Talanta 111 (2013) 39–46

Contents lists available at SciVerse ScienceDirect

Talanta

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

Potential of time series-hyperspectral imaging (TS-HSI) for non-invasive determination of microbial spoilage of salmon flesh

Di Wu, Da-Wen Sun *

Food Refrigeration and Computerised Food Technology (FRCFT), School of Biosystems Engineering, University College Dublin, National University of Ireland, Agriculture & Food Science Centre, Belfield, Dublin 4, Ireland

ARTICLE INFO

Article history:
Received 12 November 2012
Received in revised form
5 March 2013
Accepted 16 March 2013
Available online 22 March 2013

Keywords:
Time series hyperspectral imaging
Salmon
Spoilage
Total viable count (TVC)
Imaging spectroscopy
Fish

ABSTRACT

This study investigated the potential of using time series-hyperspectral imaging (TS-HSI) in visible and near infrared region (400–1700 nm) for rapid and non-invasive determination of surface total viable count (TVC) of salmon flesh during spoilage process. Hyperspectral cubes were acquired at different spoilage stages for salmon chops and their spectral data were extracted. The reference TVC values of the same samples were measured using standard plate count method and then calibrated with their corresponding spectral data based on two calibration methods of partial least square regression (PLSR) and least-squares support vector machines (LS-SVM), respectively. Competitive adaptive reweighted sampling (CARS) was conducted to identify the most important wavelengths/variables that had the greatest influence on the TVC prediction throughout the whole wavelength range. As a result, eight variables representing the wavelengths of 495 nm, 535 nm, 550 nm, 585 nm, 625 nm, 660 nm, 785 nm, and 915 nm were selected, which were used to reduce the high dimensionality of the hyperspectral data. On the basis of the selected variables, the models of PLSR and LS-SVM were established and their performances were compared. The CARS-PLSR model established using Spectral Set I (400–1000 nm) was considered to be the best for the TVC determination of salmon flesh. The model led to a coefficient of determination (r_p^2) of 0.985 and residual predictive deviation (RPD) of 5.127. At last, the best model was used to predict the TVC values of each pixel within the ROI of salmon chops for visualizing the TVC distribution of salmon flesh. The research demonstrated that TS-HSI technique has a potential for rapid and non-destructive determination of bacterial spoilage in salmon flesh during the spoilage process.

1. Introduction

Food safety is a global issue, which receives increasing attention for consumers all around the world. Over recent years, some microbial food safety incidents occurred in both developed and developing countries. As a scientific discipline, the aim of food safety is to ensure that food produced, distributed or marketed from farm to fork meets the highest standards of safety, prevent foodborne illness, and avoid potentially severe health hazards. Salmon is an important product of aquaculture that has a production of 1400,000 T in 2009 with a value of more than 7 billion US dollars [1]. Freshness is one of the most critical quality attributes of salmon. Consumers prefer the selection of fresh salmon, which is however a perishable food that has a moist and nutritious surface for the growth of spoilage bacteria conductively. The increased trade between countries makes a need for a good freshness grading system for salmon [2]. Microbial spoilage is a main source of foodborne outbreaks. About one-third of the world's food production is lost annually as a result of microbial spoilage [3]. Salmon products are described as spoiled if organoleptic changes, such as discoloration, emission of off-odours, and slime formation, reached unacceptable levels, thus unacceptable by consumers. It is a great challenge to the safety of salmon due to microbial growth because potentially harmful microorganisms can grow rapidly from very low numbers in salmon and proliferate in the human body once ingested [4]. It is generally accepted that spoilage is a result of the decomposition and formation of metabolites caused by the growth and enzymatic activity of microorganisms [5]. The characteristic of the quality deterioration of fresh fish could be perceived as an initial loss of fresh fish flavour, like sweet and seaweedy. Over a period, off-odours and flavours could be detected. Progressively, the odour and flavour become from being neutral or non-specific into being unacceptable, leading to rejection of the fish. Storage temperature and fish species are two main aspects affecting the time to spoilage [6]. In recent years, high profile food safety issues always bring social unrest, resulting in increased public concern, as well as awareness.

* Corresponding author. Tel.: +353 1 716 7342; fax: +353 1 7167 493.
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.2013.03.041

These outbreaks of food-borne illnesses have led to the requirement for advances in a rapid (available for online) and accurate inspection system for microbiologically spoiled or contaminated fish. Strict policies and effective measures should be conducted by board of governors for the superintendence of food safety. The realization of accurate and rapid determination of spoilage degree of fish is important to meet consumer demands and to meet legislative requirements. Such determination is also helpful to predict remaining shelf life of fish products.

Current methods readily available for assessment of freshness and spoilage in seafood could be sensory based, microbiological, chemical and biochemical. Sensory methods are carried out by trained or experienced assessors. The quality index method (QIM) is a commonly used grading system for the assessment of fish freshness. However, considerable skills and experience are required for the execution of sensory panels to obtain reliable results, which cause the sensory tests costly and time-consuming, especially when a lot of trained assessors are required. Another limiting factor of sensory methods is that they depend largely upon subjective evaluation of the panellist, leading to results that are subjective and data from different batches of samples or different laboratories cannot be compared directly. Microbiological test procedures have been developed, which have two main groups: enumeration and presence/absence tests. Enumeration methods generally include microscopy, ATP bioluminescence or the measurement of electrical phenomena [7]. Chemical and biochemical examination procedures are carried out based on some indices of quality or spoilage in fish during storage due to microbial activity, autolytic enzymes or chemical reactions, such as total volatile basic nitrogen (TVB-N) that primarily includes trimethylamine (TMA) and other volatile amines, biogenic amines, and volatiles [8]. Visual spoilage indicators and electronic noses are two main tools for the detection of volatile amines [9]. Besides, biochemical methods, such as immunological methods, nucleic acid-based procedures, serological and molecular approaches are also researched for detecting bacterial spoilage in muscle foods [10]. However, these methods have been described time-consuming, labor-intensive, destructive, and expensive, and therefore are inappropriate for on-line or at-line inspection in fish processing. Special personnel training is also required for these methods for sample manipulation, or sample enrichment steps to permit cell recovery and microbial growth before detection [10]. In order to overcome these shortcomings, there is also a need for advanced technology that can provide rapid, accurate and non-destructive detection of bacterially spoilage or contamination in fresh salmon as part of on-line process control. In addition, with such a technology, the producers of salmon products can inspect quality changes during storage and predict shelf life.

