Talanta 109 (2013) 74–83

Contents lists available at SciVerse ScienceDirect

Talanta

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

Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of *Pseudomonas* loads in chicken fillets

Yao-Ze Feng, Da-Wen Sun *

FRCFT Group, School of Biosystems Engineering, University College Dublin, National University of Ireland, Agriculture and Food Science Centre, Belfield, Dublin 4, Ireland

ARTICLE INFO

Available online 4 February 2013

Keywords:
Chemometrics
Preprocessing
Chemical imaging
Chicken breast fillets
Bacterial pathogen
Spoilage
Prediction map
Standard normal variate (SNV)

ABSTRACT

Hyperspectral imaging was exploited for its potential in direct and fast determination of *Pseudomonas* loads in raw chicken breast fillets. A line-scan hyperspectral imaging system (900–1700 nm) was employed to obtain sample images, which were then further corrected, modified and processed. The prepared images were correlated with the true *Pseudomonas* counts of these samples using partial least squares (PLS) regression. To enhance model performance, different spectral extraction approaches, spectral preprocessing methods as well as wavelength selection schemes based on genetic algorithm were investigated. The results revealed that extraction of mean spectra is more efficient for representation of sample spectra than computation of median spectra. The best full wavelength model was attained based on spectral images preprocessed with standard normal variate, and the correlation coefficients (R) and root mean squared errors (RMSEs) for the model were above 0.81 and below 0.80 $\log_{10}$ CFU g^{-1}, respectively. In development of simplified models, wavelengths were selected by using a proposed two-step method based on genetic algorithm. The best model utilized only 14 bands in five segments and produced R and RMSEs of 0.91 and 0.55 $\log_{10}$ CFU g^{-1}, 0.87 and 0.65 $\log_{10}$ CFU g^{-1} as well as 0.88 and 0.64 $\log_{10}$ CFU g^{-1} for calibration, cross-validation and prediction, respectively. Moreover, the prediction maps offered a novel way for visualizing the gradient of *Pseudomonas* loads on meat surface. Hyperspectral imaging is demonstrated to be an effective tool for nondestructive measurement of *Pseudomonas* in raw chicken breast fillets.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Chicken fillets are one of the popular meat products valuable for human nutrition and health. Due to their high perishability, there is a need to develop innovative techniques such as refrigeration methods [1-6] and processes [7-11] to enhance their quality and safety, there is also a need to develop novel detection methods to ensure their food safety. Pseudomonas is one of the bacterial genera most isolated in high numbers on spoiled poultry. *Pseudomonas* is a genus of rod-shaped and Gram-negative bacteria that require only simple nutrition for growth and such easy-to-survive characteristic has contributed to its wide distribution in the environment [12]. *Pseudomonas* is closely associated with food waste due to spoilage species causing food spoilage [13,14]. *Pseudomonas* is greatly suitable for utilizing low-molecular-weight compounds such as glucose which was metabolized into 2-oxo-gluconate or gluconate via the pathway of Entner–Doudoroff metabolism [15]. With the progressing of interaction between meat and microbes, pseudomonades start to use amino acids which can then lead to increase of ammonia and pH. Due to its fast growth rate, at the time of spoilage, *Pseudomonas* is the predominant bacterial genus on animal carcasses, modified atmosphere packed meat, meat products and even occasionally on the vacuum stored meat [16]. Consequently, it is very important to determine *Pseudomonas* in the food matrix either to verify the spoilage stage or to identify the safety condition of food. Current methods for determination of *Pseudomonas*, including standard plate count techniques, polymerase chain reaction methods as well as enzyme-linked immunosorbent assay approaches, are normally destructive, tedious, time-consuming and procedure-complicated [12]. Therefore, a fast, non-destructive and precise method is greatly needed for better food quality and microbiological safety control.

By combining the advantages of computer vision [17–21] and spectroscopic technique, Hyperspectral imaging (HSI), or chemical imaging, is a promising technology that is being intensively explored in many fields [22–26]. Unlike traditional spectroscopic or imaging technologies that can only record spectral or spatial information, HSI attains both spectral and spatial knowledge of the objects under

* Corresponding author. Tel.: +353 1 7167342; fax: +353 1 7167493.
E-mail address: dawen.sun@ucd.ie (D.-W. Sun).
URLS: http://www.ucd.ie/refrig, http://www.ucd.ie/sun (D.-W. Sun).

0039-9140/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2013.01.057

study and therefore provides versatile opportunities for data analyses so that the features of interest can be better interpreted. However, since HSI is the combination (or compromise) of spectroscopy and traditional computer vision, its signal quality is not always warranted to be as high as the above two individual technologies [23]. To solve this problem, spectral preprocessing methods, including standard normal variate (SNV), multiplicative scatter correction (MSC), detrending (Detrend) as well as first and second derivatives (Der1 and Der2) can be utilized [27]. Another disadvantage for HSI is the huge image size which demands high-quality hardware for storage and processing. Solutions for this problem can be bifurcated: either to employ good equipment at high costs or to reduce image sizes without losing significant variance in the image (or hypercube). The latter can be achieved by retaining single-band images at certain wavelengths and these particular bands can be selected by diverse variable selection methods. For example, the regression coefficients from partial least squares regression (PLSR) can be utilized to identify important wavelengths by allocating dominant peaks and valleys in the curve [28–31]. Among others are methods based on PLS (partial least squares) variable importance metric [32], stepwise regression [33], bound and boundary [34], successive projection algorithm [35], and others [31,36,37]. Use of these variable selection methods can possibly result in better model performance and meanwhile enable the development of multispectral imaging systems so that detection can meet industrial requirements.

In terms of applications in food analyses, hyperspectral imaging has been successfully employed to inspect fecal contamination and wholesomeness of chicken carcasses [33,37,38], to determine meat quality [39–41] as well as to detect defects and contamination of fruits and vegetables [42–44], and so on. However, its efficiency in direct determination of microbial loads in food is still not well explored though a limited number of studies have demonstrated its feasibility [31,45,46]. Up to the knowledge of the authors, no research has been carried out on the use of hyperspectral imaging for detecting *Pseudomonas* in food. With respects to hyperspectral image analyses, spectral preprocessing methods have been concluded as useful in improving final predictions in some studies [47], however, there are also some studies reporting the inefficiency of these methods [28]. Moreover, most of the wavelength selection methods precipitate allocation of separate individual wavelengths and use of these selected variables for predicting new samples in new systems may produce more errors than when several wavelength ranges are utilized. This is because the former spectral parameters are more susceptible to the disturbance of either random noise or instrumental shift than the latter, which has a smoothing effect on such changes. Therefore, the objective of the current study was (1) to investigate the potential of hyperspectral imaging for nondestructive quantification of *Pseudomonas* loads in raw chicken fillets; (2) to apply different spectral preprocessing methods to test its usefulness in enhancing model performance; (3) to select vital wavelengths using proposed two-step schemes based on genetic algorithm (GA), and (4) to establish and validate full-wavelength and simplified PLS calibration models for predicting *Pseudomonas* counts.

2. Materials and methods

2.1. Chicken samples and microbiological experiment

Chicken breast fillet samples (12 packages) were purchased at a local supermarket and prepared in the same way as mentioned in an earlier study [31]. The fillet blocks were stored in a fridge at 4 °C for up to 9 days. In order to capture large variations of bacterial loads during the spoilage process of meat, imaging and microbiological experiments were carried out at daily basis during the first six days and every 12 h during the last three days. To determine the reference *Pseudomonas* counts in chicken samples, each sample (weighing approximately 10 g) was homogenized with ca 90 ml of buffered peptone water (CM1049, Oxoid, Basingstoke, Hampshire, England) in a stomach bag. The prepared initial solution was then subsequently diluted at a factor of 10^{-1}. Duplicate aliquots of 0.1 ml appropriate dilutions were then transferred to and spread evenly on *Pseudomonas* agar, which was prepared with Cetrimide agar (105284, Merck, Darmstadt, Germany) and *Pseudomonas* CFC Selective Supplement (107627, Merck, Darmstadt, Germany). The inoculated plates were then inverted and incubated at 30 °C for 48 h. The *Pseudomonas* counts were calculated based on plates where the number of typical colonies fell in 25 and 250. Finally, a total of 52 samples were obtained.

