging tasks in processing multi-dimensional hyperspectral images with high spectral and spatial resolutions is to extract useful information from the vast amount of data volume of numerous spectral bands (ElMasry et al., 2007). The quality of the acquired hyperspectral images, the way of extracting information from hyperspectral images, and its transformation into a useful representation, enables the description of intrinsic characteristics of the sample in the scene by relevant calibration models (Khan, Thomas & Hardeberg, 2017). When the sample is heterogeneous (*e.g.* non-uniform meat cuts having different muscles), a reliable method for selecting representative regions from the sample is a critical

step before building such calibration models. Because the reference chemical measurements are not available in each individual pixel in the image, one global reference value for the whole sample is usually used to represent the content of the whole sample (Garrido-Novell et al., 2015). Thus, one major problem reported in all attempts at pixel-based prediction using models developed from hyperspectral images is posed by constructing models from individual spectra extracted from unrepresentative regions with the attendant risk of artificiality or overfitting.

In case of packaged dry-cured ham slices, the problem is manifold because the ham slices contain different muscles (horizontal heterogeneity) with different composition of water, fat, protein and salt contents (Arnau, Guerrero, Casademont & Gou, 1995; Boadas, Gou, Valero & Arnau, 2001; Gou et al., 2013). In addition, for marketing reasons, the slices themselves are arranged inside the package in a telescopic form where the slices are slightly shifted from the slices underneath (vertical heterogeneity). This means that muscle portions appeared in the first (upper) slice are not necessarily corresponding to the same muscle portions in the rest of the slices in the package. This poses a great problem during selecting representative regions from the hyperspectral images of the dry-cured ham slices because a region of interest that appears representative in the first slice is entirely different from the rest of the slices. Previous studies conducted on hyperspectral experiments were performed either on completely homogeneous ham slices (ElMasry, Iqbal, Sun, Allen & Ward, 2011; Talens, Mora, Morsy, Barbin, ElMasry & Sun, 2013) or on thick horizontally heterogeneous ham samples (Garrido-Novell et al., 2015; Gou et al., 2013). In such studies, spectral data were extracted from the whole surface of the samples. However, when the samples are heterogeneous in composition in both horizontal and vertical profiles, other spectral selection options should be examined.

Thus, the main aim of this study was to define the optimal protocol for pretreating hyperspectral images and selecting hyperspectral data required for developing accurate predictive models in excessively heterogeneous samples using sliced dry-cured ham as a case study. Three different options were used for selecting regions of interest (ROI) from the dry-cured ham samples: (1) from the whole packaged slices (WholeROI), (2) from a small representative area of the packaged ham slices (SmallROI) and (3) from unsliced dry-cured muscles (MuscleROI). The WholeROI is heterogeneous in both horizontal and vertical profiles (Fully heterogeneous ROI), SmallROI is heterogeneous in vertical profile (semi-homogeneous ROI) and the MuscleROI is completely homogeneous in the horizontal and

vertical profiles (Fully homogeneous ROI). The proposed strategy brings a comprehensive output for deciding the best protocol for analysing the packaged sliced hams that provides the highest and balanced performance. The possibility of using spectral data extracted from either semi-homogeneous portions (SmallROI) or unsliced fully-homogeneous cured muscles (MuscleROI) to build calibration models to predict the composition of the whole package of the sliced ham was also investigated.

2. MATERIALS AND METHODS

2.1. Dry-cured ham samples

A total of 267 commercial packages of sliced dry-cured ham of approximately 250 g each were obtained from different producers in Spain. Inside each package, there were 12 slices of 2 mm in thickness arranged one above another in a telescopic form where the slices were slightly shifted from the slices underneath. Each slice contains *Semimembranous* (SM), *Semitendinosus* (ST) and *Biceps femoris* (BF) muscles (Muñoz, Gou & Fulladosa, 2019; Muñoz, Rubio-Celorio, Garcia-Gil, Guàrdia & Fulladosa, 2015). Moreover, additional 52 samples of 25 mm in thickness of fully unsliced dry-cured muscles (SM, ST and BF) were also collected from different ham producers.

2.2. Acquisition of hyperspectral images

Dry-cured ham was removed from its package, intimately wiped by using tissue paper to remove surficial water and residues and then immediately scanned as one unit in the hyperspectral imaging system in the reflectance sensing mode. By using a fully calibrated hyperspectral imaging system (Resonon Inc., Bozeman, MT, USA), a hyperspectral image was acquired for the whole content of the ham package (12 slices) at once without separating the slices from one another. It was important to scan the sample as soon as possible in order to ensure that the temperature did not significantly affect the sample during the acquisition. As the number of hydrogen bonds in the water molecules is temperature dependent, it is well documented (Büning-Pfaue, 2003; Maeda, Ozaki, Tanaka, Hayashi & Kojima, 1995) that water absorption bands in the NIR region may shift in both peak position and amplitude according to sample temperature. Thus, to maintain spectral consistency from the ham samples, the temperatures of the samples were kept around 10ºC during all image acquisitions.

The main configuration of the hyperspectral imaging system used in image acquisition is shown in **Fig. 1**. The system consisted of a camera, a spectrograph, a conveying platform, a computer supported with data acquisition software (SpectrononPro, Resonon Inc., Bozeman, MT, USA). The conveying platform set at a distance of 500 mm from the camera lens was driven by a stepping motor at a speed of 0.03 m s^{-1}. The field of view of the camera was illuminated by four 50 W tungsten-halogen lamps. Using this push-broom hyperspectral imaging system, the spectral images were collected line by line in a wavelength range of 400-1000 nm with 300 wavebands at 2 nm intervals in the spectral domain. The acquired hyperspectral image is a three-dimensional image called 'hypercube' having two spatial dimensions (*x, y*) and one spectral dimension (λ).

2.3. Pre-processing of hyperspectral images

To overcome the problem of spectral non-uniformity of the illumination and to make the acquired images independent from the spectral power distribution of the illumination source and from the spectral sensitivity of the camera sensor, all acquired images were dynamically corrected using two additional images. The first image is called a dark image (0% reflectance) taken when the light lamps were switched off, and the second image called a white image was taken for a white reference object of 80×200 mm made from Teflon (SpectrononPro, Resonon Inc., Bozeman, MT, USA) with reflectance values of 99.9% and using the following formula for correction:

$$R(\lambda) = \frac{R_0(\lambda) - R_{Dark}(\lambda)}{R_{White}(\lambda) - R_{Dark}(\lambda)} \quad (1)$$

where $R(\lambda)$, $R_0(\lambda)$, $R_{Dark}(\lambda)$and $R_{White}(\lambda)$ are the corrected, raw, dark and white images at wavelength λ, respectively.