Recently, optical sensing techniques have been investigated as a potential tool for the automatic quality and safety evaluation of food products, typically spectroscopy [11] and computer vision [12,13]. However, spectroscopy is lack of spatial information of sample, which greatly limits its application to quantify spatial-related distribution. On the other hand, computer vision [14-20] that operate at visible wavelengths in the forms of monochromatic or color images is lack of spectral information. Therefore, hyperspectral imaging technique is proposed, which integrates both spectroscopic and computer vision techniques into one system, resulting in acquiring a spatial map of spectral variation of sample. A typical hyperspectral image contains the spectrum at each pixel location in a two dimensional spatial matrix, resulting a three-dimensional datacube, which has one spectral dimension and two spatial dimensions. The hyperspectral imaging approach is powerful to determine the distribution of different quality attributes in a sample rapidly and non-invasively, resulting in many successful applications in the quality control of many food products [21,22], such as beef [23], pork [24], lamb [25], prawn [26], citrus [27], and cucumber [28]. Especially in the quality determination of salmon, the attributes like fat [29], colour [30], ice fraction [31], pH [32], moisture [33], texture [34], and NaCl content [35] have been successfully assessed using hyperspectral imaging, showing that hyperspectral imaging is a preferable analytical tool in inspecting qualities of salmon, and is believed to be dominant in the future.

A thorough understanding of the spoilage process of fish flesh is necessary for designing an optimal, product specific quality assurance programme or for predicting the shelf-life of a product at specific conditions (temperature, packaging) if microbiological data are to be utilized [6]. Especially, the inspection of spoilage process using a time-resolution method is important to gain the knowledge of spoilage. Recently, Van Der Weerd and Kazarian [36] presented a comprehensive introduction to time series-hyperspectral imaging (TS-HSI). TS-HSI is defined as a series of hyperspectral images of the same or similar samples acquired at different times [37]. TS-HSI, also called multivariate imaging time series or multivariate movies, is needed when both spatial and temporal resolutions are required simultaneously. TS-HSI is useful for achieving complete understanding of processes in a sample in cases where a heterogeneous sample changes with time [38] and it is thus expected that the application of TS-HSI will become more popular for analyzing quality changes of foods over time. Unfortunately, study on using TS-HIS is limited. Recently, Wu et al., [38] applied TS-HSI for rapid and non-invasive determination of water distribution within beef during dehydration. However, to the best of our knowledge, no research endeavourers have been reported yet for predicting bacterial spoilage of salmon or other fish during spoilage process using the TS-HSI technique.

Given the limited effort on the investigation of rapid techniques for determination spoilage levels of salmon flesh, the major objective of this study was to investigate the potential of TS-HSI for assessing bacterial spoilage of salmon flesh and demonstrating spoilage distribution within salmon flesh during the spoilage process. The spoilage degree was quantified using the total viable count (TVC) per gram. The specific objectives of this study were to:

- acquire TS-HSI images of salmon chops in the visible and NIR spectral region (400–1700 nm) during the spoilage process using two hyperspectral imaging systems;
- compare the prediction abilities of TS-HSI in 400–1000 nm and 900–1700 nm to find the best imaging spectrograph for TVC determination;
- identify important wavelengths that are most informative for the assessment of TVC of salmon by using competitive adaptive reweighted sampling (CARS) method;
- establish multivariate calibration models using partial least squares regression (PLSR) and least-squares support vector machines (LS-SVM), respectively, based on the whole range spectra or spectra at the identified wavelengths; and
- develop image processing algorithms for visualizing TVC distribution of salmon chops at a pixel-wise level at different spoilage stages.

2. Materials and methods

2.1. Samples preparation

Two groups (22 fillets in all) of fresh farmed Atlantic salmon fillets (*Salmo salar*) were purchased from local supermarkets in Dublin, Ireland. In specific, eleven fillets were from Norway and the other eleven fillets were from Ireland. The salmon fillets were fresh and of superior quality. The fillets were labeled and then transported to laboratories of Food Refrigeration & Computerized

Food Technology (FRCFT), University College Dublin (UCD), Ireland. Each fillet was first cut into three chops with the size of 3 cm × 3 cm × 1 cm (length × width × thickness). Then the samples were packed and left to spoil in a refrigerator at 4 °C for further analysis.

2.2. Hyperspectral imaging system and data acquisition

There were two hyperspectral imaging systems for measuring the hyperspectral images at 400–1000 nm (measured by System I) and 880–1720 nm (measured by System II), respectively. Each chop was placed on the moving table of each system and then was scanned line by line. System I (DV optics, Padua, Italy) mainly consisted of a Specim V10E spectrograph (Spectral Imaging Ltd., Oulu, Finland) covering the spectral range of 400–1000 nm (spectroscopic resolution of 5 nm), a high performance CCD camera (Basler A312f, effective resolution of 580 × 580 pixels by 12 bits), an objective lens (25 mm focal length), an illumination source (150 W halogen lamp source attached to a fibre optic line light positioned at an angle of 48° to the moving table), a mirror, a moving table, a computer system equipped with an acquisition software (Spectral Scanner, DV Optics, Padua, Italy). The main components of System II (DV Optics Ltd., Padua, Italy) included a Specim ImSpector N17E spectrograph (Spectral Imaging Ltd., Oulu, Finland) covering the spectral range of 880–1720 nm (spectroscopic resolution of 7 nm), an InGaAs camera (SUI Goodrich SU320M-1.7RT, effective resolution of 320 × 280 pixels), a lens, a tungsten halogen light source, a mirror, a moving table, and a scanner software (SpectralScanner, DV Optics, Padua, Italy).

At 2 day intervals, 13 chops were taken from the refrigerator and placed on the moving table for acquisition of hyperspectral images. The collected images were called 'hypercube' with a dimension of (x, y, λ). In details, a 2-D image (x, λ) with the whole spectral dimension (λ) with one spatial dimension (x) was acquired at a time. A complete hyperspectral cube was generated as the chop was scanned along the direction of y dimension. As a result, hyperspectral image data of 66 samples were obtained. Among them, 50 samples were used for the calibration and the remaining 16 samples were used for prediction. After the imaging, the TVC values of each chop sample was measured with standard plate count method according to BS EN ISO 4833:2003 [39] as quickly as possible and recorded as colony-forming units (CFUs).