2.2. Hyperspectral imaging system

The hyperspectral imaging system utilized in this study was a laboratory-based pushbroom imaging equipment, and its configuration has been introduced in detail in previous studies [48–51]. Briefly, the main components include: a spectrograph with PGP optical structure (ImSpector, N17E, Spectral Imaging Ltd, Finland), a 12-bit CCD camera (V-light, Lowel Light Inc, USA) cooled by a peltier cooling device, an illumination unit with two 500 W tungsten-halogen lamps (V-light, Lowel Light Inc, NY, USA) and a translation stage controlled by a step motor (GPL-DZTSA-1000-X, Zolix Instrument Co, China). The harmonious work of the integral system was assured by using control software (SpectralCube, Spectral Imaging Ltd., Finland). Parameters for image acquisition are summarized in Table 1.

2.3. Image processing

The raw images acquired by the HSI system were initially in radiance unit. To improve the comparability and robustness of models to be subsequently built, *as per* the following equation, the raw images (I) were corrected as reflectance images (I_C) by using two external reference images, a white image (W) with ~100% reflectance acquired on a tile and a dark current image (D) with ~0% reflectance obtained when the lens of camera were completely covered [52].

$$I_C = \frac{I-D}{W-D} \quad (1)$$

Based on the corrected images, the meat portions should be isolated from the background so that meat spectra can be extracted and calculated to represent the sample. In this study, the isolation was conducted by constructing a mask produced by thresholding (global threshold value of 0.22) a difference image, which was a subtraction of a sub-band image of lower reflectance (at 944 nm) from another single-band image of higher reflectance (at 1395 nm) in the same hypercube. Image open was also

Table 1
Parameters of hyperspectral imaging system for acquisition of images.

Parameters	Values
Spectral range	897–1752 nm
Spectral increment	3.34 nm
Spectral resolution	6 nm
Image resolution	0.575 mm pixel^{-1}
Moving speed of the translation stage	~46 mm s^{-1}
Spatial dimension of images	280 × 320 pixels
Frame rate	98 fps
Exposure time	0.02 s

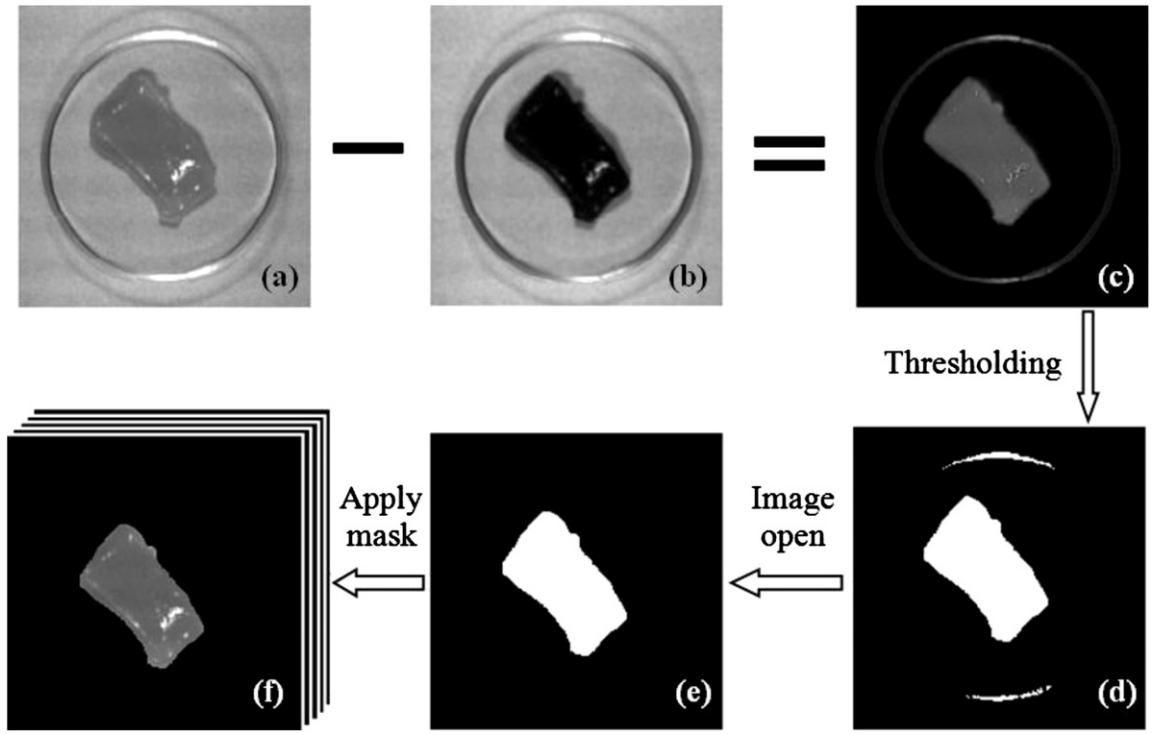

Fig. 1. Procedure for image masking. (a) sing-band image at 944 nm; (b) sing-band image at 1395 nm; (c) difference image of (a) and (b); (d) segmented image (threshold = 0.22); (e) final meat mask after image open operation; (f) masked hypercube.

implemented to remove the redundant parts due to Petri dishes. The resulting masks characterized the meat portions as 1 s and the background as 0 s. The whole procedure is illustrated in Fig. 1. When the developed masks were applied to the original hypercubes, new hypercubes were produced, where the reflectance values in meat portions remained the same while the values for background were suppressed to 0 s.

Furthermore, a representative spectrum was calculated as either the mean or median of the spectra (or preprocessed spectra) in the identified meat portions in the corresponding sample and all these sample spectra were further correlated with measured *Pseudomonas* counts in order to establish quantitative models. Since two different spectra (mean and median) were computed to stand for the same sample, the two spectral extraction methods were compared on their capability for representation of sample spectra in terms of contributing to a better prediction of the microbial loads.

2.4. Spectral preprocessing

The purpose of spectral processing is to enhance the quality of instrumental measurements as well as to reduce or eradicate information that is present in the spectra but has trivial correlation with the target values to be modelled (e.g., *Pseudomonas* loads in chicken fillets). In this study, standard normal variate (SNV), multiplicative scatter correction (MSC), detrending (Detrend), derivatives and their combinations were investigated. SNV and MSC are capable of correcting multiplicative noise (e.g., particle size influence, scatter interference and path length effect) arising from the physical structure of samples [53, 54]. Detrend is a polynomial baseline correction method for suppressing the baseline shifting and curvilinearity in spectra [55]. Moreover, derivative methods, normally first and second derivatives, are powerful to eliminate background noise as well as to separate overlapped peaks, thus enhancing spectral resolution [56, 57]. All the preprocessing methods were implemented in Matlab (The Mathworks Inc., Natick, MA, USA).

2.4.1. Standard normal variate (SNV)

SNV transformation is performed based on individual spectrum, which requires the calculation of the mean and standard deviation (SD) values of the specific spectrum. The corrected spectrum is obtained as:

$$X_{\mathrm{SNV}} = (X-\overline{X})/\sqrt{\frac{\sum_i^n (X_i-\overline{X})^2}{n-1}} \tag{2}$$

where X and X_{SNV} are the original and SNV-corrected spectra, respectively; $\overline{X}$ is the mean of the n spectral values along the whole wavelength range of X.