Because poor image quality negatively affects not only the subsequent data processing steps but also building robust calibration models, noise and specular highlights that may appear in some spots of the acquired images were completely obviated. Because ham samples are inhomogeneous objects, the reflectance spectrum at any pixel in the image is a linear sum of diffuse and specular reflections (Koirala, Pant, Hauta-Kasari & Parkkinen, 2011; Shen & Zheng, 2013). The diffuse component shows the real reflectance related to the physicochemical properties of the ham sample in the scene. To keep only the diffuse reflection, it was necessary to apply a method to remove specular component from the acquired images (Akashi & Okatani, 2016). As the hyperspectral imaging system used in this study was not supported with a polarising filter either in front of the camera or at the light

source, a method based on dichromatic reflection model (Koirala et al., 2011) was utilised for wiping off the specular components from all pixels of the acquired hyperspectral images before spectral data being extracted. The dichromatic model applied for a hyperspectral image with Ω contiguous narrow bands centred at λ_1, λ_2....., λ_Ω, was used to calculate the reflectance spectrum $I(x)$ or the response of the camera receptor at a geometric pixel location (x) as described in equation 2:

$$I(x) = \alpha(x) \int_{\Omega} R_d(x,\lambda) E(\lambda) S(\lambda)\, d\lambda + \beta(x) R_s(x) \int_{\Omega} E(\lambda) S(\lambda)\, d\lambda \qquad (2)$$

where $I(x)$ is the reflectance vector of image intensity at a pixel (x) having a spatial coordinates of $x = \{x, y\}$ representing its 2D location. The factor α is a factor for diffuse reflection and β is the weighting specular factor. The terms $R_d(x,\lambda)$ and $R_S(x)$ represent the diffuse and specular reflectance values at a pixel position (x) and wavelength λ, E(λ) is the spectral power distribution (SPD) of the illuminant at a wavelength λ and S(λ) is the camera sensor sensitivity at a wavelength λ.

2.4. Extraction of hyperspectral data

Due to its capability to provide both spatial and spectral information, the hyperspectral imaging system enables the flexibly to extract spectra of any spatial locations (*i.e.* regions of interest) in the sample. Three different options were tested for sampling representative spectral data. The selection of different regions of interests (ROIs) from the image was a critical step because it subsequently affects the performance of the developed prediction models. The first option was to select the whole imaged sample as the main region of interest from which spectral data were collected (Fully heterogeneous ROI). For isolating the whole sample from the background (the conveying platform), two grayscale images at bands 550 nm and 690 nm were first picked up from the spectral space and then subtracted from each other. Because the reflectance values of the conveying platform were consistently stable at these particular bands as well as throughout the entire spectrum, the subtraction process lead to zero values at all pixels belonging to the conveying stage. Thus, the resulting grey image from the subtraction operation was segmented by a simple thresholding method to give a binary image in which the whole ham sample was appeared as white pixels (ones) representing the main region of interest (ROI) in a dark background (zeros). The resulting binary image was then treated by a median filter with a size of 5×5 pixels to fill all holes in the ROI. This extracted ROI called 'WholeROI' was acted as a mask from which spectral

signatures of all pixels within this area were collected and averaged to represent the spectral fingerprint of the whole ham sample.

Selecting a small representative portion of the ham sample as the main region of interest was the second option used in extracting spectral signatures of ham samples. Because the ham sample contains different muscles (SM, ST and BF) and each of them has its own uniform characteristics (Muñoz et al., 2019), this option depends on collecting spectral data from only one portion from each sample instead of the whole ham to guarantee horizontal homogeneity within the selected area (Semi-Homogeneous ROI). The selected region of interest in this option was manually selected and called 'SmallROI' from which the spectral signature was extracted as the average spectrum of all pixels within the selected region. To avoid the problem of vertical heterogeneity between both sides of the same sliced ham sample observed in the abovementioned two options (WholeROI and SmallROI), the third option depended on using individual unsliced dry-cured muscles (SM, ST and BF)) for extracting spectral signatures as a fully homogeneous ROI. The regions of interest from these unsliced muscles were individually collected and called 'MuscleROI'.

2.5. Reference measurements of chemical composition

A total of 319 samples were collected from different regions of interest (178 WholeROIs, 89 SmallROIs and 52 MuscleROIs) from the dry-cured ham samples. After image acquisition, each sample was individually minced in a mincing machine (La Picadoro, Moulinex, Spain) at its higher speed for 30 seconds, homogenised and its reference values of fat, water, salt and protein contents were determined in the laboratory. Total fat content was measured in duplicate by Soxtec extraction (Soxtec HT 6-1043 and Service Unit 1046). Water content was determined by the standard drying method; meanwhile protein and salt contents were determined by FoodScan spectral system (FOSS, Electric A/S, Hillerød, Denmark) in near infrared transmittance (NIT) mode by AOAC method 2007.04 (Anderson, 2007).

2.6. Data modelling

The average spectra of the ham samples extracted from each region of interest (WholeROI, SmallROI and MuscleROI) were arranged in three different matrixes (X1, X2 and X3) along with their reference measurements of the major chemical attributes. The rows in each data matrix of a specific region of interest (X1, X2 or X3) represent the number of samples involved and the columns represent the average reflectance magnitudes at 300 wavelengths (variables or predictors) in the range of 400-1000 nm. Each spectral data matrix (X) of the

ham samples and their corresponding reference chemical composition (Y) were concatenated and then divided randomly into two datasets: a training group with 2/3 of the initial data and a testing group with 1/3 of the initial data. To ensure fair partitioning of the data, t-test was carried for each group to ensure that there was no significant difference between the two groups in all examined chemical features. The main aim of data partitioning is to develop calibration models on the training group and then using such developed models in predicting the chemical compositions in the testing/prediction group.

Partial least square (PLS) regression models were developed on the training dataset under full cross validation routine. The PLS Regression is a preferable modeling method in case of great number of independent variables (300 wavelebands in this case) because it has mean advantage of combining features from principal component analysis (PCA) and linear regression. As the spectral data are very noisy and great collinearity exists among the 300 wavebands (predictors) involved in the test, PLS regression is suitable for this kind of data to predict the dependent variable (the chemical composition of the ham arranged in the vector Y) from the predictors (spectral data arranged in X). Thus, the PLS applying least square principle provides a solution to obtain regression coefficients of the predictors and by decomposing spectral data (X) and the reference chemical values (Y), the PLS modelling extracts a new set of orthogonal variables called principle components or latent factors (LFs) that have the best predictive power and removes noises from both of these matrices. The more LFs included in the model, the more complex the PLS model will be. Therefore, selecting the ideal number of latent factors in the model is critical for minimising the expected error and to avoid under-fitting and overfitting of the prediction process. Using a large number of latent factors may provide good performance in fitting the current attribute, but it usually leads to overfitting because the model considers significant amount of noise rather than the real spectral information. On the other hand, the under-fitting means the model does not have enough information for accurate prediction. In full cross validation using one-leave-out method, one sample was left out at a time and the PLS model was built for the remaining samples. The model was then used to predict the chemical attribute of the sample left out, and the same routine was repeated until all samples removed once.

For each chemical constituent, three different PLS regression models were developed for predicting this composition in the dry-cured ham samples, one PLS model for the data of each ROI. Besides the number of LFs used in building the model, PLS models were evaluated in terms of coefficient of determination in calibration (R_c^2), coefficient of the

289 determination in cross validation (R^2_{cv}), the root mean squared error of calibration (RMSEC)
290 and the root mean squared error under cross validation (RMSECV). Evaluating the model
291 based on RMSEC only is not advised because a portion of the noise in the reference values is
292 inadvertently modelled by the estimated parameters. Thus, applying cross-validation during
293 the development of the calibration model provides a better estimate of its predictive ability.
294 Moreover, the performance of the three developed PLS models were compared in predicting
295 the same chemical attributes in an independent dataset (the testing set) in which the models
296 were evaluated in terms of the coefficient of determination in prediction (R^2_p) and the
297 standard error of prediction (SEP). **Fig. 2** shows the full scheme followed in this study to
298 evaluate the three different models developed from different ROIs (WholeROI, SmallROI
299 and MuscleROI) in predicting the main chemical composition of the dry-cured hams.