2.3. Spectral data extraction

The CCD device in the hyperspectral imaging system collects the detector signal intensity, rather than actual reflectance spectra. Therefore, the acquired images were first corrected into reflectance images according to the method used by Wu et al., [30]. To conduct spectral data extraction from each chop in the hyperspectral image, the Region of Interests (ROI) Function of ENVI v4.6 software was used to isolate the chop from the background. The isolated region had the same shape of the chops whose TVC was measured by the standard plate count method, so that quantitative spectral models established using the mean spectra over the isolated region and the reference TVC value were meaningful. Although each pixel in the identified ROI has its own spectrum with its spatial position, it is practically impossible to measure the reference TVC values for every pixel within the ROI using the standard plate count method. To solve this problem, the TVC quantitative model was established based on the average spectrum of all pixels within the ROI where the corresponding reference TVC value could be measured using the standard plate count method. In specific, the first step was to extract the spectrum of each pixel within the ROIs, and the spectra of all pixels within ROI were then averaged at each wavelength to obtain one mean value representing the ROI. Due to decreased CCD detector sensitivity in the wavelength regions of 880–950 nm and 1601–1720 nm, subsequent analysis of hyperspectral images acquired by System II was performed only on data in the 950–1601 nm spectral range. Finally, two sets of spectral data were obtained, which were the one within the wavelength ranges of 400–1000 nm (Spectral Set I, 121 wavelength variables) and the one within the wavelength ranges of 950–1601 nm (Spectral Set II, 94 wavelength variables). In addition, because most spectral preprocessing algorithms are conducted based on the full range spectra and it is difficult to obtain the preprocessed spectra at only several important wavelengths by using a multispectral imaging system, no preprocessing treatments were applied to the spectral data during the selection of important wavelengths and the development of the calibration model in this study.

2.4. Multivariate data calibration

Multivariate data analysis was conducted to determine TVC of salmon chops using their corresponding spectral information. The quantitative models between the TVC and the spectral data extracted from salmon chops at different spoilage stages were established using a classical linear calibration algorithm named PLSR and a non-linear calibration algorithm named LS-SVM, respectively. Both PLSR and LS-SVM are methods to model a response variable when there are a large number of predictor variables, and those predictors are highly correlated or even collinear. PLSR is a classic spectral calibration technique that has been commonly used for spectral analysis [40-44]. PLSR projects the spectral data onto a set of orthogonal factors called latent variables (LVs), and explores the optimal function by minimizing the error of sum squares (finding the optimal LVs), which is typically done by cross-validation [45]. The process of extracting the LVs does take the response variable into account, which is an important advantage of PLSR against principal component regression (PCR). LS-SVM is an optimized version of the standard support vector machine (SVM) and has been applied for spectral calibration [46,47]. It employs nonlinear map function and maps the input features to a high dimensional space, thus changing the optimal problem into equality constraint condition. Lagrange multiplier is utilized to calculate the partial differentiation of each feature to obtain the optimal solution. A grid-search technique is usually applied to determine the optimal parameter values, that is, the regularization parameter γ and the RBF kernel function parameter σ^2. The details of LS-SVM description is shown in the literature [48]. The performances of PLSR and LS-SVM models based on full range spectra were compared to choose the best calibration algorithm. For the analysis of hyperspectral image data, the calibrated quantitative models were evaluated for their performances by using a prediction process with a new set of samples. The performance of a calibration model is usually evaluated in terms of root mean square error (RMSE) of calibration (RMSEC) and coefficients of determination (r^2) of calibration (r_C^2) in the calibration process, and root mean square error (RMSE) of prediction (RMSEP), coefficients of determination of prediction (r_P^2), and residual predictive deviation (RPD) in the prediction process. Generally, a good model should have higher values of r_C^2, r_P^2, and RPD, lower values of RMSEC and RMSEP, and a small difference between RMSEC and RMSEP.

2.5. Spectral variable selection

As providing rich spectral and spatial information, TS-HSI technique has unprecedented analysis capability for comprehensively understanding the dynamics of processes of food samples.

However, this technique faces a problem of containing large volumes of data to analyze, which is more serious than common hyperspectral imaging technique. Data mining strategies are required to cope the large multivariate data structures generated in the TS-HSI experiments [38]. Typical algorithms used for TS-HSI analysis include pixel binning, spectral averaging, k-means clustering, parallel factor analysis (PARAFAC), multivariate curve resolution (MCR), and principal components analysis (PCA). However, pixel binning and spectral averaging lose much spatial information that is important for analyzing heterogeneous food products; k-means clustering and PARAFAC need the image sequence to be congruent; and MCR and PCA require the TS-HSI images with whole spectral range for data mining, making the data acquisition process time-consuming [38]. In this study, the large number of the TS-HSI data was reduced by selecting the most informative and useful wavelengths/variables. Besides reducing the computational burden of the TS-HSI data and predigesting calibration modeling, the variable selection is also important for designing an optimized multispectral system imaging for practical on-line purpose [30].

Competitive adaptive reweighted sampling (CARS) is a novel variable selection algorithm proposed by Li et al., [49]. CARS selects an optimal combination of the variables from the full range variables coupled with PLSR according to a standard, which is that the absolute coefficients of variables in PLSR model are set as an index for evaluating the importance of each variable and the variable with large absolute coefficients have more contribution to the model calibration and prediction. In the CARS process, N subsets of wavelengths are selected from N Monte Carlo (MC) sampling runs in an iterative and competitive manner. In each sampling run, there are four successive steps, namely Monte Carlo model sampling, enforced wavelength reduction by exponentially decreasing function (EDF), competitive wavelength reduction by adaptive reweighted sampling (ARS) and root mean square error of cross validation (RMSECV) calculation for each subset. In EDF process, there are two stages, a 'fast selection' and a 'refined selection'. The ARS step is a mimicry of the 'survival of the fittest' originating in Darwin's Evolution Theory [50]. Details of the CARS algorithm can be found in the literature [49]. In this study, the calibration in the way that uses the full range spectra (121 wavelengths for Spectral Set I and 94 wavelengths for Spectral Set II) was called Method I; while utilizing only the important wavelengths selected by CARS for TVC determination was called Method II throughout this paper.

2.6. Visualization of TVC distribution

To observe the variances of TVC in salmon from sample to sample at different spoilage stages and even within the same sample, distribution maps of TVC in salmon chops are required instead of the measurement of one TVC value for the whole chop. In the visualization process, only the spectra at the selected variables were used for the establishment of the quantitative model in order to reduce the calculation burden of the TS-HSI images. Each pixel in the hyperspectral image has a spectral profile with its spatial position. The TVC values of each pixel were first calculated by inputting the spectrum of each pixel into the established quantitative model. A distribution map of TVC within the whole chop was then generated according to the spatial position of every pixel and their predicted TVC values. All of the computations, chemometric analyses and visualization process were operated with ENVI 4.6 (Research System, Inc., USA), "The Unscrambler V9.7" (CAMO PROCESS AS, Oslo, Norway), and programs developed by the authors in the Matlab 2011a software (The Mathworks, Inc., Natick, MA, USA).