2.4.2. Multiplicative scatter correction (MSC)

To apply multiplicative scatter correction, a reference spectrum is needed. Rather than in the case of near-infrared spectroscopic analysis where the reference spectrum is normally obtained as the mean spectrum of all spectra in a calibration set, in preprocessing of hyperspectral images with MSC, the reference spectrum (X_{ref}) is calculated as the average of the meat spectra for each image (a sample). Given the reference, a hypercube can be corrected by transforming individual spectrum using the following equation:

$$X_{\mathrm{MSC}} = \frac{X-a}{b} \tag{3}$$

where a and b are the regression coefficients obtained from the following regression function:

$$X = a + bX_{\mathrm{ref}} \tag{4}$$

2.4.3. Detrend

Spectral detrending is a process that a linear pattern is removed from the original spectral signal. This pattern usually indicates a systematic shift due to sensor drift. In the present study, the trend was computed as best least-squares fit of

respective spectra [58]. Detrend was also combined with SNV and MSC for pretreatment of spectral images.

2.4.4. Derivatives

Derivative methods (e.g., first (Der1) and second (Der2) derivative of spectra) can facilitate distinguishing overlapped spectral peaks while removing baseline offsets and/or slops. Considering derivative methods are prone to exaggeration of noise, Savitzky–Golay algorithm was employed with a third-order polynomial fitting within a moving window of a width of 15. Moreover, SNV and MSC were also performed on the raw spectra prior to the application of derivative methods.

2.5. Chemometrics

Prior to establishment of calibration models, the attained samples ($n=52$) were divided into two sets, namely calibration set ($n=42$) and prediction set ($n=10$). The method employed for data partitioning was SPXY, which considered both spectra and reference values [59]. In this study, partial least squares regression was utilized to predict *Pseudomonas* counts in raw chicken breast fillets.

2.5.1. PLS calibration

PLS regression is a powerful method for revealing the linear relationship between spectra (X) and parameters under investigation (Y). During PLS analysis, both X and Y matrices are first transformed into new spaces, and the obtained data called X scores and Y scores are then carefully selected and correlated in an attempt to maximize the interpretation of Y scores by X scores. Afterwards, the predicted Y scores are subsequently employed to produce the prediction of Y. Details about PLS regression can be found elsewhere [60]. Consequently, one can use the following simple function to feature the best linear relationship between X and Y.

$$Y = Xb + e \tag{5}$$

where b is the regression coefficients and e is the prediction error.

As indicated above, when regression coefficients b is finally determined for a specific problem, the Y values for new samples can be predicted with reasonable errors by multiplying the spectra (X) of these new samples with the obtained regression coefficients.

2.5.2. Genetic algorithm for variable selection

Inspired by the evolution theory, genetic algorithm aims to reach the global optimum for a problem by sifting 'best' individuals in the population using mutation and cross-over operations [61]. A detailed description of GA can be found elsewhere [62,63] and the parameters adopted in this study are given in Table 2.

The nature of randomness in GA leads to the variation in the selected wavelengths during different implementations, therefore, GA programs were executed repeatedly to select the initial wavelength candidates [63]. It is assumed that the common wavelengths selected by most of the different runs of GA are of exceptionally great importance in accounting for the targets of interest. In this study, some majority consensus rules, as shown in Fig. 2, are proposed. Specifically, after five runs of GA programs, the five wavelength sets produced were compared and the common wavelengths agreed by no less than 3, 4 and all the five sets were allocated for further investigation, and the corresponding models established were denoted as GA-PLS-3, GA-PLS-4 and GA-PLS-5 models, respectively. In addition, all the candidate wavelengths and individual wavelength sets initially selected were also used to explore their potential for establishing stable and precise calibration model to predict the loads of *Pseudomonas*. The model based on spectral data with all possible selected wavelengths was called GA-PLS-Any model and GA-PLS-BT was noted for the best model gained from different wavelength combinations during trials.

Table 2
Parameter settings for genetic algorithm.

Parameters	Description
Population	30
Number of runs	100
Number of variables selected in the same chromosomes	≤ 30
Deletion groups	5
Probability of mutation	0.01
Probability of cross-over	0.5
Window width for smoothing	3
Data	Autoscaled
Regression method	PLS regression

2.5.3. Model assessment

Statistical criteria, including correlation coefficients (R) and root mean squared errors (RMSEs), were computed to evaluate the performance of established models. These parameters are defined as follows:

$$R = \sqrt{1 - \frac{\sum (y_{i,\mathrm{pred}} - y_{i,\mathrm{act}})^2}{\sum (y_{i,\mathrm{pred}} - \bar{y})^2}} \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i}^{m} (y_{i,\mathrm{pred}} - y_{i,\mathrm{act}})^2} \tag{7}$$

where $y_{i,\mathrm{pred}}$ is the predicted *Pseudomonas* loads and $y_{i,\mathrm{act}}$ is the actual *Pseudomonas* count for the ith sample; n is the number of samples; $\bar{y}(\,)$ stands for the average of the actual loads of all samples.

Besides, prediction errors of sum of squares (PRESS) for leave-one-out full cross validation were calculated to optimize the number of latent variables (LVs) that should be employed in PLS regression. It was defined as follows:

$$\mathrm{PRESS} = \sum_{i}^{m} (y_{i,\mathrm{pred}} - y_{i,\mathrm{act}})^2 \tag{8}$$

The best number of latent variables was determined in light of the work by Haaland and Thomas [64] where the smallest number of LVs whose corresponding PRESS was not significantly different from the global minimum PRESS was selected. In this study, the significance level for F-test was set as 0.35.

2.6. Prediction visualization

The visualization of *Pseudomonas* loads on chicken breast fillets comprised of three steps as illustrated in Fig. 3. First, if full wavelength models were to be used, the original hypercubes (x, y, λ) were preprocessed using spectral pretreatment methods described above. If simplified models were utilized, the original hypercubes were prepared into multi-band images (x, y, λ') in light of the wavelengths selected. Second, to facilitate model application, the transformed images were reshaped into a two-dimension matrix ($x \times y$, λ'), which was then multiplied by the vector of regression coefficients derived from the calibration model. After model application, the predicted bacterial loads were recorded in a new vector that can be reshaped back to a two-dimensional image. By masking the background with previously created masks and endowing the response values in the meat portion with corresponding colors, a prediction map for the

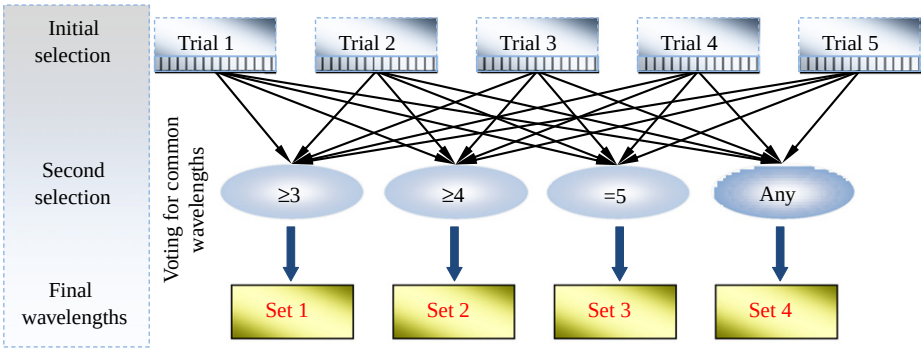

Fig. 2. Wavelength selection scheme based on genetic algorithm. Five trials are the five initial wavelength sets selected by different implementations of GA. The middle layer shows the strategies for second selection, i.e., the common wavelengths agreed by no less than 3, 4 or all 5 trials or any of the candidate wavelengths selected by the five runs of GA. The four sets in the bottom layer are the four new wavelength sets after second selection.

