

Developing robust food composition models: strategies for handling temperature and packaging variations in dry-cured ham using near infrared spectrometry

Fulladosa, E^{*1}, Chong, M. W. S.², Parrott, A. J.², Dos Santos, R¹, Russell, J.², Nordon, A².

¹IRTA. Food Quality and Technology. Finca Camps i Armet, s/n, 17121 Monells, Girona, Spain.

²WestCHEM, Department of Pure and Applied Chemistry and Centre for Process Analytics and Control Technology (CPACT), University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL, UK

*Corresponding author: elena.fulladosa@irta.cat

ABSTRACT

Low-cost near infrared devices intended for consumers able to easily determine composition and quality of food products may boost adoption of sustainable healthy diets. However, predictive algorithms robust to external variations are needed. The aim of this work was to evaluate different data analysis strategies to develop robust predictive models for food composition when using spectrometric data subjected to external variations, specifically temperature and packaging material, acquired using low-cost sensors. Usefulness of global modelling (GM), Generalised least squares weighting (GLSW), Loading space standardisation (LSS), Multiplicative Effects Model (MEM) were explored, and the effect of samples heterogeneity evaluated. To do so, two low-cost handheld NIR-based devices with different spectral ranges and resolutions were used. The food matrix samples were obtained from different anatomical muscles of commercial dry-cured ham. Spectra were acquired on two types of packaging films at different temperatures to further explore the usefulness of global modelling (GM), generalised least squares weighting (GLSW), loading space standardisation (LSS), and multiplicative effects model (MEM) to retrieve these effects. Results show that the inherent food sample heterogeneity produces as much spectral variability as temperature and packaging materials. For temperature compensation, LSS did not decrease the predictive error caused by this factor probably due to the heterogeneity of the samples used. In contrast, the GLSW method decreased the predictive errors from 0.52% to 0.46% for salt and from 2.10% to 1.40% for water. Only a slight effect of packaging was observed, and GM models were found to be the best strategy to compensate it, showing a decrease of bias from -1.35 to 0.012. The examined compensation strategies could facilitate the deployment of low-cost spectrometers for consumer use, as they offer an effective means to mitigate or eliminate variations from any source in the data that are unrelated to the properties of interest.

Keywords

Temperature, packaging, low-cost spectrometers, near infrared spectroscopy, global modelling, Generalised least squares weighting, Loading space standardisation, Multiplicative Effects Model

Highlights

- GLSW was the best strategy to deal with temperature variations
- Sample heterogeneity can interfere with the compensation strategies
- GM is the best strategy to compensate for scattering caused by packaging
- Compensation strategies are useful to improve in situ composition predictions on food

1. Introduction

Food safety and implementation of sustainable healthy diets is a challenge for all European governments. In a society increasingly focused on ICT (Information and communication technologies), the improvement and adaptation of comprehensive digital tools is crucial to facilitate the achievement of the mentioned challenges. In this sense, the development of new digital devices that provide real-time information on food composition and quality could help consumers make informed, healthier, and more sustainable purchasing decisions; assist retail stores in managing stocks and preventing food waste; and boost consumer confidence in food origin by increasing transparency and reducing the risk of food fraud.

Non-invasive technologies based on infrared and other spectrometric technologies through digital devices could help to achieve these challenges. It has been demonstrated that composition of food [1-5], nutritional claims and mislabelling [6-10] and quality parameters [11-13] can be determined using these technologies. The continuous instrumental developments and applications observed over the last few years have launched near infrared (NIR) spectroscopy into a new era of on-site and in-the field analysis [14]. Besides, manufacturing cost have been considerably reduced and the progress in miniaturization has been accompanied by software development aimed to facilitate its use by the non-expert consumer community [15]. However, caution should be applied with the instruments intended for consumers [16] since reliability of these devices to characterize foods depends on the technical factors such as lighting and the configuration of the equipment but also on the effect of external conditions when acquiring spectra [17]. Near infrared (NIR) spectra are affected by the variation in the physical properties of samples such as surface characteristics or heterogeneity, and external factors such as environmental light, temperature or packaging, which may need to be removed from the spectra. Traditional thermoplastic resin materials, such as polypropylene and polyethylene, used for food packaging may produce different absorption profiles and reflectance interference [18, 19]. In this regard, advanced pre-processing methods to remove the effects of temperature [20, 21] and particle size [22] from spectra have been used before in other fields. These include several filtering techniques such as External Parameter Orthogonalization (EPO) [23], Orthogonal Signal Correction (OSC) [24], Generalised Least Squares Weighting (GLSW) [25], and Loading Space Standardization (LSS) [26]. Other approaches for temperature compensation are augmentation of the calibration matrix through adding simulated noise on the spectra or correction methods that removes the non-relevant variation from new spectra [27].

Despite the significant amount of work done, the application of these approaches in low-cost sensors of various types requires further investigation. This is particularly important due to the different environmental conditions they may encounter, such as temperature fluctuations and

packaging material variations, which are common in the food industry and at the supermarket and consumer's level.

The aim of this work was to evaluate different data analysis strategies to develop robust predictive models for food composition when using spectrometric data subjected to external variations, specifically temperature and packaging materials, acquired using low-cost NIR spectrometers. Usefulness of global modelling (GM), Generalised least squares weighting (GLSW), Loading space standardisation (LSS), Multiplicative Effects Model (MEM) were explored, and the effect of samples heterogeneity evaluated.

2. Material and Methods

2.1 Food matrix and packaging materials

Food matrix samples used in this study were obtained from different anatomical muscles of 50 commercial dry-cured hams (Monte Nevado, Segovia, Spain) named *Biceps femoris*, *Semimembranosus* and *Semitendinosus*. Two types of packaging films commonly used at the food industry were chosen to study spectral variation produced by packaging. Specifically, Plastic film 1 (P1) was a 50 $\mu\text{m} \pm 10\%$ Polyvinylidene Chloride (OSB Cryovac, Barcelona, Spain) and Plastic film 2 (P2) was a 120 $\mu\text{m} \pm 10\%$ layered Polyamine/Polyethylene films (P2) (Estudi Graf, Barcelona, Spain).

