Food Microbiology, Volume 28, Issue 4, June 2011, Pages 782–790

A comparison of artificial neural networks and partial least squares modelling for the rapid detection of the microbial spoilage of beef fillets based on Fourier transform infrared spectral fingerprints

Efstathios Z. Panagou [a] Fady R. Mohareb [b], Anthoula A. Argyri [a,c], Conrad M. Bessant [b], George-John E. Nychas [a,*]

[a] *Agricultural University of Athens, Department of Food Science, Technology and Human Nutrition, Laboratory of Microbiology and Biotechnology of Foods, Iera Odos 75, 118 55 Athens, Greece*

[b] *Bioinformatics Group, Cranfield Health, Cranfield University, College Road, Cranfield, Bedfordshire MK43 0AL, United Kingdom*

[c] *Applied Mycology Group, Cranfield Health, Cranfield University, College Road, Cranfield, Bedfordshire MK43 0AL, United Kingdom*

* Corresponding author: Department of Food Science, Technology and Human Nutrition, Agricultural University of Athens, Iera Odos 75, Athens, Greece, GR-118 55. Tel/fax: +30-210-5294693. E-mail address: gjn@aua.gr

Abstract

A series of partial least squares (PLS) models were employed to correlate spectral data from FTIR analysis with beef fillet spoilage during aerobic storage at different temperatures (0, 5, 10, 15, and 20°C) using the dataset presented by Argyri et al. (2009). The performance of the PLS models was compared with a three-layer feed-forward artificial neural network (ANN) developed using the same dataset. FTIR spectra were collected from the surface of meat samples in parallel with microbiological analyses to enumerate total viable counts. Sensory evaluation was based on a three point hedonic scale classifying meat samples as fresh, semi-fresh, and spoiled. The purpose of the modelling approach employed in this work was to classify beef samples in the respective quality class as well as to predict their total viable counts directly from FTIR spectra. The results obtained demonstrated that both approaches showed good performance in discriminating meat samples in one of the three predefined sensory classes. The PLS classification models showed performances ranging from 72.0 to 98.2% using the training dataset, and from 63.1 to 94.7% using independent testing dataset. The ANN classification model performed equally well in discriminating meat samples, with correct classification rates from 98.2 to 100% and 63.1 to 73.7% in the train and test sessions, respectively. PLS and ANN approaches were also applied to create models for the prediction of microbial counts. The performance of these was based on graphical plots and statistical indices (bias factor, accuracy factor, root mean square error). Furthermore, results demonstrated reasonably good correlation of total viable counts on meat surface with FTIR spectral data with PLS models presenting better performance indices compared to ANN.

Keywords: artificial neural networks, aerobic storage, beef fillets, FTIR, machine learning, meat spoilage, partial least squares regression, pattern recognition

1. Introduction

One of the most commonly consumed food commodities on a global basis is meat, due to its high nutritional value in the human diet. In the USA alone the retail market of beef industry amounted to $76 billion in 2008 with an overall consumption of approximately 27.3 billion pounds in that year (USDA, 2008). During meat production/processing quality assurance is difficult due to the heterogeneous nature of the raw material, since the chemical composition, technological and sensory attributes are highly influenced by pre-slaughter (e.g., breed, age, environment) intrinsic (e.g. pH, available nutrients) and extrinsic (e.g., storage method, period and temperature of storage) factors (Damez and Clerjon, 2008; Nychas et al., 2008; Prieto et al., 2009). Consequently, in order to keep the quality standards as close as possible to the preference of the consumer, control procedures must be undertaken including sensory, microbiological and physico-chemical analysis. Today, more than 50 such methods have been employed for the characterization of microbiologically spoiled or contaminated meat (Ellis and Goodacre, 2001; Nychas et al., 2008). However, these methods suffer certain disadvantages as they are time-consuming, destructive, require highly trained personnel, provide retrospective information, and hence they are unsuitable for on-line monitoring (Dainty, 1996; Nychas et al., 1998, 2008; Ellis et al., 2002, 2004; Liu et al., 2004).

Nowadays, various rapid, non-invasive methods based on analytical instrumental techniques, such as Fourier transform infrared spectroscopy (FTIR), Raman spectroscopy, near infrared spectroscopy, and electronic nose technology are being researched for their potential as reliable meat quality sensors (Ellis et al., 2005; Rajamäki et al., 2006; Damez and Clerjon, 2008; Ammor et al., 2009; Argyri et al., 2009; Balasubramanian et al., 2009; Prieto et al., 2009). The principle underlying this approach is based on the assumption that the metabolic activity of microorganisms on meat results in biochemical changes with the concurrent formation of metabolic by-products which may indicate or may contribute to spoilage. The quantification of these metabolites constitutes a characteristic fingerprint providing information about the type and rate of spoilage (Ellis and Goodacre, 2001; Nychas et al., 2008).

The introduction of converging technologies in the food industry is among the priorities of the 7th Framework Programme and they are anticipated to predominate in the future and result in substantial changes in the manner in which researchers design their research (Hair et al., 1998; NBIC report USA 2002). This can be achieved thorough the integration of modern analytical and high throughput platforms with computational and chemometric techniques. Multivariate statistical analyses (e.g., partial least square regression, discriminant function analysis, cluster analysis) and intelligent methodologies (e.g., artificial neural networks), can result in the development of a decision support system for timely determination of safety/quality of meat products, and also prevent unnecessary economic losses (Mataragas et al., 2007; Nychas et al., 2008; Guillén et al., 2010). Furthermore, the development of computational research platforms and online experimental databases such as Combase (Baranyi and Tamplin, 2004) and Sym'Previus (Leporq et al., 2005), provide research scientists with fast and efficient means of storing and exchanging knowledge despite their geographic distribution.

Partial least squares discriminant analysis (PLS-DA) and artificial neural networks (ANNs) are widely employed modelling approaches due to their ability to relate the input and output variables without having any prior knowledge on the system under study, provided that an accurate and adequate amount of data on the system variables is available (Singh et al., 2009). Compared to other areas, the application of ANNs in the field of food science is still in the early development stage. Nevertheless, interest in using ANNs as secondary models in food microbiology is increasing as they have shown promising results in several applications such as growth parameter estimation of microorganisms (Geeraerd et al., 1998; Hervás et al., 2001; García-Gimeno et al., 2005), bacterial heat resistance (Lou and Nakai, 2001; Esnoz et al., 2006), production of metabolites (Poirazi et al., 2007), and simulation of survival curves (Palanichamy et al., 2008; Panagou, 2008). The multi-layer perceptron (MLP) is the most frequently used type of neural network in practical applications (Siripatrawan et al., 2006). The basic structure is comprised of three distinctive layers, the input layer where the data are introduced to the model and computation of the weighted sum of the input is performed, the hidden layer or layers where data processing takes place, and the output layer where the results of the neural network are produced (Bishop, 2004; Huang et al., 2007).

The purpose of the present study was to compare the performance of a multilayer perceptron (MLP) neural network and partial least squares (PLS) regression models in order to (i) classify beef fillets stored aerobically at different temperatures (0, 5, 10, 15, and 20°C) in terms of quality classes (i.e., fresh, semi-fresh, spoiled), and (ii) predict the total viable counts on the surface of meat samples directly from FTIR data.