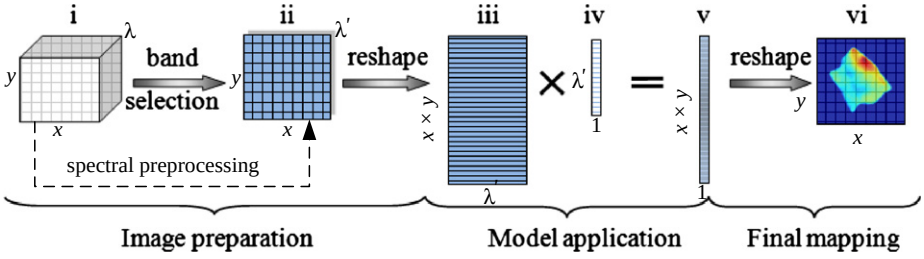

Fig. 3. The procedure for visualizing *Pseudomonas* loads in chicken using hyperspectral imaging.

sample is thereafter generated. This visualization map enables insights into concentration distribution of targets under study, and therefore it provides foundation for quality and safety control in the food industry.

3. Results and discussion

3.1. Sample and spectra

The chicken samples were stored for nine consecutive days during the experiment and this allowed for the acquisition of good variance of samples specifically in terms of chemical changes due to microbial spoilage. Such variations were confirmed by large standard deviation ($1.36\ \log_{10}\mathrm{CFU\ g^{-1}}$) of measured microbial counts.

Fig. 4a depicts the raw spectra of all samples in the calibration set, where two evident absorption peaks were found at near 960 and 1200 nm. The 960 nm band, known as water absorption band, is due to combination transition of symmetric and antisymmetric O–H stretch [65], while the band at 1200 nm could be ascribed to fatty acids or protein content that arises from C–H second overtone stretch [66]. Despite the similarity, the studied original spectra were different from each other due to great scatter effect as indicated by the large distance between spectral plots [67]. To correct the scatter effect, different spectral pretreatment methods, namely, SNV, MSC, Detrend, Der1, Der2 and their combinations, were employed and the resulting spectra are shown in Fig. 4(b-l). It is apparent that all the methods successfully suppressed the scatter effect except MSC, which hardly produced any improvement by presenting a similar pattern to that in the original spectra. Several studies reported that SNV and MSC worked similarly in data preprocessing [68,69], however, this conclusion was normally drawn in spectroscopic analyses, where the transform was carried out on the final averaged sample spectra. In this study, the individual spectrum at each pixel of a given sample image was firstly preprocessed and the average of all the pretreated meat spectra was then used to represent the sample. In other words, the preprocessing in traditional spectroscopic analyses is an among-samples method and the method applied in this study is within-sample-based. Since the reference spectrum for each implementation of MSC was chosen as the average of the spectra of only one sample image, the scatter effect corrected by using MSC is the local 'noise' within the sample other than among samples. Besides, MSC was also found to tend to induce spectral outliers [70]. The application of SNV does not need any reference spectrum and produces spectra with zero-mean and unit variance thus can work among samples and show better stability and generalization. Nevertheless, when Detrend, Der1 and Der2 were used following MSC, the scatter effect was also greatly diminished and the performance was quite similar to the corresponding cases when SNV and these methods were combined. It should be noted that the spectra shown in Fig. 4 are the mean-version spectra, however, sample spectra calculated as median spectra exhibit a similar form, and therefore they are not shown here.

In Fig. 4, it is also very easy to discover that derivative methods are the most effective in reducing scatter effects on sample spectra as illustrated by the concentrated plots, and this agreed well with some previous study [67]. Besides, derivative methods are also useful methods for revealing some new bands that may be important indicators of chemical information within sample spectra. Two sharp peaks around 930 and 1140 nm revealed from derivative spectra may be ascribed to C–H stretching third overtone and the overlap of the second overtone stretching modes of C–H and N–H groups [71], respectively.

3.2. PLS full-wavelength models

The performance of calibration models based on full wavelengths and different spectral preprocessing methods is shown in Table 3. Generally, models derived from mean spectra tended to have better performance than those based on median spectra with less number of latent variables involved. Such advantageous

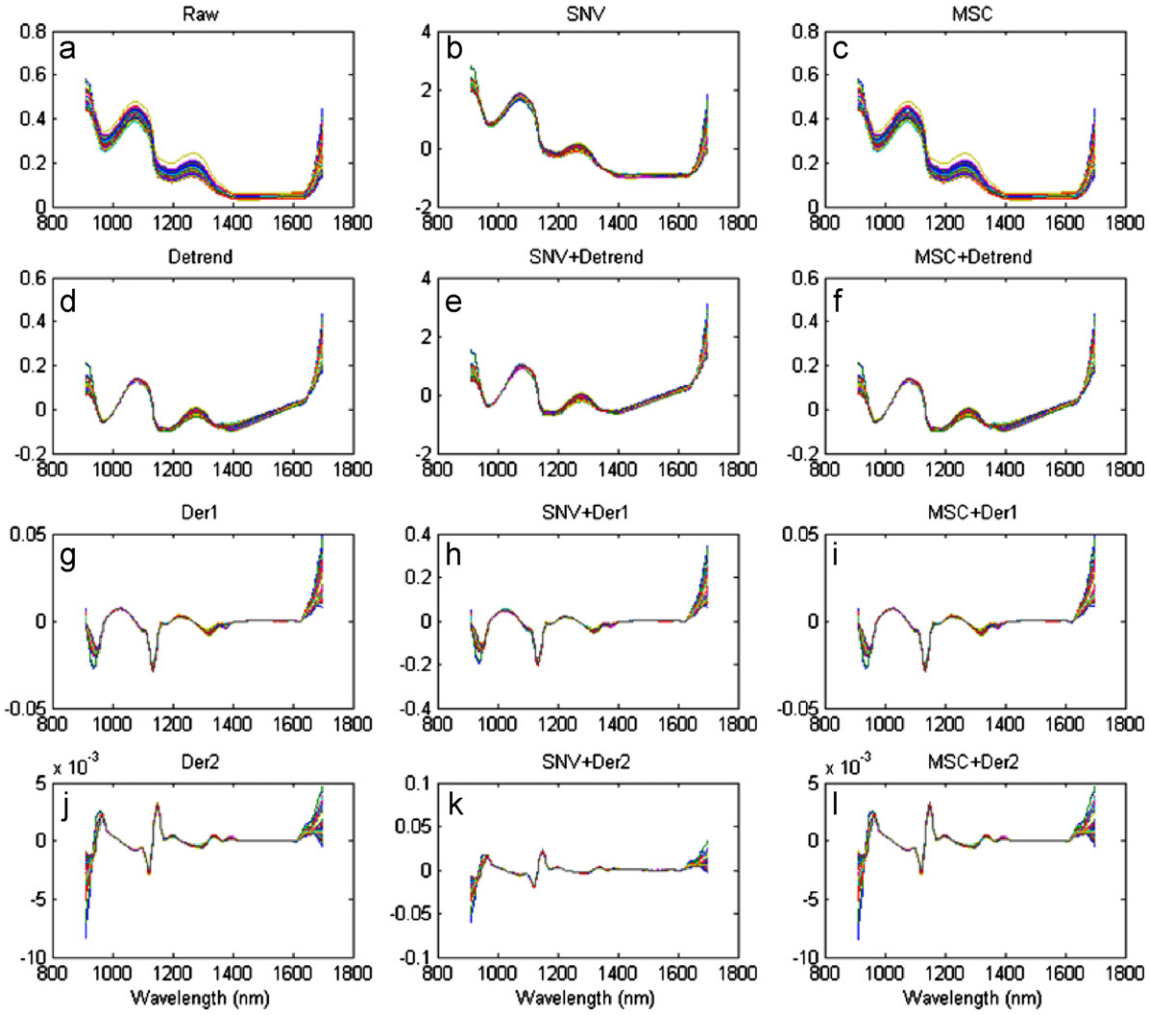

Fig. 4. Raw and preprocessed spectra of chicken meat at 910–1700 nm.