2.2 NIR spectrometers

Spectra were acquired using two low-cost handheld NIR-based devices with different spectral ranges and resolutions. Spectrometer A was the smartphone-based SCiO spectrometer (Consumer Physics Inc., Tel Aviv, Israel) that can acquire 331-points reflection spectra that are ranging from 740 to 1070 nm. A shade accessory was used to avoid external's light influence and to keep the same 10 mm distance in all the collected spectra. Device A was calibrated using the optical head containing a bakelite reference plate and the automatic internal calibration of the handheld spectrometer. Afterwards, spectra were collected using the SCiO smartphone app (The Lab, version 1.3.1.81), sent to a smartphone via Bluetooth and then uploaded to the Consumer Physics Cloud database. Spectrometer B was a handheld low-cost NIR setup developed in house (NIRTA 3.0) that uses a Hamamatsu sensor in reflectance mode (Hamamatsu Photonics C14384MA-01; Hamamatsu, Japan) and an internal source of halogen light of 0.55 W of power to acquired spectra in a range between 600 nm and 1160 nm with a 2.18 nm resolution. Spectrometer B was calibrated using a white reference object of 80 x 200 mm made from Teflon (SpectronPro, Resonon Inc., Bozeman, MT, USA) with reflectance values of 99.9% and internal procedure of the system. Afterwards, spectra were acquired at a working distance of 50 mm. For both devices, the spectral signals were downloaded and imported into Matlab®, version 7.1 (The MathWorks, Inc) to further develop the predictive models.

2.3 Experimental procedure

The experimental procedure is represented in Figure 1A. In brief, a total of 155 regions of interest (ROIs) were selected on 2 cm thick slices of the 50 dry-cured hams described above and sampled (Sample set A). Depending on the slice, 0, 1 or 2 ROIs were obtained for a given muscle. Two thirds of the samples were randomly assigned to the calibration set and 1/3 to the validation set. Next, a total of 5 spectra on different positions were acquired directly on the muscle as received (Sample set A, Figure 1b) of non-packed samples (NP) at different temperatures $-7\pm 1.5^{\circ}\text{C}\pm$, $4\pm 1.5^{\circ}\text{C}\pm$ and $10\pm 1.5^{\circ}\text{C}\pm$ or $15\pm 1.5^{\circ}\text{C}$ for spectrometer B and A, respectively. The temperature of the samples while acquiring the spectra was measured using a non-contact infrared thermometer (Testo 830-T2, Testo SE & Co. KGaA, Germany). Additionally, twenty additional samples were sampled and a total of 5 spectra on different positions were acquired at $-15\pm 0.5^{\circ}\text{C}$, $-5\pm 0.5^{\circ}\text{C}$, $5\pm 0.5^{\circ}\text{C}$ and $15\pm 0.5^{\circ}\text{C}$ (non-homogenized LSS set) and temperature measured. This set of samples was used when applying the LSS compensation method described below. Next, all the samples were vacuum packed first using P1, taken out of the package and packaged again using P2, and spectra acquired on the packaging at a constant temperature ($10\pm 1.5^{\circ}\text{C}$ or $15\pm 1.5^{\circ}\text{C}$ for spectrometer B and A, respectively).

In order to study the effect of samples heterogeneity, samples were homogenised using a conventional mincer (Moulinex, Barcelona, Spain) (Sample set B, Figure 1c) and spectra acquired again following the same procedure and temperatures described before. First, spectra were acquired on NP samples at different temperatures ($-7\pm 1.5^{\circ}\text{C}\pm$, $4\pm 1.5^{\circ}\text{C}\pm$ and $10\pm 1.5^{\circ}\text{C}\pm$ or $15\pm 1.5^{\circ}\text{C}$ for spectrometer B and A, respectively) using the two spectrometers, and then vacuum package again first on P1 and later on P2 for spectral acquisition at a constant temperature ($10\pm 1.5^{\circ}\text{C}$ or $15\pm 1.5^{\circ}\text{C}$ for spectrometer B and A, respectively). Additional spectra at $-15\pm 0.5^{\circ}\text{C}$, $-5\pm 0.5^{\circ}\text{C}$, $+5\pm 0.5^{\circ}\text{C}$ and $+15\pm 0.5^{\circ}\text{C}$ on LSS set once homogenized were also acquired to explore the effect of homogenisation when applying LSS compensation method (homogenized LSS set) and temperature measured. Finally, salt and water contents were analytically determined in triplicate as described below.

2.4 Compensation methods

Several compensation methods were investigated as strategies against temperature and/or packaging material variations [38].

Global modelling (GM) involves incorporating in the calibration sample data collected at different conditions (such as temperature variation or packaging materials), covering the expected range for the application. This typically results in a more complex model due to the increase in the number of latent variables needed to describe these additional variances. A major downfall of this

Developing robust food composition models: strategies for handling temperature and packaging variations in dry-cured ham using near infrared spectrometry

method is the need to collect many more spectra at various conditions to ensure that the model space covers the range of expected variation.

Global modelling - measured variable as an independent variable (GM-X) adds a measured variable of the sample (such as temperature), at the time spectra were collected, to the spectral data matrix. Therefore, independent variables used to develop the calibration are made up of the spectral absorbances of interest and the measured variable at which they were collected.

Global modelling-measured variable as a dependant variable (GM-Y) adds the measured variable of the sample (such as temperature), at the time spectra were collected, to the measured data matrix. Therefore, the PLS2 algorithm is used to simultaneously develop calibration models for both variables (i.e salt or water content, and the temperature). By having to predict the two variables simultaneously, it is thought that that the model can identify which spectral regions are affected by the studied variables.