300 All multivariate analyses for building and testing the calibration models were carried out by
301 using The Unscrambler v9.7 (CAMO Software AS, OSLO, Norway). The open-source
302 programming in Matlab® (The Mathworks Inc., Natick, Massachusetts, USA) was used to
303 develop in-house script for image correction, spectral data extraction and for all subsequent
304 processing regimes of the hyperspectral images.

305 3. RESULTS AND DISCUSSION

306 3.1. Characteristics of the selected regions of interests

307 As shown in **Fig. 3a**, the selected area within the fully heterogeneous WholeROI was a
308 mixture of lean and fat portions (*i.e.* horizontal heterogeneity). Also, when the sample was
309 flipped into the other side (**Fig. 3d**), its appearance was entirely different and the sample
310 looked full of fat portions compared to the upper side of the sample (*i.e.* vertical
311 homogeneity) meaning that the concentrations of the chemical constituents through the
312 vertical profile of the sample are not evenly distributed between the upper side, the sample
313 core and the bottom side of the sample. Because the amount of fat and its distribution are
314 different from both sides of the sample as shown in **Fig. 3d**, the spectral signatures of both
315 sides were totally different from each other (**Fig. 3e**). In essence, when the sample was
316 homogeneous in the vertical direction, its spectrum should not differ despite the imaged side
317 of the sample as long as the chemical composition of a ham sample does not change during
318 acquisition. This was not the case here of the fully heterogeneous WholeROI because the
319 reflectance magnitudes of the upper side of WholeROI were lower than those of the bottom
320 side of WholeROI because the second side had more fat that exhibited higher reflectance.

Because the tested samples are normally images from one side during inspection process, the spectrum of the upper side of WholeROI was used as the main spectral fingerprint of the whole ham sample that will be used for the subsequent model development.

As shown in **Fig. 3b**, the semi-homogeneous SmallROI seems homogeneous from the upper side; meanwhile the other side of the same region was extremely heterogeneous due to the presence of fat edges within the selected area (vertical heterogeneity). This was resulted from the telescopic arrangement of the ham slices in the package. Due to vertical heterogeneity between both sides of SmallROI, the spectrum of the upper side of SmallROI was also different from the spectrum of the other bottom of the same region (**Fig. 3f**). The difference of reflectance magnitudes between both sides of the SmallROI (**Fig. 3f**) was much higher than that observed in both sides of the WholeROI (**Fig. 3e**).

In case of fully homogeneous MuscleROI, both sides of the selected region were quite homogeneous (**Fig. 3c**) and exhibited similar spectral fingerprints as shown in **Fig. 3g**. The main variation in the average spectrum between the upper and bottom sides of the MuscleROI depends only on the relative amount and distribution of intramuscular fat and lean within the selected area. In this sense, it was assumed that the spectral signatures of the raw unsliced cured muscles could be used in estimating the major composition of the whole sliced ham package. However, the pattern of the spectral fingerprints of unsliced muscles (**Fig. 3g**) are substantially different from that of the sliced samples (**Fig. 3e** and **Fig. 3f**) in terms of reflectance magnitudes as well as the remarkable absorption peaks.

3.2. Chemical characterization of dry-cured ham samples

Table 1 shows the proximate composition from different regions of interest (WholeROI, SmallROI and MuscleROI) of dry-cured ham samples. As shown in Table 1, the salt content in the three different ROIs had consistent values in all ROIs with average values of 5.64 ± 0.94%, 5.24 ± 0.38% and 4.67 ± 0.50% for WholeROI, SmallROI and MuscleROI, respectively. For fat, water and protein contents in the tested ROIs, there was a significant difference ($p < 0.05$) in the values of these constituents among the extracted ROIs. The average fat contents for the three ROIs were 8.55 ± 3.43, 6.77 ± 3.60 and 7.04 ± 4.35% and the average water contents were 50.78 ± 3.43, 53.37 ± 3.64 and 48.09 ± 4.14% and the average protein contents were 32.65 ± 2.51, 30.45 ± 2.39 and 38.23 ± 5.82% for WholeROI, SmallROI and MuscleROI, respectively.

Generally, the composition estimates of the ham samples in this study are in agreement with the existing literature (Bou, Llauger, Arnau & Fulladosa, 2018; Parolari, Aguzzoni & Toscani, 2016) for standard composition of the dry-cured hams. In general, the wide range of fat, water and protein contents found in the examined samples implies the possibility of obtaining good prediction models of these attributes using the proposed multivariate analysis. Because it is more desirable to make interpolations rather than extrapolations when making predictions from a calibration model, the range of concentrations in the calibration samples should have a fairly uniform coverage across the range of interest.

Table 1 Proximate composition (%) of the major constituents of the dry-cured hams determined from different regions of interest (WholeROI, SmallROI and MuscleROI).

ROI	Attribute	Min	Max	Mean* ± SD
WholeROI $n = 178$	Fat	3.10	18.70	$8.55^{a} \pm 3.43$
	Water	41.35	57.58	$50.78^{a} \pm 3.43$
	Protein	26.25	38.59	$32.65^{a} \pm 2.51$
	Salt	3.95	5.98	$5.64^{a} \pm 0.94$
SmallROI $n = 89$	Fat	2.16	19.44	$6.77^{b} \pm 3.60$
	Water	40.98	59.06	$53.37^{b} \pm 3.64$
	Protein	25.28	37.22	$30.45^{b} \pm 2.39$
	Salt	3.99	6.24	$5.24^{b} \pm 0.38$
MuscleROI $n = 52$	Fat	1.48	20.52	$7.04^{b} \pm 4.35$
	Water	36.61	54.29	$48.09^{c} \pm 4.14$
	Protein	27.76	50.42	$38.23^{c} \pm 5.82$
	Salt	3.38	5.87	$4.67^{c} \pm 0.50$

** Different subscripted letters beside the mean value of a constituent indicate significant difference among the regions of interest ($p < 0.05$). SD: standard deviation.*

3.3. Effect of specularity correction

The quality of the acquired hyperspectral images, the way of extracting spectral fingerprints from the images and the methods of data modelling have substantial impacts on the outcomes of the subsequent data analyses. Because the hyperspectral images were acquired in the line-scan reflectance mode at numerous contiguous wavelengths within the visible and NIR regions (400-1000 nm) of the electromagnetic spectrum, the hyperspectral images came also with a phenomenon of specular highlights in which some zones in the acquired images exhibit extreme reflectance values due to the relative arrangement between the illumination units and the ham samples (Khan et al., 2017; Washburn, Stormo, Skjelvareid & Heia, 2017);. The assumption of ignoring this specularity problem (where it clearly exists) may reduce the robustness of the developed models (Khan et al., 2017). Therefore, it was extremely important to correct the acquired images for the specularity highlights. According to the process used for specularity correction routine implemented in this study, all pixels having extreme reflectance values in all wavelengths either in the ham sample itself or even in the background area (the conveying stage) were treated to exclude the specular component from the image and keep only the diffuse component. Consequently, all spectral data extracted from any regions of interest in the treated hyperspectral images will be specular-free and contain only the diffuse reflectance values. In fact, this step was substantially useful for the next processing step because extreme values can lead to inaccurate results, false segmentations, deceptive object measurements, recognition errors or even calculation overflow. **Fig.** 4 shows an example of a hyperspectral image with this correction step, and it also illustrates the resulting spectra for all pixels in the image after being corrected.