Table 3
Performance of full wavelength models based on different preprocessing methods.

Spectra	Preprocessing	R_C	R_{CV}	R_P	RMSEC ($\log_{10}$ CFU g^{-1})	RMSECV ($\log_{10}$ CFU g^{-1})	RMSEP ($\log_{10}$ CFU g^{-1})	LVs
Mean	Raw	0.89	0.81	0.74	0.59	0.75	0.91	5
	SNV	**0.88**	**0.83**	**0.81**	**0.60**	**0.71**	**0.80**	**4**
	MSC	0.89	0.82	0.75	0.58	0.73	0.90	5
	Detrend	0.88	0.83	0.81	0.59	0.71	0.78	4
	SNV+Detrend	0.88	0.83	0.80	0.60	0.71	0.79	4
	MSC+Detrend	0.89	0.83	0.68	0.59	0.72	1.01	4
	Der1	0.85	0.77	0.62	0.64	0.77	1.09	3
	SNV+Der1	0.93	0.86	0.83	0.48	0.65	0.75	5
	MSC+Der1	0.85	0.77	0.62	0.64	0.77	1.09	3
	Der2	0.96	0.82	0.79	0.37	0.83	0.93	11
	SNV+Der2	0.91	0.83	0.87	0.54	0.74	0.64	5
	MSC+Der2	0.96	0.85	0.87	0.38	0.75	0.83	11
Median	Raw	0.93	0.85	0.53	0.47	0.68	1.25	6
	SNV	0.87	0.81	0.75	0.62	0.74	0.92	4
	MSC	1.00	0.83	0.61	0.11	0.73	1.20	15
	Detrend	0.88	0.81	0.78	0.59	0.74	0.86	4
	SNV+Detrend	0.88	0.83	0.70	0.59	0.72	1.02	4
	MSC+Detrend	0.94	0.87	0.63	0.42	0.65	1.13	6
	Der1	0.84	0.78	0.62	0.64	0.75	1.10	3
	SNV+Der1	0.78	0.74	0.60	0.76	0.82	1.13	2
	MSC+Der1	0.85	0.79	0.61	0.63	0.74	1.12	3
	Der2	1.00	0.86	0.84	0.09	0.69	0.84	15
	SNV+Der2	1.00	0.81	0.82	0.06	0.81	1.18	17
	MSC+Der2	1.00	0.81	0.82	0.05	0.78	1.01	17

trend was tenable when considering various spectral preprocessing methods though when the spectra were pretreated with first and second derivative alone, the prediction performance of models based on mean spectra was not as good as the corresponding models based on median spectra. In addition, it is clearly shown in Table 3 that the models based on median spectra in this study were generally poor or tend to be severely overfitted. Of particular attention were MSC-PLS, Der2-PLS, MSC+Der2-PLS and SNV+Der2-PLS models where R_C were as close as 1.00. Since these models utilized too many LVs (≥ 15), it was believed that the good calibration performance was a result of modeling the noise that was not eliminated by the corresponding pretreatments [72], therefore, these models were not reliable and the good performance was meaningless. Only the models based on mean spectra are to be discussed in the following sections.

In terms of models based on mean spectra, using most of the spectral preprocessing methods would lead to calibration models of better prediction capacity than that when raw spectra were used. However, application of certain pretreatment methods, e.g., Der1, MSC+Detrend and MSC+Der1, degraded the prediction precision. This was probably because the information removed by applying these methods may be useful indicators for measuring the target of interest [73]. Moreover, as expected from the spectral plots in Fig. 4 where similar patterns were found between MSC-processed spectra and the original spectra, the performance of models based on these two spectral matrixes also gave comparable results. Much better models were found when the spectral images were preprocessed with SNV, Detrend, SNV+Detrend, SNV+Der1 and SNV+Der2 where R_P were all larger than or equal to 0.80. This finding demonstrated that SNV is a useful tool for improving model performance by correcting scatter effects, which agreed well with the study by Taghizadeh et al. [74]. Though preprocessing using SNV combined with either Der1 or Der2 would result in higher correlation coefficients, the relatively large difference between RMSEC and RMSECV would indicate less stability of such models. Considering SNV is a simple method to apply, it was favoured as an efficient pretreatment method in prediction of *Pseudomonas* loads on chicken fillets using hyperspectral imaging. Its corresponding full wavelength model was denoted as SNV-PLS model.

3.3. GA-PLS models

Simplified models were developed by using a few wavelengths that were identified by the proposed selection scheme based on GA. Wavelengths were only selected based on the raw images rather than preprocessed hypercubes considering that statistics (e.g., mean, SD and/or fitting parameters) across the whole wavelengths will be required but absent when dealing with multispectral images. The performance of established simplified calibration models is given in Table 4. It is clear that most of the simplified models outperformed the one based on full-wavelengths especially in regards of prediction. The only exception was found when a '≥ 3' rule was applied during secondary selection, where R_P and RMSEP were 0.71 and 1.01 $\log_{10}$ CFU g^{-1}, respectively. The noticeable improvement may be ascribed to the elimination of chemical information that is not related to the detection of *Pseudomonas* loads as well as to the good combination of chemical information at the selected wavelengths. However, despite the reasonable calibration performance of GA-PLS-3 model, the degradation of prediction indicated its lack of adaptability for new unknown samples. The best prediction was gained after applying the GA-PLS-4 model, where the correlation coefficients and RMSEs were 0.91 and 0.55 $\log_{10}$ CFU g^{-1}, 0.87 and 0.65 $\log_{10}$ CFU g^{-1} as well as 0.88 and 0.64 $\log_{10}$ CFU g^{-1} for calibration, cross-validation and prediction, respectively. Although the GA-PLS-BT model produced a comparable result, it utilized more wavelengths (24 wavelengths in nine discontinues segments) than GA-PLS-4, where only 14 wavelengths in five discontinues segments were employed (Fig. 5a). Moreover, GA-PLS-4 model was also considered to be better in terms of smaller difference in RMSEs between calibration and cross validation or prediction when compared with that for the GA-PLS-BT model.

Fig. 5 gives an overview of the GA-PLS-4 model. The red bar in Fig. 5a illustrates the selected wavelengths that fall in five regions, i.e., 1138–1155, 1195–1198, 1392–1395, 1452–1455 and 1525–1529 nm. As mentioned above, the region of 1195–1198 nm can be assigned to fatty acids or protein and it can be easily found in the original spectra. In the meantime, the region between 1138 and 1155 nm, which could be associated with some aromatic groups due to C–H stretching second overtone as well as amino acids due to second overtone N–H stretch [71,75], were also proved to be important in analyzing *Pseudomonas* loads in chicken. The wavelength segment near 1450 nm can be ascribed to water content and 1525–1529 nm is closely related to proteins [71,76]. In addition, the two bands (1392 and 1395 nm) are due to combination of C–H stretching and deformation vibration modes [75]. Fig. 5b shows that 4 LVs were used for the best model. However, it should be noted that it is a coincidence that the number of LVs corresponding to the global minimum PRESS was selected since when less LVs were used, their corresponding PRESS values were all significantly different from the global minimum ($p < 0.35$). The remaining two plots in Fig. 5 further elucidated the good precision of the GA-PLS-4 model in both calibration and prediction by showing sample plots scattered around the ideal line (dotted). The added ideal line suggests that the predicted and measured parameters are perfectly the same. GA-PLS-4 model was thus recognized as the best simplified model and it was used to visualize the distribution of *Pseudomonas* loads on chicken fillets.