Multivariate filter—generalised least squares weighting (GLSW) down weights spectral variables using a covariance matrix generated from the differences between similar samples defined by their measured value. The extent of the weighting depends on the variable parameter alpha (α). In this study, values for α ranged from 0.000001 to 1.0. Low α values result in increases in the weighting of the filter, whereas high values reduce the effect of the filter. A few different values of alpha were trialled (0.1, 0.01, 0.001, 0.0001 and 0.00001. 0.000001, 0.0000001 and 0.00000001) as to get the best performance from the filter without removing information related to the aimed variable to predict. The use of the optimal α value results in the reduction in the complexity of the model as well as improvements to the predictive ability of the model.

Loading space standardisation (LSS) assumes that the absorbance of each chemical species in every wavelength follows simple polynomials with respect to temperature. Therefore, the corresponding elements of the loading vectors for spectral data measured at different temperatures can also be described by simple polynomials. In LSS method, the polynomials estimated from a specific set of samples (LSS set) measured at specific temperatures (Step 1) can be used to predict the loading vectors at the test temperature. This step plays an important role in correcting the spectral variations resulting from the temperature differences between the multivariate calibration model and the validation samples [26]. Two parameters require to be estimated to build a LSS correction model, the degree of the polynomials (dependent on the nonlinearity of the temperature effects) and the number of factors describing the spectral information. In this study, LSS compensation method was applied using sample set A and B to explore the effect of samples homogenisation. Besides, preprocessing before and after LSS was explored.

Multiplicative effects model (MEM) uses two-step procedure for the correction of multiplicative effects in spectral measurements. First, the multiplicative parameters accounting for

multiplicative effects in the spectral measurements of the calibration samples are estimated by the optical pathlength estimation and correction (OPLEC) algorithm [28] and then the multiplicative effects in the spectral measurements are efficiently removed by a dual calibration strategy. In this work a modified version of the OPLEC algorithm is used which has been found to be faster and more robust [29, 30].

2.5 Model construction and performance

Partial least squares (PLS) regression was used to build the calibration models using PLS Toolbox chemometrics software (PLS_Toolbox_8.0.1, Eigenvector Research Inc., Manson, WA, USA). Models were developed using 2/3 of the samples for calibration and 1/3 for validation. Samples from different muscles and different hams were assigned to calibration and validation data sets to obtain a similar variability in terms of composition in both data sets. In all the cases, optimising the model involved trialling various pre-processing techniques and reviewing different spectral regions. The scatter correction methods evaluated included Mean centering, Multiplicative Scatter Correction (MSC), Extended MSC and Standard Normal Variate (SNV). First and second derivative were also explored to improve models' accuracy. The best combination of pre-treatments was selected based on the lowest cross validation error using venetian blinds (RMSECV) and the lowest number of latent variables (LV). A maximum of 10 LV was fixed for all the models to avoid overfitting. The goodness of fit of the models was assessed using the coefficient of determination (R_p^2), and the root mean square error of Prediction (RMSEP). Spectral shooter software (Spectral shooter v.3.0, CPACT, UK) was used for LSS and MEM model development. The models predictive ability obtained using the different combinations of pre-treatments were compared using a Fisher-Snedecor distribution (or F-distribution) in Microsoft Excel 2021 (Microsoft Corporation, USA). A p-value < 0.05 in this test shows a significant difference between the errors of each model [31].

2.6 Physicochemical determinations

Water content was analysed by drying at 103 ± 2 °C until reaching a constant weight [32]. Chloride content was determined according to [33] using a potentiometric titrator 785 DMP Titrino (Metrohm AG, Herisau, Switzerland) and expressed as salt content.

3. Results and discussion

3.1 Samples characterization

Table 1 shows the compositional characteristics of the different sets of samples used as food matrix for the evaluation of different compensation methods of spectral variations caused by temperature and packaging materials. Similar salt and water content ranges were obtained ensuring the appropriate application of spectral compensation methods and development of predictive models.

3.2 Effect of samples heterogeneity on spectral variability

Although visually homogeneous regions of interest (ROI) on dry-cured ham were selected (sample set A, Figure 1b), obtained spectra from these samples were found to have high to moderate variability. For this reason, five spectra for each sample were acquired and averaged before preprocessing and developing the models when using both spectrometers A and B. Thin fat streaks, tyrosine precipitates, colour variations and even slight variations in salt and water composition inside the ROI can produce an important variability. In fact, variability caused by sample's heterogeneity can be in some cases as high as the variability caused by external factors such as temperature (see Figure 2). This fact could be aggravated to the small measurement area of the used sensors, having 15- and 12-mm diameter for spectrometer A and B, respectively. Similar results were found when analysing ROIs on sample set B (homogenized samples) since still some heterogeneity can be found. Therefore, samples heterogeneity might be a big challenge for the development of bespoke sensors for the compositional prediction of food at a supermarket level. Number of spectra acquired for each kind of food should be optimized [21, 34, 35]. Use of hyperspectral imaging systems that allows visualisation and averaging of the whole sample and allow the segmentation of unwanted parts that might be not representatives for the prediction of a given component, might also be a solution to solve sample's heterogeneity problem.

3.3 Effect of temperature variation and packaging materials on spectral variability.

Temperature and packaging materials are common external factors occurring when measuring food products on site which produce important spectral variations. Figure 3 shows the spectral variation caused by temperature in spectrometer A and B using no packaging. Increases in temperature produce molecules to move up to higher energy levels resulting in spectral shifts and changes in absorption intensity [21, 36, 37]. Consequently, these variations must be considered inside the predictive model. Another factor increasing predictive error is the presence of packaging films. Figure 4 shows spectral variations caused by packaging materials of different composition and thickness at a constant temperature. In the case of spectrometer A, P1 and P2 produce a similar absorbance to non-packaged samples. In contrast, in the case of spectrometer B, the observed shift is higher which was attributed to the higher working distance between the sample and the sensor, that can increase scattering and reflectance of light.