Because it is not possible to visualise the hyperspectral image in its current 3-D hypercube, a pseudo-colour image could be created to see the effect of the specular removal method on the overall appearance of the hyperspectral image. The pseudo-colour image could be built by gathering three different bands from the hyperspectral images across the spectrum to represent the red, green and blue channels. In the example shown in **Fig. 4**, the pseudo-colour image rendered from a hyperspectral image was formed by concatenating three bands at 640, 550 and 460 nm. In general, the specular regions appeared in the image are characterised by their maximum intensity along all wavebands in the spectrum compared to the other normal (diffuse) regions. Thus, specular zones appeared in the pseudo-colour image shown in **Fig. 4a** are characterised by extreme reflectance intensities. The correction process was able to identify those pixels and isolate them as shown in the binary image depicted in **Fig. 4b**. It can

be seen that the process of removing specularity highlighting zones from hyperspectral images helps in removing all extreme values of any pixels in the image and retain only the useful diffuse reflectance as shown in **Fig. 4c**. The difference in spectral fingerprints of all pixels in the image before and after specular correction could be visualised as shown in **Fig. 4d** and **Fig. 4e**, respectively. It is quite clear to notice that the specular problem appeared in some zones of the original image shown in **Fig. 4** (as indicated by the arrows) were completely remedied, leaving diffuse only reflections. Instead of deleting specular pixels from the acquired raw hyperspectral images or even treating them as outliers, all pixels with peculiar fingerprints were corrected by keeping only their diffuse reflectance values and get rid of their specular components as shown in **Fig. 4e**.

3.4. Spectral features of ham samples

The specular correction step resulted in the separation of the specular component at every single pixel position to get only the diffuse component. The corrected spectra of any individual pixels in the hyperspectral image could be illustrated as a plot between wavelengths and the corresponding magnitudes of reflectance at the spatial location of these pixels. Thus, **Fig. 5a** shows the spectral signatures of some individual pixels before and after specular correction. It is very obvious to notice that the applied specular correction process preserved the shape and spectral patterns of these pixels by keeping the location of the absorption bands in the spectrum without any spatial shift. The specular correction operation only reduces the reflectance magnitudes for those pixels that exhibited specularity due to removing the specular components from such pixels. Once the specular correction was performed over all pixels in the acquired hyperspectral image, the average spectrum of any group of pixels (regions of interested) could be easily extracted for further investigations. **Fig. 5b** shows the raw average diffuse spectra of intermuscular fat, intramuscular fat and lean portions as the major areas appeared in the hyperspectral image of the dry-cured hams. In addition, the average spectrum of the whole ham sample (intermuscular fat + intramuscular fat + lean pixels) was also illustrated in the same figure.

The average spectrum of all pixels belonging to one ham sample could be considered as the unique spectral signature of such a sample that depends basically on the physicochemical properties of this sample. The remarkable absorption bands noticed in all spectra are ascribed to some functional chemical groups due to bending and combination motions of different molecules (Morsy & Sun, 2013). In the visible region of the spectrum (400-680 nm), the

absorption bands are mostly related to different pigments presented in the sample. Thus, in all spectra illustrated in **Fig. 5b**, the sharp absorption band in the blue region of the spectrum at 420 nm was related to Soret absorption due to erythrocyte haemoglobin or deoxymyoglobin, and the bands at 550 and 580 nm were associated with myoglobin or oxymyoglobin species (Cozzolino, Murray, Scaife & Paterson, 2000; Ortiz-Somovilla, España-España, Gaitán-Jurado, Pérez-Aparicio & De Pedro-Sanz, 2007). Those bands in the visible range of the spectrum can be used efficiently for colour variation among ham samples. However, samples having the same colour cannot be discriminated only using the bands in the visible range only. In the NIR range, a weak band at 760 nm was from the third overtone of O-H vibration but it was very difficult to discern and an absorption band at 978 nm related to OH second stretching overtone is attributed to water content in the sample (Cozzolino, De Mattos & Martins, 2002; Ortiz-Somovilla et al., 2007). Therefore, this particular absorption band was clear in the lean spectrum compared to the spectra extracted from fat portions as lean portion contains more water content than fat portions. On the contrary, the absorption band at 930 nm related to the CH third stretching overtones (ElMasry, Sun & Allen, 2011; Osborne, Fearn & Hindle, 1993) is ascribed to the fat content in the ham sample. Thus, this specific band was very sharp in the spectra extracted from pure intermuscular or intramuscular fat portions, but it was not noticeable in the spectrum of the pure lean portion in the sample.

Due to averaging the spectra of all pixels within a region of interest in the ham image (*i.e.* pixels from fat and lean portions altogether), some of the remarkable absorption bands were negatively affected. For instance, the sharp absorption bands appeared in the spectra of both intermuscular and intramuscular fat portions especially at 420 and 930 nm became either very weak or completely disappeared in the average spectrum of the sample as shown in **Fig. 5b**. This problem depends basically on the number of pixels of a particular component within the region of interest. For example, when the number of lean pixels within a region of interest was significantly higher than the number of fat pixels, the remarkable absorption bands of fat are washed out from the average spectrum and may completely disappear. The average spectrum of the ham sample drawn as a solid bold line in **Fig. 5b** represents this case, in which the absorption bands at 420 nm and 930 nm became very weak although the ham sample already contains fat pixels.

3.5. Effect of hyperspectral data selection on the prediction of major constituents in dry-cured ham

Besides being redundant and collinear at contiguous wavebands, spectral data extracted from ham samples are very complex to interpret and much care should be taken during choosing the right multivariate modelling routines. In this study, PLS regression was chosen to model spectral data of the dry-cured samples with their reference chemical composition. The spectral data (X1, X2 and X3) being modelled were extracted from either the whole sliced ham sample (Fully heterogeneous WholeROI), from a small region of sliced ham sample (Semi-homogeneous SmallROI) or from unsliced cured muscles (Fully homogeneous MuscleROI). The first two data sets (X1 and X2) were extracted directly from sliced ham samples without separating the slices from each other; therefore, both sides of a sample were substantially heterogeneous (**Fig. 3**). Meanwhile, the third dataset (X3) was built from spectral data extracted from unsliced ham muscles in which both sides of the sample were vertically homogeneous. A separate PLS regression model was developed under cross validation from each dataset in the training samples and then tested in independent validation samples. Table 2 demonstrates the performance of the PLS regression models (Model I, Model II and Model III) developed in the training samples for different hyperspectral data extracted from different regions of interest (WholeROI, SmallROI and MuscleROI) in predicting major constituents (fat, water, salt and protein) in the dry-cured ham samples for calibration and cross validation conditions. The results revealed that the worst PLS model was obtained by using spectral data extracted from semi-homogeneous regions (Model II) with coefficient of determination under cross validation of 0.72, 0.71, 0.54 and 0.58 for the prediction of fat, water, salt and protein, respectively. This could be ascribed to the great variability between the two sides of the small ROIs in terms of spectral signatures and reference chemical composition.