3. Results and discussion

3.1. Spectral features of salmon chops and statistics of measured TVC

The typical spectral patterns of the tested salmon chops with different TVC values are shown in Fig. 1. The spectral data extracted from the pixels within the ROI of the chop contains the information of chemical molecules in the chop that was related to the chop's composition. The spectral profiles of salmon are quite even with some broadband peaks in whole spectral region. In the visible region, spectra have low reflectance at around 500 nm and high reflectance at 650–700 nm, which cause salmon flesh red colour. The absorption appeared around 500 nm was due to the presence of astaxanthin [51]. There are also some broadband peaks occurring in the near infrared region related to the overtone and combination vibrations of hydrogen containing bonds, such as O–H, C–H, and N–H. The presence of water in the salmon flesh showed three feature wavelengths at 760 nm (O–H stretching third overtone) in Fig. 1a and 980 nm and 1450 nm (O–H stretching second and first overtones) in Fig. 1b. Also, the absorption peak around 1200 nm (C–H stretching second overtone) was due to fat content. In general, samples with different TVC values shared the similar spectral patterns over the entire visible and near infrared region (400–1600 nm). There were some differences in the magnitudes of spectral reflectance as shown in Fig. 1. This was caused by the changing of the major chemical compositions in salmon flesh during the spoilage process. The sample with high TVC value had a lower reflectance in both

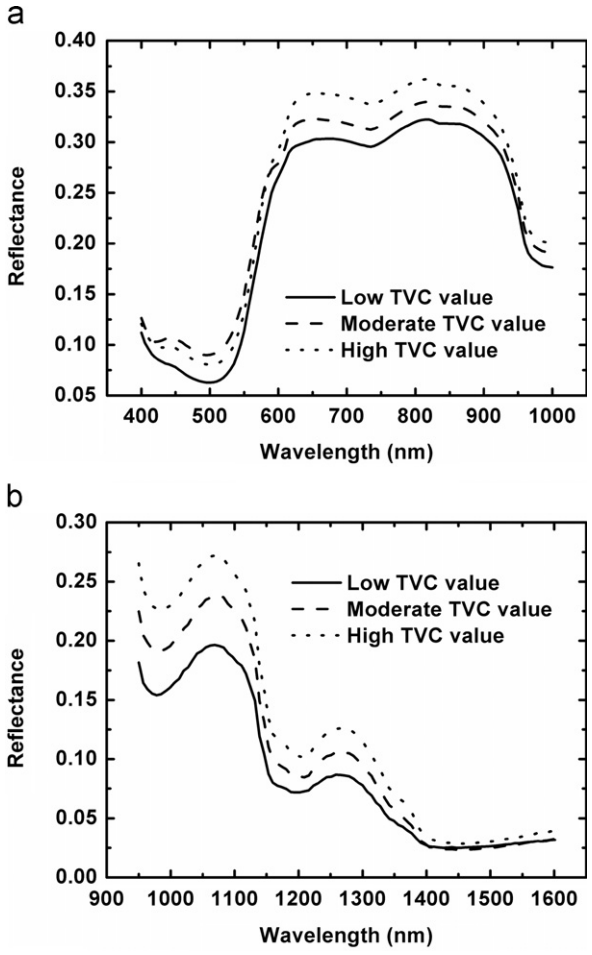

Fig. 1. Extracted mean spectral patterns of the tested salmon chops with different TVC values. (a) 400–1000 nm, (b) 950–1601 nm.

Spectral Sets, while a higher reflectance was found for the sample with low TVC value. However, when more samples were considered, the spectral profiles of the tested samples showed various magnitudes, which was caused by the uneven physical structure of the salmon samples as well as unfixed scatter of the surface [52]. Meanwhile, there was no feature peak related TVC values directly for the reflectance spectral profiles. Therefore, instead of observing the spectral profiles with naked eyes, multivariate data analysis was applied for the data mining and model calibration. In addition, in order to build a robust calibration model, a reasonable range of TVC variation should be guaranteed for the tested samples covering both the healthy samples and the spoilage samples. In general, the indicator level used for bacterial spoilage is often set at 10^6 colony-forming units (CFU)/g for salmon flesh [53]. Table 1 summarizes the descriptive statistics of TVC values for both the calibration set and prediction set, where there was a wide range of TVC variability covering both sides of the indicator level to establish reliable calibration models.

3.2. Spectral analysis

3.2.1. Calibration of TVC models based on full range spectra

The calibration of multivariate models was performed by PLSR and LS-SVM algorithms, respectively using the calibration set with full range spectra. In the calibration process, the quantitative relationship between the reflectance spectra of samples and their corresponding TVC values was established based on the samples from the calibration set. The prediction set was then analyzed as a new test set based on the established calibration models to estimate the actual predictive capability of the established models, to minimise the concrete risk of overfitting and to avoid chance correlations. Both Spectral Sets I and II were considered and compared in the processes of the model establishment using full range spectra, in order to choose the best hyperspectral imaging system for TVC determination of salmon flesh. Table 2 shows the statistical parameters of PLSR and LS-SVM models developed using the full spectral range for the TVC determination of salmon flesh in the calibration and prediction processes. The selection of the ideal number of LVs is the key step in establishing a robust PLSR model with the best calibration result. The smaller number of LVs will lead to poor prediction ability, while the larger number will result in overfitting. The lowest value of prediction residual errors sum of squares (PRESS) is commonly considered as the indicator of the optimal number of LVs. When Spectral Set I was considered, good results were achieved in both calibration and prediction conditions with an average r_C^2 of 0.963, r_P^2 of 0.924, RMSEC of 0.257, and RMSEP of 0.375 for the PLSR and LS-SVM models. When Spectral Set II was considered, the average RMSEC and RMSEP increased by 19.05% and 13.57%, respectively, showing that Spectral Set I was more suitable for the TVC determination of salmon flesh than Spectral Set II. In general, long-wave near infrared spectra (Spectral Set II) contain more information relevant to hydrogen containing bonds than visible and short-wave near infrared spectra (Spectral Set I). However, because there is high water content in salmon flesh, the information of other components is immersed in the absorption of water in Spectral Set II, which might be the reason to cause its poorer prediction ability. Considering the imaging spectrograph to acquire Spectral Set II was more expensive than the one to acquire Spectral Set I, it is suggested using the imaging spectrograph in the wavelength range of 400–1000 nm for the TVC determination in salmon flesh, rather than the one in the long-wave near infrared region.