3.4 Comparison of Temperature compensation methods

Temperature compensation methods were investigated in spectra acquired using spectrometer A. Table 2 shows the model characteristics and the predictive errors when using different strategies to correct temperature-induced spectral variations using this spectrometer on non-homogenized samples. These kinds of samples were used to be the real scenario found in a supermarket. However, as presented in this table, LSS correction was performed using both homogenized and

Developing robust food composition models: strategies for handling temperature and packaging variations in dry-cured ham using near infrared spectrometry

non-homogenized samples to evaluate the increase of performance of this method when using more homogeneous samples to create the compensation LSS model. Results in Table 2 show the use of temperature in GM-X and GM-Y do not help to increase performance of salt and water content predictive models in comparison to the GM. Food matrixes are complex and heterogeneous matrixes and including temperature information in X or Y variables did not decrease predictive error. Haroon et al [38] found that GM also performed better than GM-X and GM-Y when predicting viscosity in micellar liquids by recognising temperature effects more readily. LSS, when predicting salt, seems to slightly decrease RMSEP from 0.530% to 0.501% when sample set B was used to develop the LSS compensation algorithm. This fact could be attributed to the higher sample's homogeneity. However, similar prediction errors using sample set A and B were obtained when applied to water content prediction, thus we cannot say that the use of homogenized samples to develop the LSS model helps to increase performance of predictive models. LSS compensation method has been demonstrated to work when there is a constant deviation of the spectra due to temperature in synthetic samples [39]. However, in this study, heterogeneity of both sample sets A and B, might produce a spectral variation that can interfere with the constant spectral variations caused by temperature. This fact could lead to an insufficient correction of temperature effects and thus not providing an improvement of the model's performance. GLSW showed to be the best compensation method to retrieve temperature effect on the studied food model since decreases both salt and water content predictive errors to those found when using only spectra acquired at $15^{\circ}\text{C}\pm 2.55\%$ (which were 0.450% for salt and 1.961% for water). Salt content predictive errors using GLSW pretreatment were found to be significantly lower than other pre-treatments tested whereas water content predictive errors were lower but not significantly different than LSS pre-treatment ($p < 0.05$). The optimal α value was 0.000001 for salt and 0.0001 for water since further decreases in α resulted in increases of model error suggesting that variance in the data associated with composition was being removed. A possible explanation for this could be the fact that GLSW can retrieve simultaneously not only temperature effects, but also other interferences present in food complex matrixes.

Additionally, for comparison of heterogeneity effect, GM and GLSW methods were also used to develop predictive models using spectra acquired on homogenized samples (sample set B) (Table 3) which, although having some heterogeneity, are more homogeneous than non-homogenized samples. Salt and water content predictive errors were lower after GLSW pretreatment but only significantly different in the case of water ($p < 0.05$). Results show that there is an important difference on using sample set A and B for the performance of predictive models, especially in the case of water content prediction, (see Table 2 for comparison). This fact suggests that samples heterogeneity on samples non-homogenized might produce disturbance, but it is not relevant for compensation strategies correction.

3.5 Comparison of compensation methods for packaging films-induced variations

Since spectral variations caused by packaging were mainly observed in spectra acquired using spectrometer B (see Figure 4b), packaging compensation methods were investigated in this data set. First, it must be remarked that prediction ability of spectrometer B is lower than spectrometer A probably due to the higher noise found in these spectra. Spectrometer B showed RMSEP values of 0.703% for salt and 2.932% for water when using only NP samples and conventional spectral pretreatments (Mean scattering correction and First derivative) whereas spectrometer A provided RMSEP values of 0.496% for salt and 2.118% for water in the same conditions. In the case of salt, correlation coefficient was very low ($R^2=0.529$) indicating that salt content might be difficult to be predicted using spectrometer B. For this reason, compensation methods for retrieving packaging effect using NIRTA were only evaluated on water content parameter. Performance of GM, GLSW and MEM methods for the subtraction of packaging materials variations when developing predictive models for water content at a constant temperature using spectrometer B is presented in Table 4. GM show similar values to those obtained when using spectra from NP samples. MEM method it is not correcting the variation caused by plastic even worsen the results in the validation dataset, showing significantly higher predictive errors ($p>0.05$). GLSW did neither significantly reduce the predictive error.

To ascertain if plastic material has a real effect on the measurement, the developed predictive model using only NP samples was used to predict water content in packaged and non-packaged samples. The obtained RMSEP value was 2.668% with a bias of -1.35 %. In contrast, when using models developed using NP, P1 and P2, RMSEP was lower (2.019%) and the bias was reduced to 0.012 (Figure 6). Therefore, GM including samples acquired on different packaging materials is preferable to ensure performance and robustness of the model. Ding et al [18] also found that different packaging materials (polyvinyl chloride, polyethylene and expandable polyethylene) used for mango fruit could result in intensive spectral interference at 1150–1250 nm and 2320–2400 nm, and significantly decreased the NIR prediction accuracy of mango quality attributes. Similarly, Zhang et al [19] reported that reflectance spectra of beef in the presence of PE and PP films were affected by the light scattering, transmission and interference of films, leading to a certain effect on the prediction of beef total volatile basic nitrogen content, which could be properly eliminated by using different spectral preprocessing and modelling methods.

In overall, it must be remarked that factors such as the studied temperature and packaging film and other factors could affect the robustness of the models simultaneously. Therefore, the synergistic effect of several factors should be studied in future works.

4. Conclusions

Developing robust food composition models: strategies for handling temperature and packaging variations in dry-cured ham using near infrared spectrometry

The use of compensation strategies to overcome spectral variations caused by temperature and packaging materials when using affordable low-cost spectrometers intended for consumers can help to provide more robust and reliable predictions. GLSW method was found to be the most convenient compensation method for temperature when using spectrometer A whereas GM together with conventional spectral pretreatments was enough for retrieving packaging effect when using spectrometer B, although including spectra acquired on different packaging is recommended. Furthermore, sample heterogeneity must be also considered since inherent sample variation could produce as much spectral variation as external factors. More experimental work on compensation strategies for different types of food matrixes would help to increase predictability of models developed on food products.