The performance of the PLS models developed from either large heterogeneous regions (Model I) or from fully homogenous muscles (Model III) was comparable to each other. The coefficient of determination under cross validation (R^2_{cv}) in Model I and Model III was 0.84 and 0.92 for fat, 0.89 and 0.88 for water, 0.83 and 0.83 for salt and 0.74 and 0.88 for protein, respectively. These results are in agreement with those reported by Gou et al. (2013) and for predicting fat, water and salt in dry-cured ham using a hyperspectral imaging system employed in the interactance mode and with Garrido-Novell et al. (2015) for predicting salt

in dry-cured ham using a traditional line-scan hyperspectral imaging system in reflectance mode. In similar studies, ElMasry et al. (2013), Talens et al. (2013) and Iqbal et al. (2013) predicted the major chemical constituents (fat, water and protein) as well as colour and pH of sliced cooked hams with a degree of accuracy similar to that reported in this study. In the current study, the number of latent factors was higher in the models developed using large heterogeneous regions (WholeROI) compared to those models developed using the unsliced muscles (MuscleROI) for predicting water, salt and protein contents.

Table 2 Performance statistics of the PLS regression models (Model I, Model II and Model III) developed in distinct training dataset for hyperspectral data from different regions on interests (WholeROI, SmallROI and MuscleROI) in predicting major constituents in dry-cured ham samples.

	WholeROI 'X1' (n=119) Model I			SmallROI 'X2'(n=60) Model II			MuscleROI 'X3' (n=34) Model III		
Attribute	R_c^2	R_{cv}^2	LF	R_c^2	R_{cv}^2	LF	R_c^2	R_{cv}^2	LF
Fat	0.87	0.84	7	0.81	0.72	7	0.97	0.92	8
Water	0.93	0.89	12	0.80	0.71	7	0.96	0.88	9
Salt	0.92	0.83	15	0.85	0.54	12	0.97	0.83	9
Protein	0.80	0.74	9	0.65	0.58	4	0.94	0.88	6

The main criterion usually used by chemometricians to evaluate the overall accuracy of a certain multivariate model in predicting a specific attribute is usually identified as the capacity of this model for providing good prediction in an independent dataset that has not been 'seen' by the model during the training step. The validation step of the developed calibration model is critically important before implementing such a model for routine assays. Thus, the purpose of model validation is to determine the reproducibility of the developed calibration model and its long-term ruggedness. Accordingly, the three models developed from different hyperspectral data extracted from different regions of interest were tested in predicting the same chemical constituents (fat, water, salt and protein) in an independent validation dataset and the performance statistics of the three models are tabulated in Table 3. Similar to the results obtained in the calibration and cross validation conditions, the

performance statistics of the developed models in the validation data set indicated that the PLS model built using spectral data of semi-heterogeneous regions of interest (SmallROI) was the worst model compared to the other two models developed using either large heterogeneous regions (WholeROI) or unsliced fully-homogenous muscles (MuscleROI).

Table 3 Performance statistics of the PLS regression models (Model I, Model II and Model III) in the validation dataset developed by different spectral data extracted from different regions of interest (WholeROI, SmallROI and MuscleROI) in predicting major constituents in dry-cured ham samples.

Attribute	WholeROI 'X1' (n=59) Model I			SmallROI 'X2'(n=29) Model II			MuscleROI 'X3' (n=18) Model III		
	R_p^2	SEP(%)	LF	R_p^2	SEP(%)	LF	R_p^2	SEP(%)	LF
Fat	**0.85**	**1.43**	**7**	**0.58**	**1.60**	**7**	**0.96**	**1.18**	**8**
Water	**0.89**	**1.21**	**12**	**0.67**	**2.57**	**7**	**0.77**	**2.91**	**9**
Salt	**0.83**	**0.51**	**15**	**0.49**	**0.26**	**12**	**0.35**	**0.58**	**9**
Protein	**0.63**	**1.57**	**9**	**0.33**	**1.94**	**4**	**0.87**	**2.73**	**6**

The difference in the model performance might be ascribed to the number of samples used for each kind of ROIs (Tange *et al.*, 2017). During the experiments of this study, the WholeROI samples were scanned and analysed in two rounds of $n = 89$ samples each. To investigate the influence of sample size, the models developed from these 89 samples were compared with those models developed from SmallROI data ($n = 89$) and MuscelROI data ($n = 52$). In addition, when models developed from all WholeROI samples ($n = 178$) were compared with those ones developed from the SmallROIs ($n = 89$) and MuscelROIs ($n = 52$), the results did not change. This indicates the consistency of the obtained results of the examined samples. In general, the PLS models developed for predicting fat content using spectral data extracted from unsliced fully-homogeneous muscles (MuscleROI) was comparable with that one developed from the whole sliced ham sample (WholeROI) when tested in the independent validation datasets with coefficient of determination in prediction (R_p^2), standard error in prediction (SEP) and number of latent factors of 0.96, 1.18% and 8, respectively. However, the PLS models developed using WholeROI in predicting water and

salt contents were much better than those ones developed from MuscleROI with higher coefficient of determination in prediction (R_p^2), and lower standard error in prediction (SEP). These results indicated that to analyse a full package of sliced dry-cured ham directly using a hyperspectral imaging system, it is advised to use the whole sliced ham (WholeROI) as the main region of interest from which the spectral data should be extracted. By this way, the spectral fingerprints of all portions of the ham sample will have contributions on the average spectrum of the analysed sliced ham sample. These results are comparable with those reported by Liu et al. (2013) for predicting water and salt contents in different cuts of fresh pork at different stages in the salting process with coefficients of determination of 0.9 and 0.9 and SEPs of 0.682 for water and 0.007 for salt. **Fig. 6** shows the measured vs. predicted values of the four constituents (fat, water, salt and protein) in the validation data set using PLS model (Model I) developed from spectral data extracted from the whole sliced hams (WholeROI).

The point that should be further investigated now in this context is the possibility of using the model developed from semi-homogeneous regions (SmallROI) and from unsliced fully-homogenous cured muscles (MuscleROI) to predict the composition of the fully-heterogeneous packages of the whole sliced ham. Therefore, the PLS calibration models developed from SmallROI data (Model II) and from MuscleROI data (Model III) were used to predict fat, water, salt and protein contents of the WholeROI samples by considering these samples as a validation dataset for these models. As expected, Model II developed from semi-homogeneous regions was not accurate in predicting fat, water, salt and protein contents in the WholeROI samples. Although PLS models developed from MuscleROI was really efficient in predicting major compositions in the homogeneous unsliced muscles, the performance of these models was rather poor in predicting fat, water, salt and protein contents in the fully heterogeneous whole sliced ham packages with coefficient of determination of 0.29, 0.52, 0.37 and 0.36 and standard error of prediction of 4.01%, 6.19%, 0.74 and 2.58%, respectively (Table 4). The bad behaviour of these models could be attributed to the difference in scattering patterns between the unsliced and sliced hams leading to different spectral signatures between the unsliced (MuscleROI) and sliced (WholeROI) samples as shown in **Fig. 2g** and **Fig. 2e**, respectively. The obtained results revealed that the models developed using spectral data from homogenous, unsliced muscles (Model III) cannot be used for the prediction of the main composition of the packaged sliced

hams due to difference in scattering pattern and difference in compositional heterogeneity between both types of datasets.