3.4. Prediction visualization

After calibration models were established, only the best models were used for producing prediction maps where the distribution of *Pseudomonas* loads on chicken meat surface became visible. In this study, both optimal models based on full wavelengths (SNV-PLS model) and GA-selected wavelengths (GA-PLS-4) were used. The resultant visualization maps on three samples representing different levels of microbial loads are given in Fig. 6. For comparison, conventional colour images (RGB

Table 4
Performance of PLS models based on selected wavelengths. Wavelengths were selected by using the proposed two-step selection scheme based on GA.

	R_C	R_{CV}	R_P	RMSEC ($\log_{10}$ CFU g^{-1})	RMSECV ($\log_{10}$ CFU g^{-1})	RMSEP ($\log_{10}$ CFU g^{-1})	Bands
GA-PLS-BT	0.93	0.87	0.88	0.48	0.64	0.67	24
GA-PLS-3	0.91	0.84	0.71	0.53	0.68	1.01	20
GA-PLS-4	**0.91**	**0.87**	**0.88**	**0.55**	**0.65**	**0.64**	**14**
GA-PLS-5	0.87	0.82	0.87	0.62	0.74	0.65	11
GA-PLS-Any	0.88	0.82	0.85	0.65	0.79	0.74	47

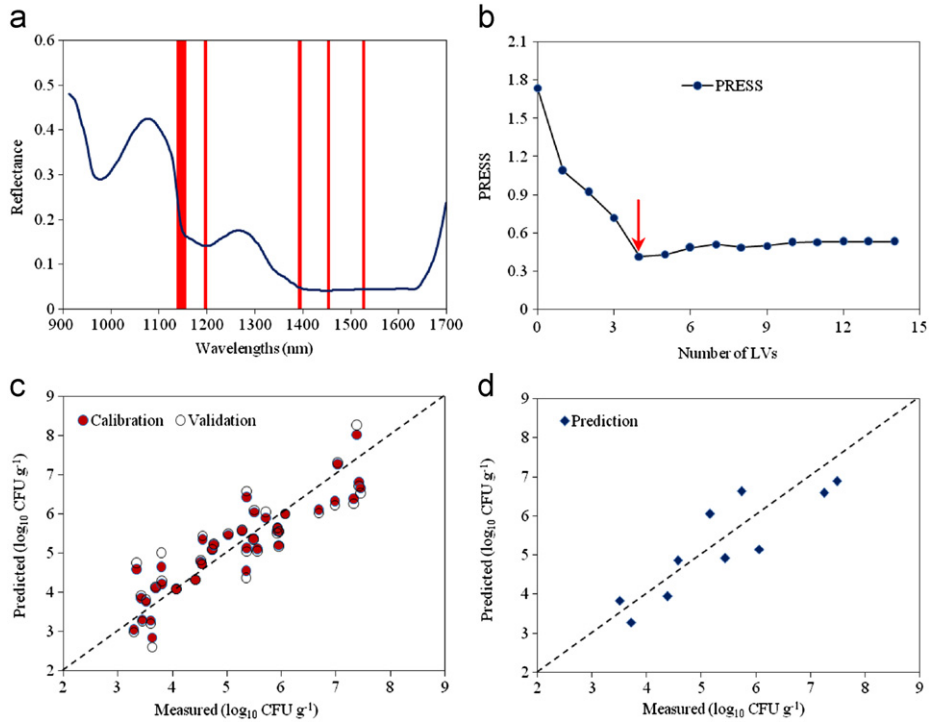

Fig. 5. The performance of the best simplified models (GA-PLS-4). (a) indication of selected wavelengths; (b) Scree plot for selection of optimal number of LVs; (c) measured vs. predicted plot for calibration and cross-validation and (d) measured vs. predicted plot for prediction.

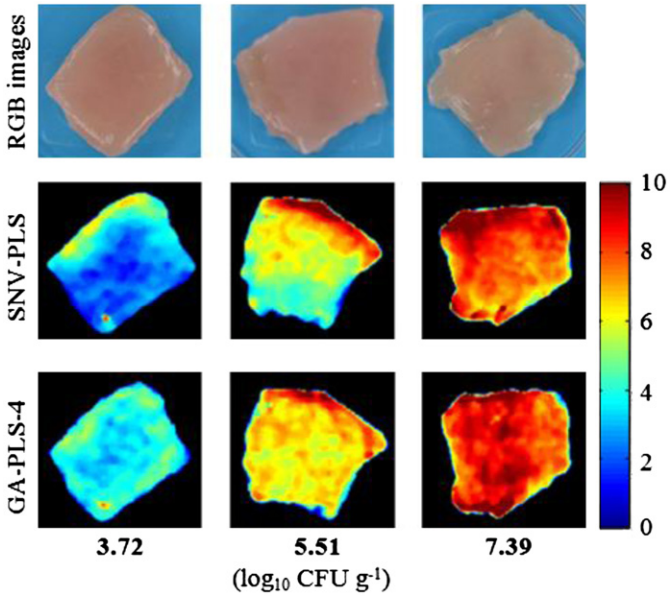

Fig. 6. Traditional colour images (RGB images) of samples and their corresponding prediction maps by applying the optimal full wavelength model (SNV-PLS) and the best simplified model (GA-PLS-4). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

images) of the same samples taken by a conventional camera as described by Valous et al. [77] are also given in the first row of Fig. 6. By checking RGB images, one can easily read the colour and shape information but can hardly obtain any message about the microbial condition of the samples. In marked contrast, the prediction maps in the following two rows of Fig. 6, resulting from the SNV-PLS model and the GA-PLS-4 model, respectively, clearly exhibit the distribution gradient of *Pseudomonas* loads in the samples. Such visualization should be owed to the superb merit of hyperspectral imaging for retaining both spectral and spatial information of samples [78–80]. For the prediction maps, guided by the colour scale on the right-hand side of Fig. 6, the changes of bacterial loads could be easily accessed by scrutinizing the colour variations, where evolution of colour from dark blue to dark red would indicate the transition of bacterial loads from very low (0 $\log_{10}$ CFU g^{-1}) to very high (10 $\log_{10}$ CFU g^{-1}) levels. The background was artificially set to black colour to avoid any confusing information. By using HSI for bacterial determination, it was supposed that the system was recording the chemical changes in meat other than the compositional components of bacteria since the mass of bacteria was too trivial to be detected (approximately 0.01 mg for 10^7 cells). The chemical changes of meat were mostly caused by the bacteria metabolizing meat compounds and therefore the chemical variations of meat should, to some extent, reflect the growth of bacteria (i.e., changes in number). Moreover, the calibration models were established to find out the relationship between spectral images and *Pseudomonas* counts, therefore, when a good model was established and applied for prediction, it should lead to maps illustrating the predicted bacterial loads rather than certain chemical components.