Author Contributions

Elena Fulladosa: Investigation, formal analysis, writing-original draft preparation, conceptualization; Magdalene Chong, Andrew Parrot, Alison Nordon: formal analysis, writing-review and editing, conceptualization. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

This work was supported by CCLabel project (RTI-2018-096883-R-C41) funded by the Spanish Ministry of Science and Innovation of the Spanish Government (MCIN/AEI/10.13039/501100011033). Ricardo dos Santos received a grant [PRE2022-103798] funded by MICIU/AEI/10.13039/501100011033 and ESF+. Acknowledgements are extended to the consolidated Research Group (2021 SGR 00461), CERCA programme from Generalitat de Catalunya, Centre for Process Analytics and Control Technology (CPACT) and Engineering and Physical Sciences research Council (EPSRC; EP/P006965/1). Elena Fulladosa acknowledges the receipt of a fellowship from the OECD Co-operative Research Programme: Sustainable Agricultural and Food Systems in 2022 and a mobility grant from the Institute of Agrifood Research and Technology (IRTA) within the Incentives For Research Program 2023.

Conflicts of Interest

The authors declare no conflict of interest.

5. References

1. Prevolnik, M., et al., *Accuracy of near infrared spectroscopy for prediction of chemical composition, salt content and free amino acids in dry-cured ham*. Meat Science, 2011. **88**(2): p. 299-304.
2. Muller, M. and M.R.L. Scheeder, *Determination of fatty acid composition and consistency of raw pig fat with near infrared spectroscopy*. Journal of Near Infrared Spectroscopy, 2008. **16**: p. 305-309.

3. Wold, J.P., et al., *Non-contact transreflectance near infrared imaging for representative on-line sampling of fried salted coalfish (bacalao)*. Journal of Near Infrared Spectroscopy, 2006. **14**: p. 59-66.
4. Pullanagari, R.R., I.J. Yule, and M. Agnew, *On-line prediction of lamb fatty acid composition by visible near infrared spectroscopy*. Meat Science, 2015. **100**: p. 156-163.
5. Gou, P., et al., *Feasibility of NIR interactance hyperspectral imaging for on-line measurement of crude composition in vacuum packed dry-cured ham slices*. Meat Science, 2013. **95**(2): p. 250-255.
6. Cayuela-Sánchez, J.A., et al., *Compositional method for measuring the nutritional label components of industrial pastries and biscuits based on Vis/NIR spectroscopy*. Journal of Food Composition and Analysis, 2020. **92**: p. 103572.
7. Pérez-Marín, D.C., et al., *Near-infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs: chemical composition and open-declaration*. Animal Feed Science and Technology, 2004. **116**: p. 333-349.
8. Fulladosa, E., et al., *Near infrared sensors for the precise characterization of salt content in canned tuna fish*. Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 2023. **303**.
9. Giro-Candanedo, M., et al., *Differentiation between fresh and frozen-thawed mackerel fish using low-cost portable near infrared spectrometry devices*. Journal of Food Engineering, 2024. **379**.
10. Giro-Candanedo, M., et al., *Precise nutritional labelling of sliced packaged dry-cured ham using multi-energy X-ray absorptiometry*. Meat Science, 2023. **204**.
11. He, H.J., D. Wu, and D.W. Sun, *Nondestructive Spectroscopic and Imaging Techniques for Quality Evaluation and Assessment of Fish and Fish Products*. Critical Reviews in Food Science and Nutrition, 2015. **55**(6): p. 864-886.
12. Li, M., et al., *Evaluating the performance of a consumer scale SCiO™ molecular sensor to predict quality of horticultural products*. Postharvest Biology and Technology, 2018. **145**: p. 183-192.
13. Reis, M.M., et al., *Non-invasive differentiation between fresh and frozen/thawed tuna fillets using near infrared spectroscopy (Vis-NIRS)*. LWT - Food Science and Technology, 2017. **78**: p. 129-137.
14. Bec, K.B., J. Grabska, and C.W. Huck, *Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers*. Chemistry-a European Journal, 2021. **27**(5): p. 1514-1532.
15. Yang, Z.Y., et al., *Miniaturization of optical spectrometers*. Science, 2021. **371**(6528): p. 480-+.
16. Beć, K.B., et al., *Handheld near-infrared spectrometers: Where are we heading?* NIR news, 2020. **31**(3-4): p. 28-35.
17. Fodor, M., et al., *The Role of Near-Infrared Spectroscopy in Food Quality Assurance: A Review of the Past Two Decades*. Foods, 2024. **13**(21): p. 3501.
18. Ding, F., et al., *Non-invasive prediction of mango quality using near-infrared spectroscopy: Assessment on spectral interferences of different packaging materials*. Journal of Food Engineering, 2023. **357**: p. 111653.
19. Zhang, W., L. Pan, and L. Lu, *Prediction of TVB-N content in beef with packaging films using visible-near infrared hyperspectral imaging*. Food Control, 2023. **147**: p. 109562.
20. Chen, Z.-P., et al., *Calibration of Multiplexed Fiber-Optic Spectroscopy*. Analytical Chemistry, 2011. **83**(7): p. 2655-2659.
21. Kartakoullis, A., et al., *Feasibility study of smartphone-based Near Infrared Spectroscopy (NIRS) for salted minced meat composition diagnostics at different temperatures*. Food Chemistry, 2019. **278**: p. 314-321.
22. Chen, Z.-P., et al., *Quantitative Analysis of Powder Mixtures by Raman Spectrometry: the influence of particle size and its correction*. Analytical Chemistry, 2012. **84**(9): p. 4088-4094.