Table 4 Performance statistics of the PLS regression models developed from semi-homogeneous regions (SmallROI) and from fully homogeneous, unsliced cured muscles (MuscleROI) data to predict the composition of the fully-heterogeneous packages of the sliced ham.

Chemical attribute	**SmallROI Model (Model II)**		**MuscleROI Model (Model III)**	
	R_p^2	SEP (%)	R_p^2	SEP (%)
Fat	0.36	2.81	0.29	4.01
Water	0.42	2.70	0.52	6.19
Salt	0.05	1.12	0.37	0.74
Protein	0.53	1.77	0.36	2.58

3.6. Mapping and identification of the ham samples

Due to the availability of both spatial and spectral information in the image, the power of the hyperspectral imaging could be extended to visualise the distribution of certain constituent by showing its concentrations and distribution in all spots of the ham samples. This is mathematically performed by applying the PLS models in every single pixels in the image resulting in a distribution map called the chemical image. The major problem found in all trails of pixel-based prediction using the PLS prediction models developed from spectral data of hyperspectral images of meat and meat products is posed by developing such models from the mean spectra of the whole sample along with its mean reference chemicagl measurements (Garrido-Novell et al., 2015) because it is practically impossible to have reference determinations at the pixel level. The chemical images are pseudo-colour images in which each colour corresponds to a certain concentration of the mapped constituent. In the presented chemical images shown in **Fig. 7** and **Fig. 8**, the blue colour indicates low concentrations and red colour denotes high concentrations of the constituent. Based on the number of calibration models developed from the same dataset, a number of chemical images are expected to be generated. In this regards, the PLS models developed using the whole sliced ham samples was used to produce chemical images to show the distribution of fat, water, salt and protein

contents in the whole sliced dry-cured ham samples. Because there was an independent PLS model for each constituent, four different chemical images were generated for each ham sample showing their spatial distributions throughout the whole sample as shown in **Fig. 7**. As the distributions of the constituents are visualised in all pixels within a ham sample to show the difference from portion to portion in the same image, it was also possible to demonstrate the difference among samples having different concentrations of these constituents. As demonstrated in **Fig. 8**, it is possible to recognise ham samples with low, intermediate and high concentrations of a certain constituent. The arrow drawn at the top of the figure indicates the direction of increasing the content of the estimated constituents in the ham samples. Accordingly, the chemical images shown in the left-hand side in **Fig. 8** represent ham samples with low contents of fat, water, salt and protein contents compared to the chemical images appeared in the right-hand side of **Fig. 8** that show ham samples having high contents of these constituents.

It is quite important to emphasize that all examined samples were commercial ham samples that are available in the market as final products from different producers without any prior treatments practiced in the laboratory. Also, the telescopic arrangement of the slices inside the scanned ham sample may explain why the chemical images shown in **Fig. 7** and **Fig. 8** looked inhomogeneous in the constituents appeared in the chemical images compared to those ones illustrated by Liu et al (2013) and Liu et al. (2014) who analysed dry-cured ham samples at different periods after being salted in the lab and scanned these samples directly after preparation. In essence, the visualised forms of the distribution maps of the essential chemical constituents are very important for the developers and manufacturers to take the suitable action during ham processing to add better control of salting and drying processes during processing and production. This great capacity of hyperspectral image could not have been achieved by using either the point-scan spectrometer or traditional colour imaging alone.

4. CONCLUSION

The proposed models were extensively tested using three options of data extraction with promising performance; in which the highest performance was achieved by using the data extracted from the whole ham sample (WholeROI). The results revealed that when packed dry-cured ham slices are heterogeneous in the horizontal and vertical profiles, the size of the selected region of interest (ROI) should include the whole ham slices for better prediction

accuracy. Based on the good predictability of the multivariate models developed in this study, the practical applications of hyperspectral imaging for ham composition authentication seems to be possible especially for the retailing industry to ensure the compliance with the information reported in the produced packages. Besides being non-destructive and rapid technique, the application of four different predictive models (one for each chemical constituent) will help more in understanding the progress of certain process during ham processing and to guarantee the quality of the final product. The results could be also examined under important wavelengths instead of the whole spectral range by excluding redundant wavelengths that do not carry reliable spectral information. By selecting the proper wavelengths and in the scope of the promising results obtained, this study opens an opportunity to develop a simple state-of-the art spectral imaging module supported with relevant machine learning tool to contribute in the progression of non-invasive quality evolution of raw and processed meat products in real-time applications. Such a module can be implemented across every meat processing stages from preparation until counter.

Acknowledgements

This work was partially financed by CCLabel project (RTI-2018-096883-R-C41) and CERCA programme from Generalitat de Catalunya. Authors significantly acknowledge the financial support from the Marie Sklodowska-Curie COFUND P-SPHERE project under the European Union's Horizon 2020 research and innovation programme with grant agreement No 665919 and the Distinguished Scientist Fellowship Program (DSFP) King Saud University.

5. REFERENCES

Akashi, Y., & Okatani, T. (2016). Separation of reflection components by sparse non-negative matrix factorization. *Computer Vision and Image Understanding*, 146(1), 77-85.

Alander, J. T., Bochko, V., Martinkauppi, B., Saranwong, S., & Mantere, T. (2013). A review of optical nondestructive visual and near-infrared methods for food quality and safety. *International Journal of Spectroscopy*, 2013, 1-36.

Anderson, S. (2007). Determination of fat, moisture, and protein in meat and meat products by using the FOSS FoodScan near-infrared spectrophotometer with FOSS artificial

660 neural network calibration model and associated database: collaborative study.
661 *Journal of AOAC International*, *90*(4), 1073-1083.

662 Arnau, J., Guerrero, L., Casademont, G. & Gou, P. (1995). Physical and chemical changes in
663 different zones of normal and PSE dry cured ham during processing. *Food Chemistry*,
664 52(1), 63-69.

665 Barbin, D. F., ElMasry, G., Sun, D. W., Allen, P., & Morsy, N. (2013). Non-destructive
666 assessment of microbial contamination in porcine meat using NIR hyperspectral
667 imaging. *Innovative Food Science & Emerging Technologies*, 17, 180-191.

668 Boadas, C., Gou, P., Valero, A. & Arnau, J. (2001). Changes in different zones of dry-cured
669 ham during drying: Moisture and sodium chloride content. Fleischwirtschaft
670 (Frankfurt), 81(1), 91-93.

671 Bou, R., Llauger, M., Arnau, J., & Fulladosa, E. (2018). Zinc-protoporphyrin content in
672 commercial Parma hams is affected by proteolysis index and marbling. *Meat Science*,
673 139, 192-200.

674 Büning-Pfaue, H. (2003). Analysis of water in food by near infrared spectroscopy. *Food*
675 *Chemistry*, 82(1), 107-115.

676 Cozzolino, D., De Mattos, D., & Vaz Martins, D. (2002). Visible/near infrared reflectance
677 spectroscopy for predicting composition and tracing system of production of beef
678 muscle. *Animal Science*, 74, 477–484.