2. Materials and methods

2.1 Experimental design

A detailed description of the methodology employed in this work is presented elsewhere (Argyri et al., 2009). In brief, fresh deboned pieces of beef were purchased from a local butcher shop and transported under refrigeration to the laboratory within 30 min. The samples were prepared by cutting meat pieces into portions (40 mm wide x 50 mm long x 10 mm thick) that were subsequently placed into 90 mm Petri dishes and stored at 0, 5, 10, 15, and 20˚C in high-precision (±0.5°C) incubation chambers until spoilage was evident.

For the FTIR measurements, a thin slice (0.5 cm thickness) of the aerobic upper surface of the fillet was excised and used for further analysis. Spectra were collected using a ZnSe 45° ATR (Attenuated Total Reflectance) crystal on a Nicolet 6700 FT-IR Spectrometer, collecting spectra over the wavenumber range of 4,000 to 400 cm^{-1}, by accumulating 100 scans with a resolution of 4 cm^{-1}. The collection time for each sample spectrum was 2 min. Spectra collected between 1800 and 1000 cm^{-1} were initially subjected to smoothing according to the Savitzky-Golay algorithm prior to further analysis.

For microbiological analysis a portion (40 mm wide x 50 mm long x 10 mm thick) was added to 150 ml sterile quarter strength Ringer's solution, and homogenized in a stomacher for 60 s at room temperature. Further decimal dilutions were prepared with the same diluent, and duplicate 0.1 ml samples of three appropriate dilutions were spread in triplicate on plate count agar for counts of total viable bacteria, incubated at 30ºC for 48 h.

Sensory evaluation of meat samples was performed during storage, based on the perception of colour and smell before and after cooking (20 min at 180˚C in preheated oven) (Gill and Jeremiah, 1991). Each sensory attribute was scored on a three-point hedonic scale corresponding to: 1=Fresh; 2=Marginal; and 3=Spoiled. Score of 1.5 was characterized as Semi-fresh and it was considered as the early detection of meat spoilage. Overall, 76 meat samples were evaluated by the sensory panel and classified into the selected groups as fresh (n = 26), semi-fresh (n = 16), and spoiled (n = 34).

2.2 Partial least squares (PLS) modelling

The partial least squares regression (PLS-R) derives its usefulness from its ability to analyze data with strongly collinear, noisy and numerous variables in the predictor matrix X (i.e., independent variables) and responses Y (i.e., dependent variables) (Eriksson et al., 2001). The PLS-R method projects the initial input-output data down into a latent space, extracting a number of principal factors (also known as latent variables) with an orthogonal structure, while capturing most of the variance in the original data. In brief, it can be expressed as a bilinear decomposition of both X and Y as:

$$X = \boldsymbol{TW}^T + E_X \quad (1)$$

and

$$Y = \boldsymbol{UQ}^T + E_Y \quad (2)$$

Therefore, the scores in the *X*-matrix and the scores of the yet unexplained part of *Y* have maximum covariance. In equations (1) and (2), ***T*** and ***W***, ***U*** and ***Q*** are the vectors of *X* and *Y* PLS scores and loadings, respectively, and E_X, E_Y are the *X* and *Y* residuals (Singh et al., 2009). The aim of PLS method is to find a linear (or polynomial) relationship between *X* and *Y* matrices, so that:

$$Y = bX + E \quad (3)$$

where *b* is the regression coefficient. The PLS models are developed in two stages; the initial dataset is divided into training and testing subsets. The former dataset is used to build the models and compute a set of regression coefficients ($\boldsymbol{b}_{PLS}$), which are subsequently used to make a prediction of the dependent variable in the test subset. The initial dataset consisted of 74 beef fillet spectral patterns corresponding to different storage temperatures (0, 5, 10, 15, and 20°C) and storage times (up to 350 hours depending on storage temperature). The database was randomly partitioned into a training and testing subset representing 75% (n = 57) and 25% (n = 19) of the data, respectively. Test data were not employed in any step of training the PLS model but they were used exclusively to determine its performance. A series of PLS models were created using a number of latent variables ranging from 1 to 25, hence 25 models were developed in total. The performance of each generated model was calculated using leave-one-out cross validation. The optimum numbers of components were used to build the final model. The resulting model was then tested with the independent data set. This procedure was repeated two times for predicting the predefined sensory class: *i*) based on storage time and temperature as two input variables in addition to the FTIR dataset, and *ii*) based entirely on the FTIR data where no storage condition data was included to build the models. Similarly, two sets of models were developed to predict the total viable counts (TVC), firstly based on including the storage conditions as additional input variables, and secondly based entirely on the FTIR data. Therefore four sets of models were developed in total.

2.3 Artificial neural networks modelling

Mean-centered and standardized spectral data were initially subjected to principal components analysis (PCA) for dimensionality reduction, and the variables (wavenumbers) for which communality values were less than 0.6 were excluded from further analysis, as they were considered to contain not enough information to explain the variance of spectral data. The remaining wavenumbers (from 1718 to 1203 cm^{-1} and 1020 to 1001 cm^{-1}) were subjected to a second PCA, where the total variance (100%) of the dataset was cumulatively explained by 37 principal components (PCs). The scores of the first five PCs were extracted and used in further analysis as they explained a cumulative variance of 98.08% of the dataset.

The selected network was a multilayer perceptron (MLP) based on backpropagation. The basic element in an MLP is the "neuron" that receives a set of input signals (x_i) with weight (w_i), calculates their impact using the summation function ($I = \sum x_i \cdot w_i$), and finally produces an output using some activation function ($y = f\left(I\right)$). The determination of the weights is achieved through training of the system. Normally, supervised training is performed in such a way as to minimize the difference between the network output and the measured value:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{predicted,i} - y_{observed,i}\right)^2 \qquad (4)$$

where, $y_{predicted,i}$ and $y_{observed,i}$ represent the predicted and observed values of the variable, respectively, and n is the number of observations. Back propagation (BP) is the most commonly used training algorithm in neural networks, also employed in this work. It works on the principle that after the information has gone through the network in a forward direction and an output has been produced, the error associated with this output is redistributed backwards through the model and weights are adjusted accordingly. Minimization of the error occurs through several iterations (training cycles) (Ham and Kostanic, 2001).

Two separate networks were developed in this work comprising of an input layer with seven nodes, one for temperature and storage time, respectively, and the remaining five for each one of the five PCs. The output layer contained one node for the prediction of either meat quality class (i.e., F, SF, S) or total viable counts on the surface of meat samples ($\log_{10}$ cfu cm^{-2}). In addition two other similar neural networks were developed in which storage time and temperature were excluded from the input layer as dependent variables, in an attempt to investigate the performance of the network to discriminate meat samples based only on FTIR data. Therefore four neural networks were developed in total. Based on previous work (Argyri et al., 2009) the best performance of the network was obtained with 10 neurons in the hidden layer. To facilitate comparison between the two models, the database was also randomly divided into a training subset with 75% of the data, and a test subset with the remaining 25%. These data were not employed at all in the training session of the network but they were used to assess its capability to foresee for unknown cases. The MLP network was developed using NeuralTools version 1.0 (Palisade Corp., Ithaca, NY, USA).

2.5 Evaluation of model performance

The classification accuracy of the neural network and PLS model was determined by the number of correctly classified meat samples in each sensory class divided by the total number of samples in the class. The overall correct classification (accuracy, %) of the model was determined as the number of correct classifications in all classes divided by the total number of samples analyzed (Panigrahi et al., 2006). For the prediction of total viable counts (TVC) in each meat sample three performance indices were calculated, namely the bias (B_f) and accuracy (A_f) factors (Ross, 1996) and the root mean squared error (*RMSE*).