Moreover, the best calibration algorithm changes for different spectral analysis. In order to choose the optimal calibration method for the TVC determination, the performances of PLSR and LS-SVM algorithms were compared for both Spectral Sets I and II. When LS-SVM was considered, an average RMSEP of 0.316 for two Spectral Sets was obtained, which was only 65% of that of PLS models. The average r_P^2 was also improved from 0.873 (PLSR models) to 0.955 (LS-SVM models). According to the standard of model evaluation, when full range spectra were considered, LS-SVM was better than PLSR for the determination of TVC in salmon flesh.

3.2.2. Variable selection

As high-dimensional data, it takes time for acquiring hyperspectral images, which is one of the reasons why the current hyperspectral imaging system has not been available for implementation in salmon processing lines but only used in laboratory-based analysis. The selection of the most important wavelengths/variables is an important step of analyzing hyperspectral images, which leads to reducing the high dimensionality of the hyperspectral cubes used for the attribute determination of samples. According to the selected variables, an optimized multispectral imaging inspection system could be further developed, resulting in speeding up the acquisition of spectral images. Therefore, only the most important variables for the prediction should be kept in the final spectral model to satisfy the requirement of online application. The full range spectra were delivered to the calibration models in Method I, and there was no consideration given to select important variables contributed for the TVC determination. In this study, CARS was carried out to select the important variables by using the simple but effective principle 'survival of the fittest' on which Darwin's Evolution Theory is based. Only the spectral range of 400–1000 nm (Spectral Set I) was considered in the identification of optimum wavelengths, as it achieved better results than Spectral Set II did.

Fig. 2 shows the changing trend of the number of sampled variables (Fig. 2a), 5-fold RMSECV values (Fig. 2b) and the regression coefficient path of each variable (Fig. 2c) with the increasing

Table 1
Reference TVC values of salmon samples measured by the standard plate count method.

	Maximum	Minimum	Mean	Standard deviation	Range
Calibration set	7.538	3.900	5.745	1.401	3.637
Prediction set	7.525	3.984	5.671	1.411	3.540

Table 2
Results of TVC prediction of salmon samples by using the full range spectra.

Spectral set	Calibration model	Variable number	Number of latent variables	Calibration		Prediction		
				r_C^2	RMSEC	r_P^2	RMSEP	RPD
I	PLSR	121	6	0.941	0.335	0.887	0.460	2.978
I	LS-SVM	121	/	0.984	0.178	0.961	0.290	5.088
II	PLSR	94	6	0.939	0.343	0.860	0.511	2.685
II	LS-SVM	94	/	0.963	0.269	0.949	0.341	4.256

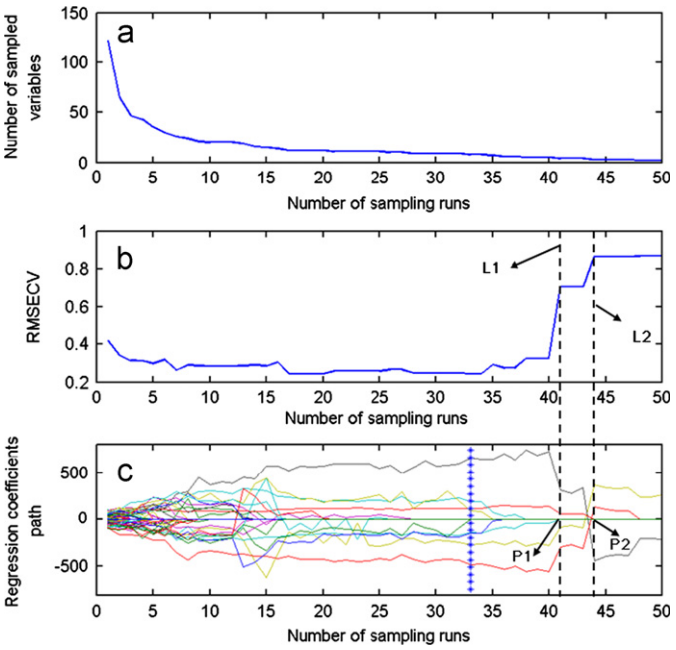

Fig. 2. Changing trend of the number of sampled variables (plot a), 5-fold RMSECV values (plot b) and regression coefficients of each variables (plot c) with the increasing of sampling runs. The line (marked by asterisk) denotes the optimal point where 5-fold RMSECV values achieve the lowest.

of sampling runs in the CARS calculation on the spectra of samples in the calibration set in Spectral Set I. During the CARS process, some incompetent variables were eliminated while the key variables were retained. In Fig. 2a, it is shown that the number of sampled variables decreased fast at the first stage of EDF (fast selection) and then slowly at the second stage of EDF (refined selection). On the other hand, as shown in Fig. 2b, along with the number of sampling runs increased, RMSECV values first descend in a gentle way from sampling runs 1–17, in which the uninformative variables were eliminated, then almost no change of RMSECV value happened in the sampling runs 18–34, and finally the RMSECV increased which should be ascribed to the loss of some key variables. The optimal variable subset was determined corresponding to the minimal 5-fold RMSECV value, and its position is marked by the vertical asterisk line. Moreover, it is noteworthy to analyse the regression coefficient path of each wavelength shown in Fig. 2c. Each line here recorded the changing of coefficient values of each variable at different sampling runs. The absolute value of regression coefficient of each wavelength was very small at the beginning of the sampling run. Later, the coefficients became larger and larger for some variables while others became smaller and smaller. It should be noticed that the coefficients of the corresponding wavelengths eliminated by CARS were set as zero because of their incompetence. In other words, the larger the absolute coefficient is, the more probable the corresponding wavelength can survive. Such selection mechanism in CARS is similar to the 'survival of the fittest' in Darwin's Evolution Theory [49]. After the critical point denoted by the asterisk line, the RMSECV values began to increase because some key wavelengths were removed. For example, the RMSECV sharply rose up much higher stage denoted by dot line L1, when one variable (marked as P1) was eliminated as its coefficient dropped to zero. Another case is that the coefficient of another variable denoted by P2 drops to zero, resulting in the sharp rising of RMSECV value (L2). These variables were termed as key variables as the model's prediction ability would be reduced dramatically without considering these variables. As a result of the CARS calculation, eight variables at 495 nm, 535 nm, 550 nm, 585 nm, 625 nm, 660 nm, 785 nm, and 915 nm were identified as the important wavelengths for determining TVC of salmon flesh. Instead of the full range spectra, these selected wavelengths were then used as the inputs of two calibration algorithms of PLSR and LS-SVM, respectively, because the most relevant information of TVC in the salmon flesh was contained in these variables.