679 Cozzolino, D., Murray, I., Scaife, J. R., & Paterson, R. (2000). Study of dissected lamb
680 muscles by visible and near infrared reflectance spectroscopy for composition
681 assessment. *Animal Science*, 70, 417–425,

682 Craigie, C. R., Johnson, P. L., Shorten, P. R., Charteris, A., Maclennan, G., Tate, M. L. &
683 Reis, M. M. (2017). Application of Hyperspectral imaging to predict the pH,
684 intramuscular fatty acid content and composition of lamb M. longissimus lumborum at
685 24 h post mortem. *Meat Science*, 132, 19-28.

686 Crichton, S. O., Kirchner, S. M., Porley, V., Retz, S., von Gersdorff, G., Hensel, O. & Sturm,
687 B. (2017). Classification of organic beef freshness using VNIR hyperspectral imaging.
688 *Meat Science*, 129, 20-27.

689 Dumont, J., Hirvonen, T., Heikkinen, V., Mistretta, M., Granlund, L., Himanen, K. &
690 Nygren, M. (2015). Thermal and hyperspectral imaging for Norway spruce (*Picea*
691 *abies*) seeds screening. *Computers and Electronics in Agriculture*, 116, 118-124.

692 ElMasry, G., Iqbal, A., Sun, D. W., Allen, P., & Ward, P. (2011). Quality classification of
693 cooked, sliced turkey hams using NIR hyperspectral imaging system. *Journal of Food*
694 *Engineering*, 103(3), 333-344.

695 ElMasry, G., Mandour, N., Al-Rejaie, S., Belin, E., & Rousseau, D. (2019a). Recent
696 applications of multispectral imaging in seed phenotyping and quality monitoring-An
697 overview. *Sensors*, 19(5), 1090.

698 ElMasry, G., Mandour, N., Wagner, M. H., Demilly, D., Verdier, J., Belin, E., & Rousseau,
699 D. (2019b). Utilization of computer vision and multispectral imaging techniques for
700 classification of cowpea (Vigna unguiculata) seeds. *Plant Methods*, 15(1), 24.

701 ElMasry, G., Nassar, A., Wang, N., Vigneault, C. (2008). Spectral methods for measuring
702 quality changes of fresh fruits and vegetables. *Stewart Postharvest Review*, 4(4): 1-13.

703 ElMasry, G., Sun, D. W., & Allen, P. (2011). Non-destructive determination of water holding
704 capacity in fresh beef by using NIR hyperspectral imaging. *Food Research*
705 *International*, 44(9), 2624–2633.

706 ElMasry, G., Sun, D. W., & Allen, P. (2013). Chemical-free assessment and mapping of
707 major constituents in beef using hyperspectral imaging. *Journal of Food Engineering*,
708 117(2), 235-246.

709 ElMasry, G., Sun, S-W (2010). Meat Quality Assessment Using a Hyperspectral Imaging
710 System. In *Hyperspectral Imaging for Food Quality Analysis and Control* (pp. 175-
711 240), Edited by Da-Wen Sun, Academic Press/Elsevier, San Diego, California, USA.

712 ElMasry, G., Wang, N., ElSayed, A., & Ngadi, M. (2007). Hyperspectral imaging for
713 nondestructive determination of some quality attributes for strawberry. *Journal of Food*
714 *Engineering*, 81(1), 98-107.

715 ElMasry, G., Wang, N., Vigneault, C. m., Qiao, J., & ElSayed, A. (2008). Early detection of
716 apple bruises on different background colours using hyperspectral imaging. *LWT -*
717 *Food Science and Technology*, *41*(2), 337-345.

718 Foca, G., Ferrari, C., Ulrici, A., Sciutto, G., Prati, S., Morandi, S. & Oliveri, P. (2016). The
719 potential of spectral and hyperspectral-imaging techniques for bacterial detection in
720 food: A case study on lactic acid bacteria. *Talanta*, 153, 111-119.

721 Garrido-Novell, C., Garrido-Varo, A., Pérez-Marín, D., Guerrero-Ginel, J. E., & Kim, M.
722 (2015). Quantification and spatial characterization of moisture and NaCl content of
723 Iberian dry-cured ham slices using NIR hyperspectral imaging. *Journal of Food*
724 *Engineering*, 153, 117-123.

725 Gou, P., Santos-Garces, E., Høy, M., Wold, J. P., Liland, K. H., & Fulladosa, E. (2013).
726 Feasibility of NIR interactance hyperspectral imaging for on-line measurement of crude
727 composition in vacuum packed dry-cured ham slices. *Meat Science*, 95(2), 250-255.

728 Hersleth, M., Lengard, V., Verbeke, W., Guerrero, L., & Næs, T. (2011). Consumers'
729 acceptance of innovations in dry-cured ham: Impact of reduced salt content, prolonged
730 aging time and new origin. *Food Quality and Preference*, 22(1), 31-41.

731 Iqbal, A., Sun, D. W. & Allen, P. (2013). Prediction of moisture, color and pH in cooked,
732 pre-sliced turkey hams by NIR hyperspectral imaging system. *Journal of Food*
733 *Engineering*, 117(1), 42-51.

734 Kamruzamman, M., Nakauchi, S., & ElMasry, G. (2015). Online screening of meat and
735 poultry product quality and safety using hyperspectral imaging. In: High Throughput
736 Screening for Food Safety Assessment (pp. 425-466). Woodhead Publishing.

737 Kaur, A., Hieke, S. & Rayner, M. (2016). Do health and nutrition claims meet consumers'
738 health needs? *Agro Food Industry Hi-Tech*, 27(3): p. 22-24.

739 Khan, H. A., Thomas, J.-B., & Hardeberg, J. Y. (2017). Analytical survey of highlight
740 detection in colour and spectral images. In International Workshop on Computational
741 Colour Imaging (pp. 197-208). March 2017, Milan, Italy.

742 Koirala, P., Pant, P., Hauta-Kasari, M., & Parkkinen, J. (2011). Highlight detection and
743 removal from spectral image. *Journal of the Optical Society of America A (JOSA A)*,
744 28(11), 2284-2291.

745 Liu, D., Qu, J., Sun, D. W., Pu, H., & Zeng, X. A. (2013). Non-destructive prediction of salt
746 contents and water activity of porcine meat slices by hyperspectral imaging in a salting
747 process. *Innovative Food Science & Emerging Technologies*, 20, 316-323.

748 Liu, D., Sun, D. W., Qu, J., Zeng, X. A., Pu, H., & Ma, J. (2014). Feasibility of using
749 hyperspectral imaging to predict moisture content of porcine meat during salting
750 process. *Food Chemistry*, 152, 197-204.

751 Maeda, H., Ozaki, Y., Tanaka, M., Hayashi, N., & Kojima, T. (1995). Near infrared
752 spectroscopy and chemometrics studies of temperature-dependent spectral variations of
753 water: relationship between spectral changes and hydrogen bonds. *Journal of Near*
754 *Infrared Spectroscopy*, 3(4), 191-201.

755 Morsy, N. & Sun, D. W. (2013). Robust linear and non-linear models of NIR spectroscopy
756 for detection and quantification of adulterants in fresh and frozen-thawed minced beef.
757 *Meat Science*, 93(2), 292-302.