The bias factor (B_f) indicates whether, on average, the observed TVC counts are above or below the line of equity ($y = x$), and if so, by how much. The index is defined as:

$$B_f = 10^{\left(\frac{\sum \mathbf{log}\left(\frac{\mathbf{log}\, N(t)_{predicted}}{\mathbf{log}\, N(t)_{observed}}\right)}{n}\right)} \qquad (5)$$

where n is the number of observations. A bias factor = 1 indicates a perfect model where the predictions are in full agreement with observations. Values < 1 indicate that the observed total viable counts are larger than predicted ones.

The accuracy factor is a measure of the average deviation between predictions and observations, i.e. how close predictions are to observations.

$$A_f = 10^{\left(\frac{\sum \mathbf{log}\left(\left|\frac{\mathbf{log}\, N(t)_{predicted}}{\mathbf{log}\, N(t)_{observed}}\right|\right)}{n}\right)} \qquad (6)$$

The values of this index are ≥ 1. The larger the value the less accurate is the average estimate.

The goodness of fit of the modelling approach was also evaluated by the root mean square error (RMSE), which measures the average deviation between observed and predicted values (Ratkowsky, 2004). The smaller the value of this index the better the fit of the model to the experimental data:

$$RMSE = \sqrt{\frac{\sum\left(\mathbf{log}\, N(t)_{predicted} - \mathbf{log}\, N(t)_{observed}\right)^2}{n}} \qquad (7)$$

where n is the number of observations.

3. Results

Typical FTIR spectral data from 1000 to 1800 cm^{-1} collected from beef fillets stored at 0°C for different storage times are presented in Figure 1. The selected spectra correspond to each one of the three quality classes (i.e., fresh, semi-fresh, spoiled) employed in this work. Based on Figure 1, a major peak at 1640 cm^{-1} was apparent in the meat sample due to the presence of moisture (O-H stretch) with an underlying contribution from amide I, whereas a second peak at 1550 cm^{-1} appeared due to the absorbance of amide II (N-H bend, C–N stretch). A second amide vibration was observed at 1400 cm$^{−1}$ (C–N stretch), followed by amide III peaks at 1315 and at 1240 (C-N stretch, N-H bend, C-O stretch, O=C-N bend). The peaks at 1460, 1240 and 1175 cm^{-1} can be attributed also to fat. Finally, the peaks arising from 1025 to 1140 could be absorbance due to amines (C-N stretch) (Chen et al., 1998; Ellis et al., 2002, 2004; Ammor et al., 2009; Argyri et al., 2009).

A PLS model performance evaluation was performed using leave-one-out cross validation for the prediction of sensory class of beef samples. The number of latent variables (LVs) was selected on the basis of the highest number of correctly classified samples of the testing subset. For this reason, different models were developed with the LVs ranging from 1 to 25. For each model, the number of correctly classified samples in both the training and test dataset was calculated. When the PLS models were built based entirely on the FTIR data (*i.e* no storage time and temperature was included), a number of 21 LVs was finally selected presenting the highest correct classification (%) in the training (98.2%) and test (68.4%) subsets (Fig. 2, Table 1). For the training subset, the PLS approach provided 100% correct classification for fresh and semi-fresh meat samples, whereas for spoiled samples the respective number was 96.1%, representing 1 misclassification out of 26 spoiled samples (Table 1). However, for the testing subset the relative percentages were lower, which is not unusual as these data were not involved at all in model development but provided as unknown cases for prediction. Specifically, the highest correct classification was observed in spoiled (71.4%) and fresh (75%) samples, with 2 samples misclassified as semi-fresh out of 7 and 8 samples, respectively. The lowest performance was obtained in semi-fresh samples with 2 misclassifications out of 4 samples. However, the performance was slightly improved when storage time and temperature were associated with the training data prior to building the model. The best performance in this case was monitored when 20 LVs (Fig. 2), showing a performance of 94.7% on the training and 70.0% on the independent testing dataset. For the training dataset, the PLS approach provided 18 out of 20 correct classification for fresh meat samples (Table 2), whereas for semi-fresh and spoiled samples, the respective numbers were 5 and 6 misclassifications out of 15 semi-fresh and 22 spoiled samples, respectively.

Similar performance was obtained for the ANN model developed entirely on the FTIR dataset (*i.e.* storage time and temperature were excluded from model development as dependent variables). The obtained correct classifications were 98.2% and 63.1% for the training and test datasets, respectively (Table 1). Within each sensory class in the training dataset, the ANN model provided 100% correct discrimination for fresh and semi-fresh samples, whereas for spoiled samples there was 1 misclassification out of 27 meat samples (96.3%). However, for the test dataset the performance of the ANN was lower but still comparable with the PLS model. Specifically, the highest correct classification was obtained for the fresh and spoiled sensory class where 2 samples were misclassified as spoiled and fresh, respectively (Table 1). Less consistent results were obtained for the semi-fresh class with 3 misclassifications out of 5 samples which is quite reasonable taking into account that sensorial discrimination of this class is rather difficult and requires highly trained taste panels. The performance of the ANN model was slightly improved when storage time and temperature were included as additional inputs in model development (Table 2). The obtained results indicated that correct classification increased by approximately 2% and 10% for the training and test datasets, respectively. In this case, the ANN provided 100% correct classification for all sensory classes in the training dataset. With regard to the test dataset, classification performance was improved by approximately 14% for the spoiled meat samples, compared with the ANN model developed on FTIR data only, with 1 misclassification out of 7 samples. For fresh and semi-fresh meat samples, the calculated correct classifications were 71.4% and 60.0%, representing 2 misclassifications out of 5 semi-fresh and 7 spoiled meat samples, respectively (Table 2).

The PLS approach was also used to associate spectral data with total viable counts (TVC) on the surface of meat samples. The model was developed on the assumption that when the difference between individual predictions and observations was higher than a threshold value of 1 log unit, then the prediction was false. When PLS was applied using only the FTIR data (i.e. no storage time and temperature was included within the input matrix), the model correctly predicted 87.7% of the training data, and 60% of the independent testing data. In the case of including the storage time and temperature within the input dataset, the model showed an increase in performance, reaching 100% and 84.2% for the training and testing, respectively.

For models developed on FTIR data only, the calculated value of the bias factor for the ANN training dataset was close to 1 indicating no systematic bias (under or overprediction) (Table 3), whereas for PLS model a slight underestimation was evident (B_f 0.967). The values of bias factor were improved when storage time and temperature were included as inputs in model development, especially for the PLS approach (Table 4). For the test datasets, underprediction ($B_f < 1$) was observed for the PLS models whereas overprediction ($B_f > 1$) was evident in ANN models, regardless of the approach employed in model development (i.e., inclusion or not of storage time and temperature as inputs). These calculations were also graphically verified by the comparison of the observed vs. predicted total viable counts (TVC) plots (Figs. 3 and 4).

Moreover, based on the calculated indices for the test datasets between ANN and PLS models that were developed on FTIR data only, it can be concluded that the PLS model presented a comparatively better performance as it yielded lower values for accuracy factor (1.321) and root mean square error (1.993) (Table 3). However, when storage time and temperature were included as input parameters to the models, then the best performance was obtained for ANN based on the comparison of the same indices (Table 4).