23. Roger, J.M., F. Chauchard, and V. Bellon-Maurel, *EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits*. Chemometrics and Intelligent Laboratory Systems, 2003. **66**(2): p. 191-204.
24. Fearn, T., *On orthogonal signal correction*. Chemometrics and Intelligent Laboratory Systems, 2000. **50**(1): p. 47-52.
25. Sun, X.D., P. Subedi, and K.B. Walsh, *Achieving robustness to temperature change of a NIRS-PLSR model for intact mango fruit dry matter content*. Postharvest Biology and Technology, 2020. **162**.
26. Chen, Z.P., J. Morris, and E. Martin, *Correction of temperature-induced spectral variations by loading space standardization*. Analytical Chemistry, 2005. **77**(5): p. 1376-1384.
27. Segtnan, V., et al., *Low-Cost Approaches to Robust Temperature Compensation in Near-Infrared Calibration and Prediction Situations*. Applied spectroscopy, 2005. **59**: p. 816-25.
28. Chen, Z.P., J. Morris, and E. Martin, *Extracting chemical information from spectral data with multiplicative light scattering effects by optical path-length estimation and correction*. Analytical Chemistry, 2006. **78**(22): p. 7674-7681.
29. Jin, J.W., et al., *Quantitative Spectroscopic Analysis of Heterogeneous Mixtures: The Correction of Multiplicative Effects Caused by Variations in Physical Properties of Samples*. Analytical Chemistry, 2012. **84**(1): p. 320-326.
30. Du, W., et al., *Maintaining the predictive abilities of multivariate calibration models by spectral space transformation*. Analytica Chimica Acta, 2011. **690**(1): p. 64-70.
31. Mohr, D.L., W.J. Wilson, and R.J. Freund, *Statistical methods*. 2021: Academic Press.
32. AOAC, *Official method 950.46, Moisture in meat, B. Air drying*. In K. Helrich (Ed.) (15th ed.). *Official methods of analysis of the association of official analytical chemists (Vol. II, pp. 931)*. 1990: Arlington: Association of Official Analytical Chemists, Inc.
33. ISO 1841-2, *Meat and meat products. Determination of chloride content — Part 2: Potentiometric method (Reference method)*. Geneva: International Organization for Standardization., in *ISO 1841-2 (1996)*. 1996.
34. Moon, E.J., et al., *Evaluation of Salmon, Tuna, and Beef Freshness Using a Portable Spectrometer*. Sensors (Basel), 2020. **20**(15).
35. Lörchner, C., et al., *Towards common useable spectra in non-targeted analysis - A feasibility study by mid-infrared spectroscopy, transfer and correction approaches*. Chemometrics and Intelligent Laboratory Systems, 2023. **240**: p. 104904.
36. Hansen, W.G., et al., *Tolerance of near infrared calibrations to temperature variations; a practical evaluation*. Journal of near Infrared Spectroscopy, 2000. **8**(2): p. 125-132.
37. Blanco, M. and D. Valdes, *Influence of temperature on the predictive ability of near infrared spectroscopy models*. Journal of near Infrared Spectroscopy, 2004. **12**(2): p. 121-126.
38. Haroon, K., et al., *Evaluation of temperature compensation methods for a near-infrared calibration to predict the viscosity of micellar liquids*. Journal of Chemometrics, 2020. **34**(11).
39. Chong, M.W.S., et al., *Temperature Correction of Spectra to Improve Solute Concentration Monitoring by In Situ Ultraviolet and Mid-Infrared Spectrometries toward Isothermal Local Model Performance*. Organic Process Research & Development, 2022. **26**(11): p. 3096-3105.

FIGURES

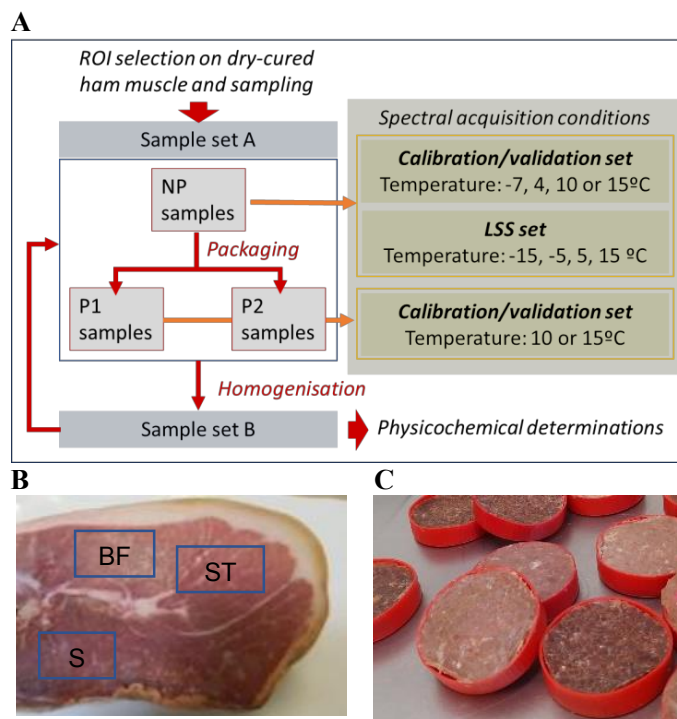


Figure 1. Experimental procedure (A) and description of the samples used in this study. Sample set A refers to the samples obtained from different regions of interest (ROIs) corresponding to *Biceps femoris*, *Semitendinosus* and *Semimembranosus* muscles (B), and Sample set B refers to the same samples after homogenization (C). NP: non-packaged samples, P1: packaged samples in plastic film 1, P2: packaged samples in plastic film 2.

Developing robust food composition models: strategies for handling temperature and packaging variations in dry-cured ham using near infrared spectrometry

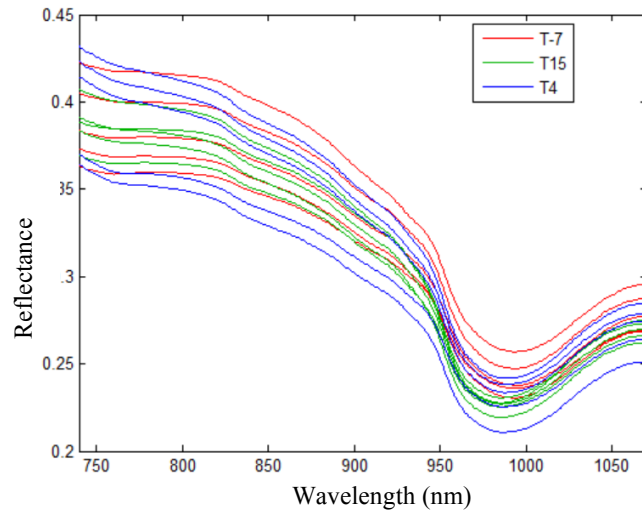


Figure 2. Spectral variability caused by sample's heterogeneity at three different temperatures. Spectra with the same colour represent replicates acquired on the same sample (from samples set A) using SCIO spectrometer.