Based on the selected variables, the spectral data were then reduced to two new matrices with dimension of $c \times v$ and $p \times v$, where c and p are the number of samples in the calibration ($c=50$) and prediction sets ($p=16$), the number of variables v was reduced from 121 to 8 variables (the number of important variables). The new calibration processes were then conducted based on the new matrix ($c \times v$) using PLSR and LS-SVM, respectively, and another new matrix was used as an external test set for evaluating the performance of the new established calibration models (Method II). Table 3 presents the main statistics achieved in the calibration and prediction of the PLSR and LS-SVM models developed by using the abovementioned important variables. When LS-SVM was considered, the effective wavelength-based models had lower RMSEP (decreased by 8.7%) than that of the full spectral model for Spectral Set I. More significant improvement was achieved for the new PLSR model, where the RMSEC and RMSEP were decreased by 21.4% and 39.1%, respectively. It is noticed that when full range spectra were used to build the PLSR model, the obtained RPD was only 2.978; and the RPD value increased dramatically by 72.2% to 5.127% when the selected variables were used for the PLSR model establishment. In addition, the absolute differences between RMSEC and RMSEP of the CARS-PLSR and CARS-LS-SVM models decreased 86.8% and 36.4%, respectively, showing that the established models using the important variables selected by CARS were more robust than those with the wavelengths in the full spectral range. The results indicated that the use of CARS for variable selection was helpful to improve the prediction accuracy and robustness of the spectral models for the TVC determination in salmon flesh, and such improvement was even achieved by using only 6.6% of variables in the full range spectra (6 vs. 121).

Although LS-SVM showed a better prediction ability than PLSR when full range spectra were chosen as the input variables (shown in Table 2), with the help of CARS calculation, the new PLSR model (Method II) obtained a similar performance to the new LS-SVM model (Method II). Compared with the CARS-LS-SVM model, the r_P^2, RMSEP, and RPD of the CARS-PLSR model only decreased by 1.0%, 5.6%, and 3.0%, respectively. Moreover, it should be noticed that the absolute difference between RMSEC and RMSEP was only 0.016 for the new PLSR model, which was only 23.0% of that of the new LS-SVM model (0.071). The small absolute difference indicated that the new PLSR model was more robust than the new LS-SVM model. Therefore, the best quantitative model for the assessment of TVC values in salmon flesh was the one established using the CARS-PLSR method with Spectral Set I. According to the regression coefficients of the CARS-PLSR model with Spectral Set I, the quantitative function for the TVC measurement of salmon flesh was obtained and is shown as follows:

$$Y_{TVC} = 2.07 - 140.25X_{495\text{nm}} + 393.64X_{535\text{nm}}$$

Table 3
Results of TVC prediction of salmon samples by using only the important wavelengths.

Calibration model	Variable number	Number of latent variables	Calibration		Prediction		
			r_C^2	RMSEC	r_P^2	RMSEP	RPD
PLSR	8	7	0.964	0.264	0.958	0.280	5.127
LS-SVM	8	/	0.981	0.194	0.967	0.265	5.288

$$-308.05X_{550nm}+87.90X_{585nm}-209.83X_{625nm}+310.17X_{660nm} -230.51X_{785nm}+119.87X_{915nm} \quad (1)$$

where X_{inm} is the reflectance spectra at the wavelength of i nm and Y_{TVC} is the predicted TVC value. For the best quantitative model for the assessment of TVC values in salmon flesh (CARS-PLSR method with Spectral Set I), its r_P^2 and RMSEP were 0.958 and 0.280, respectively, showing that the CARS-PLSR method with Spectral Set I had a good predictive ability for a new sample set. The results shown in Tables 2 and 3 confirmed the suitability of TS-HSI for TVC determination of salmon flesh in a rapid and nondestructive manner, and therefore the distribution of TVC within salmon chops during spoilage process was visualized based on the CARS-PLSR model with Spectral Set I. It should be noticed that the models in this study were established for the spoilage determination of TVC. Specific bacterial for salmon and other fish should be considered in future works. Moreover, to improve the applicability and robustness of the model, more samples are required in the calibration process to account for more variability of TVC values in salmon flesh.

3.2.3. Distribution map of TVC

The final and one of the important steps of analyzing hyperspectral images is to visualize the distribution of the predicted values. Such visualization in a pixel-wise manner is the main advantage of hyperspectral imaging against the traditional spectroscopy. In this study, the visualization process was conducted based on the new TS-HSI matrix with the images only at the selected variables/wavelengths in order to speeding up the visualization process and making it easier to establish a multispectral imaging system. The dot product was then calculated between the new matrix and the PLSR coefficients shown in Eq. (1), resulting in the TVC values of every pixel within the ROI of salmon chop. According to the information of two spatial dimensions contained in each pixel, the predicted TVC values were easily transferred into the distribution map of TVC of the salmon chop. A linear colour scale was generated with the different TVC values from small to large shown in different colour from blue to red (the colour bar in Fig. 3), leading to similar predicted TVC values having a similar scale in the generated distribution map.

Examples of TVC distribution maps of some tested salmon chops with different TVC values are shown in Fig. 3. The maps indicate how the magnitude of TVC varies from chop to chop or from pixel to pixel within the same chop. Although a composite of different colours were observed across the salmon chop, there is a general trend of increment in overall TVC values from blue to red, which is a clear indication of TVC status during the spoilage process of salmon chops. Many pixels within the salmon flesh at the early stages of spoilage process were blue, because of the low TVC values of fresh salmon chops. During the spoilage process, the color of pixels changed from blue, green, and to red. There are only a few pixels in dark blue within the salmon chop when its average TVC was over five. Pixels in red were found after a further spoilage, and most pixels turned into red color when the average TVC was 7.61 as shown in the lower right corner of Fig. 3. None of these phenomena could be observed by the naked eyes, let alone the quantitative TVC prediction.