4. Discussion

So far the assessment of meat quality and safety is based on sensory and retrospective microbiological analyses (Nychas et al., 2008). Sensory analysis is an important and common method to evaluate quality of food commodities since the consumer is the ultimate judge of quality of a product (Lee and O'Mahony, 2005). However, the method has certain disadvantages as it relies on highly trained taste panels, a procedure which makes it costly and unattractive for daily analysis. In addition, a limited number of samples can be analysed daily due to the fatigue of the senses of the panellists. Finally, sensory evaluation has a subjective connotation, although this effect could be reduced by applying scientific protocols under carefully controlled conditions. On the other hand, microbiological analyses are laborious, time-consuming, costly and highly technical (molecular tools), as well as destructive to products analysed, requiring in most cases a complex process of sample preparation, while not able to give the 'immediate answer required' (McMeekin et al., 2007).

A major challenge of the meat industry in the 21st century is to obtain reliable information on meat quality and safety throughout the production, processing, and distribution chain, and finally turn this information into practical management support systems to ensure high quality final products for the consumer (Damez and Clerjon, 2008; Sofos, 2008). These systems must be readily available to the industry, and easy-to-use without requiring special expertise form the end-users. Certain databases are available today, such as the Combase (www.combase.cc) and Sym'Previus (www.symprevius.net) providing information on growth/death kinetics of microorganisms in order to define the shelf-life of various foods incorporating mathematical models (Baranyi and Tamplin, 2004; Leporq et al., 2005). It must be stressed however, that the existing predictive microbiology spoilage models tend to underestimate important factors such as microbial interaction among the members of the microbial association as well as with the food matrix (Wilson et al., 2002; Koutsoumanis et al., 2004). In the latter case the changes in the concentration of microbial metabolites on meat surface due to microbial activity can be used to monitor quality deterioration. There is thus a need to replace, or at least limit, the number and extent of microbiological analyses, with (bio)chemical analyses in an attempt to define metabolic indices as potential indicators of spoilage. The concept is not new and it was proposed as a promising alternative to monitor meat spoilage in the late 80s and 90s (McMeekin, 1982; Gill, 1986; Nychas et al., 1988; Kakouri and Nychas, 1994; Dainty, 1996). However, the idea of a single biochemical substance as spoilage indicator put forward at that time, has been replaced today by the metabolomic concept which is based on a holistic approach of spoilage profile (Goodacre et al., 2004; Nychas et al., 2008).

Recent developments in sensor technologies and data analysis procedures have stimulated interest in developing rapid and non-invasive techniques to monitor changes in meat quality. Among these, spectroscopic methods are widely used for muscle food quality assessment and control, in both laboratory and meat industry installations (Hildrum et al., 2006). In contrast to conventional methods for the determination of meat quality parameters, Fourier transform infrared spectroscopy (FTIR) is a sensitive, rapid and non-destructive analytical technique, with simplicity in sample preparation, allowing simultaneous assessment of numerous meat properties. This technique has found numerous applications in foods such as olive oil (Maggio et al., 2010), honey (Kelly et al., 2006), wine (Versari et al., 2010), coffee (Briandet et al., 1996). Ellis et al. (2002, 2004) have been the pioneers to report that FTIR spectral data collected directly from the surface of meat could be used as biochemical interpretable “fingerprints” to provide information on early detection of microbial spoilage of chicken breast and rump steaks. However, the amount of information provided by spectral data require special data mining techniques based on multivariate statistical analysis (e.g. cluster analysis, principal components analysis, discriminant function analysis, partial least squares regression) and/or soft computing methodologies (e.g. artificial neural networks, genetic algorithms, support vector machines) to provide information related to (a) the responses of specific spoilage microorganisms in meat and (b) the discrimination of meat samples in quality classes (Goodacre, 2000; Mataragas et al., 2007; Verouden et al., 2009).

In the present work, FTIR spectral data from beef fillets stored under aerobic conditions at five different storage temperatures were analyzed by partial least squares regression in an effort to classify meat samples in three sensorial categories (fresh, semi-fresh, spoiled) as defined by a taste panel. The performance of the PLS approach was compared with a multi-layer perceptron (MLP) neural network. Two different approaches were followed in model development. Firstly, storage time and temperature were treated as input variables and associated with FTIR spectral data during model development. However, in practice, the history of a meat sample in terms of storage temperature and time is not always known, and hence meat quality must be assessed by spectral data only. To cope with this issue separate models were developed based on the FTIR data only and the two approaches were compared.

Results showed relatively better performance when storage time and temperature were included as inputs in model development, as a more precise dataset was used for the training of models. Good classification accuracies were obtained for fresh and spoiled meat samples, demonstrating the effectiveness of the method to discriminate samples between these two classes (Table 1 and 2). The high classification rate of both models (i.e., PLS and ANN) could be associated to the beginning of proteolysis in meat (Nychas and Tassou, 1997) resulting in changes in the concentration of amides and amines (Ellis and Goodacre, 2001), as well as to glucose consumption and the resulting changes in the levels of organic acids (Dainty, 1996; Nychas et al., 1998). It must be emphasized however that the number of examined samples within each class was not equal due to the different spoilage rate of beef samples at different storage temperatures resulting in variable number of samples in each class. This may have affected the training process which is basically a data driven approach (Basheer and Hajmeer, 2000), and could thus account for the lower classification accuracies observed in certain classes (e.g. fresh and semi-fresh) (Table 1 and 2). Finally, the lower accuracies observed in the semi-fresh class could also be attributed to the performance of the taste panel, as the difference between “fresh/semi-fresh” and “semi-fresh/spoiled” is sometimes subjective and affects the overall classification, as the developed models are based on supervised training for parameter optimization.

Another interesting perspective from a microbiological point of view would be the correlation of FTIR spectra to bacterial population counts on the surface of meat samples. In this way laborious and time consuming microbiological analyses could be replaced in the long term by spectral data in order to provide rapid, low cost and non-invasive microbiological analyses (Nychas et al., 2008). The graphical plots between observed and predicted total viable counts as well as the calculated performance indices showed that for models developed on FTIR spectral data alone better performance was obtained by the PLS model (Table 3; Fig. 3) although the model had a tendency to underestimate total viable counts. However, when storage time and temperature were included in model development together with FTIR data the best performance was obtained by ANN (Table 4; Fig. 4). Generally, ANN models tended to overestimate microbial counts (B_f > 1) in contrast to PLS models where underestimation of total viable counts was evident (B_f < 1). An interesting alternative approach to evaluate the effectiveness of FTIR spectral data in the determination of sensory rating and total viable counts prediction in meat samples, would be the implementation of experimental studies in which meat samples would have been artificially contaminated with spoilage bacteria at different initial populations. Further research is needed in this direction as results from such studies would be valuable in the evaluation of the robustness of the FTIR approach.

In conclusion, the correlation between microbial growth and chemical changes during storage has been recognized as a way to identify indicators that could be employed to quantify quality as well as the degree of spoilage. Spectral data collected from FTIR analysis combined with an appropriate machine learning strategy (partial least squares regression, artificial neural networks) could become an interesting tool to monitor beef fillets spoilage through the measurement of biochemical changes occurring in meat substrate. Future work should also focus on the association of specific microbial groups (e.g. lactic acid bacteria, pseudomonads, enterobacteria) with FTIR spectral data in an attempt to increase the prediction performance of the models.