Developing robust food composition models: strategies for handling temperature and packaging variations in dry-cured ham using near infrared spectrometry

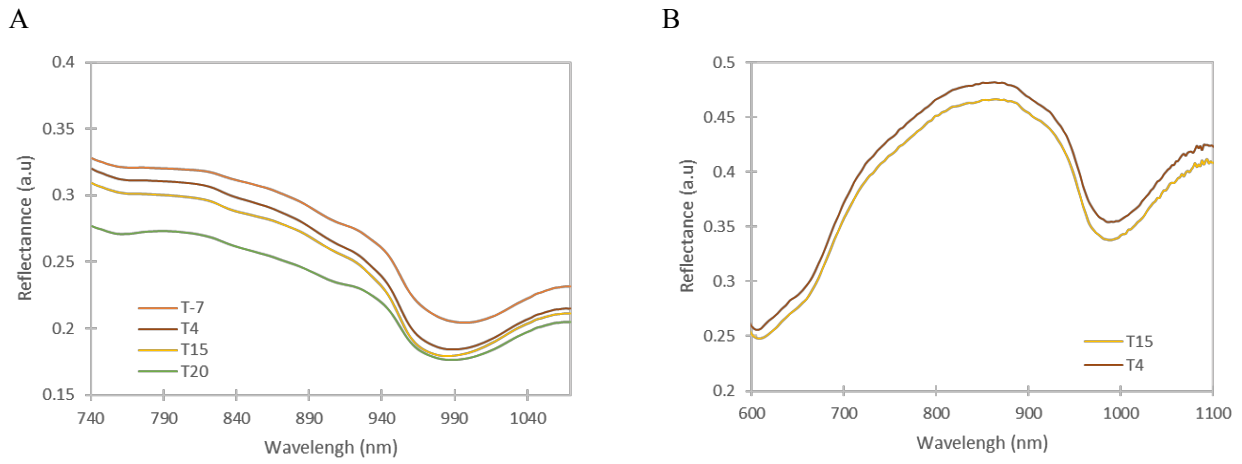


Figure 3. Spectral variations of the mean raw spectra acquired on sample set A using SCIO (A) and NIRTA (B) spectrometers at different temperatures without packaging.

Developing robust food composition models: strategies for handling temperature and packaging variations in dry-cured ham using near infrared spectrometry

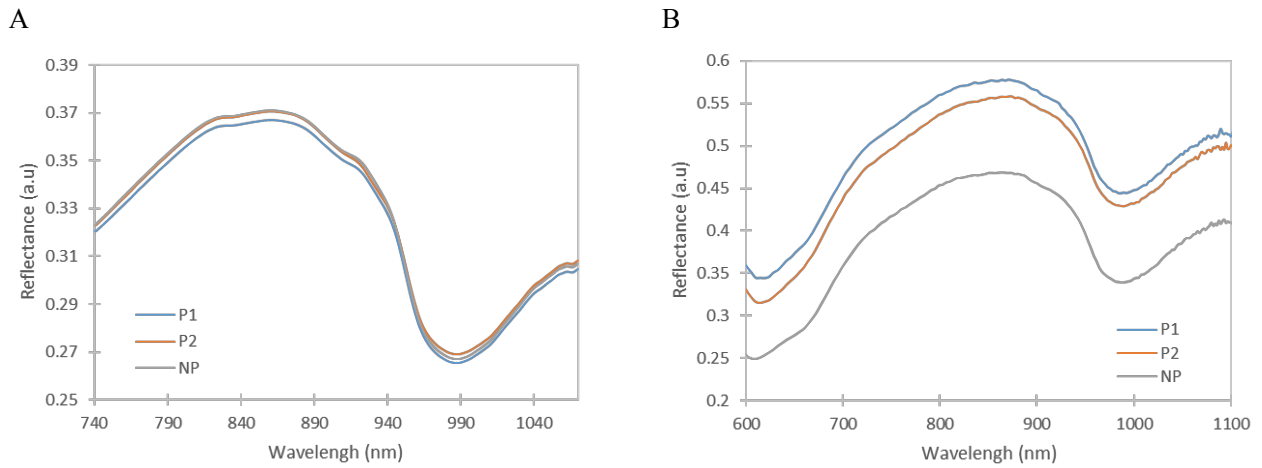


Figure 4. Spectral variations of mean raw spectra acquired on non-packaged (NP) and packaged using different packaging films (P1 and P2) using SCIO (A) and NIRTA (B) spectrometers at a constant temperature.