758 Muñoz, I., Gou, P., & Fulladosa, E. (2019). Computer image analysis for intramuscular fat
759 segmentation in dry-cured ham slices using convolutional neural networks. *Food*
760 *Control*, 106, 106693.

761 Muñoz, I., Rubio-Celorio, M., Garcia-Gil, N., Guàrdia, M. D., & Fulladosa, E. (2015).
762 Computer image analysis as a tool for classifying marbling: A case study in dry-cured
763 ham. *Journal of food engineering*, 166, 148-155.

764 Ortiz-Somovilla, V., España-España, F., Gaitán-Jurado, A. J., Pérez-Aparicio, J., & De
765 Pedro-Sanz, E. J. (2007). Proximate analysis of homogenized and minced mass of pork
766 sausages by NIRS. *Food Chemistry*, 101, 1031–1040.

767 Osborne, B. G., Fearn, T., & Hindle, P. H. (1993). Practical NIR spectroscopy: With
768 applications in food and beverage analysis. Harlow: Longman Scientific & Technical.

769 Parolari, G., Aguzzoni, A., & Toscani, T. (2016). Effects of Processing Temperature on
770 Colour Properties of Dry-Cured Hams Made without Nitrite. *Foods*, 5(2), 33.

771 Pathmanaban, P., Gnanavel, B. K. & Anandan, S. S. (2019). Recent Application of imaging
772 techniques for Fruit quality Assessment. *Trends in Food Science & Technology*, 94,
773 32.42.

774 Pérez-Santaescolástica, C., Fraeye, I., Barba, F. J., Gómez, B., Tomasevic, I., Romero, A. &
775 Lorenzo, J. M. (2019). Application of non-invasive technologies in dry-cured ham: An
776 overview. *Trends in Food Science & Technology*, 86, 360-374.

Reis, M. M., Van Beers, R., Al-Sarayreh, M., Shorten, P., Yan, W. Q., Saeys, W. & Craigie, C. (2018). Chemometrics and hyperspectral imaging applied to assessment of chemical, textural and structural characteristics of meat. *Meat Science*, 144, 100-109.

Resano, H., Sanjuán, A. I., Cilla, I., Roncalés, P., & Albisu, L. M. (2010). Sensory attributes that drive consumer acceptability of dry-cured ham and convergence with trained sensory data. *Meat Science*, 84(3), 344-351.

Shen, H.-L. & Zheng, Z.-H. (2013). Real-time highlight removal using intensity ratio. *Applied optics*, 52(19), 4483-4493.

Siripatrawan, U. (2018). Hyperspectral imaging for rapid evaluation and visualization of quality deterioration index of vacuum packaged dry-cured sausages. *Sensors and Actuators B: Chemical*, 254, 1025-1032.

Talens, P., Mora, L., Morsy, N., Barbin, D. F., ElMasry, G., & Sun, D. W. (2013). Prediction of water and protein contents and quality classification of Spanish cooked ham using NIR hyperspectral imaging. *Journal of Food Engineering*, 117(3), 272-280.

Tange, R. I., Rasmussen, M. A., Taira, E. & Bro, R. (2017). Benchmarking support vector regression against partial least squares regression and artificial neural network: Effect of sample size on model performance. *Journal of Near Infrared Spectroscopy*, 25(6), 381-390.

Velásquez, L., Cruz-Tirado, J. P., Siche, R., & Quevedo, R. (2017). An application based on the decision tree to classify the marbling of beef by hyperspectral imaging. *Meat Science*, 133, 43-50.

Wakholi, C., Kandpal, L. M., Lee, H., Bae, H., Park, E., Kim, M. S. & Cho, B. K. (2018). Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics. *Sensors and Actuators B: Chemical*, 255, 498-507.

Washburn, K. E., Stormo, S. K., Skjelvareid, M. H., & Heia, K. (2017). Non-invasive assessment of packaged cod freeze-thaw history by hyperspectral imaging. *Journal of Food Engineering*, 205, 64-73.

Figure Captions

Fig. 1 Configuration of the lab-based spectral imaging system used in acquiring spectral images of ham samples in the NIR spectral range of 400-1000 nm.

Fig. 2 Procedure of predicting major chemical composition of dry-cured hams using spectra extracted from different regions of interest. (a) locating the main region of interest in the original images, (b) isolating different regions and extracting spectral data from these particular regions, (c) arranging spectral data for each ROI in a separate data matrix (X1, X2 & X3) along with their corresponding reference chemical measurements (Y), (d) development of PLS regression models for the data of each ROI, (e) plotting the measured *vs.* predicted values of each chemical attribute resulting from the PLS model for each ROI and (f) evaluating the performance of the three PLS models in predicting the chemical attributes under investigation for choosing the best model.

Fig. 3 Regions of interests from hyperspectral images of dry cured-ham samples. (a) 'WholeROI' includes the whole area of the sample (Fully heterogeneous ROI), (b) 'SmallROI' including a small homogeneous lean region in one side but homogeneous fat from the other side (Semi-homogeneous ROI), (c) 'MuscleROI' includes a selected lean region in both sides of unsliced muscles (Fully homogeneous ROI), (d) both sides of the selected regions of interest and (e-g) spectral signatures of both sides of ROIs in the three cases ('WholeROI', 'SmallROI' and 'MuscleROI').

Fig. 4 Removal of specular zones from hyperspectral images. (a) Pseudo-colour image rendered at 640 nm, 550 nm and 460 nm from the original hyperspectral image having a specular zone marked with yellow arrows, (b) locating specular zones in the raw image (in white pixels), (c) pseudo-colour image of the hyperspectral image after specularity removal treatment in which all specular zones were significantly mitigated, (d) 3D visualization of the spectral fingerprints of all pixels in the original hyperspectral image with extreme reflectance values at the specular zones marked with red arrows and (e) spectral fingerprints of the same image after specularity removal treatment (the extreme reflectance value were mediated in this treated image).

Fig. 5 (a) Spectral signatures of some individual lean pixels before and after specular correction that significantly decreases the magnitudes of reflectance of the pixels throughout the spectrum, (b) Corrected average spectra of different regions (intermuscular fat, intramuscular fat and lean) in the hyperspectral image of a dry-cured ham sample. The solid bold line stands for the overall mean spectrum of the whole ham sample including pixels from all regions (intermuscular fat + intramuscular fat + lean pixels). Arrows point to the remarkable absorption bands (420, 550, 580, 930 and 970 nm) of some essential functional chemical groups related to different pigment derivatives, fat and water contents in the ham sample.

Fig. 6 Measured *vs.* predicted values of (a) fat, (b) water, (c) salt and (d) protein contents in the validation dataset ($n = 59$) resulting from the PLS calibration models (Model I) developed from spectral data extracted from the whole sliced hams (WholeROI).

Fig. 7 Chemical images from two independent samples produced with the aid of Models I. The values written at the bottom of each chemical image designate the average measured value of a constituent in the whole sliced ham.

Fig. 8 Chemical images produced from four different PLS models developed for predicting fat, water, salt and protein contents in packages of sliced ham samples having different contents of these attributes. The arrow at the top of the figure indicates the direction of increasing the concentration of the constituents in the sliced ham samples.