Acknowledgements

The authors acknowledge the Symbiosis-EU (www.symbiosis-eu.net) project (no 211638) financed by the European Commission under the 7th Framework Programme for RTD. The information in this document reflects only the authors' views and the Community is not liable for any use that may be made of the information contained therein.

References

Ammor, M.S., Argyri, A., Nychas, G.-J.E., 2009. Rapid monitoring of the spoilage of minced beef stored under conventionally and active packaging conditions using Fourier transform infrared spectroscopy in tandem with chemometrics. Meat Sci. 81, 507-514.

Argyri, A.A., Panagou, E.Z., Tarantilis, P.A., Polysiou, M., Nychas, G.-J.E., 2009. Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks. Sens. Actuators B: Chem. doi:10.1016/j.snb.2009.11.052.

Balasubramanian, S., Panigrahi, S., Logue, C.M., Gu, H., Marchello, M., 2009. Neural networks-integrated metal oxide-based artificial olfactory system for meat spoilage identification. J. Food Eng. 91, 91-98.

Baranyi, J., Tamplin, L.M., 2004. Combase: a common database on microbial responses in food environments. J. Food Prot. 67, 1967-1971.

Basheer, I.A., Hajmeer, M., 2000. Artificial neural networks: fundamentals, computing, design, and application. J. Microbiol. Methods 43, 3-31.

Bishop, C.M., 2004. Neural networks for pattern recognition. Oxford University Press, Oxford.

Briandet, R., Wilson, R.H., Kemsley, E.K., 1996. Discrimination of Arabica and Robusta in instant coffee by Fourier transform infrared spectroscopy and chemometrics. J. Agric. Food Chem. 44, 170-174.

Chen, M., Irudayaraj, J., McMahon, D.J., 1998. Examination of full fat and reduced fat cheddar cheese during ripening by Fourier Transform Infrared Spectroscopy. J. Dairy Sci. 81, 2791–2797.

Dainty, R.H., 1996. Chemical/biochemical detection of spoilage. Int. J. Food Microbiol. 33, 19-34.

Damez, J.L., Clerjon, S., 2008. Meat quality assessment using biophysical methods related to meat structure. Meat Sci. 80, 132-149.

Ellis, D.I., Broadhurst, D., Clarke, S.J., Goodacre, R., 2005. Rapid identification of closely related muscle foods by vibrational spectroscopy and machine learning. Analyst 130, 1648-1654.

Ellis, D.I., Broadhurst, D., Goodacre, R., 2004. Rapid and quantitative detection of the microbial spoilage of beef by Fourier transform infrared spectroscopy and machine learning. Anal. Chim. Acta 514, 193-201.

Ellis, D.I., Broadhurst, D., Kell, D.B., Rowland, J.J., Goodacre, R., 2002. Rapid and quantitative detection of the microbial spoilage of meat by Fourier transform infrared spectroscopy and machine learning. Appl. Environ. Microbiol. 68, 2822-2828.

Ellis, D.I., Goodacre, R., 2001. Rapid and quantitative detection of the microbial spoilage of muscle foods: current status and future trends. Trends Food Sci. Technol. 12, 414-424.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 2001. Multi- and megavariate data analysis: Principles and applications. Umetrics AB, Sweden.

Esnoz, A., Periago, P.M., Conesa, R., Palop, A., 2006. Application of artificial neural networks to describe the combined effect of pH and NaCl on the heat resistance of *Bacillus stearothermophilus.* Int. J. Food Microbiol. 106, 153-158.

García-Gimeno, R.M., Hervás-Martínez, C., Rodríguez-Pérez, R., Zurera-Cosano, G., 2005. Modelling the growth of *Leuconostoc mesenteroides* by artificial neural networks. Int. J. Food Microbiol. 3, 317-332.

Geeraerd, A.H., Herremans, C.H., Cenens, C., Van Impe, J.F., 1998. Application of artificial neural networks as a non-linear modular modelling technique to describe bacterial growth in chilled food products. Int. J. Food Microbiol. 44, 49-68.

Gill, C.O., Jeremiah, L.E., 1991. The storage life of non-muscle offals packaged under vacuum or carbon dioxide. Food Microbiol. 8, 339-353.

Gill, C.O., 1986. The control of microbial spoilage in fresh meats. In: Pearson, A.M., Dutson, T.R. (Eds.), Advances in Meat Research: Meat Poultry Microbiology. AVI Publishing, Co., Westport, CT, pp. 49-88.

Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G., Kell, D.B., 2004. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol. 22, 245-252.

Goodacre, R., 2000. Applications of artificial neural networks to the analysis of multivariate data. In: Cartwright, H.M. (Ed.), Intelligent Data Analysis in Science: a Handbook. Oxford University Press, Oxford, pp. 123-152.

Guillén, A., del Moral, F.G., Herrera, L.J., Rubio, G., Rojas, I., Valenzuela, O., Pomares, H., 2010. Using near-infrared spectroscopy in the classification of white and Iberian pork with neural networks. Neural Comput. Applic. 19, 465-470.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate data analysis with readings. Prentice-Hall, New Jersey.

Ham, F.M., Kostanic, I., 2001. Fundamental neurocomputing concepts. In: Ham, F.M., Kostanic, I. (Eds.), Principles of Neurocomputing for Science and Engineering. Arnold Publishers, London, pp. 24-91.

Hervás, C., Zurera, G., García, R.M., Martínez, J., 2001. Optimization of computational neural network for its application to the prediction of microbial growth in foods. Food Sci. Technol. Int. 7, 159-163.

Huang, Y., Kangas, L.J., Rasco, B.A., 2007. Applications of artificial neural networks (ANNs) in food science. Crit. Rev. Food Sci. Nutr. 47, 113-126.

Hildrum, K.I., Wold, J.P., Vegard, H.S., Renou, J.P., Dufour, E., 2006. New spectroscopic techniques for on-line monitoring of meat quality. In: Nollet, L.M.L., Toldra, F. (Eds.), Advanced Technologies for Meat Processing. CRC Press, Boca Raton, FL.

Kakouri, A., Nychas, G.-J.E., 1994. Storage of poultry meat under modified atmospheres or vacuum packs: possible role of microbial metabolites as indicator of spoilage. J. Appl. Bacteriol. 76, 163-172.

Kelly, J.D., Petisco, C., Downey, G., 2006. Application of Fourier transform midinfrared spectroscopy to the discrimination between Irish artisanal honey and such honey adulterated with various sugar syrups. J. Agric. Food Chem. 54, 6166-6171.

Koutsoumanis, K.P., Kendall, P.A., Sofos, J.N., 2004. A comparative study on growth limits of *Listeria monocytogenes* as affected by temperature, pH and a_w when grown in suspension or on a solid surface. Food Microbiol. 21, 415-422.

Lee, H.S., O'Mahony, M., 2005. Sensory evaluation and marketing: Measurement of a consumer concept. Food Qual. Prefer. 16, 227-235.

Leporq, B., Membré, J.M., Dervin, C., Buche, P., Guyonnet, J.P., 2005. The "Sym'Previus" software, a tool to support decisions to the foodstuff safety. Int. J. Food Microbiol. 100, 231-237.