The ability of TS-HSI in determining TVC of salmon flesh quantitatively lies in its provided spectral information related to the composition changes during the spoilage period. Meanwhile, the contained spatial information makes TS-HSI available to visualize the TVC distribution within salmon chops for understanding the TVC variation during the spoilage process. Therefore, both spectral and spatial information contained in TS-HSI data is critical to determine TVC distribution within salmon chops quantitatively. The results provided by TS-HSI analysis are important to determine whether the salmon flesh is suitable for consumers or not, leading to improving the safety assurance of the final product. Knowing the variation of TVC in salmon flesh is also useful to reflect the influence of some storage regimes and to further optimize them. In addition, TS-HSI could be further applied to assess the variation of other quality attributes of salmon flesh such

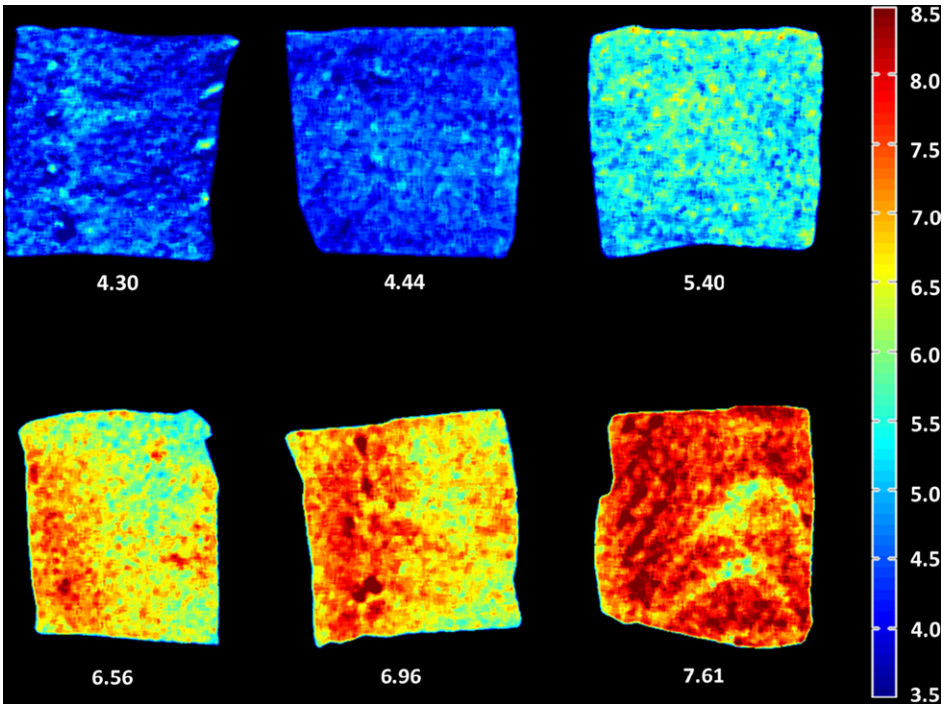

Fig. 3. TVC distribution maps of some tested salmon chops with different TVC values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as fat and tenderness that are known to differ from salmon's bulk compositions during storage and other processes.

4. Conclusions

The rapid microbiological testing is important in the modern and highly automated food processing and storage environment. This is the first reported work to use TS-HSI technique for determining TVC distribution in salmon flesh during spoilage process. The results and illustrations presented in this work suggest that TS-HSI was satisfied as a rapid and non-invasive tool instead of the standard plate count method for assessing the TVC distribution of salmon flesh during the spoilage process. Two spectrographs were used to acquire the hyperspectral images of salmon chops at different spoilage stages, and the results show that the one working in the visible and short-wave near infrared range of 400–1000 nm (Spectral Set I) had better performance than the other one working in the long-wave near infrared range of 950–1601 nm (Spectral Set II). CARS algorithm was conducted for the first time to analyse hyperspectral images, resulting in reducing the high dimensionality with redundancy and colinearity among the visible and near infrared spectra extracted from the TS-HSI images. As a result, eight wavelengths (495 nm, 535 nm, 550 nm, 585 nm, 625 nm, 660 nm, 785 nm, and 915 nm), which carried the most relevant information of TVC, were chosen by CARS as important wavelengths to build a new spectral matrix for the further calibration. The best quantitative model was determined as the CARS-PLSR model with Spectral Set I, which had high prediction ability with r_P^2 of 0.985 and RPD of 5.127. At last, the regression coefficients obtained from the CARS-PLSR model were applied for each pixel in the ROI of salmon chop by using image process algorithms to realize the visualization of the TVC distribution, which is important for providing more detailed information of spoilage development in salmon flesh. Such information cannot be provided by using either conventional computer vision or spectroscopy. According to the results obtained in this study, simple, fast and low cost multispectral imaging instruments can be designed to refine this technique for TVC assessment in salmon flesh during the spoilage process without additional laborious biological analysis. This study opens up an attractive prospect that microbial spoilage in salmon could be determined quantitatively within seconds as opposed to hours or days. As the first research on rapid and non-invasive determination of TVC distribution of salmon flesh, the results obtained here are very promising and will promote more efforts on investigating TS-HSI for assessing bacterial contamination of salmon and other food products during the spoilage process.

Acknowledgements

The authors would like to acknowledge the financial support provided by the Irish Research Council for Science, Engineering and Technology under the Government of Ireland Postdoctoral Fellowship scheme. We also thank Professor Colm O'Donnell for lending us the hyperspectral imaging equipment and Dr Carlos Esquerre Feranandez for excellent technical assistance.