Generally speaking, applying the two models has led to two consistent visualization maps. The colour for the maps of the sample with low bacterial load (3.72 $\log_{10}$ CFU g^{-1}) is mainly blue or cyan, while yellow and red are leading in the maps for samples with bacterial loads of 5.51 and 7.39 $\log_{10}$ CFU g^{-1}, respectively. Moreover, unanimous trend was found between two groups of maps where the predictions in specific areas of samples in one group could normally be well reflected in the maps of the other group. Of particular interest are two small red or orange spots with yellow halos in the prediction maps for the sample with microbial loads of 3.72 $\log_{10}$ CFU g^{-1}. Another interesting phenomenon is that the edges of meat are always entrusted with much higher or much lower predicted loads (in comparison with the measured values). By checking the original images, the red spots were actually the predictions of bacterial

loads on chicken fat tissues rather on the muscle. Since the spectra of fat and muscle are quite different [49], it may be interesting to investigate the effect of fat on detecting bacterial loads on meat in future studies. Another point, i.e., how edge effect is affecting the prediction of microbial loads, can also be investigated. In spite of similarities, difference was found between maps for the same samples, and such difference was believed to be the result of difference in models, where GA-PLS-4 outperformed SNV-PLS as indicated by Tables 3 and 4.

4. Conclusions

Near-infrared hyperspectral imaging is a useful tool for quantification of *Pseudomonas* loads in raw chicken breast fillets. In development of full wavelength models, different approaches for extracting spectra and various spectral preprocessing methods were employed to enhance model performance. It was demonstrated that use of mean spectra as feature spectra for model calibration was a better choice when predicting *Pseudomonas* loads in raw chicken breast fillets. The optimal and most stable model was gained when the spectral images were pretreated with SNV and average spectra were computed to represent samples. The correlation coefficients and RMSEs of the resultant SNV-PLS model were 0.88 and 0.60 $\log_{10}$ CFU g^{-1}, 0.83 and 0.71 $\log_{10}$ CFU g^{-1} as well as 0.81 and 0.80 $\log_{10}$ CFU g^{-1} for calibration, cross-validation and prediction, respectively. In terms of the simplified models where wavelengths were selected by a two-step scheme based on genetic algorithm, an even better model was gained. It utilized only 14 wavelengths in 5 narrow discontinues wavelength ranges and gave correlation coefficients higher than 0.87 and RMSEs lower than 0.65 $\log_{10}$ CFU g^{-1}. The two consistent prediction maps produced by applying the two optimal models illustrated the promising capability of hyperspectral imaging and feasibility of multispectral imaging for visualizing the distribution of *Pseudomonas* on meat surface, which cannot be realized by either conventional imaging or spectroscopic technologies. However, future research should be carried out to further verify the effectiveness of hyperspectral imaging and to investigate the feasibility of multispectral imaging systems for real-time predicting *Pseudomonas* loads in chicken or other meat.

Acknowledgements

This study is funded by the UCD-CSC scholarship awarded to Mr. Yaoze Feng. Dr. Noha Morcy, Dr. Amalia Scannell, Dr. Gamal ElMasry and Dr. Des Walsh are gratefully acknowledged for kind assistance.

References

[1] K. Mc Donald, D.-W. Sun, J. Food Eng. 48 (3) (2001) 195–202, http://dx.doi.org/10.1016/S0260-8774(00)00158-8.
[2] L.J. Wang, D.-W. Sun, Int. J. Refrig. — Rev. Int. Froid 25 (7) (2002) 854–861, http://dx.doi.org/10.1016/S0140-7007(01)00094-9, PII S0140-7007(01)00094-9.
[3] B. Li, D.-W. Sun, J. Food Eng. 55 (3) (2002) 277–282, http://dx.doi.org/10.1016/S0260-8774(02)00102-4, PII S0260-8774(02)00102-4.
[4] D.-W. Sun, Z.H. Hu, Int. J. Refrig. — Rev. Int. Froid 26 (1) (2003) 19–27, http://dx.doi.org/10.1016/S0140-7007(02)00038-5.
[5] D.-W. Sun, B. Li, J. Food Eng. 57 (4) (2003) 337–345, http://dx.doi.org/10.1016/S0260-8774(02)00354-0.
[6] D.-W. Sun, L.Y. Zheng, J. Food Eng. 77 (2) (2006) 203–214, http://dx.doi.org/10.1016/j.jfoodeng.2005.06.023.
[7] D.-W. Sun, I.W. Eames, Int. J. Energy Resl. 20 (10) (1996) 871–885, http://dx.doi.org/10.1002/(SICI)1099-114X(199610)20:10 < 871::AID-ER201 > 3.3.CO;2-W.
[8] D.-W. Sun, I.W. Eames, S. Aphornratana, Int. J. Refrig. — Int. Food 19 (3) (1996) 172–180, http://dx.doi.org/10.1016/0140-7007(96)00010-2.
[9] D.-W. Sun, Energy Convers 38 (5) (1997) 479–491, http://dx.doi.org/10.1016/S0196-8904(96)00063-5.
[10] D.-W. Sun, Energy Convers. Mange 39 (5-6) (1998) 357–368, http://dx.doi.org/10.1016/S0196-8904(97)00027-7.
[11] D.-W. Sun, Energy Coners. Manage. 40 (8) (1999) 873–884, http://dx.doi.org/10.1016/S0196-8904(98)00151-4.
[12] B. Ray, Fundamental Food Microbiology, second ed., CRC press, 2000.
[13] A.I. Doulgeraki, D. Ercolini, F. Villani, G.-J.E. Nychas, Int. J. Food Microbiol. 157 (2012) 130–141.
[14] S. Bruckner, A. Albrecht, B. Petersen, J. Kreyenschmidt, J. Food Qual. 35 (2012) 372–382.
[15] G. Nychas, V.M. Dillon, R. Board, Biotechnol. Appl. Biochem. 10 (1988) 203–231.
[16] A. Davies, R. Board, The Microbiology of Meat and Poultry, Blackie Academic & Professional, London, 1998.
[17] D.-W. Sun, J. Food Engi. 44 (4) (2000) 245–249, http://dx.doi.org/10.1016/S0260-8774(00)00024-8.
[18] D.-W. Sun, T. Brosnan, J. Food Engi. 57 (1) (2003) 81–89, http://dx.doi.org/10.1016/S0260-8774(02)00275-3.
[19] C.J. Du, D.-W. Sun, J. Food Engi. 72 (1) (2006) 39–55, http://dx.doi.org/10.1016/j.jfoodeng.2004.11.017.
[20] C.X. Zheng, D.-W. Sun, L.Y. Zheng, Trends Food Sci Technol 17 (3) (2006) 113–128, http://dx.doi.org/10.1016/j.tifs.2005.11.006.
[21] C.X. Zheng, D.-W. Sun, L.Y. Zheng, Trends Food sci Technol 17 (12) (2006) 642–655, http://dx.doi.org/10.1016/j.tifs.2006.06.005.
[22] J.M. Amigo, Anal. Bioanal.Chem. 398 (2010) 93–109.
[23] Y.-Z. Feng, D.-W. Sun, Crit. Rev. Food Sci. Nutr. 52 (2012) 1039–1058.
[24] H. Fenniri, O. Terreau, S.K. Chun, S.J. Oh, W.F. Finney, M.D. Morris, J. Comb. Chem. 8 (2006) 192–198.
[25] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, Trends Food Sci. Technol. 18 (2007) 590–598.
[26] T. Vo-Dinh, D.L. Stokes, M.B. Wabuyele, M.E. Martin, J.M. Song, R. Jagannathan, E. Michaud, R.J. Lee, X.G. Pan, IEEE Eng. Med. Biol. Mag. 23 (2004) 40–49.
[27] C. Esquerre, A. Gowen, J. Burger, G. Downey, C. O'Donnell, Chemom. Intell. Lab. Syst. 117 (2012) 129–137.
[28] D.F. Barbin, G. ElMasry, D.-W. Sun, P. Allen, Anal. Chim. Acta 719 (2012) 30–42.
[29] M. Kamruzzaman, G. ElMasry, D.-W. Sun, P. Allen, Anal. Chim. Acta 714 (2012) 57–67.
[30] G. ElMasry, D.-W. Sun, P. Allen, Food Res. Int. 44 (2011) 2624–2633.
[31] Y.-Z. Feng, G. ElMasry, D.-W. Sun, A. Scannell, D. Walsh, N. Morcy, Food Chem. 138 (2013) 1829–1836.
[32] G. ElMasry, N. Wang, C. Vigneault, J. Qiao, A. ElSayed, LWT— Food Sci. Technol. Int. 41 (2008) 337–345.
[33] K. Chao, Y.R. Chen, W.R. Hruschka, B. Park, Appl. Eng. Agric. 17 (2001) 99–106.
[34] S. Nakariyakul, D.P. Casasent, J. Food Eng. 94 (2009) 358–365.
[35] D. Wu, S. Wang, N. Wang, P. Nie, Y. He, D.W. Sun, J. Yao, Food Bioprocess Technol. (2012) 1–16.
[36] P.M. Mehl, Y.R. Chen, M.S. Kim, D.E. Chan, J. Food Eng. 61 (2004) 67–81.
[37] B. Park, K.C. Lawrence, W.R. Windham, D.P. Smith, J. Food Eng. 75 (2006) 340–348.
[38] I. Kim, M. Kim, Y. Chen, S. Kong, Trans. ASAE 47 (2004) 1785–1792.
[39] M. Kamruzzaman, D. Barbin, G. ElMasry, D.-W. Sun, P. Allen, Innov. Food Sci. Emerg. 16 (2012) 316–325.
[40] G.K. Naganathan, L.M. Grimes, J. Subbiah, C.R. Calkins, A. Samal, G.E. Meyer, Comput. Electron. Agric. 64 (2008) 225–233.
[41] J. Qiao, M.O. Ngadi, N. Wang, C. Gariépy, S.O. Prasher, J. Food Eng. 83 (2007) 10–16.
[42] J. Gómez-Sanchis, J.D. Martín-Guerrero, E. Soria-Olivas, M. Martínez-Sober, R. Magdalena-Benedito, J. Blasco, Expert Syst. Appl. 39 (2012) 780–785.
[43] J. Qin, T.F. Burks, M.A. Ritenour, W.G. Bonn, J. Food Eng. 93 (2009) 183–191.
[44] A.M. Lefcourt, M.S. Kim, Y.-R. Chen, Comput. Electron. Agric. 48 (2005) 63–74.
[45] R. Grau, A.J. Sánchez, J. Girón, E. Iborra, A. Fuentes, J.M. Barat, Food Res. Int. 44 (2011) 331–337.
[46] Y. Peng, J. Zhang, W. Wang, Y. Li, J. Wu, H. Huang, X. Gao, W. Jiang, J. Food Eng. 102 (2011) 163–169.
[47] P. Williams, P. Geladi, G. Fox, M. Manley, Anal. Chim. Acta 653 (2009) 121–130.
[48] D. Wu, D.-W. Sun, Y. He, Innov. Food Sci. Emerg. 16 (2012) 361–372.
[49] M. Kamruzzaman, G. ElMasry, D.-W. Sun, P. Allen, J. Food Eng. 104 (2011) 332–340.
[50] G. ElMasry, D.-W. Sun, P. Allen, J. Food Eng. 110 (2012) 127–140.
[51] D. Barbin, G. Elmasry, D.-W. Sun, P. Allen, Meat Sci. 90 (2012) 259–268.
[52] R. Lu, Y. Chen, in: Y.-R. Chen (Ed.), Pathoaen Detection and Remediation for Safe Eating, Boston. Massachusetts, 1998, pp. 121–133.
[53] F. Shen, D. Yang, Y. Ying, B. Li, Y. Zheng, T. Jiang, Food Bioprocess Technol. 5 (2012) 786–795.
[54] Å. Rinnan, F v.d. Berg, S.B. Engelsen, TrAC, Trends Anal. Chem. 28 (2009) 1201–1222.
[55] J. Luypaert, M. Zhang, D. Massart, Anal. Chim. Acta 478 (2003) 303–312.
[56] F. Liu, Y. He, L. Wang, G. Sun, Food Bioprocess Technol. 4 (2011) 1331–1340.
[57] V. Sinija, H. Mishra, Food Bioprocess Technol. 4 (2011) 136–141.
[58] The Mathworks, Matlab Version 7 User Guide, Natick, MA, USA., 2004.