Liu, Y., Lyon, B.G., Windham, W.R., Lyon, C.E., Savage, E.M., 2004. Prediction of physical, color, and sensory characteristics of broiler breasts by visible/near infrared reflectance spectroscopy. Poultry Sci. 83, 1467-1474.

Lou, W., Nakai, S., 2001. Application of artificial neural networks for predicting the thermal inactivation of bacteria: a combined effect of temperature, pH and water activity. Food Res. Int. 34, 573-579.

Maggio, R.M., Cerretani, L., Chiavaro, E., Kaufman, T.S., Bendini, A., 2010. A novel chemometric strategy for the estimation of extra virgin olive oil adulteration with edible oils. Food Control 21, 890-895.

Mataragas, M., Skandamis, P., Nychas, G.-J.E., Drosinos, E.H., 2007. Modeling and predicting spoilage of cooked, cured meat products by multivariate analysis. Meat Sci. 77, 348-356.

McMeekin, T.A., 1982. Microbial spoilage of meats. In: Davies, R. (Ed.), Developments in Food Microbiology. Applied Science Publishers, London, pp. 1-40.

McMeekin, T.A., Bowman, J., Dobson, S., Mellefont, I., Ross, T., Tamplin, M., 2007. The future of predictive microbiology: Innovative applications and great expectations. In: Nychas, G.-J.E., Taoukis, P., Koutsoumanis, K., Van Impe, J., Geeraerd, A. (Eds.), Proceedings of the 5th international conference of predictive modelling in foods “Fundamentals, state of the art and new horizons”. Athens, Greece, pp. 1-4.

NBIC (2002) Converging Technologies for Improving Human Performance. Nanotechnology, Biotechnology, information technology and cognitive science (Eds) Mihail C. Roco and William Sims Bainbridge, National Science Foundation www.wtec.org/ConvergingTechnologies/Report/NBIC_report.pdf

Nychas, G.-J.E., Skandamis, P.N., Tassou, C.C., Koutsoumanis, K.P., 2008. Meat spoilage during distribution. Meat Sci. 78, 77-89.

Nychas, G.-J.E., Drosinos, E.H., Board, R.G., 1998. Chemical changes in stored meat. In: Board, R.G., Davies, A.R. (Eds.), The Microbiology of Meat and Poultry. Blackie Academic and Professional, London, pp. 288-326.

Nychas, G.-J.E., Tassou, C.C., 1997. Spoilage process and proteolysis in chicken as noted by HPLC method. J. Sci. Food Agric. 74, 199-208.

Nychas, G.-J.E., Dillon, D.M., Board, R.G., 1988. Glucose: a key substrate in microbial spoilage of meat and meat products. Biotechnol. Appl. Biochem. 10, 203-231.

Palanichamy, A., Jayas, D.S., Holley, R.A., 2008. Predicting survival of *Escherichia coli* O157:H7 in dry fermented sausage using artificial neural networks, J. Food Prot. 71, 6-12.

Panagou, E.Z., 2008. A radial basis function neural network approach to determine the survival of *Listeria monocytogenes* in katiki, a traditional Greek soft cheese. Journal of Food Protection, 71(4):750-759.

Panigrahi, S., Balasubramanian, S., Gu, H., Logue, C., Marchello, M., 2006. Neural-network-integrated electronic nose system for identification of spoiled beef. Lebensm.-Wiss. Technol. 39, 135-145.

Poirazi, P., Leroy, F., Georgalaki, M.D., Aktypis, A., Vuyst, L., Tsakalidou, E., 2007. Use of artificial neural networks and a gamma-concept-based approach to model growth of and bacteriocin production by *Streptococcus macedonicus* ACA-DC 198 under simulated conditions of kasseri cheese production, Appl. Environ. Microbiol. 73, 768-776.

Prieto, N., Roehe, R., Lavín, P., Batten, G., Andrés, S., 2009. Application of near infrared reflectance spectroscopy to predict meat and meat products quality: a review. Meat Sci. 83, 175-186.

Rajamäki, T., Alakomi, H.L., Ritvanen, T., Skyttä, E., Smolander, M., Ahvenainen, R., 2006. Application of an electronic nose for quality assessment of modified atmosphere packaged poultry meat. Food Control 17, 5-13.

Ratkowsky, D.A., 2004. Model fitting and uncertainty. In: McKellar, R.C., Lu, X. (Eds.), Modeling Microbial Responses in Food. CRC Press, Boca Raton, FL, pp. 152-196.

Ross, T., 1996. Indices for performance evaluation of predictive model in food microbiology. J. Appl. Microbiol. 81, 501-508.

Singh, K.P., Ojha, P., Malik, A., Jain, G., 2009. Partial least squares and artificial neural networks modeling for predicting chlorophenol removal from aqueous solution. Chemom. Intell. Lab. Syst. 99, 150-160.

Siripatrawan, U., Linz, J.E., Harte, B.R., 2006. Detection of *Escherichia coli* in packaged alfalfa sprouts with an electronic nose and an artificial neural network. J. Food Prot. 69, 1844-1850.

Sofos, J.N., 2008. Challenges to meat safety in the 21st centrury. Meat Sci. 78, 3-13.

United States Department of Agriculture (USDA), 2008. U.S. Beef and Cattle Industry: Background Statistics and Information (http://www.ers.usda.gov/news/BSECoverage.htm, assessed 16.02.2010).

Verouden, M.P.H., Westerhius, J.A., Werf, M.J.V., Smilde, A.K., 2009. Exploring the analysis of structured metabolomics data. Chemometr. Intell. Lab. 98, 88-96.

Versari, A., Parpinello, G.P., Scazzina, F., Rio, D.D., 2010. Prediction of total antioxidant capacity of red wine by Fourier transform infrared spectroscopy. Food Control 21, 786-789.

Wilson, P.D.G., Brocklehurst, T.F., Arino, S., Thuault, D., Jakobsen, M., Lange, M., Farkas, J., Wimpenny, J.W.T., Van Impe, J.F., 2002. Modelling microbial growth in structured foods: towards a unique approach. Int. J. Food Microbiol. 73, 275-289.

Fig. 1. Typical FTIR spectra in the range of 1800 to 1000 cm^{-1} collected from beef fillets stored at 0°C at the beginning of storage (A; Fresh), after 96 h (B; Semi-fresh), and 216 h (C; Spoiled).

Fig. 2. Optimization of the PLS-DA classification models using latent variables ranging from 1 to 25 for the training (grey line) and test (black line) subsets after leave-one-out cross validation. (A) sensory class prediction based on FTIR data; (B) total viable counts prediction based on FTIR data; (C) sensory class prediction based on FTIR data plus storage time and temperature as additional inputs; (D) total viable counts prediction based on FTIR data plus storage time and temperature as additional inputs

Fig. 3. Comparison between observed and predicted total viable counts (TVC) of beef fillets by the ANN (a) and the PLS-DA (b) model based on FTIR spectral data (open symbols: training data; solid symbols: test data; dotted lines are ± 1 log units area).

Fig. 4. Comparison between observed and predicted total viable counts (TVC) of beef fillets by the ANN (a) and the PLS-DA (b) model based on FTIR spectral data with storage time and temperature as additional inputs to the models (open symbols: training data; solid symbols: test data; dotted lines are ± 1 log units area).