Developing robust food composition models: strategies for handling temperature and packaging variations in dry-cured ham using near infrared spectrometry

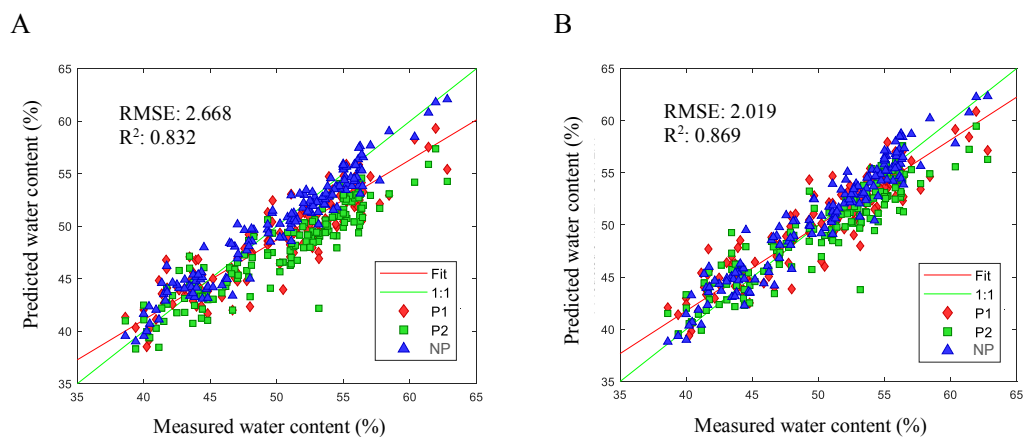


Figure 5. Measured vs predicted water content for a data set containing NP, P1 and P2 samples when using a model developed with spectra acquired only on NP samples (A) and a model developed with spectra acquired on NP, P1 and P2 samples (B). NP: non-packaged samples, P1: packaged samples in plastic film 1, P2: packaged samples in plastic film 2.

TABLES

Table 1. Chemical composition of the different data sets used in the study.

Data set	n	Salt content (%)				Water content (%)			
		Mean	std	min	max	Mean	std	min	max
LSS	20	5.09	0.98	3.39	7.11	50.09	5.64	40.39	56.23
Calibration	104	5.17	0.90	3.14	7.50	49.94	5.60	38.61	62.80
Validation	51	5.17	0.85	3.42	7.33	50.49	5.50	40.22	61.39

LSS: Loading Space Standardisation , n is the number of samples

Developing robust food composition models: strategies for handling temperature and packaging variations in dry-cured ham using near infrared spectrometry

Table 2. Characteristics and Prediction Errors of models developed applying different temperature compensation strategies on spectra acquired using SCIO spectrometer on non-homogenized samples.

Temperature compensation strategies	Calibration				Validation		
	n* spectra	LV	R ²	RMSEC	n* spectra	R ²	RMSEP**
Salt content %							
GM ¹	364	10	0.665	0.530	185	0.638	0.619 a
GM-X	364	10	0.623	0.562	185	0.586	0.607 a
GM-Y	364	10	0.623	0.562	185	0.586	0.607 a
LSS-non homogenized samples ²	364	10	0.667	0.528	183	0.617	0.642 a
LSS-homogenized samples ³	364	10	0.700	0.501	185	0.664	0.598 a
GLSW ⁴	364	3	0.787	0.422	183	0.719	0.461 b
Water content %							
GM ¹	364	10	0.929	1.955	185	0.914	2.192 a
GM-X	364	6	0.913	2.161	185	0.899	2.242 a
GM-Y	364	10	0.901	2.036	185	0.899	2.277 a
LSS-non-homogenized samples ²	364	6	0.916	2.118	183	0.909	2.100 ab
LSS-homogenised samples ³	364	5	0.911	2.185	185	0.911	2.108 ab
GLSW ⁴	364	3	0.929	1.948	183	0.925	1.935 b

¹Model developed using spectra acquired at different temperatures from -9.6 to 18.3°C.

²Model developed after application of LSS Correction Model with two significant factors and using non homogenized samples (in LSS set).

³Model developed after application of LSS Correction Model with two significant factors and using homogenised samples (in LSS set).

⁴Alpha value was 0.000001 in salt content predictive model and 0.00001 in water content predictive model.

*Number of spectra acquired at different conditions on samples used to develop or validate the model.

^{ab}Different letters indicate significant differences on predictive ability between the models according to the F-distribution (p-value < 0.05).

All the spectra were pre-processed using mean center and first derivative before model development except for models developed using GLSW that were not pre-processed.

Developing robust food composition models: strategies for handling temperature and packaging variations in dry-cured ham using near infrared spectrometry

Table 3. Characteristics and Prediction Errors of models developed applying GM and GLSW as temperature compensation strategies on spectra acquired using SCIO spectrometer on homogenized samples.

Temperature compensation strategies	Calibration				Validation		
	n * spectra	LV	R ²	RMSE C	n * spectra	R ²	RMSEP **
Salt content %							
GM ¹	324	10	0.842	0.429	156	0.649	0.519 a
GLSW ²	324	5	0.853	0.413	155	0.687	0.458 a
Water content %							
GM ¹	315	8	0.935	1.458	154	0.843	2.095 a
GLSW ²	315	3	0.935	1.462	151	0.924	1.400 b

¹Model developed using spectra acquired at different temperatures from -9.6 to 18.3°C. Spectra were pre-treated using mean center and first derivative before model development.

²Alpha values were 0.0001 in both salt and water content predictive models.

*Number of spectra acquired at different conditions on samples used to develop or validate the model.

^{ab}Different letters indicate significant differences on predictive ability between the models according to the F-distribution (p-value < 0.05).

Table 4. Characteristics and Prediction Errors of models developed applying different packaging compensation strategies on spectra acquired using NIRTA sensor on homogenized samples.

Packaging compensation strategies	Calibration				Validation		
	n * spectra	LV	R ²	RMSEC	n * spectra	R2	RMSEP **
Water content %							
GM ¹	312	7	0.868	2.028	150	0.726	2.887 a
MEM ²	312	10	0.836	1.539	150	0.637	3.382 b
GLSW ³	312	3	0.888	1.866	150	0.856	2.603 a

¹Model developed using spectra acquired at a constant temperature (10°C) on samples non packaged (NP) and vacuum packaged using different plastic films (P1 and P2).

²MEM were developed using 10 OPLEC components.

³Alpha value was 0.001 in water content predictive model.

*Number of spectra acquired at different conditions on samples used to develop or validate the model. All the spectra were pre-processed using a Mean scattering correction and First derivative before model development except for models developed using GLSW that were not pre-processed.

^{ab}Different letters indicate significant differences on predictive ability between the models according to the F-distribution (p-value < 0.05).