[59] R.K.H. Galvão, M.C.U. Araujo, G.E. José, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha, Talanta 67 (2005) 736–740.
[60] P. Geladi, B.R. Kowalski, Anal. Chim. Acta 185 (1986) 1–17.
[61] R. Leardi, R. Boggia, M. Terrile, J. Chemom. 6 (1992) 267–281.
[62] R. Leardi, J. Chemom. 14 (2000) 643–655.
[63] R. Leardi, L.A. González, Chemom. Intell. Lab. Syst. 41 (1998) 195–207.
[64] D.M. Haaland, E.V. Thomas, Anal. Chem. 60 (1988) 1193–1202.
[65] M. Lin, M. Al-Holy, M. Mousavi-Hesary, H. Al-Qadiri, A.G. Cavinato, B.A. Rasco, Lett. Appl. Microbiol. 39 (2004) 148–155.
[66] P. Ritthiruangdej, R. Ritthiron, H. Shinzawa, Y. Ozaki, Food Chem. 129 (2011) 684–692.
[67] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C.D.L. Pezuela, Appl. Spectrosc. 51 (1997) 240–246.
[68] C. Pizarro, I. Esteban-Díez, A.J. Nistal, J.M. González-Sáiz, Anal. Chim. Acta 509 (2004) 217–227.
[69] F. Zhu, S. Cheng, D. Wu, Y. He, Food Bioprocess Technol. 4 (2011) 597–602.
[70] T. Fearn, C. Riccioli, A. Garrido-Varo, J.E. Guerrero-Ginel, Chemom. Intell. Lab. Syst. 96 (2009) 22–26.
[71] C.E. Miller, Chemical principles of near infrared technology, in: P. Williams, K. Norris (Eds.), Near Infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists, Minnesota, USA, 2001, pp. 19–37.
[72] H. Martens, M. Martens, Food Qual. Preference 11 (2000) 5–16.
[73] J.S. Shenk, J.J. Workman, M.O. Westerhaus, Application of NIR spectroscopy to agricultural products, in: D.A. Burns, E.W. Ciurczak (Eds.), Handbook of Near-Infrared analysis, CRC Press, Boca Raton, Florida, USA, 2008.
[74] M. Taghizadeh, A. Gowen, P. Ward, C.P. O'Donnell, Innovative Food Sci. Emerg. Technol. 11 (2010) 423–431.
[75] B.G. Osborne, T. Fearn, P.H. Hindle, Practical Near Infrared Spectroscopy with Applications in Food and Beverage Analysis, second ed., Longman Group, Burnt Mill, Harlow, Essex, England, UK, 1993.
[76] D. Alexandrakis, G. Downey, A. Scannell, Food Bioprocess Technol. 5 (2012) 338–347.
[77] N.A. Valous, F. Mendoza, D.-W. Sun, P. Allen, Meat Sci. 81 (2009) 132–141.
[78] Y.-Z. Feng, D.-W. Sun, Talanta 105 (2013) 244–249.
[79] P. Menesatti, A. Zanella, S. D'Andrea, C. Costa, G. Paglia, F. Pallottino, Food Bioprocess Technol. 2 (2009) 308–314.
[80] F. Zhu, D. Zhang, Y. He, F. Liu, D.-W. Sun, Food Bioprocess Technol. (2012) 1–7.