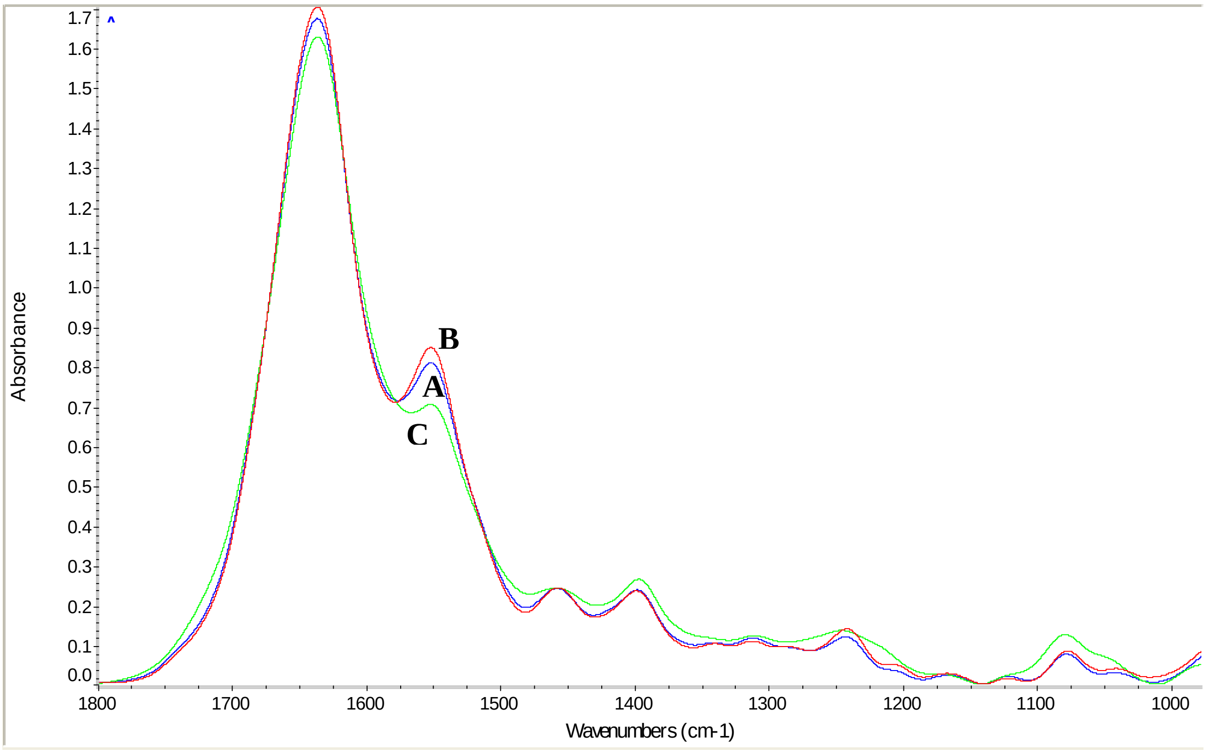
Absorbance
Wavenumbers (cm-1)
A
B
C

Fig. 1.

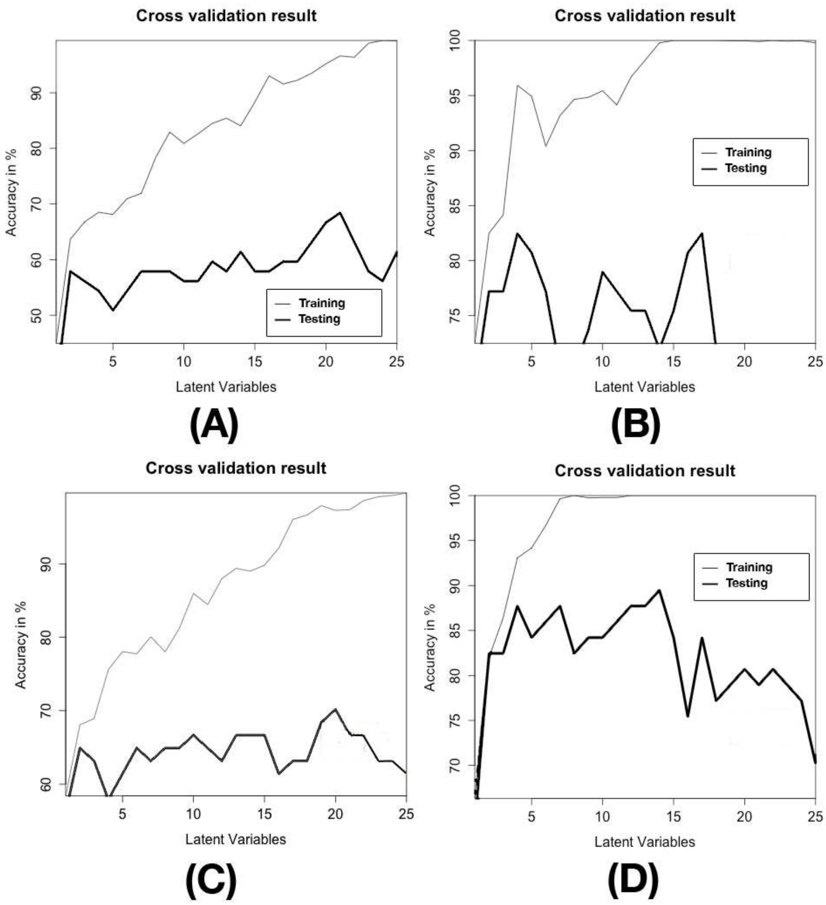
Cross validation result
Accuracy in %
Latent Variables
Training
Testing
(A)
(B)
(C)
(D)

Fig. 2.

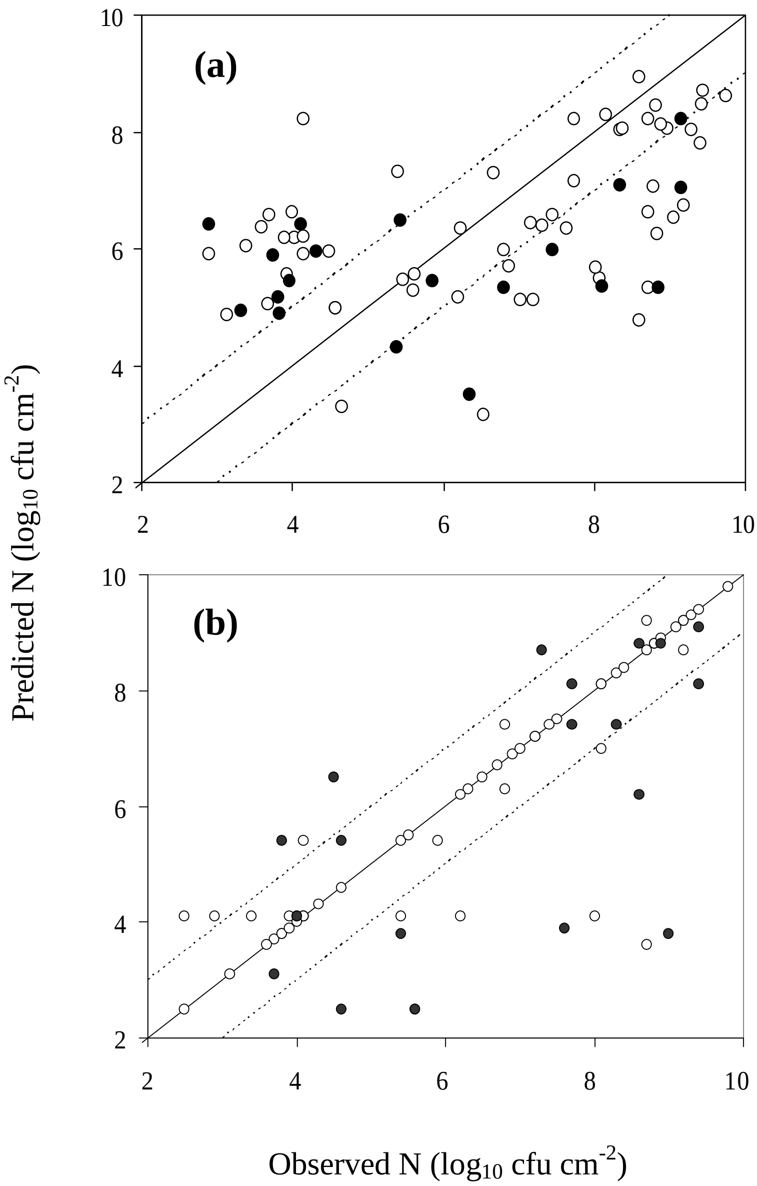
(a)
(b)
10
8
6
4
2
2
4
6
8
10
Predicted N (log10 cfu cm-2)
Observed N (log10 cfu cm-2)