References

[1] FAO, Fish. Aquaculture Inf. Stat. Serv. (2011).
[2] K. Sveinsdottir, E. Martinsdottir, G. Hyldig, B. Jorgensen, K. Kristbergsson J. Food Sci. 67 (2002) 1570–1579.
[3] B. Lund, A.C. Baird-Parker, G.W. Gould, Microbiological Safety and Quality of Food, Springer, NY, USA, 1999.
[4] H. Tent, Food Control 10 (1999) 239–241.
[5] D.I. Ellis, D. Broadhurst, D.B. Kell, J.J. Rowland, R. Goodacre, Appl. Environ. Microbiol. 68 (2002) 2822–2828.
[6] L. Gram, H.H. Huss, Int. J. Food Microbiol. 33 (1996) 121–137.
[7] D.I. Ellis, R. Goodacre, Trends Food Sci. Technol. 12 (2001) 414–424.
[8] S. Limbo, N. Sinelli, L. Torri, M. Riva, LWT Food Sci. Technol. 42 (2009) 977–984.
[9] K. Crowley, A. Pacquit, J. Hayes, L. King Tong, D. Diamond, A gas-phase colorimetric sensor for the detection of amine spoilage products in packaged fish, in: The Printing House Inc. (Eds.) Proceedings of IEEE Sensors 2005 Conference, 2005, pp. 754–757.
[10] Y.K. Peng, J. Zhang, W. Wang, Y.Y. Li, J.H. Wu, H. Huang, X.D. Gao, W.K. Jiang J. Food Eng. 102 (2011) 163–169.
[11] N. Prieto, R. Roehe, P. Lavin, G. Batten, S. Andres, Meat Sci. 83 (2009) 175–186.
[12] P. Jackman, D.-W. Sun, P. Allen, Trends Food Sci. Technol. 22 (2011) 185–197.
[13] D. Wu, D.-W. Sun, Trends Food Sci. Technol. 29 (2012) 5–20.
[14] D.-W. Sun, T. Brosnan, J Food Eng. 57 (1) (2003) 81–89, http://dx.doi.org/10.1016/S0260-8774(02)00275-3, S0260-8774(02)00275-3.
[15] C.J. Du, D.-W. Sun, J Food Eng. 68 (3) (2005) 277–287, http://dx.doi.org/10.1016/j.jfoodeng.2004.05.044.
[16] 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.
[17] J. Patrick, D.-W. Sun, C.-J. Du, P. Allen, G. Downey, Meat Sci. 80 (4) (2008) 1273–1281, http://dx.doi.org/10.1016/j.meatsci.2008.06.001.
[18] D.-W. Sun, J Food Eng. 61 (1) (2004) 1–2, http://dx.doi.org/10.1016/S0260-8774(03)00182-1.
[19] H.H. Wang, D.-W Sun, J Food Eng. 52 (3) (2002) 279–284, http://dx.doi.org/10.1016/S0260-8774(01)00116-9, S0260-8774(01)00116-9.
[20] D.-W. Sun, T. Brosnan, J Food Eng. 57 (1) (2003) 91–95, http://dx.doi.org/10.1016/S0260-8774(02)00276-5, S0260-8774(02)00276-5.
[21] D. Lorente, N. Aleixos, J. Gomez-Sanchis, S. Cubero, O.L. Garcia-Navarrete, J. Blasco, Food Bioprocess Technol. 5 (2012) 1121–1142.
[22] L.S. Magwaza, U.L. Opara, H. Nieuwoudt, P.J. Cronje, W. Saeys, B. Nicolaï, Food Bioprocess Technol. 5 (2012) 425–444.
[23] G. ElMasry, D.-W. Sun, P. Allen, J. Food Eng. 110 (2012) 127–140.
[24] D.F. Barbin, G. ElMasry, D.-W. Sun, P. Allen, Anal. Chim. Acta 719 (2012) 30–42.
[25] M. Kamruzzaman, G. ElMasry, D.-W. Sun, P. Allen, Anal. Chim. Acta 714 (2012) 57–67.
[26] D. Wu, H. Shi, S. Wang, Y. He, Y. Bao, K. Liu, Anal. Chim. Acta 726 (2012) 57–66.
[27] D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, J. Blasco, Food Bioprocess Technol. 6 (2013) 530–541.
[28] D.P. Ariana, R.F. Lu, Comput. Electron. Agric. 74 (2010) 137–144.
[29] V.H. Segtnan, M. Hoy, F. Lundby, B. Narurn, J.P. Wold, J. Near Infrared Spectrosc. 17 (2009) 247–253.
[30] D. Wu, D.-W. Sun, Y. He, Innovative Food Sci. Emerg. 16 (2012) 361–372.
[31] S. Ottestad, M. Hoy, A. Stevik, J.P. Wold, J. Near Infrared Spectrosc. 17 (2009) 77–87.
[32] H.-J. He, D. Wu, D.-W. Sun, Application of hyperspectral imaging technique for non-destructive pH prediction in salmon fillets, Proceedings of the Third CIGR International Conference of Agricultural Engineering (CIGR-AgEng2012), Valencia, Spain, 2012.
[33] H.-J. He, D. Wu, D.-W. Sun, Innovative Food Sci. Emerg. (2013), http://dx.doi.org/10.1016/j.ifset.2013.02.009, in press.
[34] D. Wu, H.-J. He, D.-W. Sun, Non-destructive texture analysis of farmed salmon using hyperspectral imaging technique, Proceedings of the Third CIGR International Conference of Agricultural Engineering (CIGR-AgEng2012), Valencia, Spain, 2012.
[35] V.H. Segtnan, M. Hoy, O. Sorheim, A. Kohler, F. Lundby, J.P. Wold, R. Ofstad J. Agric. Food Chem. 57 (2009) 1705–1710.
[36] J. Van Der Weerd, S.G. Kazarian, Multivariate movies and their applications in pharmaceutical and polymer dissolution studies, in: H.F. Grahn, P. Geladi (Eds.), Techniques and Applications of Hyperspectral Image Analysis, John Wiley & Sons Ltd., Chichester,England, 2007, pp. 221–260.
[37] A.A. Gowen, F. Marini, C. Esquerre, C. O'Donnell, G. Downey, J. Burger, Anal. Chim. Acta 705 (2011) 272–282.
[38] D. Wu, S. Wang, N. Wang, P. Nie, Y. He, D.-W. Sun, J. Yao, Food Bioprocess Technol. (2013).
[39] FSA Total, Aerobic Plate Counts, Food Standards Agency Publications, London, UK, 2009.
[40] X.J. Chen, X.X. Lei, J. Agric. Food Chem. 57 (2009) 334–340.
[41] X. Chen, D. Wu, Y. He, S. Liu, Anal. Chim. Acta 638 (2009) 16–22.
[42] X. Zhu, G. Huang, S. Luo, X. Guan, X. Chen, Anal. Lett. 46 (2013) 671–681.
[43] V. Sinija, H. Mishra, Food Bioprocess Technol. 4 (2011) 136–141.
[44] F. Antonucci, F. Pallottino, G. Paglia, A. Palma, S. D'Aquino, P. Menesatti, Food Bioprocess Technol. 4 (2011) 809–813.
[45] D. Wu, J. Chen, B. Lu, L. Xiong, Y. He, Y. Zhang, Food Chem. 135 (2012) 2147–2156.
[46] X.J. Chen, D. Wu, Y. He, S. Liu, Food Bioprocess Technol. 4 (2011) 753–761.
[47] Y.D. Liu, R.J. Gao, Y. Hao, X.D. Sun, A.G. Ouyang, Food Bioprocess Technol. 5 (2012) 1106–1112.
[48] D. Wu, Y. He, S.J. Feng, D.-W. Sun, J. Food Eng. 84 (2008) 124–131.
[49] H. Li, Y. Liang, Q. Xu, D. Cao, Anal. Chim. Acta 648 (2009) 77–84.
[50] C. Darwin, G. Beer, The Origin of Species, Oxford University Press, 1998.
[51] I. Sone, R.L. Olsen, A.H. Sivertsen, G. Eilertsen, K. Heia, J. Food Eng. 109 (2012) 482–489.
[52] G. ElMasry, D.-W. Sun, P. Allen, Food Res. Int. 44 (2011) 2624–2633.
[53] G. Olafsdottir, E. Chanie, F. Westad, R. Jonsdottir, C.R. Thalmann, S. Bazzo, S. Labreche, P. Marcq, F. Lundby, J.E. Haugen, J. Food Sci. 70 (2005) S563–S574.