Fig. 3

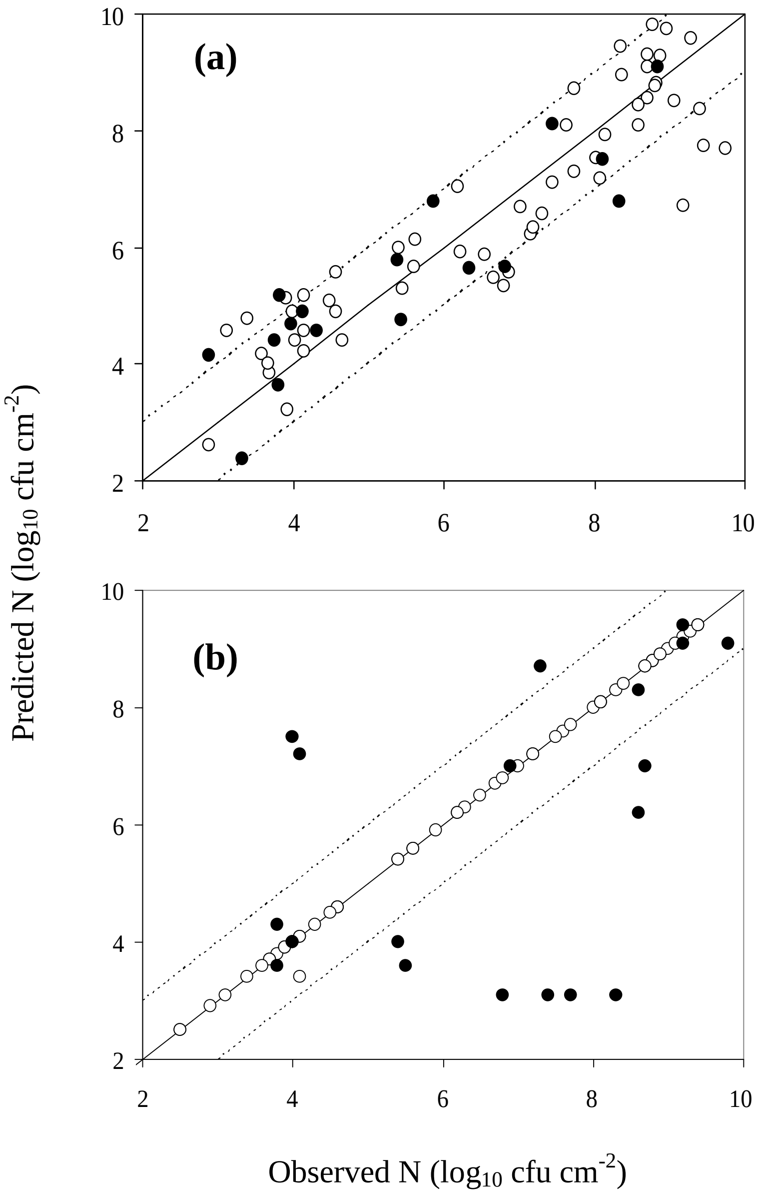

Fig. 4

Table 1. Confusion matrix of the ANN classifier and the PLS model regarding sensory quality discrimination of beef fillets based on FTIR spectral data.

From/to	ANN training (n = 57)				
	Fresh	Semi-fresh	Spoiled	Total	Correct (%)
Fresh	19	0	0	19	100
Semi-fresh	0	11	0	11	100
Spoiled	1	0	26	27	96.3
	ANN testing (n = 19)				
Fresh	5	0	2	7	71.4
Semi-fresh	2	2	1	5	40.0
Spoiled	2	0	5	7	71.4
	PLS training (n = 57)				
Fresh	18	0	0	18	100
Semi-fresh	0	13	0	13	100
Spoiled	0	1	25	26	96.1
	PLS testing (n = 19)				
Fresh	6	2	0	8	75.0
Semi-fresh	2	2	0	4	50.0
Spoiled	0	2	5	7	71.4

Overall correct classification (accuracy) for ANN train and test datasets: 98.2% and 63.1%, respectively.

Overall correct classification (accuracy) for PLS train and test datasets: 98.2% and 68.4%, respectively.

Table 2. Confusion matrix of the ANN classifier and the PLS model regarding sensory quality discrimination of beef fillets based on FTIR spectral data together with storage time and temperature as additional inputs to the models.

From/to	ANN training (n = 57)				
	Fresh	Semi-fresh	Spoiled	Total	Correct (%)
Fresh	19	0	0	19	100
Semi-fresh	0	11	0	11	100
Spoiled	0	0	27	27	100
	ANN testing (n = 19)				
Fresh	5	0	2	7	71.4
Semi-fresh	2	3	0	5	60.0
Spoiled	1	0	6	7	85.7
	PLS training (n = 57)				
Fresh	18	2	0	20	90.0
Semi-fresh	2	10	3	15	66.7
Spoiled	1	5	16	22	72.7
	PLS testing (n = 19)				
Fresh	5	1	0	6	83.4
Semi-fresh	0	1	1	2	50.0
Spoiled	2	1	8	11	72.7

Overall correct classification (accuracy) for ANN train and test datasets: 100.0% and 73.7%, respectively.

Overall correct classification (accuracy) for PLS train and test datasets: 77.2% and 73.6%, respectively.

Table 3. Comparison of validation indices between the PLS and ANN models for total viable counts (TVC) predictions in meat samples based on FTIR spectral data.

Parameter	ANN		PLS model	
	Train	Test	Train	Test
Bias factor (B_f)	1.002	1.034	0.967	0.854
Accuracy factor (A_f)	1.291	1.390	1.090	1.321
RMSE	1.821	1.978	1.073	1.993

Table 4. Comparison of validation indices between the PLS and ANN models for total viable counts (TVC) prediction in meat samples based on FTIR spectral data together with storage time and temperature as additional inputs to the model.

Parameter	ANN		PLS model	
	Train	Test	Train	Test
Bias factor (B_f)	1.008	1.038	0.996	0.833
Accuracy factor (A_f)	1.118	1.166	1.003	1.409
RMSE	0.852	0.921	0.092	2.501